Univerza v Ljubljani
Fakulteta za računalništvo in informatiko

Jure Žabkar

Učenje kvalitativnih odvisnosti

DOKTORSKA DISERTACIJA

Ljubljana, 2010

Univerza v Ljubljani
Fakulteta za računalništvo in informatiko

Jure Žabkar

Učenje kvalitativnih odvisnosti

DOKTORSKA DISERTACIJA

Mentor: akad. prof. dr. Ivan Bratko
Somentor: doc. dr. Janez Demšar

Ljubljana, 2010

Povzetek

V tem delu opisujemo učenje kvalitativnih modelov iz podatkov, ki sodi v presek področji kvalitativnega sklepanja in strojnega učenja. Kvalitativni modeli so abstrakcija oziroma poenostavitev numeričnih modelov. Cilj kvalitativnega modeliranja je strojno učenje kvalitativnih modelov iz podatkov, pri čemer naj bi bili ti modeli enostavni in človeku razumljivi. Kvalitativni modeli so koristni, ker so blizu človeškemu načinu razmišljanja. Avtomatsko učenje takih modelov iz podatkov sledi enemu temeljnih ciljev področja umetne inteligence – narediti stroj, ki bo razmišljal kot človek. Še bolj je koristno kot orodje, ki človeku pomaga pri razumevanju in odkrivanju novega znanja iz podatkov, ki jih dandanes ljudje shranjujemo na vseh področjih.

Pri učenju kvalitativnih modelov izhajamo iz dejstva, da ljudje večinoma razmišljamo o odvisnostih med dvema količinama hkrati in čeprav tega eksplicitno ne izrazimo, predpostavljamo, da se ostale količine v istem kontekstu pri tem ne spreminjajo. V matematiki se to načelo imenuje parcialni odvod. Od odkritja (Newton in Leibniz ob koncu 17. stol.) dalje velja za nepogrešljivo orodje v matematiki in fiziki in je del osnovnega izreka matematične analize. Algoritmi za računanje parcialnih odvodov iz podatkov ter njihova uporaba pri učenju kvalitativnih modelov so osrednja tema te disertacije.

Izhajajoč iz matematične definicije parcialnega odvoda razvijemo šest metod s skupnim imenom Padé za računanje parcialnih odvodov iz regresijskih podatkov ter metodo za računanje parcialnih odvodov v klasifikaciji, algoritem Qube. Za učenje kvalitativnih modelov na podlagi parcialnih odvodov predlagamo dvo-stopenjsko shemo učenja, kjer v prvi fazi izračunamo parcialne odvode za vse učne primere, v drugi fazi pa uporabimo poljuben klasifikacijski algoritem strojnega učenja za gradnjo kvalitativnega modela. V sklopu teh raziskav razvijemo še štiri druge algoritme za gradnjo kvalitativnih modelov in Q^2 učenje, ki zaokrožajo to disertacijo.

Omenjene algoritme smo implementirali v programskem okolju za strojno učenje Orange in jih preizkusili na številnih umetnih podatkih, kjer smo analizirali njihove lastnosti. Nakazali smo nekaj primerov uporabe kvalitativnih parcialnih odvodov pri odkrivanju znanja iz podatkov. Opis eksperimentalnega dela zaokrožujemo s štirimi realnimi primeri – dvema manjšima s področja robotike in dvema večjima študijama primera uporabe v praksi. Padé smo preizkusili na pol-realni fizikalni domeni *biljard*, algoritem Qube pa na realni medicinski

domeni s področja infektologije. V obeh primerih smo za razlago modelov prosili strokovnjake z omenjenih področji. Oba algoritma sta se izkazala kot dobra in koristna pri konceptualizaciji domen. Delo zaključimo s kratko diskusijo rezultatov in možnostmi za nadaljnje delo.

Ključne besede

- umetna inteligenca, kvalitativno modeliranje, strojno učenje
- kvalitativno učenje, parcialni odvodi, okolice, sosednost
- Padé, Qube, izbirni nomogrami, Edgar, Strudel, Qing

Abstract

The thesis presents novel approaches to learning qualitative models from given data. Learning qualitative models lies in the intersection of qualitative reasoning and machine learning. Qualitative models are abstractions and simplifications of numerical models. Qualitative modelling thus tends to learn simple and comprehensible models. Qualitative models are useful because they support human way of reasoning. Automatic learning of such models from data follows the main goal of artificial intelligence which is to make machines reason like humans. Another useful aspect of automatic learning of qualitative models is to make software tools that assist humans in understanding and discovering new knowledge from data.

The motivation behind the methods that we develop in this thesis comes from the fact that, when inspecting the relation between two quantities, people most often consider only two quantities at a time. Although they do not make it explicit, they assume other quantities in the context constant. In mathematics, this principle is known as partial derivative and has been, since its discovery by Newton and Leibniz at the end of 17^{th} century, an indispensable tool in mathematics and physics. The core of this thesis deals with the algorithms for computation of partial derivatives from data and learning qualitative models from partial derivatives.

Taking mathematical definition of partial derivative as a foundation, we have developed six methods (with a common name Padé) for computation of partial derivatives in regression domains and a method (Qube) for computation of probabilistic partial derivatives in classification. We proposed a novel two-phase method for learning qualitative models from precomputed partial derivatives. In the first phase, we compute qualitative partial derivatives for each learning example and in the second phase we use an appropriate machine learning algorithm for classification to induce a qualitative model. As a part of our research, we have also developed four other algorithms for learning qualitative models and Q^2 learning. We shortly describe these algorithms at the end of the thesis.

We have implemented the above mentioned methods in Orange, an open source machine learning framework. We tested and evaluated them in controlled environment, a set of artificial domains, where we studied their properties. Further, we demonstrated how qualitative partial derivatives can be used in knowledge discovery from data. We conclude the experimental section with

four realistic domains – two smaller robotic domains and two case studies. We used Padé in realistically simulated domain *billiards* and Qube in a real medical domain from infectology. In both cases we asked domain experts for explaining the induced qualitative models. Both algorithms induced simple, accurate and comprehensible models and proved useful in the conceptualization of the domains. We conclude the thesis with a short discussion of the results and possible further work.

Keywords

- artificial intelligence, qualitative modelling, machine learning
- qualitative learning, partial derivatives, neighbourhood
- Padé, Qube, selective nomograms, Edgar, Strudel, Qing

IZJAVA O AVTORSTVU
doktorske disertacije

Spodaj podpisani *Jure Žabkar*,
z vpisno številko *63990371*,

sem avtor doktorske disertacije z naslovom

Učenje kvalitativnih odvisnosti

S svojim podpisom zagotavljam, da:

- sem doktorsko disertacijo izdelal samostojno pod vodstvom mentorja *akad. prof. dr. Ivana Bratka* in somentorja *doc. dr. Janeza Demšarja*
- so elektronska oblika doktorske disertacije, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko doktorske disertacije
- in soglašam z javno objavo elektronske oblike doktorske disertacije v zbirki "Dela FRI".

V Ljubljani, dne 10. septembra 2010 Podpis avtorja:

Zahvala

Najprej bi se rad zahvalil mentorjema, akad. prof. dr. Ivanu Bratku in doc. dr. Janezu Demšarju. Vsak na svoj način sta mi pokazala kakšno je raziskovalno delo na zelo visokem nivoju in obema sem za izkušnjo dela z vama zelo hvaležen. Skupno delo na različnih projektih je popestrilo moje raziskovalno delo in me dodatno motiviralo, pri čemer sem se neprestano učil od vaju. Rad bi se zahvalil tudi ostalima članoma komisije, prof. dr. Neži Mramor-Kosta in prof. dr. Sašu Džeroskemu, za koristne pripombe k tej disertaciji. Neža, hvala za neštete matematične debate in prijazno pomoč pri reševanju problemov.

V laboratoriju za umetno inteligenco sem se že od nekdaj dobro počutil – hvala vsem kolegom, ki so prispevali k temu. Še posebno sem hvaležen dr. Martinu Možini, s katerim sva se skupaj prebijala čez doktorski študij, evropske projekte in neštete bolj ali manj zanimive ideje. Rad bi se zahvalil prof. dr. Blažu Zupanu, ki me je že kot študenta navdušil za strojno učenje in me sprejel med programerje najboljšega orodja za strojno učenje na svetu. Čeprav se najina delovna časa le bežno sekata, Saša Sadikov vedno najde čas za bolj ali manj učeno kramljanje. Saša, hvala za prijaznost in pomoč pri mnogih stvareh. Hvala tudi Blažu Strletu, Damjanu Kužnarju, Aljažu Košmerlju in Tadeju Janežu, s katerimi smo sodelovali na evropskih projektih. Veselilo me je tudi delo z dr. Gregorjem Jeršetom, s katerim sva modrovala o računski topologiji. Hvala tudi ostalim članom laboratorija, ki so se ukvarjali z drugimi področji, a so bili vedno pripravljeni prisluhniti tudi mojim idejam: Tomažu Curku, Gregorju Lebanu, Minci Mramor, Mateju Guidu, Lanu Umeku, Mihi Štajdoharju, Marku Toplaku, Juretu Žbontarju, Lanu Žagarju in Vidi Groznik.

Za pomoč pri delu na projektih, ki so posredno vključeni tudi v to disertacijo bi se rad zahvalil Ashoku Mohanu, Sašu Moškonu, Simonu Kozini ter ostalim sodelavcem projektov ASPIC, XMEDIA in XPERO. Za pomoč bi se rad zahvalil prof. dr. Francu Strletu in dr. Jerneji Videčnik s klinike za infekcijske bolezni in vročinska stanja v Ljubljani, s katerima sem sodeloval pri analizi podatkov o bakterijskih okužbah pri starostnikih. Njuni so tudi komentarji in razlage modelov, ki sem jih vključil v to disertacijo. Čeprav sem se od njiju naučil veliko novega, je moje znanje in razumevanje njunega področja še vedno tako majhno, da se mi je pri pisanju utegnila pripetiti kakšna napaka.

S posebno hvaležnostjo, bi se rad zahvalil mami Nuši in očetu Antonu, ki sta me po svojih najboljših močeh vzgojila in mi pomagala na moji življenski poti.

Hvala tudi bratoma, Janezu in Blažu, za čase, ko smo znanost na svoj način odkrivali skupaj. Hvala tudi teti Idi za vse, kar je dobrega storila zame.

Nazadnje bi se rad zahvalil svoji družini, ženi Beti in najinim trem "doktoratom": Klari, Antonu in Martinu – vsak od vas je nekaj posebnega. Hvala za potrpežljivost, ko sem bil z znanostjo namesto z vami, četudi smo pri tem sedeli za isto mizo ali se igrali z istimi kockami. Verjamem, da je velikokrat videti drugače, ampak zame ste vi na prvem mestu. Hvala, da ste mi ves čas stali ob strani. Hvala za ljubezen, veselje in družinsko srečo, ki jo doživljam z vami.

Jure Žabkar

Kazalo

Poglavje 1

Uvod

Učenje kvalitativnih odvisnosti oziroma kvalitativnih modelov spada na področje kvalitativnega sklepanja (*angl. Qualitative Reasoning*). Kvalitativno sklepanje je področje umetne inteligence, ki se tradicionalno ukvarja z modeliranjem zveznih količin (npr. prostor, čas, energija). Kvalitativno sklepanje namesto natančnih funkcijskih zvez med spremenljivkami (matematičnih enačb) išče trende naraščanja in padanja ter pomembne odlikovane vrednosti spremenljivk (*angl. landmarks*). Področje sicer obsega še druge oblike sklepanja in modeliranja, ki pa se neposredno ne navezujejo na pričujoče delo in jih zato ne omenjamo.

Kvalitativno sklepanje poskuša posnemati človeški način sklepanja. Slednje predstavlja glavno gonilo razvoja tega področja, ki izhaja iz dejstva, da ljudje v vsakodnevnem življenju večinoma sklepamo kvalitativno. S tem mislimo na sklepanje na podlagi majhne količine relativno nezanesljivih podatkov glede na podatke, ki bi bili potrebni za kvantitativne metode. Večina sklepanja, celo v naravoslovni znanosti, poteka v naravnem jeziku, brez matematičnih enačb [Kalagnanam et al., 1991].

Primer kvalitativnega opisa je izjava: "Močneje ko vržem kamen, višje leti.", ustrezen fizikalni opis pa: "$h_{max} = d_0 + v_0^2/2g$.", kjer je h_{max} višina, ki jo doseže kamen, v_0 in d_0 pa sta začetni vrednosti hitrosti in položaja. Kvalitativno lahko povemo še, da se kamnu hitrost zmanjšuje vse do najvišje točke, ko doseže vrednost 0, nato pa začne kamen ob padanju spet pridobivati na hitrosti. Če ga ujamemo na isti višini kot smo ga spustili iz roke, je njegova hitrost ob povratku enaka začetni hitrosti (ob predpostavki, da zanemarimo zračni upor). Podobno lahko razmišljamo tudi o energiji kamna. Razmeroma enostaven poskus lahko zelo natančno kvalitativno opišemo. Kvalitativni opis s hitrostjo in položajem

je tako enostaven, da je razumljiv celo predšolskim otrokom. Vsaj slednje za opis z enačbo ne drži. Enačbe na prvi pogled dajejo vtis večje natančnosti, seveda ob predpostavki, da poznamo natančne vrednosti začetnih pogojev, v_0 in d_0. Vendar že pri tako enostavnem poskusu teh vrednosti praktično nikoli ne poznamo. Večina ljudi nima niti občutka ali vrže kamen s hitrostjo $1m/s$ ali $10m/s$, pa tudi ocena d_0 je zelo nenatančna. Podrobnosti, kot so zračni upor (izbira med linearnim in kvadratnim je odvisna od hitrosti), nadmorska višina in dejstvo, da zgornja enačba velja za točkasto maso, kar kamen gotovo ni, postanejo za samo sklepanje nepomembne. Natančnost, ki jo obljublja enačba, v takih primerih zbledi.

Zato, da ljudje uporabljamo kvalitativno sklepanje, obstaja enostaven razlog. Človek svet okrog sebe dojema na način, ki je primeren za tak način sklepanja in vsekakor ni primeren za reševanje numeričnih sistemov in posledično sklepanje na njihovi osnovi. Področje umetne inteligence si prizadeva razviti stroje, ki bi čim bolj posnemali človeški način sklepanja. Toda, če človekovi "senzorji" ne delajo s številkami, za moderne robote velja ravno obratno. Praktično vsi robotski senzorji, od video kamer do lidarjev, delajo z veliko numerično natančnostjo. Čemu potem kvalitativno sklepanje? Človek je svoj življenjski prostor v veliki meri podredil sebi in skozi evolucijo je razvijal tudi ustrezen način sklepanja. Z vprašanji o primernosti robotskih senzorjev za namene opazovanja človekove okolice se tu ne bomo ukvarjali, predlagali pa bomo, po našem mnenju ustrezen način modeliranja s strojnim učenjem, ki informacijo iz numerične oblike pretvori v kvalitativno in s tem tudi robotu omogoči kvalitativno sklepanje.

Večino kvalitativnih modelov ljudje zgradimo na podlagi izkušenj oziroma predhodnega znanja. Ob razmahu računalniške tehnologije in velikih količin podatkov, ki jih z njeno pomočjo ljudje zbiramo, pa je smiselno razmišljati tudi o avtomatskem učenju kvalitativnih modelov iz podatkov. Pri tem si želimo, da bi bili ti modeli še vedno tako enostavni, razumljivi in uporabni kot tisti, ki jih zgradimo ljudje.

Učenje kvalitativnih modelov iz podatkov je veja kvalitativnega sklepanja, ki se ukvarja z gradnjo modelov z namenom uporabe v procesih kvalitativnega sklepanja. Tu se področje kvalitativnega sklepanja sreča s področjem strojnega učenja. Kvalitativne odvisnosti med zveznimi spremenljivkami, ki jih opisujejo kvalitativni modeli, so abstrakcije numeričnih odvisnosti med zveznimi spremenljivkami. Preprost primer kvalitativne odvisnosti med spremenljivkama x

in y, kjer velja $y(x) = x$, opišemo s pravilom: "x narašča (pada) $\Rightarrow$ y narašča (pada)". V kontekstu tega dela kvalitativni modeli opisujejo, kako kvalitativne spremembe vhodnih spremenljivk vplivajo na izhodno spremenljivko. Kvalitativni modeli so pogosto tako enostavni, da jih ne omenjamo eksplicitno. Primer takega modela v ekonomiji je zakon povpraševanja: "manjša kot je cena, daljša je vrsta (in obratno)". Obstaja veliko približnih modelov, ki so bolj ali manj natančni ob določenih predpostavkah. Splošnega numeričnega modela, ki bi opisoval zakon povpraševanja, ni, saj bi moral tak model gotovo vsebovati spremenljivke, ki jih ni možno izmeriti tako natančno, da bi bil model uporaben. Omenjeni kvalitativni model trgovci neprestano, čeprav morda podzavestno, uporabljajo pri določanju cen.

Večina obstoječih algoritmov za učenje kvalitativnih modelov iz podatkov gradi modele v obliki kvalitativnih diferencialnih enačb. V tem delu se ukvarjamo z drugačnimi kvalitativnimi modeli – takimi, ki opisujejo, kako sprememba določene vhodne spremenljivke vpliva na izhod v kontekstu, ki ga določajo ostale vhodne spremenljivke.

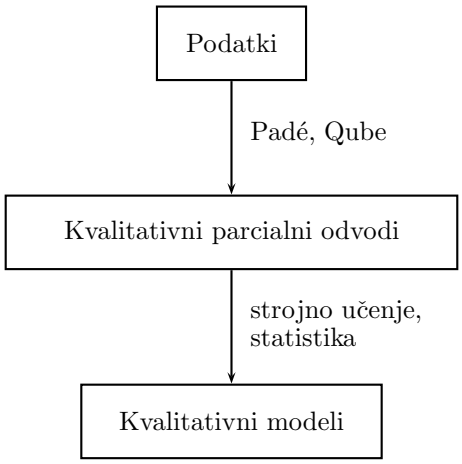
Osrednja tema disertacije so algoritmi za računanje parcialnih odvodov iz podatkov ter njihova uporaba pri učenju kvalitativnih modelov. Ločeno obravnavamo učenje kvalitativnih modelov na regresijskih (metode s skupnim imenom Padé) in klasifikacijskih podatkih (algoritem Qube). Tako Padé kot Qube sta predprocesorja, ki računata kvalitativne parcialne odvode. Iz njiju lahko z uporabo ustreznega algoritma za klasifikacijo zgradimo kvalitativni model, kot prikazuje slika 1.1.

Vsem metodam za računanje kvalitativnih parcialnih odvodov (Padé in Qube) je skupno to, da posnemajo matematično definicijo parcialnega odvoda. Parcialni odvod funkcije $f(x_1, \ldots, x_n)$ po spremenljivki x_i v točki $A = (a_1, \ldots, a_n)$ je:

$$\frac{\partial f}{\partial x_i}(a_1, \ldots, a_n) = \lim_{h \to 0} \frac{f(a_1, \ldots, a_i + h, \ldots, a_n) - f(a_1, \ldots, a_n)}{h}. \tag{1.1}$$

Parcialni odvod je sprememba funkcijske vrednosti pri spremembi izbrane spremenljivke za neskončno majhno vrednost, pri čemer vrednosti ostalih spremenljivk ostanejo fiksne. Za potrebe strojnega učenja moramo to definicijo nekoliko prilagoditi iz več razlogov:

1. V matematiki odvajamo funkcije, v strojnem učenju pa funkcij ne poznamo, še več – skriti koncepti, ki jih modeliramo, so le redko funkcije v



Slika 1.1: Splošna shema učenja kvalitativnih modelov, kot jo predlagamo v tem delu.

matematičnem pomenu. Koncept je lahko namreč večvrednostna funkcija, logična formula, odločitveno drevo ali zapis v kakšnem drugem formalnem jeziku. Za razliko od odvedljive matematične funkcije lahko koncept vsebuje tako zvezne kot diskretne atribute, logične strukture, verjetnostne porazdelitve idr. Naštete razlike se v praksi odražajo npr. pri iskanju globalnih ekstremov funkcije – matematične metode za odvajanje se lahko ujamejo v lokalnih ekstremih, če pa namesto s funkcijo delamo s podatki, lahko v linearnem času glede na število učnih primerov odkrijemo vse tipe globalnih ekstremov.

2. Izbrane neodvisne spremenljivke (atributa) ne moremo spremeniti za poljubno majhno vrednost, ker imamo na voljo samo dani vzorec podatkov. Običajno si pomagamo z nekaj najbližjimi sosedi danega primera, za katere predpostavimo, da imajo približno enake lastnosti. Tudi metode za numerično odvajanje se na nek način soočajo z istim problemom in ga tudi rešujejo na podoben način, pri čemer imajo naslednje prednosti:

   - vedno se računajo v ortogonalnih koordinatnih sistemih,
   - brez odvečnih spremenljivk
   - v razmerah brez šuma
   - poznajo funkcijo, ki jo odvajajo.

3. V izrazu 1.1 spreminjamo samo vrednost i-te spremenljivke, medtem ko vrednosti ostalih spremenljivk ostanejo iste. Ta stroga zahteva zagotavlja, da parcialni odvod izraža spremembo vrednosti funkcije glede na izbrano spremenljivko. Spreminjanje vrednosti ostalih spremenljivk bi utegnilo vpliv izbrane spremenljivke popolnoma prikriti. Pri računanju odvodov iz podatkov tej strogi zahtevi praktično ne moremo zadostiti, kar predstavlja največji problem pri posnemanju matematične definicije odvoda v strojnem učenju. Obravnavali bomo več različnih načinov, ki ta problem bolj ali manj uspešno rešujejo.

Uvodnemu poglavju, kjer smo predstavili področje in problematiko, s katero se ukvarjamo, sledi poglavje o učenju kvalitativnih odvisnosti v regresiji. V njem razvijemo algoritem Padé, ki ga sestavlja šest metod za računanje kvalitativnih parcialnih odvodov iz podatkov z zveznim razredom. V tretjem poglavju obravnavamo učenje kvalitativnih odvisnosti v klasifikaciji. Definiramo verjetnostne kvalitativne parcialne odvode in razvijemo algoritem Qube za njihovo računanje. Poglavji 3 in 4 vsebujeta tudi opis poskusov, ki smo jih opravili z razvitima algoritmoma ter diskusijo rezultatov. Metode preizkusimo na umetnih in realnih podatkih. Poskusi kažejo, da so razvite metode točne, odporne na šum v podatkih in uporabne v večrazsežnih prostorih atributov. Poskusi na realnih podatkih kažejo, da so naučeni modeli večinoma točni, enostavni in človeku razumljivi, metode pa relativno stabilne in zato zanesljive. Omenjene raziskave in rezultati predstavljajo glavne prispevke k znanosti. Poleg njih v poglavju 5 opišemo algoritme, ki se teoretično ali praktično navezujejo na Padé in Qube in nudijo priložnost za nadaljnje delo.

Prispevki k znanosti

1. Razvoj metod za izračun parcialnih odvodov na vzorčeni funkciji. Razvili smo šest novih metod (prvi trikotnik, regresija na zvezdi, regresija v cevi, τ-regresija, lokalno utežena regresija za računanje parcialnih odvodov, vzporedni pari) s skupnim imenom Padé, ki se razlikujejo po točnosti, odpornosti na šum ter časovni zahtevnosti.

2. Izvirni postopek gradnje kvalitativnih modelov, ki deluje v dveh korakih: najprej za vsak učni primer izračuna parcialne odvode, nato pa s pomo-

čjo splošnih algoritmov strojnega učenja zgradi model za napovedovanje predznakov teh odvodov.

3. Definicija diskretnega verjetnostnega kvalitativnega parcialnega odvoda na osnovi lokalnih pogojnih verjetnosti razreda.

4. Definicija kvalitativne parcialne spremembe na diskretnih domenah.

5. Razvoj metode Qube za izračun kvalitativne parcialne spremembe na vzorčeni funkciji, t.j. iz množice učnih podatkov.

6. Empirično ovrednotenje in primerjava zgoraj razvitih metod na primerno sestavljenih umetnih domenah in na realnih domenah, za katere so nam bili na voljo eksperti, ki so ocenili smiselnost in uporabnost dobljenih modelov.

Poglavje 2

Učenje kvalitativnih odvisnosti v regresiji

Osnova za učenje kvalitativnih odvisnosti v tem delu so kvalitativni parcialni odvodi (KPO) razredne spremenljivke po izbranem atributu. Odvode računamo 'po točkah', kar pomeni v vsakem učnem primeru posebej. Brez škode za splošnost vzemimo poljuben učni primer, ki ga imenujemo *referenčni*. Računanje odvoda v dani točki po definiciji zahteva premik izven te točke, saj nas zanima sprememba funkcijske vrednosti glede na spremembo v izbrani smeri v prostoru atributov. Odvod v referenčnem primeru lahko torej izračunamo le, če poznamo obnašanje funkcije v njegovi okolici. Naslednji razdelek namenimo metodam za izračun parcialnih odvodov, ki se med seboj razlikujejo po različnih tipih okolic referenčnega primera, iz česar izhaja tudi večina ostalih razlik med njimi.

2.1 Formalna definicija problema

Problem učenja kvalitativnih odvisnosti formalno definiramo takole. Dana je množica učnih primerov E. Učni primer $e = (\mathbf{x}, y) \in E$ je podan z vrednostmi atributov $\mathbf{x} = (x_1, \ldots, x_n)$ in razredom $y = f(\mathbf{x})$, kjer je f nepoznani koncept. Naloga je torej:

1. izračunati kvalitativni parcialni odvod (KPO) za vsak učni primer $e \in E$ in

2. z uporabo algoritmov strojnega učenja in izračunanih KPO zgraditi kvalitativni model, ki opisuje kvalitativne odvisnosti med razredom y in iz-

branim atributom x_i, $i \in \{1, \ldots, n\}$.

Kvalitativni parcialni odvod (KPO) razreda f po atributu x_i, $\partial_Q f/\partial_Q x$ ali krajše f_{Qx}, definiramo kot predznak parcialnega odvoda $\partial f/\partial x$: $\partial_Q f/\partial_Q x = f_{Qx} = \text{sign}(\partial f/\partial x)$. V algebrajski notaciji zapišemo: $f = Q(s\ x)$, kjer je $s \in \{+, -, \circ\}$. KPO razširi definicijo kvalitativne proporcionalnosti [Forbus, 1984] z dodatno kvalitativno vrednostjo "konstanta"($\circ$) k obstoječima "narašča"($+$) in "pada"($-$).

Kot primer vzemimo funkcijo $f(x) = x^2$. Učni podatki, E vsebujejo pare vrednosti (x, f), kjer je x atribut (neodvisna spremenljivka) in f razred (odvisna spremenljivka). Pravilni kvalitativni model, ki se ga želimo naučiti, bi bil:

$$x > 0 \Rightarrow f = Q(+x)$$

$$x < 0 \Rightarrow f = Q(-x)$$

Kvalitativna odvisnost $f = Q(+x)$ pomeni, da je f *kvalitativno proporcionalen* x. Z drugimi besedami, f narašča z x. V kontekstu Padéja pa to razumemo tudi kot $\frac{\partial f}{\partial x} > 0$.

Zapis $f = Q(-x)$ pomeni, da je f obratno kvalitativno proporcionalen x (parcialni odvod f po x je negativen).

V primeru prisotnosti diskretnega atributa (TipIzdelka), lahko kvalitativno odvisnost med ceno izdelka, tipom izdelka in njegovo velikostjo izrazimo takole:

$$\text{TipIzdelka} = \text{avto} \Rightarrow \text{Cena} = Q(+\text{VelikostIzdelka})$$

$$\text{TipIzdelka} = \text{računalnik} \Rightarrow \text{Cena} = Q(-\text{VelikostIzdelka})$$

Občasno bomo uporabljali tudi skrajšani zapis za več kvalitativnih odvisnosti skupaj, npr. $z = Q(+x)$ in $z = Q(-y)$ skrajšano zapišemo $z = Q(+x, -y)$.

2.1.1 Uvodni primer: matematično nihalo

Zamislimo si robota*, ki izvaja preprost fizikalni poskus z matematičnim nihalom. Njegova naloga je opazovati nihajni čas prvega nihaja T, v odvisnosti od dolžine vrvi l, mase uteži m, in začetnega odmika od ravnovesne lege φ, ter se naučiti čim več zakonitosti, ki jih pri tem opazi. Med izvajanjem poskusa robot shranjuje podatke v tabelo, katere vzorec podajamo v prvih štirih stolpcih tabele 2.1.

m	l	φ	T	$\partial_Q T/\partial_Q m$	$\partial_Q T/\partial_Q l$	$\partial_Q T/\partial_Q \varphi$
3.61	0.69	37.23	1.70	$\circ$	$+$	$+$
5.49	0.71	46.52	1.74	$\circ$	$+$	$+$
9.19	0.84	-48.91	1.91	$\circ$	$+$	$-$
7.17	0.33	33.89	1.17	$\circ$	$+$	$+$
6.81	0.50	65.93	1.51	$\circ$	$+$	$+$
4.64	0.69	-78.89	1.7	$\circ$	$+$	$-$

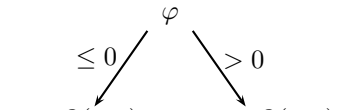
Tabela 2.1: Vzorec podatkov za poskus z matematičnim nihalom.

$T = Q(+l)$

(a) Kvalitativna odvisnost med nihajnim časom in dolžino vrvi.

$T = Q(\circ m)$

(b) Kvalitativna odvisnost med nihajnim časom in maso uteži.



(c) Kvalitativna odvisnost med nihajnim časom in začetnim odmikom.

Slika 2.1: Kvalitativni modeli, ki opisujejo odvisnosti med nihajnim časom T in opazovanimi atributi.

Ko zberemo dovolj podatkov, uporabimo Padé za izračun kvalitativnih odvisnosti $\partial_Q T/\partial_Q m$, $\partial_Q T/\partial_Q l$ in $\partial_Q T/\partial_Q \varphi$ za vsako meritev (učni primer). Izračuni so pripisani v originalno tabelo (Tabela 2.1, zadnji trije stolpci). Kvalitativne odvisnosti, ki smo jih izračunali za posamezne primere, posplošimo s kvalitativnim modelom, npr. kvalitativnim drevesom. Pri učenju kvalitativnih dreves uporabimo originalne atribute, za razred pa vzamemo kvalitativne parcialne odvode (po enega izmed zadnjih treh stolpcev iz tabele 2.1). Dobljena drevesa so prikazana na sliki 2.1. Dve drevesi sta pravzaprav samo lista: vedno velja, da nihajni čas narašča z naraščajočo dolžino vrvi, $T = Q(+l)$, in nihajni čas ni odvisen od mase uteži, $T = Q(\circ m)$. Drevo, ki opisuje odvisnost med nihajnim časom in začetnim kotom, pa pravi, da za negativne kote nihajni čas pada z naraščajočim kotom, za pozitivne kote pa narašča z njim. Skupaj si pravili lahko razlagamo kot $T = Q(+|\varphi|)$, česar pa algoritem za gradnjo dreves, ki ne more sestavljati novih konceptov, ne more izraziti v tej eksplicitni obliki.

*Del tega poskusa smo izvedli s humanoidnim robotom Nao.

2.2 Metode za izračun parcialnih odvodov

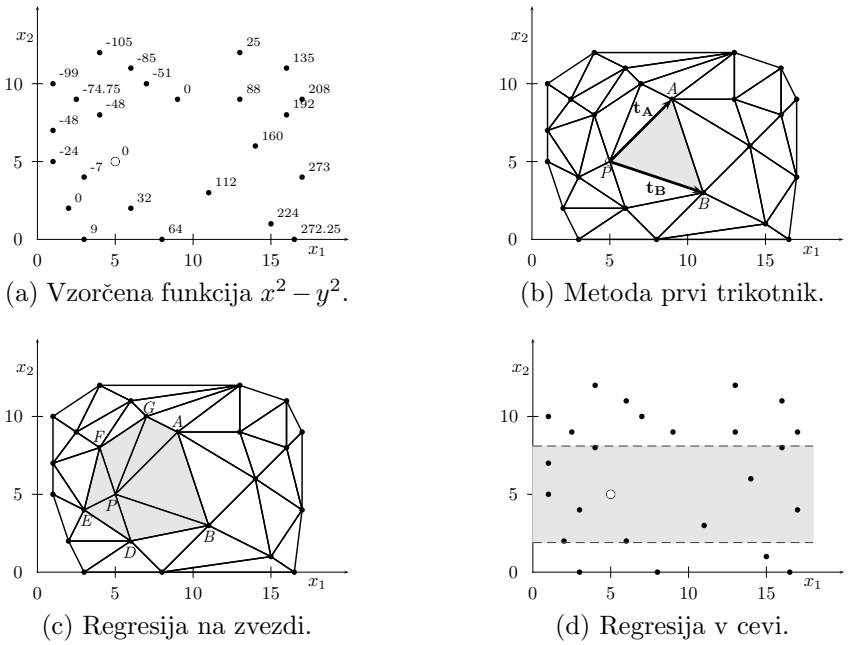
Padé je zbirka metod za izračun parcialnih odvodov iz numeričnih podatkov z zveznim razredom. Računa približke parcialnih odvodov, pri čemer nima nobene informacije o tem, kateri atributi sploh spadajo v množico neodvisnih spremenljivk. Pozna le v končno mnogo točkah vzorčeno funkcijo ter množico atributov.

Naš namen je računanje kvalitativnih parcialnih odvodov (KPO) za potrebe učenja kvalitativnih modelov. Padé vsebuje šest izvirnih metod za izračun KPO: prvi trikotnik, regresija na zvezdi, regresija v cevi, lokalno utežena regresija, τ-regresija in regresija na parih. Prvi dve metodi sta bolj matematično zanimivi, vendar manj uporabni v praksi, ker sta slabo odporni na šum in temeljita na triangulaciji. To prinaša s seboj določene omejitve, predvsem pri hitrosti in večrazsežnih prostorih. Ostale metode so bližje statistiki kot matematiki. Na račun 'mehčanja' strogih matematičnih predpostavk so metode bolj robustne in praktično uporabne. Med seboj se razlikujejo po načinu obravnavanja problema, ki smo ga opisali v točki 3 na strani 4. Vsem metodam je skupno to, da numerično izračunajo približek parcialnega odvoda, nato pa glede na njegov predznak določijo kvalitativni odvod.

2.2.1 Triangulacijske metode

Naj bo f zvezna funkcija n spremenljivk, $y = f(x_1, x_2, \ldots, x_n)$. Vrednost funkcije f je podana v N točkah, ki jim rečemo tudi učni primeri. Vsak učni primer je podan z vrednostmi atributov $(x_1, x_2, \ldots x_n)$ in razredom $f(x_1, x_2, \ldots, x_n)$. Cilj metod za izračun parcialnih odvodov je izračunati čim boljši približek $\partial f/\partial x_i$ za vsak učni primer P.

Kot primer vzemimo funkcijo dveh spremenljivk x_1 in x_2, ki jo naključno vzorčimo, kot prikazuje slika 2.2a. Vsaka točka predstavlja učni primer, koordinati točke pa vrednosti atributov. Funkcijska vrednost je izpisana ob vsaki točki. Izberimo točko $P = (5, 5)$ in spremenljivko x_1, po kateri bomo odvajali. Izračun si olajšajmo tako, da P postavimo v izhodišče koordinatnega sistema, zato označimo vektorje od P do A, B, $\ldots$s $\mathbf{t_A} = A - P$, $\mathbf{t_B} = B - P$,.... Skladno s tem definirajmo f nad krajevnimi vektorji, $f(\mathbf{t_A}) = f(A) - f(P)$, $f(\mathbf{t_B}) = f(B) - f(P)$, itn.



(a) Vzorčena funkcija $x^2 - y^2$.

(b) Metoda prvi trikotnik.

(c) Regresija na zvezdi.

(d) Regresija v cevi.

Slika 2.2: Grafični prikaz metod za računanje parcialnih odvodov vzorčene funkcije, ki temeljijo na triangulaciji, ter regresije v cevi.

Parcialni odvod $\partial f / \partial x_i$ v točki P je:

$$\frac{\partial f}{\partial x_i}(P) = \lim_{h \to 0} \frac{f(P + h\mathbf{x_i}) - f(P)}{h}, \tag{2.1}$$

kjer je $\mathbf{x_i}$ i-ti bazni vektor.

Zgornja definicija je ekvivalentna izrazu (1.1) in je še ne moremo uporabiti za računanje parcialnih odvodov na podatkih. Poleg limite nas moti predvsem to, da ne poznamo funkcije f in zato ne moremo izračunati njene vrednosti v poljubni točki. Lahko si pomagamo zgolj z danimi učnimi primeri. Taylorjev izrek nam pove, da je f v dovolj majhni okolici točke P približno linearna in se izraža takole:

$$f(x_1, \ldots x_n) = b_0 + \sum_{i=1}^{n} b_i x_i + \epsilon. \tag{2.2}$$

Ostanek Taylorjeve vrste ϵ predstavlja tako nelinearni del f kot tudi morebitni šum v podatkih. Želeni odvod $\partial f / \partial x_i$ je kar enak b_i v izrazu (2.2). Preostane nam, da definiramo ustrezno okolico točke P in izračunamo koeficient b_i. To lahko storimo na več načinov.

Metoda prvi trikotnik

Najpogostejši način za rekonstrukcijo funkcij v računski geometriji je s pomočjo triangulacije domene oz. prostora atributov. Najbolj znana je Delaunay-eva triangulacija, ki atributni prostor razdeli na disjunktna območja, t.i. simplekse, upoštevajoč razdaljo med točkami. V ravnini imajo simpleksi obliko trikotnikov, odtod ime triangulacija. Za potrebe v tem delu uporabljamo osnovno Delaunayevo triangulacijo implementirano v [Barber et al., 1996]. Slika 2.2b prikazuje triangulacijo točk s slike 2.2a. Predpostavimo, da v podatkih ni šuma in da je funkcija dovolj gosto vzorčena, tako, da je na vsakem trikotniku približno linearna.

Odvod v točki P in v smeri x_1 izračunamo po definiciji tako, da se iz P 'malo' premaknemo v smeri x_1. Tak premik nas pripelje v notranjost trikotnika PAB (slika 2.2b). Ker smo privzeli, da je funkcija na vsakem trikotniku linearna, brez težav izračunamo b_i. Formalno to zapišemo kot interpolacijo vrednosti f na oglišči trikotnika PAB in izračunamo koeficienta takole:

$$[b_1, b_2] = [f(\mathbf{t_A}), f(\mathbf{t_B})][\mathbf{t_A}, \mathbf{t_B}]^{-1},$$

oziroma

$$[b_1, b_2] = [f(A) - f(P), f(B) - f(P)][A - P, B - P]^{-1}.$$

Posplošimo naš dvorazsežni primer. Število neznanih koeficientov b_i je enako številu oglišč simpleksa. Zaradi predpostavke o linearnosti na simpleksu postavimo $\epsilon = 0$ (glej (2.2)). Z interpolacijo koeficiente izračunamo analitično. Označimo s $\mathbf{t_1}, \ldots, \mathbf{t_n}$ vektorje od točke P do ostalih oglišč izbranega simpleksa, ki leži v smeri x_i. Iščemo $b_1, \ldots, b_n$, ki zadoščajo

$$[b_1 \ldots b_n][\mathbf{t_1} \ldots \mathbf{t_n}] = [f(\mathbf{t_1}) \ldots f(\mathbf{t_n})] \tag{2.3}$$

Koeficiente b_i izračunamo po:

$$[b_1 \ldots b_n] = [f(\mathbf{t_1}) \ldots f(\mathbf{t_n})][\mathbf{t_1} \ldots \mathbf{t_n}]^{-1} \tag{2.4}$$

Zadnja izpeljava zahteva še nekaj pojasnil. Omenili smo 'izbrani simpleks' v smeri x_i. Popolnoma vseeno je ali izberemo simpleks v smeri naraščanja x_i ali v smeri padanja. Lahko se zgodi, da smer x_i ravno sovpade z robom simpleksa oziroma mejo med sosednjima simpleksoma (kar v večrazsežnem prostoru ni isto). V tem primeru spet lahko poljubno izbiramo med simpleksi, ki vsebujejo P in jih vektor $P + k\mathbf{x_i}$ kakorkoli seka (ali notranjost ali mejo).

Regresija na zvezdi

Regresija na zvezdi temelji na podobnih predpostavkah kot metoda prvi trikotnik. Ker upošteva širšo okolico točke P, je bolj odporna na šum v podatkih. Okolico P definiramo z njeno zvezdo. Zvezda točke P je v računski topologiji definirana kot množica simpleksov, ki ji točka P pripada. V našem primeru je zvezda točke P označena na sliki 2.2c. Pri izračunu parcialnega odvoda si bomo torej pomagali s točkami A, B, C, D, E in F.

Z linearno interpolacijo si ne moremo več pomagati, zato dopustimo neničelni ϵ in problem prevedemo na izračun univariatne linearne regresije na točkah v zvezdi P. Podobno kot prej označimo vektorje $\mathbf{t_1},\ldots,\mathbf{t_n}$ od točke P do ostalih točk zvezde in izračunamo parcialni odvod, ki je enak koeficientu b_i:

$$b_i = \frac{\sum_j t_{ji} f(\mathbf{t_j})}{\sum_j t_{ji}^2}, \tag{2.5}$$

kjer t_{ji} označuje i-to komponento vektorja do j-te točke v zvezdi, $\mathbf{t_j}$.

Regresija na zvezdi z upoštevanjem širše okolice kot prvi trikotnik predstavlja prvi korak v smeri robustnosti računanja parcialnih odvodov na vzorčenih funkcijah. Njena slabost ostaja odvisnost od triangulacije. Te se dokončno znebimo pri naslednji metodi, regresiji v cevi.

2.2.2 Regresija v cevi

Po pričakovanju smo na račun upoštevanja večjega števila točk pri računanju regresije povečali robustnost regresije na zvezdi glede na prvi trikotnik. Naslednji cilj, ki za zasledujemo, je dodatno povečanje robustnosti, kamor prištevamo tako izboljšanje pogojev za računanje regresije kot tudi odpravo triangulacije. Le-ta je omejujoča v višjih dimenzijah in v primerih, ko obstajata dva ali več učnih primerov z istimi vrednostmi atributov, a različnimi razredi. Pri večanju okolice točke P se torej ne bomo več ozirali na sosednost, ki jo definirajo simpleksi. Upoštevati moramo tudi osnovno idejo računanja parcialnih odvodov, t.j. opazovanje razredne spremenljivke ob spreminjanju atributa x_i, pri čemer vrednosti ostalih atributov ostajajo čim bližje tistim, ki jih ima P.

Regresija v cevi [Žabkar et al., 2007a] je metoda, ki za okolico točke P vzame (hiper)cev, katere os gre skozi P vzdolž smeri odvajanja x_i (slika 2.2d). Predpostavimo, da je iskana funkcija znotraj cevi približno linearna in izračunajmo odvod s pomočjo univariatne regresije na točkah v cevi. Cev se razteza vzdolž

celotnega definicijskega območja atributa x_i, njeno širino pa določamo s številom sosedov točke P. Cev je torej tako široka, da je v njej želeno število sosedov P, kjer razdaljo merimo v prostoru vseh atributov razen x_i. Ker cev vsebuje tudi točke, ki ležijo daleč stran od P, točke v cevi utežimo glede na njihovo razdaljo od P v smeri x_i. Utež j-te točke v cevi je torej enaka

$$w_j = e^{-t_{ji}^2/\sigma^2}, \tag{2.6}$$

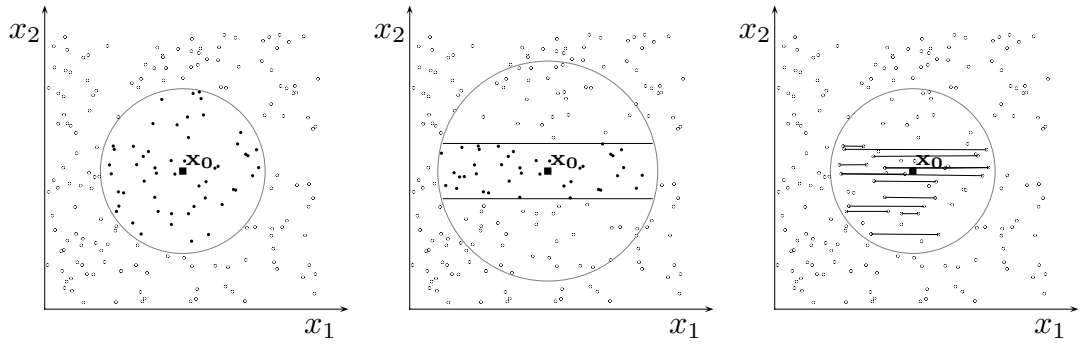
kjer je t_{ji} i-ta komponenta vektorja $\mathbf{t_j}$. Parameter σ določimo tako, da ima najbolj oddaljena točka zanemarljivo majhno utež, npr. $w = 0.001$. Z uteženo univariatno linearno regresijo izračunamo koeficient b_i,

$$b_i = \frac{\sum_j w_j t_{ji} f(\mathbf{t_j})}{\sum_j w_j t_{ji}^2}. \tag{2.7}$$

Regresijo v cevi računamo na večjem vzorcu točk (npr. 30), kar nam omogoča, da s t-testom ocenimo značilnosti izračunanih odvodov. S predznakom b_i in njegovo značilnostjo definiramo kvalitativni odvod: če je značilnost nad določenim pragom (npr. $p \leq 0.7$), je kvalitativni odvod enak predznaku b_i; sicer definiramo kvalitativni odvod $\circ$, ki ni odvisen od b_i.

2.2.3 Multivariatna lokalno utežena regresija

Regresija v cevi nakazuje smer, ki jo bomo zasledovali pri opisu naslednje trojice metod. Ukvarjali se bomo predvsem z vprašanjem, kakšna naj bo okolica točke P, da bo izračunani približek parcialnega odvoda čim boljši, ter kakšna regresija je najbolj primerna. Začeli bomo s hipersferično okolico okrog P (slika 2.3a) in multivariatno lokalno uteženo linearno regresijo (LWR). LWR hkrati izračuna vse parcialne odvode, kar jo najbolj loči od ostalih metod. Nadaljevali bomo z izpeljanko regresije v cevi, ki jo imenujemo τ-regresija. Od regresije v cevi se razlikuje po obliki okolice – neskončno dolgo cev nadomestimo s kratko cevjo, kar bolj ustreza matematični definiciji parcialnega odvoda. Cev skrajšamo tako, da jo odrežemo s hipersfero okrog P, kot kaže slika 2.3b. Odvod izračunamo z uteženo univariatno linearno regresijo na točkah v cevi. Širša kot je cev, bolj krši predpostavko definicije parcialnega odvoda o konstantnih vrednostih vseh spremenljivk razen tiste, po kateri odvajamo. V metodi vzporednih parov široko cev nadomestimo z množico neskončno ozkih cevi – vektorjev v smeri odvajanja (slika 2.3c). Izberemo jih izmed parov točk v okolici P.



(a) (b) (c)

Slika 2.3: Okolice referenčnega primera v metodah LWR (2.3a), τ-regresija (2.3b) in vzporedni pari (2.3c).

Za potrebe računanja, uvedimo nekaj novih oznak. Točko P predstavimo z vektorjem $\mathbf{x_0}$, učni primer zapišemo z $(\mathbf{x}, y)$, kjer je $\mathbf{x} = (x_1, x_2, \ldots x_n)$, y pa naj bo vrednost neznane vzorčene funkcije, $y = f(\mathbf{x})$.

Naj bo $\mathcal{N}(\mathbf{x_0})$ množica primerov $(\mathbf{x_m}, y_m)$ in naj za vsak i velja: $x_{mi} \approx x_{0i}$ (slika 2.3a). Po Taylorjevem izreku lahko vsako n-krat odvedljivo funkcijo v okolici dane točke zapišemo kot polinom, katerega koeficiente določajo odvodi funkcije v dani točki:

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_{n+1}.$$

Za naš namen bo dovolj, če privzamemo, da je funkcija v okolici točke $\mathbf{x_0}$ približno linearna, zato zapišemo:

$$f(\mathbf{x_m}) = f(\mathbf{x_0}) + \nabla f(\mathbf{x_0}) \cdot (\mathbf{x_m} - \mathbf{x_0}) + R_2, \tag{2.8}$$

kjer R_2 označuje ostanek Taylorjeve vrste. Naš cilj je izračunati vektor parcialnih odvodov $\nabla f(\mathbf{x_0})$, z drugimi besedami, parcialne odvode po vseh atributih. Izraz 2.8 zapišemo kot regresijski problem:

$$y_m = \beta_0 + \boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x_m} - \mathbf{x_0}) + \epsilon_m, \quad (\mathbf{x_m}, y_m) \in \mathcal{N}(\mathbf{x_0}). \tag{2.9}$$

Z metodo najmanjših kvadratov želimo poiskati $\boldsymbol{\beta}$ (in prosti člen β_0). Minimizirati želimo vsoto kvadratov napak ϵ_m na množici $\mathcal{N}(\mathbf{x_0})$. Napaka ϵ_m predstavlja tako ostanek Taylorjeve vrste (R_2) kot tudi šum v podatkih.

Na začetku smo privzeli okolico $\mathcal{N}(\mathbf{x_0})$, nismo pa povedali, kako jo določimo. Njena velikost je odvisna od gostote učnih primerov in od stopnje šuma v podatkih. Manj kot je šuma, manjšo okolico si lahko privoščimo. Gostoto učnih

primerov pa vpeljemo kar v definicijo $\mathcal{N}(\mathbf{x_0})$: velikost sferične okolice namesto z radijem (npr. $||\mathbf{x_m} - \mathbf{x_0}|| < \delta$), definiramo s številom k najbližjih sosedov $\mathbf{x_0}$ in jih utežimo glede na njihovo oddaljenost od referenčnega primera $\mathbf{x_0}$,

$$w_m = e^{-||\mathbf{x_m}-\mathbf{x_0}||^2/\sigma^2}. \tag{2.10}$$

Parameter σ^2 določimo tako, da ima najbolj oddaljen primer zanemarljivo majhno utež, npr. 0.001.

Začetni problem smo prevedli na računanje multivariatne lokalno utežene linearne regresije (LWR) [Atkeson et al., 1997]. Koeficienti v rešitvi predstavljajo parcialne odvode v smereh posameznih atributov,

$$\begin{bmatrix} \beta_0 \\ \boldsymbol{\beta} \end{bmatrix} = (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}WY, \tag{2.11}$$

kjer so X, W in Y:

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{k1} & \dots & x_{kn} \end{bmatrix} \qquad W = \begin{bmatrix} w_1 & 0 & \cdots \\ 0 & \ddots & \\ \vdots & & w_k \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix}. \tag{2.12}$$

Izračunani vektor $\boldsymbol{\beta}$ predstavlja približke parcialnih odvodov $\nabla f(\mathbf{x_0})$.

V praksi računanje inverza v (2.11) nadomestimo z računanjem psevdo inverza [Moore, 1920; Penrose, 1955]. S tem se izognemo težavam s singularnimi matrikami in povečamo stabilnost metode.

Definicija okolice v metodi LWR je primerna za multivariatno regresijo oz. hkratno računanje vseh parcialnih odvodov, sicer pa zelo krši predpostavko, da se vrednosti atributov, razen odvajanega, ne spreminjajo (točka 3 na strani 4). Težavo poskušamo odpraviti z drugačno definicijo okolice v naslednjem razdelku.

2.2.4 τ-regresija

τ-regresija [Žabkar et al., 2009] se od LWR bistveno razlikuje v dveh pogledih – je univaritna in ima drugačno obliko okolice referenčne točke. Univariatnost je pravzaprav posledica drugačne definicije okolice. Z novo definicijo okolice želimo združiti prednosti cevaste okolice, ki sicer spoštuje predpostavko 3 (stran 4), a z neskončno dolžino ni primerna, ter sferične okolice, kot smo jo definirali v LWR, z že omenjeno slabo lastnostjo. Kompromis je pravzaprav kar presek

obeh okolic, cevi in krogle. Z drugimi besedami, v sferični okolici, ki jo definira κ najbližjih sosedov $\mathbf{x_0}$, izberemo k ($< \kappa$) tistih sosedov, ki ležijo v cevi skozi $\mathbf{x_0}$ v smeri atributa, po katerem odvajamo (slika 2.3b). Brez izgube za splošnost predpostavimo, da odvajamo po prvem atributu x_1. Najprej izberimo κ najbližjih sosedov z metriko $||\mathbf{x_m} - \mathbf{x_0}||$, nato pa izmed teh k najbližjih sosedov z metriko $||\mathbf{x_m} - \mathbf{x_0}||_{\setminus 1}$, kjer $||\cdot||_{\setminus i}$ pomeni razdaljo po vseh dimenzijah razen i-te. Tako izbrano okolico ponovno označimo z $\mathcal{N}(\mathbf{x_0})$.

Ob ustrezni izbiri parametrov κ in k lahko predpostavimo, da za večino učnih primerov velja, da so od referenčnega veliko bolj oddaljeni v smeri odvajanja, kot v drugih smereh. Za večino primerov $\mathbf{x_m}$ torej velja $|x_{m1} - x_{01}| \gg |x_{mi} - x_{0i}|$ za $i > 1$, kar je v skladu s predpostavko 3. Če predpostavimo še, da za primere v cevi velja, da so parcialni odvodi po drugih atributih približno istega velikostnega razreda, $\partial f/\partial x_1(x_{m1} - x_{01}) \gg \partial f/\partial x_i(x_{mi} - x_{0i})$ za $i > 1$, potem lahko v skalarnem produktu v izrazu (2.8) upoštevamo samo prvo dimenzijo, kar nam da:

$$f(\mathbf{x_m}) = f(\mathbf{x_0}) + \frac{\partial f}{\partial x_1}(\mathbf{x_0})(x_{m1} - x_{01}) + R_2 \tag{2.13}$$

za $(\mathbf{x_m}, y_m) \in \mathcal{N}(\mathbf{x_0})$, oz. v naši notaciji:

$$y_m = \beta_0 + \beta_1(x_{m1} - x_{01}) + \epsilon_m, \tag{2.14}$$

kjer je β_1 približek parcialnega odvoda $\frac{\partial f}{\partial x_1}(\mathbf{x_0})$. Naša naloga je izračunati β_1, ki minimizira napako na okolici $\mathcal{N}(\mathbf{x_0})$.

Učne primere v $\mathcal{N}(\mathbf{x_0})$ utežimo glede na njihovo razdaljo od referenčnega primera $\mathbf{x_0}$, merjeno samo v smeri odvajanja,

$$w_m = e^{-(x_{m1} - x_{01})^2/\sigma^2}, \tag{2.15}$$

kjer σ nastavimo tako, da ima najbolj oddaljen primer zanemarljivo majhno utež 0.001.

Dobljeni linearni model rešimo z uteženo univariatno linearno regresijo na okolici $\mathcal{N}(\mathbf{x_0})$,

$$\frac{\partial f}{\partial x_1}(\mathbf{x_0}) = \beta_1 = \frac{\sum_{\mathbf{x_m} \in \mathcal{N}(\mathbf{x_0})} w_m x_{m1} y_m}{\sum_{\mathbf{x_m} \in \mathcal{N}(\mathbf{x_0})} w_m x_{m1}^2}. \tag{2.16}$$

Ker regresijo računamo na večjem vzorcu primerov, lahko uporabimo t-test za ugotavljanje statistične značilnosti izračunanih odvodov. Skupaj s koeficientom naklona značilnost določi predznak, s katerim definiramo kvalitativno

vrednost odvoda. Če je značilnost pod nastavljenim pragom, je vrednost kvalitativnega odvoda $\circ$, sicer pa jo določa predznak koeficienta naklona.

Predpostavka, da so parcialni odvodi po drugih atributih približno istega velikostnega razreda, ki smo jo naredili pri izpeljavi regresije τ, v realnih podatkih pogosto ne drži. Na točnost računanja parcialnega odvoda po atributu x_i imajo lahko škodljiv vpliv tisti atributi, katerih parcialni odvodi so po absolutni vrednosti znatno večji od njega. Z drugimi besedami, če funkcija bolj strmo narašča ali pada v smeri, različni od x_i, je to pri računanju odvoda po x_i zelo moteče. Vpliv ostalih atributov na računanje odvoda po x_i poskušamo odpraviti v naslednji metodi.

2.2.5 Vzporedni pari

Vplive drugih atributov na računanje odvoda po x_i bi lahko zmanjšali s tanjšo cevjo. Širina cevi v τ-regresiji je definirana s številom primerov k, zato ožanje cevi pomeni manjše število primerov za računanje regresije. Nekaj lahko pridobimo na račun podaljšanja cevi (večji κ), vendar ima to, kot smo že videli, tudi slabe lastnosti.

Z metodo Vzporedni pari poskušamo ugnati omenjeni problem. Namesto ene cevi, ki je očitno ne moremo enostavno zožati, vzamemo več cevi, ki so neskončno ozke, pri čemer pa nobena ni daljša od cevi v τ-regresiji. Dolžino cevi bomo, tako kot doslej, omejili s sferično okolico točke $\mathbf{x_0}$. Neskončno ozke cevi predstavljajo daljice med pari točk v okolici $\mathcal{N}(\mathbf{x_0})$. Da bi zmanjšali vpliv drugih atributov, moramo izbrati take daljice (pare), ki so vzporedne smeri odvajanja (slika 2.3c). Daljice uredimo glede na kot, ki ga oklepajo s smerjo odvajanja in jih ustrezno utežimo ter z linearno regresijo izračunamo odvod v smeri x_i. Opisani postopek nam bo pomagal razumeti naslednjo formalno izpeljavo.

Naj bosta primera $(\mathbf{x_m}, y_m)$ in $(\mathbf{x_j}, y_j)$ v okolici $\mathcal{N}(\mathbf{x_0})$ in naj bo daljica med njima približno vzporedna s smerjo odvajanja x_1, $||\mathbf{x_m} - \mathbf{x_j}|| \approx |x_{m1} - x_{j1}|$. Predpostavljamo lahko, da oba primera pripadata istemu linearnemu modelu (2.14) z istima koeficientoma β_0 in β_1. Vzemimo enačbo (2.14), vstavimo enkrat y_m in nato y_l ter odštejmo, da dobimo:

$$y_m - y_j = \beta_1(x_{m1} - x_{j1}) + (\epsilon_m - \epsilon_j). \tag{2.17}$$

Zapišimo $y_{(m,j)} = y_m - y_j$ in $x_{(m,j)1} = x_{m1} - y_{j1}$, kar nas pripelje do:

$$y_{(m,j)} = \beta_1 x_{(m,j)1} + \epsilon_{(m,j)}, \tag{2.18}$$

Razlika $y_m - y_j$ in razlika med vrednostmi atributov x_{m1} in x_{j1} sta povezani linearno. Koeficient β_1 je približek parcialnega odvoda $\frac{\partial f}{\partial x_1}(\mathbf{x_0})$.

Za izračun parcialnega odvoda po formuli (2.17) vzemimo sferično okolico točke $\mathbf{x_0}$, kot pri metodi LWR. Z uporabo skalarnega produkta za vsak par izračunamo njegovo ujemanje s smerjo x_1:

$$\alpha_{(m,j)} = \left| \frac{(\mathbf{x_m} - \mathbf{x_j})^{\mathrm{T}} \mathbf{e_1}}{||\mathbf{x_m} - \mathbf{x_j}||\ ||\mathbf{e_1}||} \right| = \frac{|x_{m1} - x_{j1}|}{||\mathbf{x_m} - \mathbf{x_j}||} \tag{2.19}$$

kjer je $\mathbf{e_1}$ enotski vektor v smeri x_1. Izmed κ najbližjih sosedov izberemo k najbolj poravnanih parov v okolici $\mathbf{x_0}$ (slika 2.3c) in jih utežimo glede na ujemanje s smerjo x_1:

$$w_{(m,j)} = e^{-(1-\alpha_{(m,j)})^2/\sigma^2}, \tag{2.20}$$

kjer je σ^2 tak, da je utež najslabše poravnanega para enaka 0.001.

Z univariatno linearno regresijo izračunamo parcialni odvod:

$$\frac{\partial f}{\partial x_1}(\mathbf{x_0}) = \beta_1 = \frac{\sum_{(\mathbf{x_m},\mathbf{x_j}) \in \mathcal{N}(\mathbf{x_0})} w_{(m,j)} (x_{m1} - x_{j1})(y_m - y_j)}{\sum_{(\mathbf{x_m},\mathbf{x_j}) \in \mathcal{N}(\mathbf{x_0})} w_{(m,j)} (x_{m1} - x_{j1})^2}. \tag{2.21}$$

2.2.6 Statistična značilnost KPO

Metode regresija v cevi, LWR, τ-regresija in vzporedni pari temeljijo na računanju regresije na večjem vzorcu točk. To nam omogoča, da s t-testom ocenimo značilnosti izračunanih odvodov. S predznakom b_i in njegovo značilnostjo definiramo kvalitativni odvod: če je značilnost nad določenim pragom (npr. $p \leq 0.7$), je kvalitativni odvod enak predznaku b_i. Sicer definiramo kvalitativni odvod $\circ$, ki ni odvisen od b_i.

2.2.7 Časovna zahtevnost

Naj bo N število učnih primerov (točk) in n število atributov (spremenljivk, dimenzij).

Metodi prvi trikotnik in regresija na zvezdi temeljita na Delaunayevi triangulaciji. Izračun časovne zahtevnosti Delaunyeve triangulacije je zelo zapleten saj je v marsičem odvisen od narave podatkov. Zato bomo podali le grobo oceno

uporabljene implementacije iz knjižnice qhull, ki je $O(2^n N \log N)$. Podrobna izpeljava in posebni primeri so navedeni v [Barber et al., 1996].

Za vsak učni primer mora metoda prvi trikotnik poiskati ustrezni trikotnik, kar zahteva izračun determinante n-razsežne matrike za vsak trikotnik v zvezdi učnega primera. Časovna zahtevnost je $O(Nn^3t)$, kjer je t največje število trikotnikov v zvezdi. Vrednost t je težko oceniti, v grobem pa lahko rečemo, da raste eksponentno s številom dimenzij, torej $O(Nn^3 2^n)$. Časovna zahtevnost skupaj s triangulacijo je $O(N2^n(\log N + n^3))$.

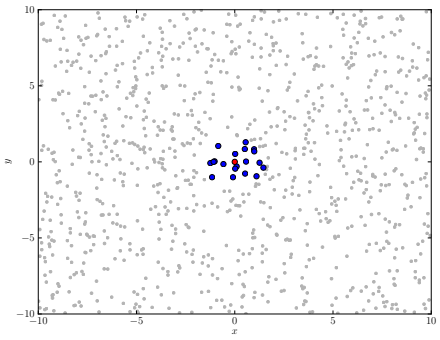
Regresija na zvezdi za vsak učni primer računa univariatno regresijo in ima časovno kompleksnost $O(Ns)$, kjer je s število primerov v zvezdi. Le-to s številom dimenzij eksponentno raste, zato je časovna zahtevnost regresije na zvezdi teoretično $O(N2^n)$, skupaj s triangulacijo pa $O(2^n N \log N)$.

Regresija v cevi, LWR, τ-regresija in Vzporedni pari imajo enako časovno zahtevnost – najprej poiščejo k najbližjih sosedov za vsak učni primer, nato pa na primerih v okolici izračunajo linearno regresijo. Časovna zahtevnost iskanja iskanja sosedov je $O(nN^2)$, linearna regresija prispeva $O(k)$, skupaj torej $O(nN^2 + k) = O(nN^2)$.

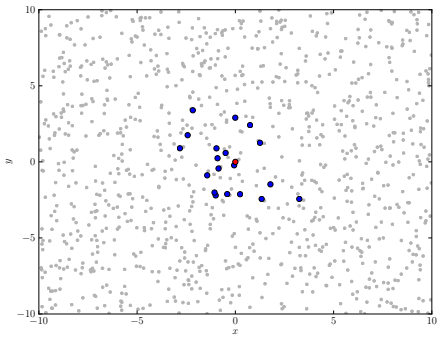
2.2.8 Sosednost v večrazsežnih prostorih

Problem najbližjih sosedov v visoko razsežnih prostorih se odraža v dejstvu, da se z naraščanjem razsežnosti razdalja do najbližjega soseda približuje razdalji do najbolj oddaljenega soseda [Beyer et al., 1999; Dasgupta, 2010]. Že v prisotnosti majhnega šuma v podatkih zato postane metoda najbližjih sosedov v višjih dimenzijah nezanesljiva. V tem razdelku bomo na praktičnih primerih orisali probleme, s katerimi se pri izbiri okolice referenčnega primera srečuje Padé.

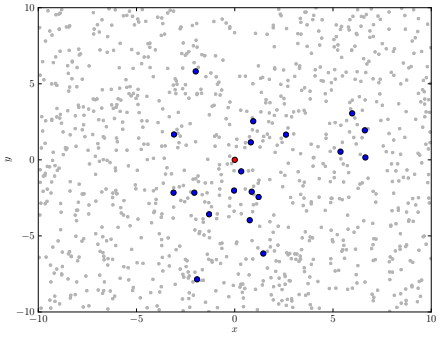
Vzemimo funkcijo $f(x, y) = x^2 - y^2$, vzorčeno v 1000 točkah na intervalu $[-10, 10] \times [-10, 10]$. Podatkom nato postopoma dodajamo nove naključne atribute. Izberimo točko $(0, 0)$ in opazujmo, kaj se dogaja z njenimi najbližjimi sosedi v projekciji na ravnino, ki jo napenjata edina pomembna atributa, x in y. Vzemimo $k = 20$ najbližjih sosedov, vrednost, ki se kasneje v poskusih izkaže za primerno standardno nastavitev Padéjevih metod. Slika 2.4 prikazuje okolice v različnih dimenzijah. V dveh dimenzijah so točke iz okolice po pričakovanju strnjene okrog referenčne točke, z višanjem dimenzije pa opazimo, da se razletijo na vse strani. Prav te točke pa denimo τ-regresija uporabi pri računanju odvoda f_x. Vidimo, da je okolica že v 5 dimenzijah vse prej kot primerna za to.



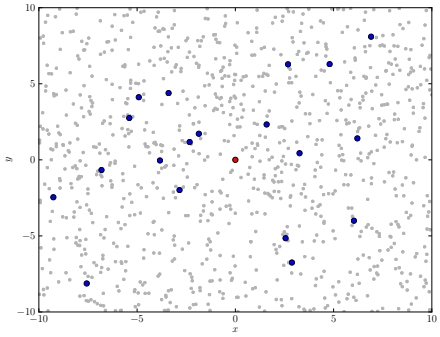
(a) Dve dimenziji.



(b) Projekcija 3-razsežnega prostora v ravnino x y.



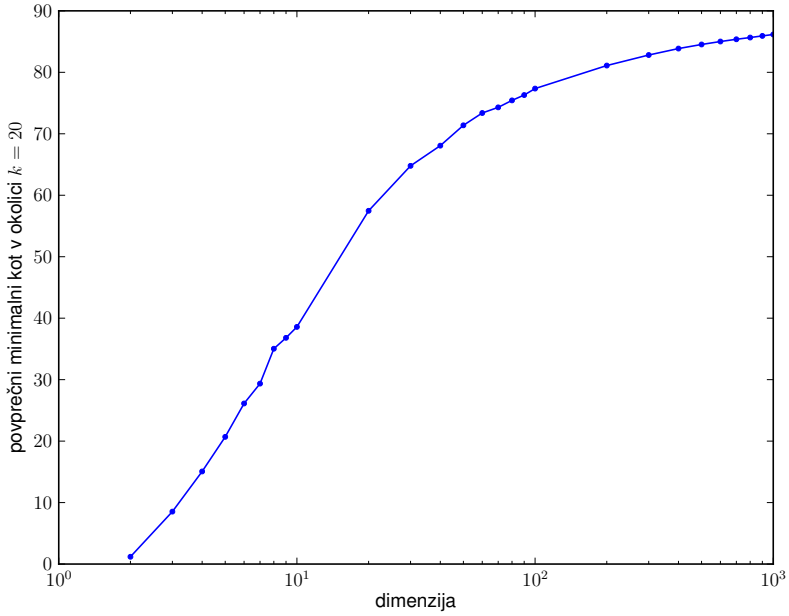
(c) Projekcija 5-razsežnega prostora v ravnino x y.



(d) Projekcija 10-razsežnega prostora v ravnino x y.

Slika 2.4: Najbližji sosedje ($k = 20$) točke $(0, 0)$ v različno razsežnih prostorih. Prikazane so projekcije večrazsežnih prostorov v ravnino x y.

Podoben problem srečamo pri vzporednih parih, kjer so pomembni koti med smerjo odvajanja in vektorji, ki jih definirajo pari. Ponovno opazujmo točko $(0, 0)$ z njeno okolico velikosti $k = 20$. Izberimo smer x, glede na katero računamo kote, ki jih ta smer oklepa s pari v okolici. Za vsakega od 100 naključnih vzorcev podatkov, v katerih se točka $(0, 0)$ vedno pojavi, izračunamo minimalni kot v njeni okolici, nato pa povprečimo čez vsa vzorčenja. Tako dobimo povprečni minimalni kot v okolici točke $(0, 0)$, torej kot med najboljšim parom in smerjo x. Poskus ponovimo pri različnem številu dimenzij, z dodajanjem naključnih atributov originalni domeni. Graf prikazuje naraščanje minimalnega kota z naraščanjem dimenzije prostora atributov.



Slika 2.5: Povprečni minimalni kot v okolici točke $(0, 0)$ prek 100 naključnih vzorčenj. Okolico velikosti $(k = 20)$ opazujemo v različno razsežnih prostorih.

2.3 Učenje kvalitativnih modelov v regresiji

Z izračunom kvalitativnih parcialnih odvodov za vse učne primere smo pripravili podatke tako, da je postopek učenja kvalitativnih modelov iz njih trivialen. Kvalitativnega modela se lahko naučimo na več načinov. Kvalitativni odvod regresijskega razreda po izbranem atributu uporabimo kot razred v klasifikacijskem problemu učenja kvalitativnega modela.

Za primer vzemimo, da odvajamo y po prvem atributu, x_1. Tabela 2.2a prikazuje originalno učno domeno z regresijskim razredom y in atributi $x_1, \ldots, x_n$. Tej domeni Padé doda dva nova atributa in sicer kvalitativni odvod $\partial_Q y/\partial_Q x_1$ in njegovo značilnost $\mathrm{sig}(\partial_Q y/\partial_Q x_1)$. Značilnosti pri učenju ne upoštevamo neposredno, zato je v novi domeni (tabela 2.2b) ta atribut označen kot meta atribut. Prav tako pri gradnji kvalitativnega modela v domeni ne potrebujemo več originalnega razreda y. Nadomestimo ga s kvalitativnim odvodom. Ker je novi razred diskretnega tipa (njegove vrednosti so $+, -, \circ$), je učni problem klasifikacijski. Za reševanje tega problema lahko uporabimo poljuben ustrezen algoritem

strojnega učenja, npr. klasifikacijska drevesa, klasifikacijska pravila, naivni Bayesov klasifikator, ipd. Rezultat učenja je kvalitativni model v izbranem jeziku hipotez.

x_1	x_2	...	x_n	y
c	c	...	c	c
				class
⋮	⋮	...	⋮	⋮

(a) Originalna domena v formatu Orange.

x_1	x_2	...	x_n	$\partial_Q y/\partial_Q x_1$	$\text{sig}(\partial_Q y/\partial_Q x_1)$
c	c	...	c	d	c
				class	*meta*
⋮	⋮	...	⋮	⋮	⋮

(b) Nova domena z atributoma, ki ju doda Padé.

Tabela 2.2: Originalna domena in domena po zagonu Padéja.

Pri učenju kvalitativnega modela si lahko pomagamo z dodatno informacijo o izračunanih kvalitativnih odvodih. Značilnost odvoda, atribut $\text{sig}(\partial_Q y/\partial_Q x_1)$, nam pove, kako zanesljiv je odvod v vsaki točki (učnem primeru). Pri učenju modela je smiselno uporabiti samo tiste učne primere, v katerih značilnost izračunanega odvoda preseže nek prag, npr. 0.8. Meta atributa $\text{sig}(\partial_Q y/\partial_Q x_1)$ torej pri učenju ne uporabljamo neposredno, z njim si pomagamo samo pri izbiri dobre podmnožice učnih primerov. Tipičen primer, ko je to koristno, je gradnja kvalitativnega modela za $\partial_Q T/\partial_Q \varphi$ pri matematičnem nihalu (glej razdelek 2.1.1). Za učne primere, ki imajo vrednost φ blizu 0, je značilnost odvoda majhna – že majhna motnja spremeni predznak. Taki primeri niso koristni za učenje, saj gre dejansko za šum.

Včasih želimo zgraditi kvalitativni model, ki prostor atributov deli upoštevajoč več parcialnih odvodov po različnih atributih hkrati. V tem primeru enostavno združimo več atributov v enega tako, da za vsak učni primer staknemo njihove vrednosti, da dobimo niz vrednosti kvalitativnih odvodov, npr. "$++--$" v primeru stika štirih kvalitativnih odvodov. Primer takega modela je kvalitativno drevo v domeni XPERO (glej razdelek 2.6). Možno je tudi večciljno (*angl. multi target*) učenje [Ženko and Džeroski, 2008].

Večina kvalitativnih modelov v tem delu je v obliki kvalitativnih dreves, pridobljenih z algoritmom C4.5.

2.4 Poskusi in rezultati

Metode, ki smo jih opisali v poglavju 2.2, smo implementirali v okolju za strojno učenje Orange [Zupan et al., 2004]. V tem poglavju bomo opisali eksperimentalno delo, pri katerem smo uporabili omenjene implementacije. Pri poskusih se osredotočimo na ocenjevanje točnosti in stabilnosti Padéja, vpliv nepomembnih atributov ter vpliv šuma v podatkih.

Točnost ocenimo na umetnih podatkih, vzorčenih matematičnih funkcijah, za katere poznamo analitične parcialne odvode, s katerimi tudi primerjamo rezultate Padéja. Primerjamo tako izračune za posamezne učne primere kot tudi napovedi kvalitativnega modela, naučenega na podatkih, ki jih izračuna Padé. Točnost definiramo kot odstotek pravilno napovedanih kvalitativnih parcialnih odvodov. Poudariti moramo, da ocenjevanje točnosti Padéja ne zahteva delitve na učne in testne podatke, saj pravi razred, s katerim primerjamo napoved, ni vključen v proces učenja.

Ocenjevanje točnosti na realnih podatkih je praktično nemogoče, saj za razpoložljive javno dostopne nabore podatkov ne poznamo pravih vrednosti odvodov razreda po atributih. V podobnih primerih (npr. na področju razvrščanja v skupine) v strojnem učenju kot mero za ocenjevanje kvalitete pogosto uporabljamo stabilnost [Ben-David et al., 2006]. Stabilnost meri vpliv spremembe vhoda na izhod sistema. Pomaga nam načrtovati robustne metode, ki jih šum v podatkih in naključnost vzorčenja ne prizadaneta.

Opisani poskusi imajo nekaj skupnih lastnosti, ki jih na tem mestu povzemamo. V poskusih z metodo vzporednih parov vedno uporabljamo nastavitev $\kappa = k$ (npr. če vzamemo $\kappa = 10$, dobimo 100 parov, a izberemo samo $k = 10$ najboljših, ostale pa zavržemo). Po prvem sklopu, kjer ocenjujemo točnost, opazimo, da so metode relativno neobčutljive na nastavitve parametrov, zato odtlej v nadaljnih poskusih, če ni posebej rečeno drugače, uporabljamo naslednje nastavitve: za LWR in vzporedne pare $k = 20$ ter $\kappa = 100$ in $k = 20$ za τ-regresijo. Za ocenjevanje stabilnosti na realnih podatkih uporabljamo $\kappa = 50$, ker podatki vsebujejo veliko manjše število učnih primerov kot umetne domene. Rezultati večine poskusov so povprečja 10 naključnih vzorcev podatkov, kjer poskus poteka drugače, pa na to opozorimo.

Pri poskusih z metodo LWR v izogib slabo pogojenim sistemom pri računanju inverza (2.11) uporabljamo psevdo inverz [Moore, 1920; Penrose, 1955]. Za gradnjo kvalitativnih modelov iz kvalitativnih parcialnih odvodov uporabljamo

klasifikacijska drevesa C4.5 iz programa Orange [Zupan et al., 2004]. Za klasifikacijska drevesa smo se odločili zaradi razumljivosti zgrajenih modelov, čeprav se zavedamo, da bi utegnili z naprednejšimi metodami, npr. SVM, doseči večje točnosti. Poleg tega je večina umetnih podatkov taka, da ne zahtevajo dreves z velikim številom listov. Algoritem C4.5 smo poganjali s privzetimi parametri, razen kjer je posebej rečeno drugače.

2.5 Točnost na umetnih podatkih

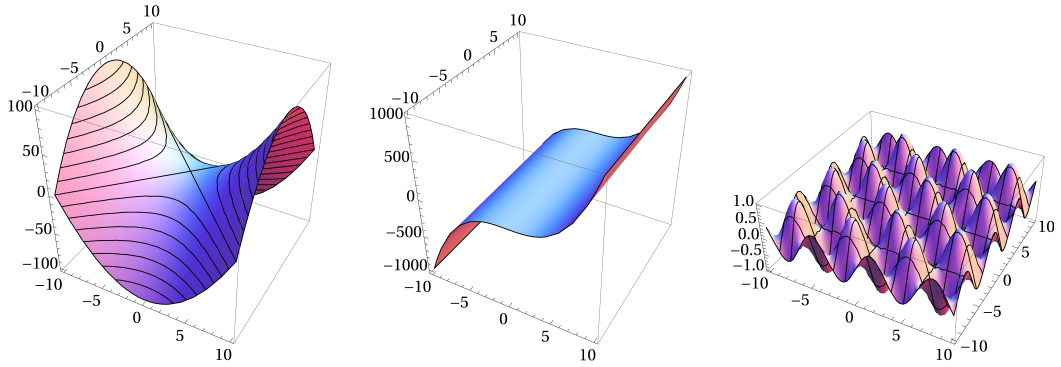
Umetne podatke pripravimo z naključnim enakomernim vzorčenjem treh funkcij: $f(x,y) = x^2 - y^2$, $f(x,y) = x^3 - y$, $f(x,y) = \sin x \sin y$. Vzorčimo jih v 1000 točkah na intervalu $[-10,10] \times [-10,10]$. Parcialne odvode po obeh spremenljivkah, s katerimi primerjamo rezultate Padéja, prikazuje tabela 2.3. Izbrane funkcije predstavljajo družine funkcij, ki imajo podobne lastnosti. Morda se na prvi pogled zdi, da so daleč od funkcij, ki jih srečamo v realnem svetu, toda kubični polinomi več spremenljivk so ravno razvoji poljubne odvedljive funkcije v Taylorjevo vrsto in v praksi najpogosteje tudi gradniki zlepkov.

$f(x,y)$	$\partial f/\partial x$	$\partial f/\partial x$
$x^2 - y^2$	$2x$	$-2y$
$x^3 - y$	$3x^2$	-1
$\sin x \sin y$	$\cos x \sin y$	$\sin x \cos y$

Tabela 2.3: Umetne funkcije in njihovi parcialni odvodi.

2.5.1 Funkcija $f(x,y) = x^2 - y^2$

Funkcija $f(x,y) = x^2 - y^2$ je standardna testna funkcija za ocenjevanje kvalitativnih modelov [Šuc, 2003]. Iz njenih parcialnih odvodov je očitno, da za $x > 0$ velja $f = Q(+x)$, za $x < 0$ pa $f = Q(-x)$, in podobno za y. Zaradi podobnosti odvodov po x in y v nadaljevanju opazujemo samo parcialne odvode po x. Točnost ocenjujemo za različne nabore parametrov κ in k. Rezultate prikazuje tabela 2.4. Različne vrednosti parametrov nimajo večjega vpliva na rezultat, točnost je vedno blizu 100%, razen pri τ-regresiji, kjer opazimo, da z večjim k (širjenje cevi) točnost pada. Prav tako pri fiksnem k točnost pada z večanjem κ



(a) $f(x,y) = x^2 - y^2$ (b) $f(x,y) = x^3 - y$ (c) $f(x,y) = \sin x \sin y$

Slika 2.6: Grafi vzorčnih funkcij.

(daljšanje cevi). Pri $\kappa = 1000$ (število vseh učnih primerov) τ-regresija preide v regresijo v cevi. Ker f_x pri $x = 0$ spremeni predznak, so dolge cevi slabe, saj vsebujejo učne primere na celotnem definicijskem območju x. Kot bomo videli pri naslednji funkciji, ki je globalno monotona, so dolge cevi tam koristne. Kako ključna je odvisnost τ-regresije od nastavitev parametrov? Kvalitativno drevo, kot vidimo v tabeli 2.4b, praktično izniči vpliv različnih nastavitev.

			τ-regresija							
	LWR	Pari	$\kappa = 30$	50	70	100	200	300	500	1000
$k = 10$	.991	.986	.948	.968	.971	.980	.982	.978	.974	.970
20	.993	.992	.929	.960	.972	.981	.986	.986	.984	.982
30	.992	.992	.909	.953	.969	.978	.986	.988	.987	.987
40	.993	.993		.950	.964	.974	.987	.989	.988	.988
50	.994	.993		.935	.961	.972	.986	.989	.988	.989

(a) Točnosti kvalitativnih odvodov.

			τ-regresija							
	LWR	Pari	$\kappa = 30$	50	70	100	200	300	500	1000
$k = 10$	.997	.993	.990	.993	.990	.991	.992	.993	.993	.993
20	.996	.995	.983	.991	.986	.991	.990	.993	.993	.992
30	.996	.994	.983	.987	.988	.993	.993	.993	.990	.991
40	.995	.995		.978	.978	.990	.994	.995	.994	.993
50	.998	.995		.977	.987	.991	.991	.992	.993	.991

(b) Točnosti kvalitativnih dreves, zgrajenih s C4.5.

Tabela 2.4: Rezultati poskusov za odvod $f(x,y) = x^2 - y^2$ po x.

2.5.2 Funkcija $f(x, y) = x^3 - y$

Funkcija $f(x, y) = x^3 - y$ je globalno monotona, naraščajoča v smeri x in padajoča v smeri y. Zanimiva je zaradi prevladujočega vpliva spremenljivke x nad y na vrednost funkcije. Z drugimi besedami, že majhna sprememba x-a močno poveča f, medtem ko relativno velika sprememba y-a še vedno ne zadošča, da bi pretehtala vpliv x-a. Kot primer vzemimo vektor od točke $(2, 0)$ do točke $(2.1, 1)$. Če vektor razstavimo na komponenti v smereh x in y ter opazujemo spremembo f po vsaki komponenti, vidimo, da se f v smeri x poveča z 8 na 9.261, v smeri y pa, čeprav naredimo 10-krat daljši korak, f pade samo za 1. Med točkama $(2, 0)$ in $(2.1, 1)$ funkcija f torej raste. Izbrani točki sta po x tako blizu, da ju pri odvajanju po y gotovo srečamo skupaj v cevi, kar vodi v napačen izračun odvoda.

Odvod po x ne dela težav nobeni metodi, točnost vseh metod je 100%. Odvod po y se izkaže za precej trši oreh. Točnost τ-regresije narašča z daljšanjem cevi in za njeno fiksno dolžino pada z večanjem širine cevi (Tabela 2.5). To potrjuje dejstvo, ki smo ga poprej opisali s primerom. Točnost LWR je prav tako nizka, okrog 70%. Le vzporedni pari nimajo večjih težav, njihova točnost pa s povečevanjem k raste.

Ponovno opazimo, da kvalitativna drevesa močno povečajo točnosti, ki jih metode dosegajo pri ocenjevanju odvodov 'po točkah' – pri LWR kvalitativni modeli dosegajo točnosti nad 95% kljub nizki točnosti 'po točkah'. To razložimo s tem, da so napake dovolj naključno porazdeljene po celem prostoru in jih algoritem C4.5 pri posploševanju odstrani kot šum.

2.5.3 Funkcija $f(x, y) = \sin x \sin y$

Parcialna odvoda funkcije $f(x, y) = \sin x \sin y$ na definicijskem območju večkrat spremenita predznak. Točnost večine metod je med 80 in 90% in se manjša z večanjem okolice (Tabela 2.6). Kljub temu pa točnost C4.5 komaj preseže 50%, kar je enako naključnemu napovedovanju. To pripisujemo nezmožnosti algoritma C4.5, da bi odkril koncept 'šahovnice'. S Padéjem izračunane parcialne odvode lahko prikažemo grafično (slika 2.7).

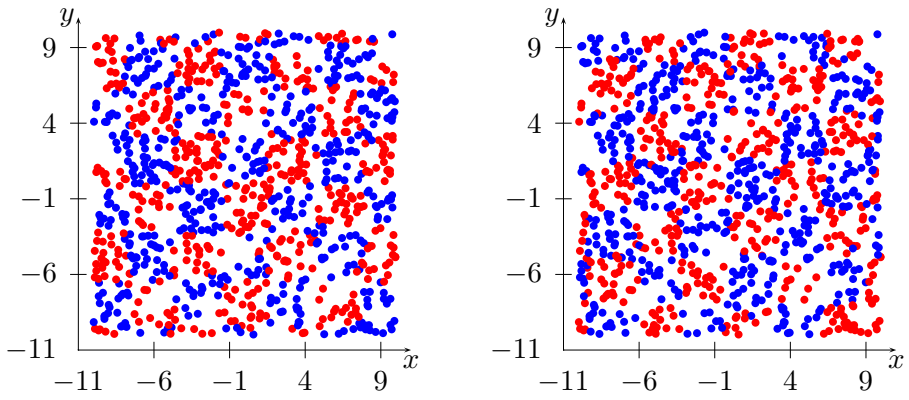
			τ-regresija							
	LWR	Pari	$\kappa = 30$	50	70	100	200	300	500	1000
$k = 10$	.727	.690	.597	.626	.639	.647	.712	.749	.834	.883
20	.725	.792	.576	.593	.619	.646	.676	.725	.795	.856
30	.751	.879	.545	.571	.609	.618	.686	.711	.774	.838
40	.740	.919		.556	.588	.613	.656	.700	.761	.821
50	.725	.966		.541	.558	.612	.656	.701	.744	.811

(a) Točnosti kvalitativnih odvodov.

			τ-regresija							
	LWR	Pari	$\kappa = 30$	50	70	100	200	300	500	1000
$k = 10$	1.00	.898	.737	.880	.890	.864	.970	.948	.967	.993
20	1.00	.930	.745	.743	.811	.860	.889	.968	.968	.978
30	.978	.954	.752	.599	.813	.777	.923	.868	.938	.964
40	.971	.964		.780	.732	.700	.827	.916	.938	.965
50	.956	.986		.906	.805	.760	.795	.905	.925	.917

(b) Točnosti kvalitativnih dreves, zgrajenih s C4.5.

Tabela 2.5: Rezultati poskusov za odvod $f(x, y) = x^3 - y$ po y.



(a) $\partial_Q f/\partial_Q x$ (b) $\partial_Q f/\partial_Q y$

Slika 2.7: Kvalitativna parcialna odvoda funkcije $f(x, y) = \sin x \sin y$ v razsevnem diagramu.

2.5.4 Primeri iz fizike

Področje kvalitativnega modeliranja se že od svojih začetkov srečuje s primeri iz fizike, zato tudi Padéja preizkusimo na treh fizikalnih domenah. Vsak fizikalni problem predstavlja enačba, ki jo uporabimo za vzorčenje, nato pa poskušamo iz podatkov rekonstruirati kvalitativne odvisnosti med odvisno in neodvisnimi

			τ-regresija							
	LWR	Pari	$\kappa = 30$	50	70	100	200	300	500	1000
$k = 10$	.882	.858	.863	.885	.890	.886	.815	.754	.629	.540
20	.870	.841	.820	.865	.880	.882	.830	.764	.640	.558
30	.862	.823	.769	.837	.861	.877	.837	.770	.656	.548
40	.844	.796		.799	.839	.853	.820	.758	.663	.548
50	.814	.754		.717	.810	.845	.812	.761	.662	.548

(a) Točnosti kvalitativnih odvodov.

			τ-regresija							
	LWR	Pari	$\kappa = 30$	50	70	100	200	300	500	1000
$k = 10$	.509	.519	.516	.512	.510	.509	.528	.524	.504	.502
20	.515	.531	.509	.518	.522	.510	.520	.547	.548	.504
30	.515	.523	.510	.513	.514	.515	.528	.545	.558	.499
40	.521	.555		.511	.507	.511	.530	.567	.570	.503
50	.516	.583		.507	.508	.515	.558	.628	.594	.520

(b) Točnosti kvalitativnih dreves, zgrajenih s C4.5.

Tabela 2.6: Rezultati poskusov za odvod $f(x, y) = \sin x \sin y$ po x.

spremenljivkami. Vsako enačbo vzorčimo v 1000 točkah.

Centripetalna sila. Vzemimo točkasto maso m, ki kroži s hitrostjo v po krožnici z radijem r. Centripetalna sila F, ki deluje na maso m, je

$$F = \frac{mv^2}{r}.$$

Definicijska območja spremenljivk v našem poskusu so: $m \in [0.1, 1]$, $r \in [0.1, 2]$ in $v \in [1, 10]$.

Najslabše se odreže τ-regresija, ki pravilno izračuna od 86% do 97% odvodov, ostali dve metodi pa sta popolnoma točni (Tabela 2.7a). Algoritem C4.5 v vseh primerih zgradi pravilne modele: $F = Q(+m)$, $F = Q(+v)$ in $F = Q(-r)$.

Gravitacija. Splošni gravitacijski zakon,

$$F = G\frac{m_1 m_2}{r^2},$$

pojasnjuje, da gravitacijska sila pojema z razdaljo ter da narašča z maso telesa (pri večji masi telesa je večja tudi njegova gravitacijska sila). V zgornji enačbi je F gravitacijska sila, G gravitacijska konstanta, m_1 in m_2 točkasti masi in r

spremenljivka	m	v	r
LWR	1.00	1.00	1.00
τ-regresija	.86	.97	.94
vzporedni pari	1.00	1.00	.98

(a) Centripetalna sila

spremenljivka	m_1	m_2	r
LWR	.964	.96	1.00
τ-regresija	.88	.87	.99
vzporedni pari	.96	.96	1.00

(b) Gravitacijska sila

spremenljivka	l	ϕ
LWR	1.00	.98
τ-regresija	1.00	.83
vzporedni pari	1.00	.97

(c) Nihalo

Tabela 2.7: Točnost izračunanih parcialnih odvodov za fizikalne primere.

razdalja med njima. Definicijska območja spremenljivk v našem primeru so: m_1 in $m_2 \in [0.1, 1]$ ter $r \in [0.1, 2]$.

Glede točnosti opazimo isto kot v primeru centripetalne sile (Tabela 2.7b), C4.5 pa zgradi naslednje pravilne modele: $F = Q(+m_1)$, $F = Q(+m_2)$ in $F = Q(-r)$.

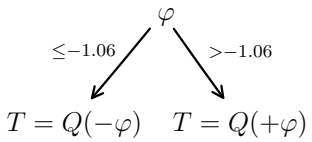
Nihalo. Nihajni čas T prvega nihaja matematičnega nihala je odvisen od dolžine vrvi l, mase uteži m in začetnega odmika od ravnovesne lege oz. začetnega kota φ[†]:

$$T = 2\pi\sqrt{\frac{l}{g}}(1 + \frac{1}{16}\varphi_0^2 + \frac{11}{3072}\varphi_0^4 + \frac{173}{737280}\varphi_0^6 + \ldots).$$

Maso uteži m smo naključno izbirali z intervala $[0.1, 1]$, dolžino l z intervala $[0.1, 2]$, kot φ pa z intervala $[-180°, 180°]$. Težni pospešek je $g = 9.81$.

Tabela 2.7c prikazuje točnost odvodov sile po dolžini in kotu. Kvalitativni model za parcialni odvod po l je $T = Q(+l)$, model za odvod nihajnega časa po kotu pa je prikazan na sliki 2.8. Odvodi zanj so izračunani z metodo LWR. Modela iz podatkov, izračunanih s τ-regresijo in vzporednimi pari sta zelo po-

[†]http://en.wikipedia.org/wiki/Pendulum



Slika 2.8: Kvalitativni model, ki opisuje odvisnost med T in φ.

dobna – edina razlika je v vrednosti, pri kateri drevo veji v korenu: $-2.7°$ za τ-regresijo in $-2.28°$ za vzporedne pare. Razlaga drevesa je preprosta: nihajni čas T pada z naraščajočim kotom, če je le-ta negativen in narašča s kotom, ko je ta pozitiven. Oba lista lahko povzamemo z enim pravilom, namreč: nihajni čas narašča z absolutno vrednostjo kota, $|\varphi|$.

2.5.5 Vpliv nepomembnih atributov

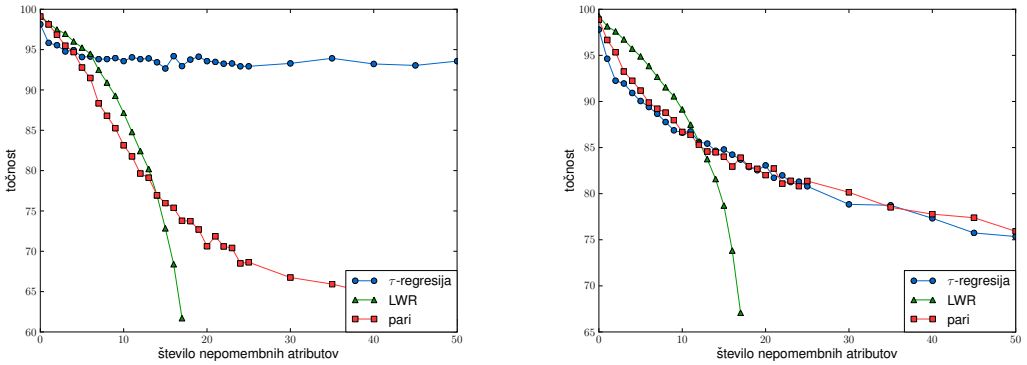
V mnogih realnih problemih, ki jih rešujemo s strojnim učenjem, se poleg nekaj pomembnih atributov pojavljajo tudi nepomembni atributi – taki, ki na iskani koncept ne vplivajo, v podatkih pa so prisotni zato, ker tega ne vemo vnaprej. Ena izmed nalog iskanja zakonitosti v podatkih je tudi odkrivanje pomembnih atributov. Pričujoči poskus ocenjuje robustnost Padéja v takih primerih.

Poskus izvedemo na funkcijah $f(x, y) = x^2 - y^2$ in $f(x, y) = xy$, točnost pa merimo kot v prejšnjem razdelku. Opazujemo pravilnost parcialnih odvodov, izračunanih s Padéjem, medtem ko domeni dodajamo med 0 in 50 naključnih atributov z istim definicijskim območjem kot x in y, $[-10, 10]$. Graf na sliki 2.9 prikazuje točnost glede na število dodanih naključnih atributov.

Točnost LWR strmo pada, kar pripisujemo dejstvu, da LWR računa multivariatno regresijo, pri čemer rešuje linearni sistem, ki mu občutljivost raste, ko se število naključnih atributov približuje k. Ko število vseh atributov preseže k, računanje regresije ni več možno oz. nima nobene podlage v podatkih, zato poskuse z metodo LWR tam prekinemo. Vzporedni pari in τ-regresija se temu problemu izogneta z računanjem univariatne regresije.

2.5.6 Odpornost na šum v podatkih

V tem poskusu ocenjujemo odpornost Padéja na šum v podatkih. Podatkom nadzorovano dodajamo šum in opazujemo točnost. Ciljna funkcija je spet



(a) Domena $x^2 - y^2$.

(b) Domena xy.

Slika 2.9: Točnost Padéja glede na število dodanih naključnih atributov.

	$\sigma = 0$	$\sigma = 10$	$\sigma = 30$	$\sigma = 50$
LWR	.993	.962	.878	.795
τ-regresija	.981	.945	.848	.760
vzporedni pari	.992	.924	.771	.680

(a) Točnost izračunanih odvodov.

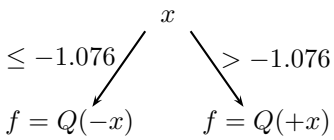
	$\sigma = 0$	$\sigma = 10$	$\sigma = 30$	$\sigma = 50$
LWR	.996	.978	.956	.922
τ-regresija	.991	.976	.949	.917
vzporedni pari	.995	.966	.949	.901

(b) Točnost kvalitativnih dreves, zgrajenih z algoritmom C4.5.

Tabela 2.8: Vpliv šuma na točnost izračunanih parcialnih odvodov funkcije $f(x, y) = x^2 - y^2$ po x z dodanim šumom $N(0, \sigma)$.

$f(x, y) = x^2 - y^2$ definirana na $[-10, 10] \times [-10, 10]$, kar pomeni, da funkcija f zavzema vrednosti na intervalu $[-100, 100]$. Dodajamo Gaussov šum s povprečno vrednostjo 0 in standardnim odklonom 0 (brez šuma), 10, 30, in 50 (šum primerljiv s signalom samim). Pri gradnji kvalitativnih dreves s C4.5 nastavimo parameter m (najmanjše število primerov v listu) na 10% števila vseh učnih primerov v podatkih, $m = 100$. Za vsak σ smo poskus ponovili 100-krat na različnih naključnih vzorcih ter izračunali povprečno točnost.

Rezultati poskusa so zbrani v tabeli 2.8. Padé se je izkazal za robustnega tudi



Slika 2.10: Primer kvalitativnega drevesa za $\partial_Q f/\partial_Q x$ pri najvišji stopnji šuma, $\sigma = 50$.

pri veliki stopnji šuma. Tako kot pri ostalih poskusih opazimo, da kvalitativni modeli znatno izboljšajo točnost navkljub nizki točnosti 'po točkah' pri večji stopnji šuma. Primer kvalitativnega drevesa, zgrajenega z algoritmom C4.5 na podatkih z največjo stopnjo šuma, $\sigma = 50$, je prikazan na sliki 2.10.

2.5.7 Izbira podmnožice koristnih atributov s Padéjem

Poskus, opisan v razdelku 2.5.5 je pokazal, da zna Padé kljub prisotnosti velikega števila naključnih atributov relativno dobro napovedati pravilen kvalitativni odvod razreda po atributu, za katerega vemo, da ni naključen. Ob tem se nam zastavlja vprašanje ali lahko Padé sam ugotovi, kateri atributi vplivajo na razred in kateri ne. Ideja za uporabo Padéja za izbiro koristnih atributov izhaja iz matematične definicije parcialnega odvoda. Če parcialni odvod ni konstanten, neodvisna spremenljivka vpliva na spremembo vrednosti funkcije. Poskus zasnujemo na funkcijah $f(x, y) = x^2 - y^2$ in $f(x, y) = xy$ takole.

Množico podatkov E naključno razdelimo na dve enako veliki množici E_1 in E_2. Za vsak atribut a naredimo naslednje: Na E_1 in E_2 uporabimo Padéja (τ-regresija, $\kappa = 100$, $k = 20$) za izračun f_{Qa}. Množico E_1 uporabimo kot učno, pri čemer originalni razred nadomestimo z f_{Qa}. Za učenje uporabimo algoritem kNN ($k = 10$). Za testno množico uporabimo E_2 in merimo AUC, t.j. ploščino pod krivuljo ROC [Fawcett, 2006]. Pričakujemo, da bo AUC zaznal 'naključnost' napovedi kNN pri tistih atributih, ki nimajo vpliva na razred in nasprotno, pričakujemo visok AUC pri atributih, ki vplivajo na razred.

Podobno kot pri analizi vpliva naključnih atributov na točnost Padéja, tudi zdaj povprečimo AUC prek 10 naključnih vzorčenj podatkov. Rezultate prikazuje tabela 2.9. Rezultate meritev AUC pri ocenjevanju koristnosti atributa x (y) označimo z AUC_x (AUC_y). Prav tako izračunamo AUC za vsak naključni atribut posebej, v tabeli pa z AUC_{rand} označimo povprečje AUC prek vseh naključnih atributov.

# naključnih atributov	AUC_x		AUC_y		AUC_{rand}	
	$x^2 - y^2$	xy	$x^2 - y^2$	xy	$x^2 - y^2$	xy
0	1.00	.99	.99	1.00	/	/
1	.99	.97	.99	.98	.520	.500
2	.98	.96	.98	.96	.505	.520
3	.97	.95	.97	.94	.508	.510
4	.97	.94	.96	.93	.503	.505
5	.96	.92	.96	.92	.502	.510
10	.91	.83	.91	.83	.498	.504
20	.85	.73	.86	.73	.498	.501
30	.79	.66	.81	.65	.501	.503
40	.77	.62	.77	.61	.498	.501
50	.76	.60	.76	.61	.491	.502

Tabela 2.9: Rezultati poskusa izbire podmnožice koristnih atributov na podlagi AUC na nešumnih podatkih.

Rezultati kažejo na značilno razliko v AUC za pomembna atributa x in y na eni in naključne atribute na drugi strani. Ker imajo naključni atributi AUC po definiciji 0.5, je za naš namen pomembno le odstopanje AUC od te vrednosti – višja kot je vrednost AUC, bolj pomemben je atribut. Ob tem se nam zastavlja vprašanje, kako dobro ta postopek deluje na podatkih, ki vsebujejo šum. Naslednji poskus nas prepriča, da je dvom odveč. Šum dodajamo kot v razdelku 2.5.6. Za vsako stopnjo šuma ponovimo zgornji poskus z dodajanjem naključnih atributov. Rezultate prikazuje tabela 2.10.

Kljub prisotnosti šuma za funkcijo $x^2 - y^2$ še vedno lahko pravilno ocenimo pomembnost atributov. Pri funkciji xy s stopnjo šuma $\sigma \geq 30$ ter 30 in več naključnih atributih z našim postopkom ne moremo več ločiti med pomembnimi in nepomembnimi atributi.

Predlagani postopek je možno razširiti v iterativno metodo, ki v vsaki iteraciji odstrani atribut z minimalnim AUC, s čimer zmanjšuje razsežnost atributnega prostora. To našemu postopku olajša delo – iz tabele 2.9 je razvidno, da so razlike med AUC pomembnih in nepomembnih atributov v nižjih dimenzijah bolj izrazite.

stopnja šuma	# naključnih	AUC_x		AUC_y		AUC_{rand}	
σ	atributov	x^2-y^2	xy	x^2-y^2	xy	x^2-y^2	xy
10	0	0.989	0.994	0.988	0.992	/	/
	3	0.959	0.940	0.972	0.920	0.524	0.492
	5	0.970	0.875	0.958	0.888	0.480	0.530
	10	0.938	0.763	0.939	0.741	0.500	0.513
	30	0.890	0.638	0.889	0.571	0.502	0.500
	50	0.858	0.556	0.835	0.589	0.498	0.501
30	0	0.980	0.951	0.972	0.954	/	/
	3	0.926	0.830	0.956	0.825	0.517	0.518
	5	0.947	0.784	0.930	0.825	0.516	0.527
	10	0.922	0.704	0.939	0.692	0.499	0.511
	30	0.861	0.598	0.889	0.551	0.503	0.502
	50	0.831	0.552	0.819	0.554	0.496	0.494
50	0	0.921	0.846	0.922	0.880	/	/
	3	0.876	0.745	0.868	0.717	0.545	0.541
	5	0.896	0.730	0.856	0.689	0.481	0.475
	10	0.905	0.608	0.858	0.605	0.508	0.509
	30	0.839	0.539	0.825	0.505	0.503	0.499
	50	0.820	0.520	0.780	0.557	0.504	0.505

Tabela 2.10: Ocene kvalitete atributov pri različnih stopnjah šuma in ob prisotnosti naključnih atributov.

2.5.8 Ocenjevanje stabilnosti na realnih podatkih

Stabilnost Padéja smo ocenjevali na šestih UCI [Asuncion and Newman, 2007] domenah: servo, imports-85, galaxy, prostate, housing, auto-mpg in štirih umetnih domenah, ki smo jih že srečali v dosedanjih poskusih: xy, x^2-y^2, x^3-y in $\sin x \sin y$. Za vsako domeno smo naredili 100 naključnih vzorcev s ponavljanjem (*angl. bootstrap*) ter za vsak vzorec izračunali kvalitativne parcialne odvode razreda po vseh atributih za vse učne primere.

Stabilnost atributa x_i glede na učni primer $\mathbf{x_m}$ označimo z $\beta_{\mathbf{x_m}}(x_i)$ in jo definiramo kot delež prevladujočih kvalitativnih parcialnih odvodov ($Q_{\mathbf{x_m}}(+x_i)$ ali $Q_{\mathbf{x_m}}(-x_i)$) med vsemi odvodi za učni primer $\mathbf{x_m}$,

$$\beta_{\mathbf{x_m}}(x_i) = \frac{max\left(\#Q_{\mathbf{x_m}}(+x_i), \#Q_{\mathbf{x_m}}(-x_i)\right)}{\#Q_{\mathbf{x_m}}(+x_i) + \#Q_{\mathbf{x_m}}(-x_i)}. \tag{2.22}$$

Stabilnost atributa x_i prek vseh učnih primerov je povprečje stabilnosti posameznih primerov. Vrednosti β so blizu 1 takrat, ko je večina kvalitativnih odvodov

enaka, medtem ko je najnižja možna vrednost β enaka 0.5. Višje vrednosti β so torej boljše.

domena/atribut	LWR	τ	pari
auto			
displacement	.76	.80	.75
horsepower	.71	.84	.75
weight	.77	.81	.78
acceleration	.75	.76	.77
imports-85			
normalized-losses	.68	.76	.72
wheel-base	.66	.85	.89
length	.72	.85	.82
width	.67	.82	.94
curb-weight	.88	.88	.96
height	.72	.79	.75
engine-size	.71	.80	.91
bore	.67	.78	.85
compression-ratio	.68	.81	.79
stroke	.65	.80	.81
horsepower	.64	.89	.93
peak-rpm	.66	.83	.75
highway-mpg	.66	.87	.84
city-mpg	.65	.87	.88
housing			
crim	.71	.80	.78
indus	.68	.85	.81
nox	.70	.80	.85
rm	.87	.88	.92
age	.82	.79	.79
dis	.71	.81	.81
tax	.70	.88	.84
ptratio	.67	.85	.84
b	.68	.76	.75
lstat	.71	.91	.93

domena/atribut	LWR	τ	pari
galaxy			
east.west	.93	.90	.88
north.south	.93	.90	.80
radial.position	.93	.90	.83
prostate			
lcavol	.92	.82	.93
lweight	.78	.80	.80
age	.69	.79	.79
lbph	.69	.78	.78
lcp	.71	.77	.71
servo			
pgain	.95	.94	1.00
vgain	.88	.72	.85
xy			
x	.99	.95	.99
y	.99	.94	.98
$x^3 - y$			
x	1.00	1.00	1.00
y	.77	.75	.78
$x^2 - y^2$			
x	.99	.95	.99
y	.99	.96	.99
$\sin x \sin y$			
x	.91	.88	.86
y	.91	.89	.86

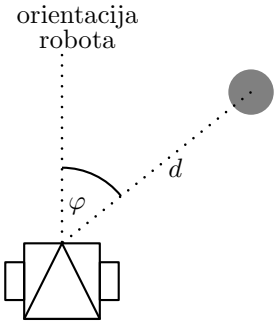
Tabela 2.11: Stabilnost na izbranih realnih in umetnih domenah.

Nastavitve parametrov metod so zaradi manjšega števila primerov v izbranih realnih domenah nekoliko drugačne (manjša vrednost parametra κ), kot v dosedanjih poskusih: $\kappa = 50$ in $k = 20$. Rezultati so zbrani v tabeli 2.11. Za vsak atribut izbrane domene podamo povprečno stabilnost.

Za vseh 46 zveznih atributov iz vseh domen uredimo metode Padéja po stabilnosti, ki jo dosežejo – za vsak atribut pripišemo vrednost 1 najboljši in vrednost 3 najslabši metodi. Povprečje rangov čez vse atribute razkrije, da so vzporedni pari najboj stabilna metoda (rang 1.8), sledi ji τ-regresija z rangom 1.84, najmanj stabilna metoda pa je LWR z znatno nižjim rangom, 2.35.

2.6 Avtonomno učenje robota

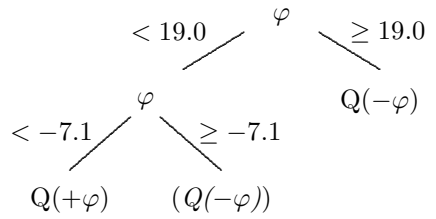
V okviru evropskega projekta XPERO (IST-29427) smo Padé uporabili pri avtonomnem učenju mobilnega robota s kamero. Robot je bil zaprt v ograjen prostor, v katerem je bila poleg njega še žoga, ki jo je robot opazoval s kamero. Njegova naloga je bila odkriti odvisnost med površino žoge P na sliki iz kamere in njegovo razdaljo ter kotom do žoge. Razdaljo in kot do žoge je robot meril s pomočjo kamere, ki je sliko zajemala s tlorisa (slika 2.11).



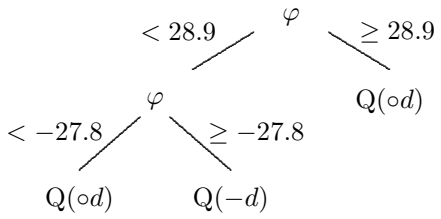
Slika 2.11: Grafični prikaz robotske domene.

Sliki 2.12a in 2.12b prikazujeta kvalitativna modela za odvisnosti med površino žoge P na sliki kamere in kotom φ ter razdaljo robota do žoge d. Drevo 2.12a si razlagamo takole: površina žoge v sliki kamere narašča, ko se robot obrača proti žogi (to je takrat, ko se negativni kot povečuje in pozitivni kot zmanjšuje). Ko je vidna cela žoga, kot nima pomena. V tem vozlišču je porazdelitev primerov približno 50% : 50%, zato drevo na tem mestu ročno porežemo.

Drevo s slike 2.12b pove, da takrat, ko je žoga vidna – približno za $\varphi \in [-28.9, 27.8]$ – površina žoge pada z razdaljo. Sicer razdalja do žoge ni pomembna.



(a) Kvalitativno drevo za $\partial_Q P/\partial_Q \varphi$. Odvisnost v oklepaju označuje ročno porezano vejo drevesa.



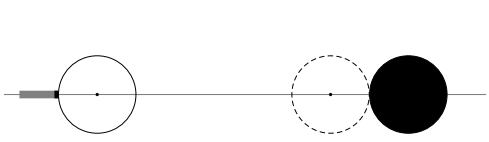
(b) Kvalitativno drevo za $\partial_Q P/\partial_Q d$.

Slika 2.12: Kvalitativna modela, ki opisujeta odvisnosti med površino žoge na sliki kamere in kotom ter razdaljo robota do žoge.

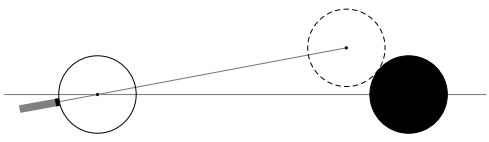
2.7 Študija primera: biljard

Praktično uporabo Padéja si oglejmo na primeru kvalitativne analize osnovnih zakonitosti pri biljardu. Biljard je skupno ime za namizne igre, ki jih igramo s palico in naborom krogel. Obstaja veliko vrst biljarda z različnimi cilji, osnovna ideja pa je skupna vsem: igralec z uporabo palice sune belo kroglo, s katero meri v želeno kroglo tako, da bi dosegel želeni cilj (izboljšal svoj položaj v igri ali oslabil položaj nasprotnika). Trenje med kroglami ter kroglami in mizo, različne rotacije in hitrosti krogel ter odboji med kroglami tvorijo zapleten fizikalni sistem [Alciatore, 2004]. Zapleteni fiziki navkljub pa amaterski igralec hitro osvoji osnovna načela in zna pravilno udariti kroglo, ne da bi poznal fizikalne zakone v ozadju.

Naš namen je zgraditi model za opis kvalitativne odvisnosti med azimutom (ki se odraža v stopnji polnosti trka; slika 2.13) in odbojnim kotom bele žoge po trku s črno žogo. *Odbojni kot* je kot med premico, ki povezuje središči obeh žog in črto, ki povezuje točko prvega trka obeh žog s točko drugega trka bele

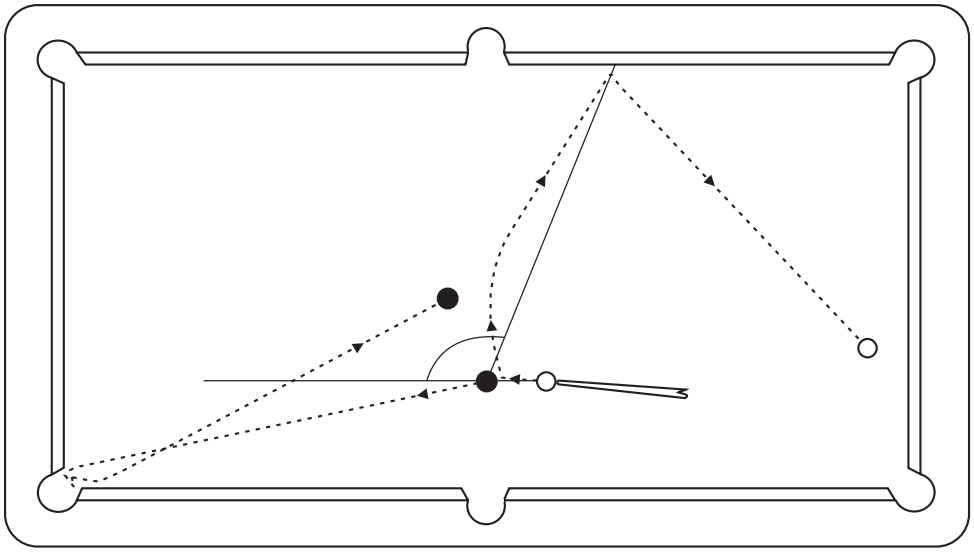


(a) Udarec z azimutom 0. Bela žoga zadane črno *na polno*.

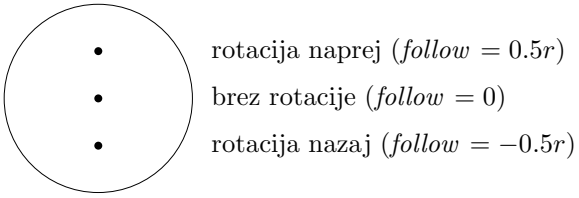


(b) Udarec z azimutom > 0 povzroči *delni trk*.

Slika 2.13: Udarca z različno *polnostjo* trka (atribut azimut).



Slika 2.14: Primer sledi gibanja po udarcu iz simulatorja.



Slika 2.15: Žogi damo pozitivno rotacijo (angl. follow) z udarcem nad središčem in negativno rotacijo z udarcem pod središčem (angl. draw). Točke na žogi predstavljajo točke udarca s palico.

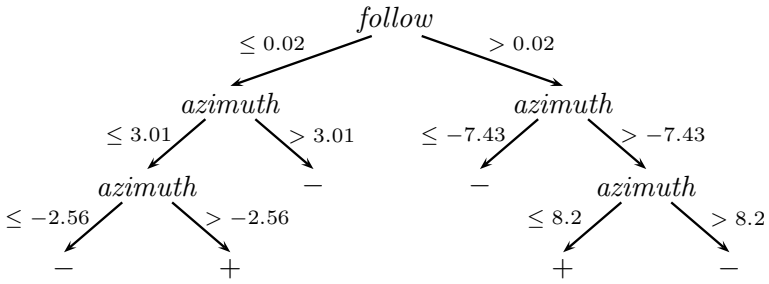
žoge (bodisi v rob mize ali v drugo žogo, slika 2.14). Kvalitativno odvisnost med odbojnim kotom in azimutom računamo v prostoru štirih atributov (tabela 2.12): vodoravni (*azimuth*) in navpični (*elevation*) kot palice ob udarcu, začetna rotacija bele žoge (*follow*) (slika 2.15) in hitrost udarca (*velocity*).

Podatke smo pridobili s simulatorjem Billiards [Papavasiliou, 2009]. Zbrali smo 5000 naključnih primerov. Za računanje odvodov smo uporabili τ-regresijo s parametroma $\kappa = 100$, $k = 20$. Kvalitativno drevo smo zgradili z algoritmom C4.5 in prosili eksperte za razlago modela.

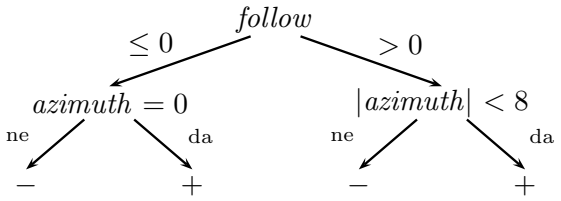
Kvalitativno drevo, naučeno z algoritmom C4.5, je prikazano na sliki 2.16a. Strokovnjaka za biljard sta model ocenila kot pravilen, enostaven in razumljiv. Zgrajeni model sta nekoliko poenostavila z uporabo absolutne vrednosti azimuta

atribut	definicijsko območje	opis
azimuth	$[-15°, 15°]$	kot palice v vodoravni ravnini
elevation	$[0°, 30°]$	kot palice v navpični ravnini
velocity	$[3m/s, 5m/s]$	hitrost palice ob udarcu
follow	$[-.5r, .5r]$	pozitivna/negativna rotacija zaradi udarca palice nad/pod središčem žoge
angle	$[-180°, 180°)$	odbojni kot bele žoge

Tabela 2.12: Atributi v domeni biljard.



(a) Kvalitativno drevo za biljard, zgrajeno s C4.5.



(b) Kvalitativno drevo, kot sta ga razložila eksperta.

Slika 2.16: Kvalitativni model, ki opisuje odvisnost med odbojnim kotom in azimutom.

(slika 2.16b). Po njunem mnenju tudi začetnikom omogoča, da pravila iz modela prenesejo v igro – veliko lažje, kot bi to storili z množico fizikalnih enačb, ki opisujejo zapleteno obnašanje pri biljardu. Dobljeni model lahko razumemo tudi kot konceptualizacijo domene.

Drevo razložimo takole: prva vejitev temelji na atributu *follow* – pri negativnih vrednostih (belo žogo udarimo pod središčem) bela žoga dobi negativno rotacijo, ki se odraža v inverzni proporcionalnosti med odbojnim kotom in azimutom, t.j. če povečamo azimut, se odbojni kot zmanjša in obratno. Izjema

je azimut pri vrednostih blizu 0, kar je posledica izbire koordinatnega sistema: odbojni kot seka poltrak nezveznosti pri $\pm 180°$. Povečanje azimuta s $179°$ na $181°$ zaznamo kot zmanjšanje z vrednosti $179°$ na $-179°$.

Pri pozitivnih vrednostih *follow* žoga dobi dodatno pozitivno rotacijo (nekaj je ima že zaradi kotaljenja). Azimut ima v tem primeru drugačen vpliv na črno žogo. Zaradi pozitivne rotacije ima bela žoga tudi po trku vztrajnost za kotaljenje naprej. Za azimut med $-8°$ in $8°$ velja, da povečanje (zmanjšanje) azimuta poveča (zmanjša) odbojni kot. Za absolutne vrednosti azimuta nad $8°$ bela žoga le delno trči v črno, kar povzroči da odbojni kot postane inverzno proporcionalen azimutu.

Pomembna je ugotovitev, da naš model pravilno prepozna oba pomembna atributa, *follow* in *azimuth*. Ostala dva, *elevation* in *velocity* v našem kontekstu nimata vpliva na odbojni kot, kar sta potrdila tudi oba strokovnjaka.

Pravilnost kvalitativnih odvisnosti iz našega modela smo preverili tudi v simulatorju. Simulirali smo majhne spremembe azimuta, vrednosti ostalih atributov pa nismo spreminjali. Pri tem smo opazovali spreminjanje odbojnega kota. Razen manjših razhajanj pri vrednostih vejitev v drevesu smo ugotovili, da je model pravilen.

Poglavje 3

Učenje kvalitativnih odvisnosti v klasifikaciji

Kvalitativno modeliranje se tradicionalno ukvarja z numeričnimi (zveznimi) domenami, kjer pojem *kvalitativno* jasno označuje posplošitev numeričnih odvisnosti. Parcialni odvod, s katerim si pomagamo pri kvalitativnem modeliranju v zveznih domenah, opisuje vpliv spremembe vrednosti neodvisne zvezne spremenljivke na odvisno zvezno spremenljivko. Kvalitativni parcialni odvod pove samo smer spremembe odvisne spremenljivke.

V domenah z diskretnim razredom lahko podobno opazujemo vplive posameznih atributov na izbrani ciljni razred. Za določen atribut nas zanima, katere vrednosti imajo večji (manjši) vpliv na ciljni razred. Pri opazovanju odvisnosti med ciljnim razredom in izbranim diskretnim atributom izhajamo iz nomograma* naivnega Bayesa [Možina et al., 2004], v katerem vplive vrednosti atributov na ciljni razred merimo s pomočjo pogojnih verjetnosti. Le-te nas pripeljejo nazaj v zvezni svet, kjer prek abstrakcije pogojnih verjetnosti definiramo parcialne kvalitativne spremembe.

Najprej predstavimo *izbirne nomograme*, ki služijo kot grafična predstavitev naivnega Bayesa na izbrani podmnožici podatkov in so izhodišče za definicijo parcialnih kvalitativnih sprememb. Za računanje le-teh in za kvalitativno modeliranje v domenah z diskretnim razredom predlagamo algoritem Qube. Podobno kot Padé, tudi Qube deluje kot predprocesor, ki pripravi podatke za gradnjo kvalitativnih modelov z ustreznim klasifikacijskim algoritmom strojnega učenja. Teoretičnemu delu sledi nekaj poskusov na umetnih domenah, kjer v nadzoro-

*V tem delu naj velja, da z nomogramom vedno mislimo na nomogram naivnega Bayesa.

vanih okoliščinah prikažemo delovanje algoritma Qube. Zaključimo z obsežno študijo, kjer Qube uporabimo na medicinskih podatkih, ki opisujejo paciente s hudimi bakterijskimi okužbami.

3.1 Izbirni nomogrami

Nomogram je grafično računsko orodje za približno (ročno) računanje vrednosti funkcij iz danih vrednosti spremenljivk. Uporablja svoj koordinatni sistem, ki ga predstavi z ravninskim diagramom, s katerim si pomagamo pri grafičnem računanju. Nomogrami se večinoma uporabljajo na področjih, kjer zadostuje že približen izračun funkcije. V statistiki in strojnem učenju najpogosteje srečamo nomograme logistične regresije in naivnega Bayesa ter nomograme podpornih vektorjev. Izbirni nomogrami, ki jih definiramo v tem razdelku, izhajajo iz nomogramov naivnega Bayesa. So interaktivno orodje, ki s pomočjo nomograma naivnega Bayesa pomagajo pri raziskovanju atributnega prostora in iskanju novega znanja iz podatkov.

Vzemimo najprej, da je okolica poljubnega primera kar množica vseh primerov, torej cel prostor. Vpliv atributa A_i na verjetnost ciljnega razreda c opazujemo tako, da v nomogramu spreminjamo vrednosti A_i, vrednosti ostalih atributov pa pustimo na apriornih vrednostih. Pri tem se spreminja verjetnost ciljnega razreda c, kar nam pove, kako izbrani atribut v kontekstu vseh podatkov vpliva na ciljni razred. Na omenjeni način opazimo globalni vpliv A_i na c, ne moremo pa opazovati njegovega vpliva v posamezni točki v prostoru.

To lahko storimo tako, da izberemo referenčni primer e ter opazujemo vpliv A_i na c v njegovi okolici. Okolico izberimo tako, da iz originalne množice podatkov izberemo tako podmnožico primerov, ki se z referenčnim primerom ujema v vrednostih vseh atributov, razen A_i. To je lastnost, ki sledi definiciji parcialnih odvodov v matematiki. Nomogram na izbrani podmnožici vsebuje samo atribut A_i. V realnih domenah tak postopek pogosto ni možen, saj je razmerje med številom atributov in primerov tako, da praktično za noben e ne obstaja dovolj velika okolica s to lastnostjo, ki bi imela dovolj primerov, da bi bilo na njej smiselno učenje z naivnim Bayesom.

Prevelika okolica v prvem primeru in premajhna v drugem, nakazujeta, da je prava okolica nekje vmes. Poiskali jo bomo z *izbirnimi nomogrami*, ki običajne nomograme razširijo z dodatno možnostjo, da po izbiri vrednosti a_k atributa A_i

iz množice podatkov, na katerih je zgrajen nomogram, izberejo podmnožico učnih primerov, ki imajo $A_i = a_k$. Postopek iterativno ponavljamo, dokler izbrana podmnožica vsebuje dovolj učnih primerov. Za izbiro želene okolice referenčnega primera e ta postopek lahko uporabimo tako, da atribute v nomogramu uredimo po njihovem vplivu na ciljni razred. Vedno izberemo najvplivnejši atribut A_{max} in tisto njegovo vrednost, ki je enaka vrednosti A_{max} v primeru e. Opisani postopek je požrešna metoda izbirnih nomogramov. Bolj napredna je *daljnovidna metoda* izbirnih nomogramov, ki lahko odkrije tudi pogojno odvisne pare ali n-terke atributov, odvisno od stopnje daljnovidnosti (npr. 2 nivoja). Daljnovidna metoda za vse atribute (ne samo najvplivnejšega) simulira požrešno metodo, dejansko pa okolico določa s tistimi atributi, ki daljnoročno najbolj povečajo verjetnost ciljnega razreda.

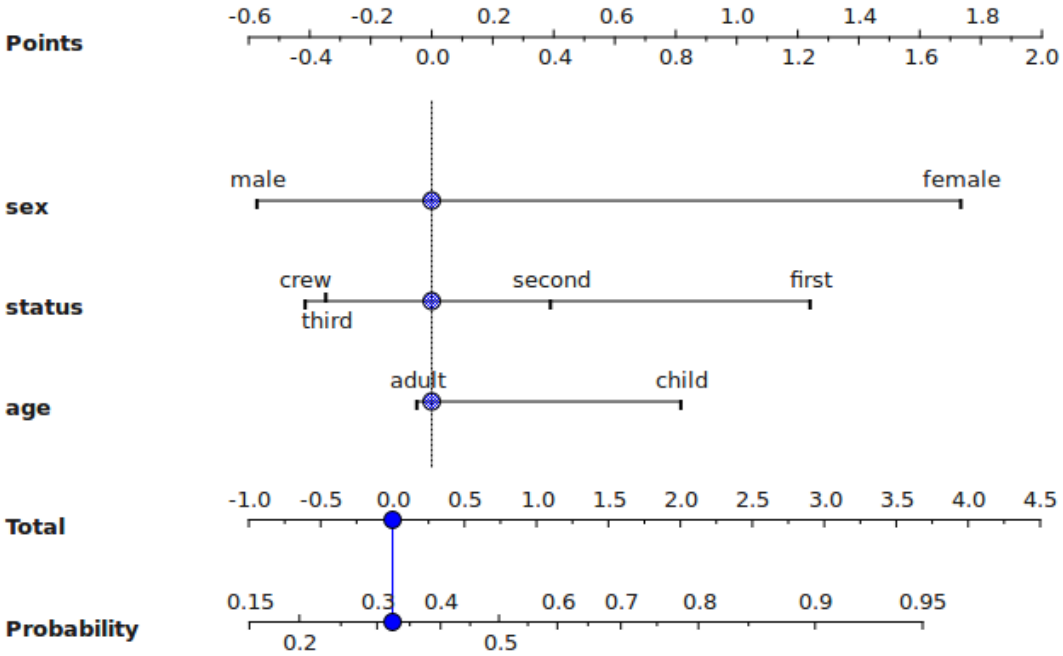
Za okolico, ki smo jo dobili s postopkom izbirnih nomogramov je značilno, da vsebuje take primere, ki so referenčnemu enaki v čim več pomembnih atributih, začenši z najvplivnejšim na ciljni razred. Najbližji sosedje so torej tisti, ki se razlikujejo po čim manj pomembnih atributih.

3.1.1 Primer: Titanic

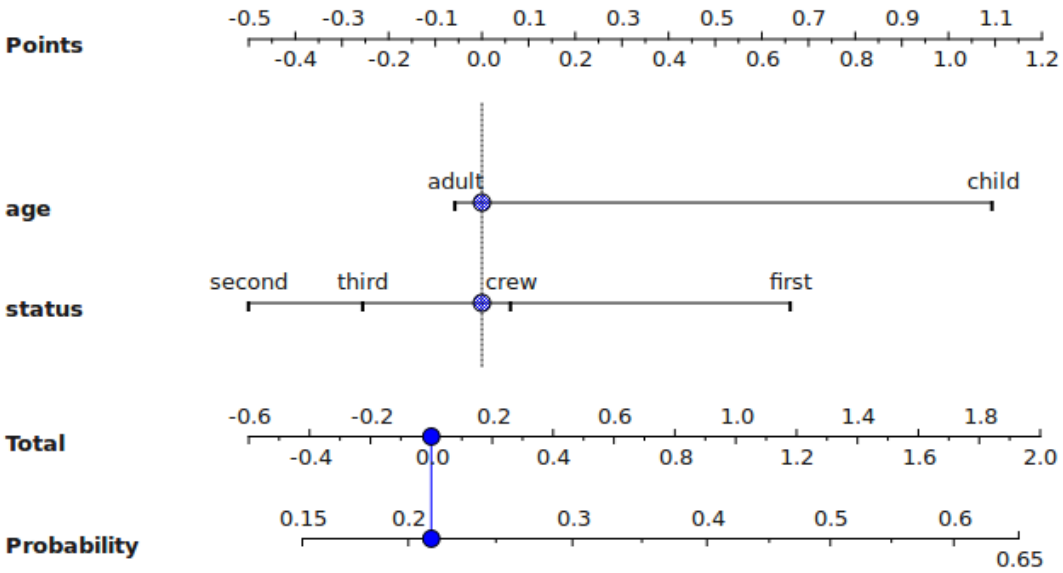
Podatki Titanic vsebujejo 2201 primer, ki je opisan s tremi atributi: *sex (male, female), status (first, second, third, crew), age (adult, child).*

Izberimo za referenčni primer e_1 odraslega moškega v prvem razredu: $e_1 = \{$*male, first, adult*$\}$. Slika 3.1a prikazuje nomogram za podatke pri izbranem ciljnem razredu *survived=yes*. Apriorna verjetnost preživetja je 32%. Gledano globalno, za vsak atribut posebej, našemu potniku spol zmanjšuje verjetnost preživetja (-11 odstotnih točk), status mu verjetnost poveča (+30 odstotnih točk), starost pa mu praktično ne škoduje (-1 odstotna točka).

Idealno okolico e_1 sestavljajo vsi primeri odraslih moških v prvem razredu. Na Titanicu je bilo takih 175, kar je dovolj velik vzorec. Recimo, da je imel e_1 s seboj sina, ki ga označimo z e_2, torej $e_2 = \{$*male, first, child*$\}$. Sin ni imel toliko družbe kot njegov oče, saj je v prvem razredu skupaj potovalo le 5 dečkov. Za potrebe statistike ali učenja je okolica e_2 premajhna. Sestavimo jo po požrešnem postopku, ki smo ga opisali zgoraj. Najvplivnejši atribut je spol, vrednost tega atributa za primer e_2 pa je *male*. Izberemo torej *sex=male* in dobimo podmnožico 1731 primerov moških vseh starosti, ne glede na status. V bližnji okolici e_2 torej gotovo ne bo ženske. Nomogram za množico moških na

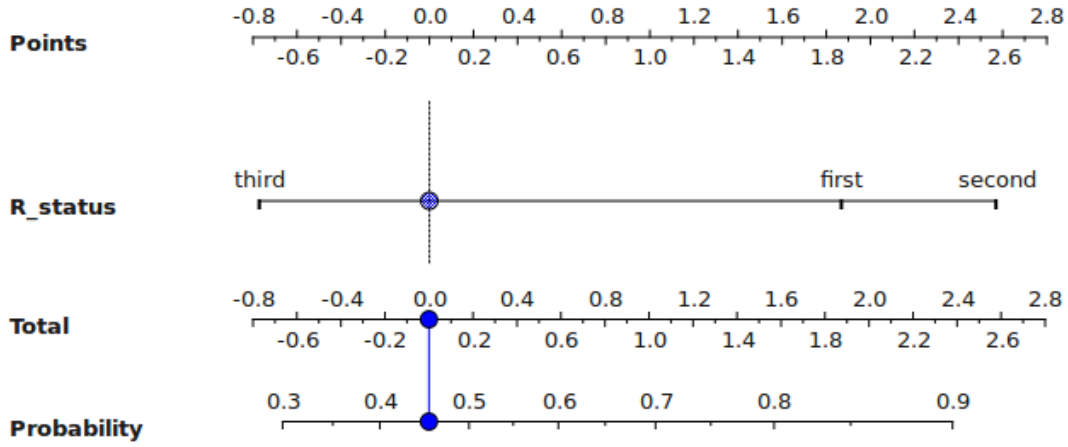


(a) Nomogram za Titanic.



(b) Nomogaam za podmnožico vseh moških na Titanicu.

Slika 3.1: Prva iteracija v izbirnem nomogramu - izbira podmnožice moških na Titanicu.



Slika 3.2: Nomogram za podmnožico vseh dečkov na Titanicu.

Titanicu je prikazan na sliki 3.1b. Najvplivnejši atribut na tej podmnožici je starost. Glede na atribut *age* pri e_2, izberemo *age=child*. Podmnožica dečkov šteje 64 primerov, novi nomogram pa je prikazan na sliki 3.2. Kot kaže nomogram na sliki 3.1b, so imeli moški potniki prvega razreda veliko večje možnosti preživetja kot potniki ostalih razredov in posadka. Nomogram na sliki 3.2, v katerem smo opazovali le dečke, pa pokaže, da je bila možnost preživetja za dečke, ki so potovali v prvem in v drugem razredu, praktično enaka. Primer nazorno kaže, kako ima lahko določen atribut (v našem primeru *status*) v nekem delu prostora (v našem primeru med dečki) lahko drugačen vpliv na razred kot v splošnem.

3.2 Parcialne kvalitativne spremembe

Izbirni nomogrami so nam služili le kot miselno orodje. V tem razdelku se vračamo k računanju parcialnih kvalitativnih sprememb (PKS), ki so podlaga za učenje kvalitativnih modelov. Idejo za izbiro okolice iz izbirnih nomogramov bomo formalno opisali in jo uporabili pri definiciji okolice za računanje PKS v domenah z diskretnim razredom in diskretnimi atributi.

Pri definiciji okolic za računanje odvodov iz numeričnih podatkov smo idealne pogoje, ki jih predpisuje matematična definicija, nekoliko omehčali, kar nas je privedlo do kratke cevne okolice. Podobnih okolic si želimo tudi v diskretnih domenah, da bi na njih lahko računali parcialne kvalitativne spremembe v pro-

storih diskretnih atributov, nato pa gradili kvalitativne modele, spet po zgledu iz zveznega sveta.

Razdelek začnemo z definicijo parcialne kvalitativne spremembe, ki temelji na pogojnih verjetnostih ciljnega razreda pri danih vrednostih atributov. Nadaljujemo z definicijo okolice referenčnega primera, ki je pogoj za izbiro ustrezne podmnožice podatkov za izračun omenjenih pogojnih verjetnosti iz podatkov. Zaključimo z računanjem parcialnih kvalitativnih sprememb s pomočjo pogojnih verjetnosti in učenjem kvalitativnih modelov iz njih.

Vzemimo multivariatno porazdelitev, ki vsakemu elementu kartezičnega produkta $\mathcal{A}_1 \times \mathcal{A}_2 \times \ldots \times \mathcal{A}_n$ priredi verjetnost y. V strojnem učenju so vrednosti $(a_1, a_2, \ldots, a_n) \in \mathcal{A}_1 \times \mathcal{A}_2 \times \ldots \times \mathcal{A}_n$ kar vrednosti diskretnih atributov $A_1, A_2, \ldots A_n$, ki opisujejo učni primer, y pa je verjetnost ciljnega razreda c za tak primer,

$$y = p(c|a_1, a_2, \ldots, a_n). \tag{3.1}$$

Spomnimo se definicije kvalitativnega odvoda f po x_i v točki $(x_1, x_2, \ldots x_n)$, ki pove, kako se sprememba x_i odraža v vrednosti f (narašča ali pada), če fiksiramo ostale vrednosti argumentov funkcije. *Parcialna kvalitativna sprememba* (PKS) ciljnega razreda po atributu A_i v $(a_1, a_2, \ldots, a_n)$ nam pove, kako sprememba vrednosti atributa A_i iz a_i v a_i' kvalitativno vpliva na verjetnost ciljnega razreda (jo poveča ali zmanjša). Formalno definicijo zapišemo takole:

$$\Delta f_{A_i:a_i \to a_i'}(a_1, \ldots, a_n) =$$

$$\begin{cases} +, & p(c|a_1, \ldots, a_i, \ldots, a_n) > p(c|a_1, \ldots, a_i', \ldots, a_n) \\ \circ, & p(c|a_1, \ldots, a_i, \ldots, a_n) = p(c|a_1, \ldots, a_i', \ldots, a_n) \\ -, & p(c|a_1, \ldots, a_i, \ldots, a_n) < p(c|a_1, \ldots, a_i', \ldots, a_n). \end{cases} \tag{3.2}$$

Definirajmo urejenost na množici $\mathcal{A}_i$ pri fiksni vrednosti a_j za $j \neq i$

$$a_i \leq a_i' \quad \Leftrightarrow \quad p(c|a_1, \ldots, a_i, \ldots, a_n) \leq p(c|a_1, \ldots, a_i', \ldots, a_n) \tag{3.3}$$

Izraz (3.2) zdaj lahko zapišemo kot

$$\Delta f_{A_i:a_i \to a_i'}(a_1, \ldots, a_n) = \begin{cases} +, & a_i \prec a_i' \\ \circ, & a_i = a_i' \\ -, & a_i \succ a_i'. \end{cases} \tag{3.4}$$

Parcialno kvalitativno spremembo f po A_i, Δf_{A_i}, definiramo kot ureditev vrednosti atributa $\mathcal{A}_i$.

3.2.1 Računanje pogojnih verjetnosti

Za izračun parcialne kvalitativne spremembe moramo za dane podatke izračunati pogojno verjetnost $p(c|a_1, \ldots, a_n)$ v točki $(a_1, a_2, \ldots, a_n)$. Te verjetnosti ne moremo izračunati neposredno, saj je v podatkih ponavadi premalo učnih primerov z vrednostmi atributov, ki ustrezajo pogojnemu delu. Tudi naivni Bayes ni uporaben, saj PKS prevede na primerjavo $p(c|a_i)$ in $p(c|a_i')$, ne upoštevajoč vrednosti vseh ostalih atributov. Iz predpostavke naivnega Bayesa o pogojni neodvisnosti atributov pri danem razredu sledi, da je PKS $\Delta f_{A_i:a_i \to a_i'}$ povsod konstantna, t.j. f se ne spremeni pri spremembi vrednosti atributa.

Problema se lotimo s postopkom, ki nekoliko spominja na delno-naivni Bayes [Kononenko, 1991]. Pogoj v $p(c|a_1, \ldots, a_n)$ nadomestimo z mehkejšim pogojem $\mathcal{D} \subseteq \{a_1, a_2, \ldots\}$, kjer $\mathcal{D}$ vsebuje le tiste vrednosti atributov, ki so pri danem razredu pogojno odvisni od a_i. Na ta način definiramo cevno okolico, podobno kot smo to storili v regresijskih domenah. Smer cevi še vedno določa atribut A_i, širino cevi pa število pogojno odvisnih atributov, ki jih zaradi pomanjkanja podatkov ne uspemo povsem odpraviti. Vrednosti pogojno neodvisnih atributov ne vplivajo na računanje PKS in jih zato lahko zanemarimo.

Urejenost verjetnosti $p(c|a_1, \ldots, a_i, \ldots, a_n)$ se ne spremeni, če v pogojnem delu izpustimo tiste vrednosti, ki so pri danem c pogojno neodvisne od a_i. Definicijo 3.2 lahko zapišemo s pomočjo logaritma razmerja pogojnih verjetnosti (*angl. log odds ratio*):

$$\begin{aligned} &\Delta f_{A_i:a_i \to a_i'}(a_1, \ldots, a_n) = \\ &\quad \operatorname{sgn}(-\ln \frac{p(c|a_1, \ldots, a_i, \ldots, a_n)/p(\bar{c}|a_1, \ldots, a_i, \ldots, a_n)}{p(c|a_1, \ldots, a_i', \ldots, a_n)/p(\bar{c}|a_1, \ldots, a_i', \ldots, a_n)}), \end{aligned} \tag{3.5}$$

kjer je $\bar{c}$ komplement ciljnega razreda c. Očitno je (3.5) ekvivalentno (3.2).

Brez izgube za splošnost predpostavimo, da so pri danem razredu vrednosti a_1 do a_k, za $k < i$ pogojno neodvisne od vrednosti a_{k+1} do a_n. Upoštevajoč predpostavko o pogojni neodvisnosti uporabimo Bayesovo formulo in iz (3.5)

dobimo

$$\Delta f_{A_i:a_i\to a_i'}(a_1,\ldots,a_n) = \\ \operatorname{sgn}(-\ln\frac{p(c|a_{k+1},\ldots,a_i,\ldots,a_n)/p(\bar{c}|a_{k+1},\ldots,a_i,\ldots,a_n)}{p(c|a_{k+1},\ldots,a_i',\ldots,a_n)/p(\bar{c}|a_{k+1},\ldots,a_i',\ldots,a_n)}). \tag{3.6}$$

Izraz je ekvivalenten izrazu (3.2) brez vrednosti a_1 do a_k. Zato lahko zapišemo:

$$\begin{aligned} &p(c|a_1,\ldots,a_i,\ldots,a_n) \le p(c|a_1,\ldots,a_i',\ldots,a_n) \iff \\ &p(c|a_{k+1},\ldots,a_i,\ldots,a_n) \le p(c|a_{k+1},\ldots,a_i',\ldots,a_n). \end{aligned} \tag{3.7}$$

Nadaljujmo s postopkom za določanje pogojev $\mathcal{D}$, ki jih bomo uporabili v požrešnem pristopu: začnemo s prazno množico $\mathcal{D}$ in iterativno dodajamo najbolj odvisne vrednosti. Naj bo e_j dogodek, da ima atribut A_j pri določenem primeru vrednost a_j, dogodek v pa naj predstavlja vrednosti atributa A_i ($v \in \mathcal{A}_i$). Preveriti želimo, ali sta e_j in v pogojno neodvisna pri danem razredu in ostalih pogojih iz množice $\mathcal{D}$. Z drugimi besedami, če za dani učni primer poznamo vrednost razreda in vemo, da primer ustreza pogojem iz $\mathcal{D}$, bi radi preverili, da verjetnostna porazdelitev vrednosti A_i ni odvisna od tega ali je vrednost j-tega atributa enaka vrednosti a_j ali ne. V vsakem koraku algoritma računamo odvisnost med e_j in v in izberemo tisti e_j, ki najbolj krši neodvisnost ter v $\mathcal{D}$ dodamo ustrezni a_j.

Neodvisnost merimo s standardnim χ^2-testom. Za ciljni razred c in njegov komplement naredimo ločeni tabeli dimenzij $2 \times |\mathcal{A}_i|$. Izračun pričakovanih absolutnih frekvenc za c in $\bar{c}$ je enak, zato ga zapišimo samo za c: pričakovani frekvenci pri dogodkih e_j in $\overline{e}_j$ sta enaki $n(c,\mathcal{D})p(a_j|c,\mathcal{D})p(v|c,\mathcal{D})$ in $n(c,\mathcal{D})(1-p(a_j|c,\mathcal{D}))p(v|c,\mathcal{D})$, kjer je $v \in \mathcal{A}_i$ in $n(c,\mathcal{D})$ število primerov s ciljnim razredom c, ki ustrezajo pogojem iz $\mathcal{D}$. Vsota χ^2 statistik za obe tabeli je porazdeljena po porazdelitvi χ^2 z $2(|\mathcal{A}_i|-1)$ prostostnimi stopnjami.

Ker imajo različni atributi različno število vrednosti, imajo pripadajoče statistike χ^2 različno število prostostnih stopenj in med seboj niso primerljive. Zato za vsak a_j izračunamo ustrezno p-vrednost in izberemo a_j z najmanjšo p-vrednostjo. Postopek izbiranja vrednosti atributov ustavimo, ko najmanjša p-vrednost preseže dani prag (npr. 0.05) oziroma, če število primerov, ki ustrezajo pogojem iz množice $\mathcal{D}$ pade pod vnaprej določeni minimum (npr. 30 primerov). Slednje je potrebno zato, da zagotovimo zanesljivost izračuna statistike χ^2 in ocene pogojnih verjetnosti. Pri uporabi p-vrednosti ne delamo popravkov za

testiranje več hipotez, saj p-vrednost uporabljamo le kot ustavitveni kriterij in ne kot mero za značilnost alternativnih hipotez.

Ko smo določili množico pogojev $\mathcal{D}$, za vse učne primere, ki ustrezajo pogojem iz $\mathcal{D}$, izračunamo $p(c|v, \mathcal{D})$ za vsak $v \in \mathcal{A}_i$ (npr. z m-oceno [Cestnik, 1990]).

χ^2-test računamo na vseh vrednostih atributa A_i, ne samo a_i in a_i'. Zato lahko množico pogojev $\mathcal{D}$ uporabimo pri izračunu vseh PKS po A_i v izbranem referenčnem primeru in zagotovimo, da so verjetnosti $p(c|a_1, \ldots, a_i, \ldots, a_n)$ za vse $a_i \in \mathcal{A}_i$ primerljive med seboj. Zaradi tega lahko za atribut A_i definiramo urejenost njegovih vrednosti a_i.

Predlagani požrešni postopek je preprost, a se v praksi izkaže za uporabnega. Dobra lastnost požrešnega pristopa je gotovo njegova hitrost, ki je zelo pomembna, ker moramo postopek ponoviti za vsak učni primer.

3.2.2 Računanje parcialnih kvalitativnih sprememb

Urejenost množice $\mathcal{A}_i$ je določena z urejenostjo ustreznih verjetnosti glede na definicijo (3.2). Zaradi robustnosti metode računanja PKS pa dve verjetnosti (in ustrezni vrednosti A_i) vzamemo za enaki, če se razlikujeta za manj kot vnaprej določen prag, ki je parameter metode.

V ta namen na vrednostih atributa A_i uporabimo hierarhično razvrščanje v skupine s povprečnim povezovanjem (*angl. average linkage*) [Sokal and Michener, 1958]. Za razdalje uporabljamo kar razlike verjetnosti. Z razvrščanjem v skupine prenehamo, ko je razdalja med najbližjima skupinama večja od npr. 0.2 (pragovna vrednost, ki jo nastavi uporabnik).

Vzemimo za primer atribut A_i s petimi vrednostmi v_1 do v_5. Verjetnosti $p(c|v_i, \mathcal{D})$ teh vrednosti naj bodo po vrsti: 0.1, 0.2, 0.3, 0.5 in 0.6. Združitveni prag naj bo 0.2. Vrednosti v_1 in v_2 združimo in jima pripišemo povprečno verjetnost $(0.1 + 0.2)/2 = 0.15$. Nato združimo v_4 in v_5, kjer je povprečna verjetnost $0.5 + 0.6 = 0.55$. Nazadnje združimo še v_1 in v_2 z v_3; povprečna verjetnost je $(0.1 + 0.2 + 0.3)/3 = 0.2$. Postopek se nato ustavi, ker nobena od razlik verjetnosti ni pod združitvenim pragom; $p = 0.2$ (za v_1, v_2 in v_3) in $p = 0.55$ (za v_4 in v_5). Končna urejenost vrednosti za atribut A_i je $v_1 = v_2 = v_3 < v_4 = v_5$.

3.2.3 Učenje kvalitativnih modelov

Podobno kot v regresiji se kvalitativnih modelov učimo v dveh fazah. Najprej izračunamo parcialne kvalitativne spremembe z algoritmom Qube, nato pa jih uporabimo kot razredno spremenljivko pri učenju s poljubnim algoritmom za učenje klasifikacije. Možnih vrednosti razredne spremenljivke je teoretično toliko kot vseh vrstnih redov vrednosti originalnega atributa, po katerem opazujemo PKS, prišteti pa moramo še nove vrednosti, ki jih konstruira Qube (združene originalne vrednosti). Poskusi kažejo, da v praksi število vrednosti novonastalega atributa, ki opisuje PKS, ni veliko. V primeru, da bi se to zgodilo, lahko z nekaj dodatnega predprocesiranja pred učenjem kvalitativnega modela pripravimo podatke tako, da je učenje smiselno. Za razliko od učenja kvalitativnih modelov iz regresijskih podatkov tu nimamo atributa, ki bi nam povedal značilnost PKS.

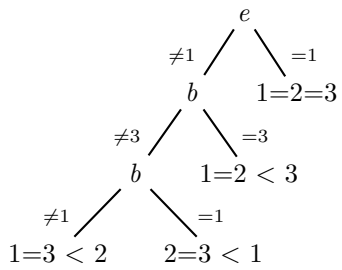
3.3 Poskusi

Poskuse z algoritmom Qube bomo predstavili na treh domenah iz zbirke podatkov UCI [Asuncion and Newman, 2007]: Monks 1, Monks 3 in Titanic. Monks 1 in Monks 3 sta zanimivi testni domeni in sta v našem primeru še posebej dobrodošli, ker zanju poznamo pravilni koncept. To nam bo omogočilo lažje ocenjevanje rezultatov. Titanic je predstavnik resničnih podatkov in nam omogoča oceniti razložljivost dobljenih modelov.

V vseh poskusih smo pogojne verjetnosti ocenjevali z m-oceno [Cestnik, 1990] in sicer za $m = 2$. Pri združevanju atributnih vrednosti smo kot pragovno vrednost uporabljali 0.2. Za gradnjo kvalitativnih modelov, ki opisujejo odvisnosti med izbranim ciljnim razredom in posameznim atributom, smo uporabljali implementacijo algoritma C4.5 v programu Orange [Zupan et al., 2004]. Zaradi preglednosti smo posamezna drevesa predstavili s seznamom pravil.

3.3.1 Monks 1

Podatki Monks vsebujejo 6 diskretnih atributov: a $(1, 2, 3)$, b $(1, 2, 3)$, c $(1, 2)$, d $(1, 2, 3)$, e $(1, 2, 3, 4)$, f $(1, 2)$ in dvojiški razred y $(1, 0)$. Ciljni koncept za Monks 1 je definiran takole: $y := (a = b) \vee (e = 1)$. Izberemo ciljni razred $y = 1$.



Slika 3.3: Model za PKS y po atributu a za Monks 1.

Parcialna kvalitativna sprememba po c je $1 = 2$ na celem prostoru, kar pomeni, da imata vrednosti atributa c 1 in 2 enak vpliv na ciljni razred. To se sklada s pravilnim konceptom, saj je c nepomemben atribut. Podobno lahko rečemo tudi za atributa d, kjer je PKS $1 = 2 = 3$ in f, kjer je PKS enak $1 = 2$.

Slika 3.3 prikazuje model za y_{Qa}; model za y_{Qb} je ekvivalenten. Drevo pravi, da pri $e = 1$ vrednosti atributov a (in b) nista pomembni; kar je res, če vemo, da je pri $e = 1$ y‘ enak 1 ne glede na vrednosti drugih atributov. Za $e \neq 1$ je verjetnost $y = 1$ večja, če sta vrednosti atributov a in b enaki. Npr., če je $b = 3$, ima razred 1 večjo verjetnost, ko je $a = 3$, kot pri $a = 1$ ali $a = 2$. Med vplivoma slednjih na razred pravzaprav ni nobene razlike, zato $1 = 2 < 3$.

Zaradi jasnosti smo model za y_{Qe} prepisali v množico pravil:

y_{Qe}:

$$
\begin{aligned}
b = 1 \ \wedge \ a = 1 \quad &\Rightarrow \quad 1 = 2 = 3 = 4 \\
b = 2 \ \wedge \ a = 2 \quad &\Rightarrow \quad 1 = 2 = 3 = 4 \\
b = 3 \ \wedge \ a = 3 \quad &\Rightarrow \quad 1 = 2 = 3 = 4 \\
b = 1 \ \wedge \ a \neq 1 \quad &\Rightarrow \quad 2 = 3 = 4 < 1 \\
b = 2 \ \wedge \ a \neq 2 \quad &\Rightarrow \quad 2 = 3 = 4 < 1 \\
b = 3 \ \wedge \ a \neq 3 \quad &\Rightarrow \quad 2 = 3 = 4 < 1
\end{aligned}
$$

Pravila opisujejo pravzaprav isto kot kvalitativno drevo za y_{Qa} na sliki 3.3, le gledano z vidika atributa e: če velja $a = b$ (prva tri pravila), je vrednost atributa e nepomembna. Sicer (če $a \neq b$) velja, da vrednost $e = 1$ poveča verjetnost ciljnega razreda, medtem ko to ne velja za ostale vrednosti atributa e. Verjetnost ciljnega razreda pri e 2, 3 in 4 je enaka.

3.3.2 Monks 3

Razred y je v Monks 3 definiran kot $y := (e = 3 \wedge d = 1) \vee (e \neq 4 \wedge b \neq 3)$, dodano pa je tudi nekaj šuma. Ponovno izberemo ciljni razred $y = 1$.

Kot v Monks 1, Qube pravilno prepozna pomembne atribute (a, c in f). Kvalitativni model y_{Qb} je skladen z definicijo originalnega koncepta.

y_{Qb}:

$e = 4 \quad \Rightarrow \quad 1 = 2 = 3$

sicer: $e = 3 \wedge d = 1 \quad \Rightarrow \quad 1 = 2 = 3$

sicer: $3 < 1 = 2$

Pri $e = 4$ je vpliv vseh vrednosti atributa b enak, ker $e = 4$ sam po sebi pomeni $y = 0$, ne glede na vrednost b. Prav tako vrednost b ni pomembna pri $(e = 3 \wedge d = 1)$, ker vedno velja $y = 1$. Sicer ima b pri vrednostih 1 in 2 večjo verjetnost za $y = 1$ kot $b = 3$.

Model za y_{Qd} ni povsem pravilen.

y_{Qd}:

$e = 3 \wedge b = 3 \quad \Rightarrow \quad 2 = 3 < 1$

$e \neq 3 \quad \Rightarrow \quad 1 = 2 = 3$

$e = 3 \wedge b \neq 3 \wedge a = 2 \wedge c = 1 \quad \Rightarrow \quad 3 < 1 = 2$

sicer: $1 = 2 = 3$

Prvo pravilo pravi, da je verjetnost za $y = 1$ večja pri $d = 1$ kot za preostali vrednosti d, če sta b in e enaka 3. To je skladno z originalnim konceptom, kot je tudi drugo pravilo, ki pravi, da so vse vrednosti d enakovredne, če $e \neq 3$. Tretje pravilo ni skladno z originalnim konceptom in ga pripisujemo vplivu šuma v podatkih.

Pravilo za y_{Qe} pravilno ugotovi, da vrednosti e 1, 2 ali 3 povečajo verjetnost ciljnega razreda $y = 1$ v primerjavi z $e = 4$. Prav tako pri $b = 3$ in $d = 1$ velja, da z $e = 3$ povečamo verjetnost $y = 1$ glede na ostale vrednosti e. Če je $b = 3$ in $d \neq 1$, vrednost e ni pomembna.

y_{Qe}:

$b \neq 3 \quad \Rightarrow \quad 4 < 1 = 2 = 3$

$b = 3 \ \wedge \ d = 1 \quad \Rightarrow \quad 1 = 2 = 4 < 3$

$b = 3 \ \wedge \ d \neq 1 \quad \Rightarrow \quad 1 = 2 = 3 = 4$

3.3.3 Titanic

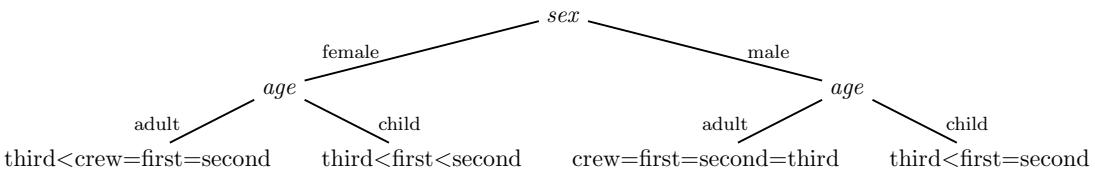
Podatki Titanic vsebujejo 2201 primer, ki ga opisujejo trije atributi: *status (first, second, third, crew)*, *age (child, adult)* and *sex (male, female)*. Razred je *survived (yes, no)*. V tem poskusu izberemo ciljni razred *survived = yes*.

Najprej opazujemo verjetnost preživetja glede na status (slika 3.4a). Verjetnost preživetja je za odrasle moške približno enaka (med 10 in 30%, kar pri

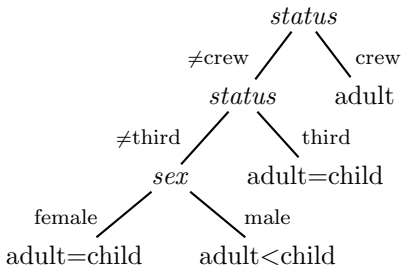
nastavljenem pragu za združevanje atributnih vrednosti, 0.2, pomeni enako). Za ostale je verjetnost preživetja najmanjša v tretjem razredu, ter večja, a enaka v ostalih, razen pri deklicah, kjer je bila verjetnost preživetja v drugem razredu večja kot v prvem.

Slika 3.4b prikazuje kvalitativni model za odvisnost med preživetjem in starostjo. Med posadko ni bilo otrok, zato je v desnem listu korena samo vrednost *adult*. V tretjem razredu je verjetnost preživetja otrok in odraslih podobna. Starost na verjetnost preživetja vpliva pravzaprav samo pri moških v prvem in drugem razredu, kjer so imeli dečki večjo verjetnost preživetja kot odrasli moški.

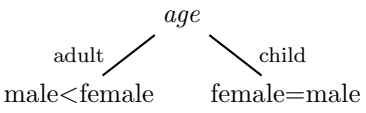
Podobno pove slika 3.4c. Spol pri otrokih ima enak vpliv, pri odraslih pa so imele ženske večjo verjetnost preživetja kot moški.



(a) $\Delta survived_status$



(b) $\Delta survived_age$



(c) $\Delta survived_sex$

Slika 3.4: Kvalitativni modeli za podatke Titanic.

3.4 Študija primera: bakterijske okužbe pri starostnikih

Starostniki, osebe stare nad 65 let, predstavljajo v razvitem svetu najhitreje rastoč segment prebivalstva [Yoshikawa, 2000]. Delež starostnikov v Sloveniji se stalno povečuje in je leta 2009 že presegel 16% [urad RS, 2010]. S staranjem

prebivalstva narašča delež bolnikov s pridruženimi kroničnimi boleznimi, ter starostnikov, ki so v oskrbi v domovih za kronično nego. Okužbe, kljub velikemu napredku v zdravljenju infekcijskih bolezni, še vedno ostajajo eden glavnih vzrokov smrti pri starostnikih. V primerjavi z mlajšo populacijo pri njih opažamo nekaj posebnosti v poteku okužbe. Večje tveganje za nastanek hude bakterijske okužbe predstavljajo motnje v delovanju imunskega sistema [Ben-Yehuda and Weksler, 1992], nepokretnost, pridružene kronične bolezni in bivanje v ustanovah za kronično nego. Pri starostnikih se okužba večkrat pokaže z neznačilnimi znaki. Pogosta je odsotnost visoke vročine [Marco et al., 1995; Castle et al., 1991; Gleckman and Hibert, 1982; Mellors et al., 1987], odsotnost kašlja ob okužbi spodnjih dihal, akuten nastanek zmedenosti ali onemoglosti [Rockwood, 1989]. To velikokrat vodi v zamudo pri postavitvi diagnoze. Učinkovito protimikrobno zdravljenje se prične (pre)pozno, kar poveča tveganje za smrt [Fontanarosa et al., 1992; Pfitzenmeyer et al., 1995]. Večanje deleža starostnikov v družbi predstavlja tudi ekonomski problem, saj se stroški zdravljenja z večanjem populacije starostnikov naglo povečujejo.

3.4.1 Podatki

Prospektivna študija je potekala na Kliniki za infekcijske bolezni in vročinska stanja Univerzitetnega kliničnega centra v Ljubljani, od 1.6.2004 do 1.6.2005. Vključevala je le bolnike z vrednostjo C-reaktivnega proteina v serumu $60mg/l$ ali več, kar je nakazovalo bakterijsko etiologijo okužbe. Bolnike so sledili 30 dni od prvega pregleda oz. do smrtnega izhoda zaradi te epizode okužbe. Upoštevali so klinične in laboratorijske pokazatelje zabeležene ob prvem pregledu v urgentni ambulanti. Naša študija obsega podatke za 602 bolnika, stara 65 let ali več, ki sta bila pregledana v urgentni ambulanti te klinike. Podatki imajo dvojiški razred SMRT (DA, NE) z naslednjo porazdelitvijo: *SMRT=DA*: $77/602 = 12.8\%$ in *SMRT=NE*: $525/602 = 87.2\%$.

Podatke opisuje 32 diskretnih atributov (tabela 3.1), ki so jih definirali zdravniki s Klinike za infekcijske bolezni in vročinska stanja, ki so vodili študijo. Podatki obsegajo spol pacienta, njegovo starost, informacijo o številu pridruženih bolezni, za vsako od njih pa še podrobnejšo informacijo o njeni prisotnosti (sladkorna bolezen, prizadeto srce, prizadeta ledvica, jetra, pluča, imunski sistem, centralni živčni sistem) ter še naslednje atribute: pokretnost pacienta, kontinenca, prisotnost dekubitusa, urinskega katetra ter vsadka (npr. srčna

atribut	vrednosti atributa
SPOL	M, Z
STAROST	A, B, C
OSKRBOVANEC	DA, NE
PRIDRUZENE	ENA, NIC, VEC
SLADKORNA	DA, NE
SRCE	DA, NE
LEDVICA	DA, NE
JETRA	DA, NE
PLJUCA	DA, NE
IMUNOST	DA, NE
CZS	DA, NE
POKRETNOST	DA, NE
KONTINENCA	DA, NE
DEKUBITUS	DA, NE
URINSKIKAT	DA, NE
VSADEK	DA, NE
TT	<=37.80, >37.80
FRDIHANJA	<=10.00, (10.00, 20.00], >20.00
SAT	<=90.00, >90.00
FRSRCA	<=60.00, (60.00, 100.00], >100.00
RR	<=90.00, >90.00
SPREMENJENA ZAVEST	DA, NE
NEZAVEST	DA, NE
OSLABELOST	DA, NE
LEVKOCITI	<=4.00, (4.00, 10.00], >10.00
PALICASTI	<=10.00, >10.00
TROMBOCITI	<=100.00, >100.00
KREATININ	<=90.00, >90.00
SECNINA	<=6.00, >6.00
GLU	<=4.00, (4.00, 7.50], >7.50
NA	<=135.00, (135.00, 145.00], >145.00
KLINICNOOK	DIHAL, DRUGO, GEA, MEHKIHTKIV, SECIL

Tabela 3.1: Atributi v domeni *bakterijske okužbe pri starostnikih*.

zaklopka, umetni kolk ipd.). Sledijo podatki o klinični sliki pacienta ob pregledu v urgentni ambulanti: telesna temperatura, frekvenca dihanja, saturacija, frekvenca srca, krvni pritisk (RR), podatki o zavesti pacienta ter njegovi oslabelosti, krvna slika (levkociti, paličasti in trombociti), kreatinin, sečnina, krvni sladkor, natrij ter podatek o klinično določeni okužbi.

3.4.2 Učenje kvalitativnih odvisnosti

Za ciljni razred izberemo *SMRT=DA*. Zanima nas torej odvisnost med smrtnostjo in posameznimi atributi oziroma, kako vrednosti posameznih atributov vplivajo na verjetnost za smrt kot posledico bakterijske okužbe. Vpliv posameznih atributov ugotavljamo z računanjem PKS z algoritmom Qube. Kvalitativne modele na podlagi izračunanih PKS zgradimo z algoritmom C4.5. Rezultate prikazuje tabela 3.2.

Dobljene modele sta komentirala in ocenila specialista infektologa. Vsi modeli so popolnoma smiselni, razen modeli za *spol*, *srce*, *ledvica*, *urinskikat*, *vsadek*, *frsrca* in *klinicnook*, ki so smiselni le delno. Modeli za *spol*, *srce*,*ledvica*, *vsadek* bi utegnili biti smiselni, a bi bilo potrebno preveriti druge dejavnike, ki ob pregledu pacientov niso bili zabeleženi. Pri ostalih delno smiselnih modelih velja, da so nekatera poddrevesa smiselna, nekatera ne. Specialista sta menila, da nekateri deli modelov predstavljajo možno odkritje novega znanja, česar pa spet ni mogoče potrditi brez nadaljnih medicinskih raziskav.

Ostalih 25 modelov je v celoti smiselnih in dobrih. Po mnenju specialistov so modeli jasni, točni in enostavni. Zaradi preobsežnosti podamo razlago samo nekaterih modelov, ki so razumljivi tudi laiku. Model za STAROST pove, da imajo mlajši pacienti s prizadetim imunskim sistemom slabše možnosti preživetja kot starejši pacienti, kar specialisti razložijo s tem, da se pri starejših telo počasneje odziva. Model za PRIDRUZENE trdi, da imajo pacienti z več pridruženimi boleznimi slabše možnosti preživetja kot tisti, ki imajo eno (pogosto je to sladkorna bolezen) ali so celo brez pridruženih bolezni. Kar smo videli že v modelu za STAROST, potrdi tudi model za IMUNOST – prizadet imunski sistem poveča verjetnost za smrt. Model za SAT pove, da nizka saturacija (prisotnost kisika v krvi) značilno poveča verjetnost za smrt (ponavadi pri bolnikih z bakterijskimi okužbami dihal). Tudi oba modela, ki govorita o zavesti pacienta, trdita, da nezavest oziroma spremenjena zavest (kot posledica okužbe) povečata verjetnost za smrt. V povezavi z zavestjo je zelo zanimiv model za

telesno temperaturo (TT), ki pove, da ob spremenjeni zavesti nižja telesna temperatura poveča verjetnost za smrt, medtem ko se pri nespremenjeni zavesti (in opazovanih tipih okužb – KLINICNOOK $\neq$ DRUGO) zgodi ravno obratno – višja temperatura pomeni večjo verjetnost za smrt. Razlaga zdravnika specialista je, da spremenjena zavest pacienta, kot posledica bakterijske okužbe, priča o zelo slabem stanju, kar v skupini starostnikov pomeni, da telo na okužbo ni več sposobno odreagirati s povečano telesno temperaturo, kot je to običajno. Prav tako je zanimiv model za VSADEK, ki trdi, da imajo pacienti s prizadetimi pljuči in bakterijsko okužbo slabše možnosti preživetja ob prisotnosti vsadka, kar zdravniki razlagajo z znanim dejstvom, da se ob vsadkih zmanjša zmožnost telesa za prepoznavanje vnetja, kar lahko privede do hujših okužb. Prav to je razlog, da je VSADEK, kot atribut, vključen v pričujočo študijo. Poudarimo, da je ta model smiseln le deloma, saj podatki ne vsebujejo dovolj informacij, da bi lahko preverili tudi pravilo za neopazovane tipe okužb.

Zaključimo lahko, da modeli potrjujejo obstoječe znanje, hkrati pa odpirajo možnosti za nadaljne raziskave ob določenih drugačnih diskretizacijah atributov. Uporabljene mejne vrednosti so namreč standardne mejne vrednosti za diskretizacijo zveznih atributov, ki veljajo za mlajšo populacijo pacientov, populacija starostnikov pa bi tu utegnila imeti drugačne lastnosti in posledično tudi druge mejne vrednosti. Omenjena analiza predstavlja priložnost za nadaljnje delo.

Tabela 3.2: Kvalitativni modeli v domeni *bakterijske okužbe pri starostnikih*.

atribut	kvalitativni model
SPOL	M=Z
STAROST	IMUNOST = DA: B=C < A IMUNOST = NE: A=B=C
OSKRBOVANEC	PRIDRUZENE = NIC: NE < DA PRIDRUZENE $\neq$ NIC: KLINICNOOK = DRUGO: NE < DA KLINICNOOK $\neq$ DRUGO: DA=NE
PRIDRUZENE	ENA=NIC < VEC
SLADKORNA	IMUNOST = DA: DA < NE IMUNOST = NE: RR $\leq$ 90: DA < NE RR > 90: DA=NE

se nadaljuje...

Tabela 3.2 – nadaljevanje

SRCE	DA=NE
LEDVICA	FRDIHANJA $\in [10, 20]$: SAT $\leq$ 90: NE < DA SAT > 90: DA=NE FRDIHANJA $\notin [10, 20]$: NE < DA
JETRA	NE<DA
PLJUCA	SAT $\leq$ 90: DA < NE SAT >90: SPREMENJENA ZAVEST = DA: DA < NE SPREMENJENA ZAVEST = NE: DA=NE
IMUNOST	NE<DA
CZS	KLINICNOOK = DIHAL: NE < DA KLINICNOOK $\neq$ DIHAL: FRSRCA > 100: NE < DA FRSRCA $\leq$ 100: IMUNOST = DA: NE < DA IMUNOST = NE: DA=NE
POKRETNOST	PALICASTI >10: DA < NE PALICASTI $\leq$ 10: NA > 145: DA < NE NA $\leq$ 145: RR > 90: DA=NE RR $\leq$ 90: DA < NE
KONTINENCA	DA < NE
DEKUBITUS	PALICASTI >10: NE < DA PALICASTI $\leq$ 10: KLINICNOOK = DRUGO: NE < DA KLINICNOOK $\neq$ DRUGO: DA=NE
URINSKIKAT	PRIDRUZENE = NIC: NE < DA PRIDRUZENE $\neq$ NIC: TROMBOCITI > 100: DA < NE DEKUBITUS = DA: NE < DA DEKUBITUS = NE: DA=NE TROMBOCITI $\leq$ 100: DA < NE
VSADEK	KLINICNOOK = DRUGO: NE < DA KLINICNOOK $\neq$ DRUGO: PLJUCA = DA: NE < DA PLJUCA = NE: DA=NE
TT	SPREMENJENA ZAVEST = DA: [>37.8] < [$\leq$ 37.8] SPREMENJENA ZAVEST = NE:

se nadaljuje...

Tabela 3.2 – nadaljevanje

	KLINICNOOK = DRUGO: [>37.8] < [≤ 37.8] KLINICNOOK ≠ DRUGO: [≤ 37.8] < [>37.8]
FRDIHANJA	[(10.00, 20.00] = >20.00] < [<=10.00]
SAT	>90.00 < <=90.00
FRSRCA	PALICASTI > 10: $(60, 100]$ < [> 100] < [≤ 60] PALICASTI ≤ 10: SAT > 90: $(60, 100]$ < [≤ 60]=[> 100] SAT ≤ 90: $(60, 100]$=[> 100] < [≤ 60]
RR	>90.00 < <=90.00
SPREMENJENA ZAVEST	NE < DA
NEZAVEST	NE < DA
OSLABELOST	NE < DA
LEVKOCITI	NA > 145: $(4, 10]$=[> 10] < [≤ 4]
PALICASTI	[<=10] < [>10]
TROMBOCITI	[>100] < [<=100]
KREATININ	URINSKIKAT = DA: [≤ 90] < [> 90] URINSKIKAT = NE: [≤ 90]=[> 90]
SECNINA	DEKUBITUS = DA: [≤ 6] < [> 6] DEKUBITUS = NE: IMUNOST = DA: [≤ 6] < [> 6] IMUNOST = NE: [≤ 6]=[> 6]
GLU	NA ≤ 135: $(4, 7.5]$=[> 7.5] < [≤ 4] NA > 135: SECNINA ≤ 6: $(4, 7.5]$=[> 7.5] < [≤ 4] SECNINA > 6: KLINICNOOK = DRUGO: $(4, 7.5]$=[> 7.5] < [≤ 4] KLINICNOOK ≠ DRUGO: $(4, 7.5]$=[> 7.5]=[≤ 4]
NA	[≤ 135.00]=$(135, 145]$ < [> 145]
KLINICNOOK	FRDIHANJA ∉ $(10, 20]$: DIHAL=DRUGO=GEA=SECIL < MEHKIHTKIV FRDIHANJA ∈ $(10, 20]$: SPREMENJENA ZAVEST = DA: SECIL < DIHAL=DRUGO=GEA=MEHKIHTKIV SPREMENJENA ZAVEST = NE: VSADEK = NE: DIHAL=DRUGO=GEA=MEHKIHTKIV=SECIL VSADEK = DA: DIHAL=MEHKIHTKIV=SECIL < DRUGO

Poglavje 4

Pregled področja

Vsebina te disertacije sodi na področje umetne inteligence, v presek strojnega učenja in kvalitativnega sklepanja. Strojno učenje kvalitativnih odvisnosti iz podatkov je bilo tradicionalno omejeno na učenje kvalitativnih diferencialnih enačb. V tem delu se ukvarjamo z alternativnimi pristopi. Na področju kvalitativnega sklepanja najdemo sorodno delo le na področju modeliranja regresijkih podatkov, kjer je Padéju najbližji algoritem Quin. Podrobno primerjavo med obema prihranimo za konec tega poglavja. Čeprav je Qube algoritem za učenje kvalitativnih modelov, sorodno delo najdemo le na področju strojnega učenja. Največ skupnega ima z algoritmi za računanje lokalnih razdalj in nekaterimi pristopi, ki skušajo odpraviti naivnost naivnega Bayesa.

V prvem razdelku tega poglavja podamo pregled področja kvalitativnega sklepanja, kamor umestimo Padé. Nadaljujemo s pregledom del, ki se navezujejo na algoritem Qube. Zadnji razdelek je namenjen primerjavi z algoritmom Quin.

4.1 Sorodna dela s področja kvalitativnega sklepanja

Področje kvalitativnega sklepanja se je začelo razvijati zunaj področja umetne inteligence. Jeffries in May [Jeffries, 1974; May, 1973] sta v zgodnjih sedemdesetih uvedla pojem kvalitativne stabilnosti v ekosistemih. Kvalitativno sta obravnavala ekosisteme, pri čemer sta Lotka-Voltera modele (plen-plenilec) za dve vrsti posplošila na več vrst. Ugotovila sta, da je za modeliranje dinamike v ekosistemu dovolj, če opazujeta kvalitativne odvisnosti med posameznimi vr-

stami. Podobno je na področju ekonomije proučeval Samuelson [Samuelson, 1983].

Na področju umetne inteligence se je kvalitativno sklepanje začelo uveljavljati z delom de Kleera [de Kleer, 1979, 1977], ki se ukvarja s kombinacijo kvalitativnega in kvantitativnega reševanja problemov v mehaniki. Med temeljna dela področja lahko štejemo Teorijo kvalitativnih procesov (*angl. Qualitative Process Theory, QPT*) [Forbus, 1984], ki je ena prvih formalizacij kvalitativne fizike. Natančno razdela obravnavano fizikalno domeno in zgradi ontologijo kvalitativnih odvisnosti med procesi in objekti v dani domeni. Področje kvalitativnega sklepanja se je od svojih začetkov pretežno ukvarjalo z modeliranjem fizikalnih problemov, zato je znano tudi pod imenom kvalitativna fizika [de Kleer and Brown, 1984]. Namen kvalitativne fizike je poenostaviti klasično fiziko do meje, ko bo še uporabna in razumljiva, a ne bo vsebovala zveznih količin in diferencialnih enačb; temeljila naj bi na vzročnih povezavah med fizikalnimi zakonitostmi in služila kot podlaga zdravorazumskemu sklepanju. V tem kontekstu so se najbolj uveljavile kvalitativne diferencialne enačbe [de Kleer and Brown, 1984] (*angl. Qualitative Differential Equations, QDE*). Kvalitativne diferencialne enačbe so kvalitativna abstrakcija navadnih diferencialnih enačb. Področje je svoja zlata leta doseglo z razvojem algoritma za kvalitativno simulacijo QSIM [Kuipers, 1986]. QSIM je namenjen kvalitativnemu sklepanju v dinamičnih sistemih, kjer rešuje sisteme QDE. Nekoliko kasneje je k matematičnim temeljem kvalitativnega sklepanja prispeval Kalagnanam [Kalagnanam et al., 1991; Kalagnanam, 1992; Kalagnanam and Simon, 1992], ki je razvil matematična orodja za sklepanje o kvalitativnih lastnostih modelov.

Kvalitativna analiza sistemov oziroma učenje kvalitativnih modelov iz podatkov je veja kvalitativnega sklepanja, ki se ukvarja z gradnjo modelov z namenom uporabe v procesih kvalitativnega sklepanja. Tu se področje kvalitativnega sklepanja sreča s področjem strojnega učenja. Večino dela na tem področju je povezanega z učenjem QDE. Eden od načinov je, da na kvalitativen način predstavimo obnašanje numeričnega sistema ter poiščemo QDE omejitve, ki zadoščajo temu kvalitativnemu obnašanju. Množica takih omejitev tvori kvalitativni model. Na ta način delujejo algoritmi GENMODEL [Coiera, 1989; Hau and Coiera, 1997], MISQ [Ramachandran et al., 1994; Richards et al., 1992] in QSI [Say and Kuru, 1996]. Podoben pristop lahko uberemo z uporabo induktivnega logičnega programiranja (ILP). Pri danem QSIM modelu ter pozitivnih

in negativnih primerih obnašanja sistema, se ILP nauči kvalitativnega modela. Primeri učenja kvalitativnih modelov z ILP so [Bratko et al., 1991; Coghill and King, 2002; Coghill et al., 2008].

Džeroski in Todorovski sta razvila nekaj novih pristopov k odkrivanju dinamike sistemov: QMN [Džeroski and Todorovski, 1995] gradi modele v obliki kvalitativnih diferencialnih enačb, medtem ko se LAGRANGE [Džeroski and Todorovski, 1993] in LAGRAMGE [Todorovski, 2003] učita modelov v obliki diferencialnih enačb. Algoritem PADLES pristop razširi na učenje parcialnih diferencialnih enačb [Todorovski et al., 2000].

Omenimo še algoritem LYQUID [Gerçeker and Say, 2006], ki s prilagajanjem polinomov numeričnim podatkom išče kvalitativne modele v dinamičnih sistemih. Čeprav avtorja tega ne omenjata, mislimo, da bi LYQUID lahko obravnaval tudi statične sisteme.

Kvalitativni modeli dinamičnih sistemov ne temeljijo nujno na QDE. Primer tega je kvalitativni model srca, KARDIO [Bratko et al., 1989], ki za opis uporablja logiko prvega reda. Algoritem QuMas [Mozetič, 1987a,b] je zmožen učenja takih modelov iz primerov kvalitativnega obnašanja v času, pri čemer uporablja ILP.

Omenjene metode učenja dinamičnih kvalitativnih modelov so v omejenem obsegu zmožne tudi učenja statičnih modelov. Algoritem QUIN [Šuc and Bratko, 2001; Šuc, 2003; Bratko and Šuc, 2003] je namenjen učenju v statičnih sistemih in je zaradi tega najpomembnejše sorodno delo algoritmoma Padé in Qube. Quin gradi kvalitativna drevesa, ki so podobna klasifikacijskim drevesom. Razlikujejo se v listih, ki pri kvalitativnih drevesih vsebujejo kvalitativne omejitve. Quin podrobneje opišemo v razdelku 4.3, kjer ga primerjamo s Padéjem.

Padé se od naštetih algoritmov razlikuje po tem, da je edini algoritem, ki lokalno računa parcialne odvode na numeričnih podatkih. Druga pomembna razlika je, da je Padé predprocesor, medtem ko ostali algoritmi gradijo modele. Danim podatkom Padé le pripiše nove atribute, ki jih lahko kasneje uporabimo v klasifikaciji ali regresiji.

V matematiki se z numeričnim računanjem odvodov ukvarja področje numeričnih metod. Tako izračunani odvodi so seveda prav tako primerni za uporabo v kvalitativnem modeliranju. Najpogosteje uporabljena metoda za numerično odvajanje je metoda končnih diferenc na mrežah točk. V primerih, ko točke

niso podane na mreži, si pomagamo z interpolacijskimi polinomi, ponavadi kubičnimi zlepki. Matematične metode niso primerne za podatke, ki vsebujejo diskretne spremenljivke.

4.2 Sorodna dela s področja strojnega učenja

Sorodnega dela s področja učenja kvalitativnih modelov v klasifikaciji nismo zasledili. V tem razdelku zato omenimo dela s področja strojnega učenja, ki se navezujejo na pristope, ki smo jih uporabili v algoritmu Qube.

Pomemben del našega algoritma se nanaša na odpravo predpostavke o pogojni neodvisnosti naivnega Bayesa. V klasifikaciji obstajajo številni postopki, ki se ukvarjajo s tem problemom. Kononenko je predlagal *delno-naivni Bayes* [Kononenko, 1991], Langley in Sage [Langley and Sage, 1994] pa *izbirni Bayesov klasifikator* - varianto naivne metode, ki pri napovedovanju uporablja le podmnožico atributov. Oba pristopa povečata klasifikacijsko točnost naivnega Bayesovega klasifikatorja. Friedman in Goldszmidt sta predlagala algoritem *tree augmented naive Bayes (TAN)* [Friedman and Goldszmidt, 1996], ki v predstavitvi z Bayesovsko mrežo atributom dovoljuje, da se povezujejo tudi prek različnih nivojev (atribut je lahko starš drugemu atributu), za razliko od naivnega Bayesa, ki ima v taki predstavitvi vse atribute na istem nivoju.

Lokalno uteženi naivni Bayes (*Locally Weighted Naive Bayes, LWNB*) [Frank et al., 2003] je leni algoritem, ki se ob klasifikaciji primera nauči lokalnega modela. LWNB prav tako lokalno omili predpostavko o pogojni neodvisnosti. Okolico primera, ki ga klasificira izbere s pomočjo metode k najbližjih sosedov, k-NN. Podoben algoritem je LBR (*angl. Lazy Bayesian Rules* [Zheng et al., 1999]. LBR za iskanje bližnje okolice ne uporablja globalne metrike ampak za vsak testni primer s požrešno metodo zgradi Bayesovsko pravilo tako, da se pogojni del pravila ujema s testnim primerom.

Pomembna razlika med naštetimi algoritmi in algoritmom Qube je v tem, da omenjene sorodne metode optimizirajo klasifikacijsko točnost, medtem ko Qube ocenjuje vpliv izbranega atributa na verjetnost izbranega ciljnega razreda.

Izbirni nomogrami temeljijo na nomogramih naivnega Bayesa [Možina et al., 2004]. V splošnem je nomogram [Wikipedia, 2010] grafična predstavitev matematične funkcije, ki omogoča grafično analizo obnašanja funkcije pri različnih naborih spremenljivk. Uporablja se tudi za ugotavljanje odvisnosti med izbrano

neodvisno in odvisno spremenljivko, pri čemer fiksiramo vrednosti ostalih neodvisnih spremenljivk. Slednje je natančno to, kar Qube potrebuje za svoj namen.

4.3 Primerjava algoritmov Quin in Padé

Algoritmu Padé je najbolj podoben algoritem Quin [Šuc and Bratko, 2001; Šuc, 2003; Bratko and Šuc, 2003]. Quin je namenjen kvalitativni analizi statičnih sistemov. Za dane numerične podatke zgradi kvalitativno drevo, ki je podobno klasifikacijskemu drevesu, razlikuje se le v listih drevesa. V listih ima kvalitativno drevo t.i. kvalitativne omejitve (npr. $z = M^{+-}(x, y)$, kar pomeni, da z narašča z x, in pada z y ter ni odvisen od drugih spremenljivk). Pri gradnji dreves si Quin pomaga z računanjem vektorjev kvalitativnih sprememb med pari učnih primerov ter rekurzivno deli atributni prostor na območja z enakimi kvalitativnimi lastnostmi. Padé in Quin sta si na prvi pogled zelo podobna predvsem zato, ker rešujeta isti osnovni problem, sicer pa se algoritma bistveno razlikujeta. Podrobni analizi razlik med njima namenjamo ta razdelek.

Algoritma bomo primerjali na pol-realni domeni biljard in umetnih domenah iz poglavja 2.4. Začnemo s funkcijo $x^2 - y^2$, ki služi kot pogosto uporabljen primer za študij učenja kvalitativnih modelov iz podatkov [Šuc and Bratko, 2001; Šuc, 2003]. Domeni dodajamo naključne atribute r_0, r_1 in r_2, ki ne vplivajo na razred. Nadaljujemo s funkcijo $x^3 - y$, ki je prejšnji zelo podobna in je celo enostavnejša, ker je globalno monotona. Je primer funkcije, v kateri imajo spremenljivke različno močne vplive na odvisno spremenljivko. Sinusna funkcija je primer funkcije, katere obnašanje je tako kompleksno, da drevo kot struktura ni primerno za modeliranje tega obnašanja. Definicijska območja atributov so, kot v poskusih doslej, $[-10, 10]$. Podatke smo pridobili z enakomernim naključnim vzorčenjem definicijskega območja v 1000 točkah. Zaključimo z biljardom, ki smo ga s Padéjem modelirali v poglavju 2.7.

Ker so primerjalne analize algoritmov odvisne od vhodnih podatkov, se lahko pojavi dvom o pristranskosti pri izbiri domen. Uporabljene umetne funkcije samo predstavljajo tipe funkcij z lastnostmi, ki vplivajo na učenje kvalitativnih modelov iz podatkov.

Zaradi bistvenih razlik med algoritmoma se zaplete že pri izbiri metode za ocenjevanje kvalitete enega in drugega. Točnost pravilno izračunanega kvalitativnega parcialnega odvoda, kot smo jo merili pri Padéju ni primerna za Quin.

Ta namreč za razliko od Padéja v enem modelu obravnava vse atribute hkrati. Kdaj je torej Quin točen? Če pravilno napove kvalitativno obnašanje za vse atribute ali samo za tiste, ki jih omeni v kvalitativni omejitvi? Če Quin atributa ne omeni v kvalitativni omejitvi, to po njeni definiciji pomeni, da razred od njega ni odvisen. Šuc je zato poleg točnosti za merjenje uspešnosti Quina uvedel še mero dvoumnosti. Pristranskosti v izogib bomo ocenjevali smiselnost modelov. Oba algoritma imata skupni cilj – graditi razumljive kvalitativne modele. Za vse testne domene pravilne kvalitativne modele poznamo, zato primerjava ne bo težka.

Morda se zdi, da sta si algoritma bolj podobna, če Padé uporabimo v kombinaciji s klasifikacijskim drevesom. Toda razlika ni nič manj izrazita – Quin namreč hkrati gradi drevo in računa vektorje kvalitativnih sprememb, kar pomeni, da za slednje 'izgublja' učne primere, ko veji drevo. Po drugi strani Padé izračuna kvalitativne parcialne odvode na vseh podatkih. Klasifikacijskemu drevesu nato preostane samo še posploševanje tega, kar je postoril Padé kot predprocesor. Učenje kvalitativnih modelov s Padéjem poteka torej v dveh fazah, medtem ko Quin vse naredi naenkrat.

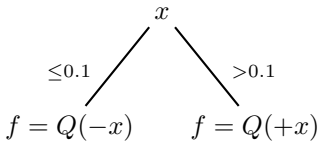
Funkcija $x^2 - y^2$

Pravi model, ki opisuje kvalitativni odvisnosti razreda f od atributov x in y lahko povzamemo z naslednjimi pravili:

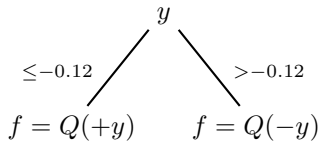
- $x > 0$: x narašča $\Rightarrow$ f narašča
- $x < 0$: x narašča $\Rightarrow$ f pada
- $y > 0$: y narašča $\Rightarrow$ f pada
- $y < 0$: y narašča $\Rightarrow$ f narašča

Quin se hkrati uči kvalitativnih odvisnosti za vse atribute in jih povzame v istem modelu, Padé pa parcialna odvoda obravnava ločeno in tudi kvalitativna modela zgradimo za vsako odvisnost posebej. Kvalitativna drevesa zgrajena s Quinom ter s C4.5 na parcialnih odvodih, izračunanih s Padéjem prikazuje slika 4.1. Quin smo uporabili z različnimi nastavitvami parametrov. Drevo na sliki 4.1c je zgrajeno pri privzetih nastavitvah, medtem ko je tisto na sliki 4.1d zgrajeno pri nastavitvah, kot jih uporablja Padé ($k = 20$ najbližjih sosedov) in

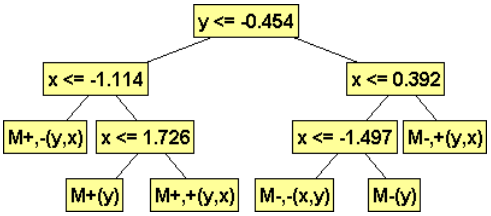
C4.5 pri gradnji drevesa (minimalno število primerov v listu drevesa je 20). V obeh Quinovih modelih opazimo, da slabše od Padéja določi vrednost vejitve po atributih x in y, ki je idealno pri 0. Quinovi drevesi imata več listov, kot bi pričakovali glede na pravi model. To pripisujemo Quinovim težavam z rezanjem kvalitativnega drevesa.



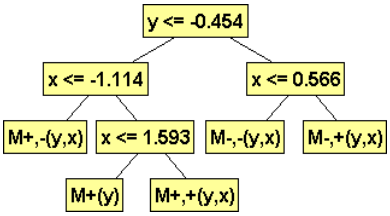
(a) Padéjevo kvalitativno drevo za odvisnost f od x.



(b) Padéjevo kvalitativno drevo za odvisnost f od y.



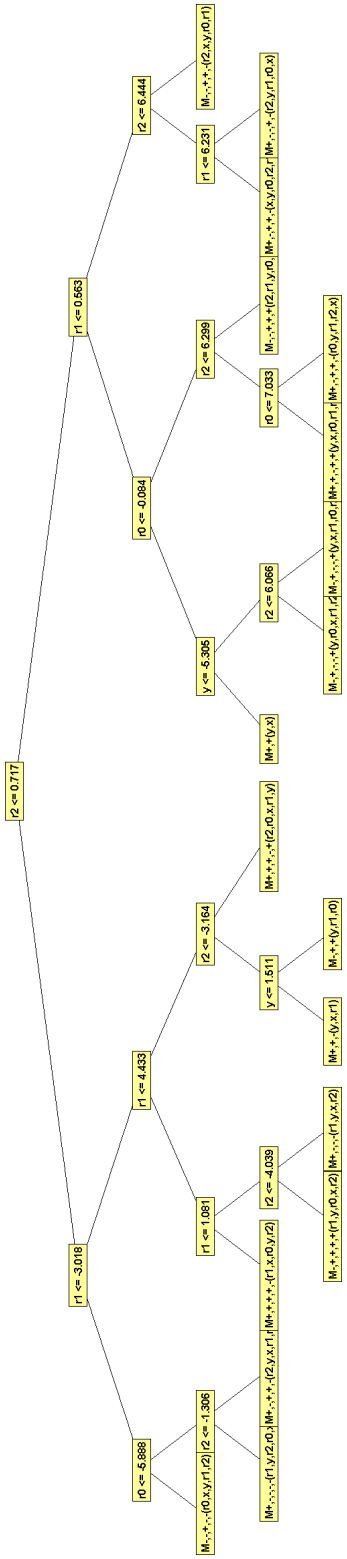
(c) Quinovo kvalitativno drevo s privzetimi nastavitvami (leva veja ustreza pogoju iz vozlišča, desna ne).



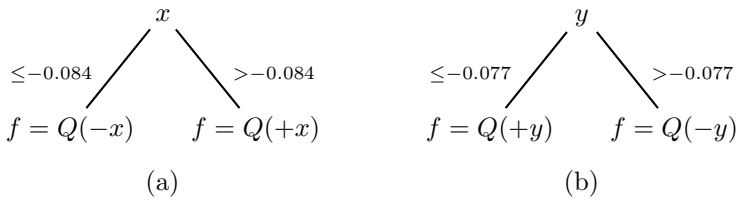
(d) Quinovo kvalitativno drevo z nastavitvami kot Padé (leva veja ustreza pogoju iz vozlišča, desna ne).

Slika 4.1: Primerjava modelov na domeni $x^2 - y^2$.

V realnih podatkih je iskani koncept le redko funkcija vseh razpoložljivih atributov, najpogosteje je funkcija le majhne podmnožice le-teh. Odkriti pravo podmnožico atributov je zato pomembna naloga algoritmov strojnega učenja, posebno tistih, ki obljubljajo razložljive modele. S tem namenom smo Padé preizkusili na domenah, kjer smo dodajali nepomembne atribute. Tu poglejmo, kako se pri tem odreže Quin. Funkciji $f(x, y) = x^2 - y^2$ dodajamo naključne atribute r_0, r_1 in r_2 in opazujemo Quinova kvalitativna drevesa. Quin se z enim dodanim naključnim atributom odreže dobro. Kvalitativno drevo je zelo podobno modeloma s slik 4.1d in 4.1c. Ko dodamo naslednji naključni atribut, r_1, Quin zaide v težave. Kvalitativno drevo nima nobenega pomena več, saj je v korenu drevesa r_1, oba naključna atributa pa se pojavita tudi v kvalitativnih omejitvah v listih. Podobno se zgodi, ko dodamo še tretji naključni atribut r_2. Padé s τ-regresijo in vzporednimi pari tudi pri treh (in več) naključnih atributih nima težav s prepoznavanjem prave podmnožice (slika 4.3).

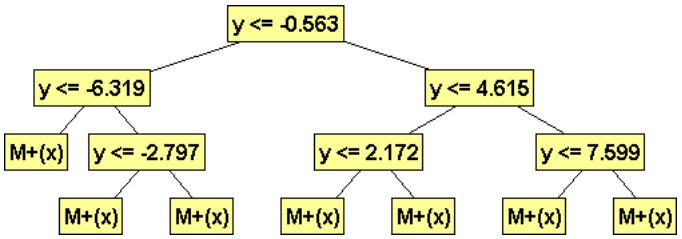


Slika 4.2: Kvalitativno drevo za $f(x, y, r_0, r_1, r_2) = x^2 - y^2$, zgrajeno s Quinom (leva veja ustreza pogoju iz vozlišča, desna ne).



(a) (b)

Slika 4.3: Kvalitativni drevesi za odvod $f(x, y, r_0, r_1, r_2) = x^2 - y^2$ po x in y. Kvalitativni parcialni odvodi so izračunani s τ-regresijo.

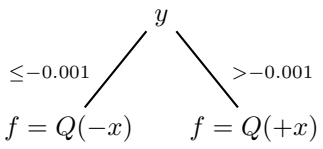


Slika 4.4: Kvalitativno drevo za $f(x, y) = x^3 - y$, zgrajeno s Quinom (leva veja ustreza pogoju iz vozlišča, desna ne).

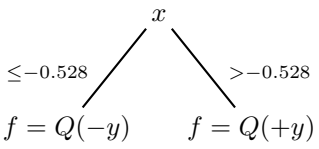
Funkcija $x^3 - y$

Padé računa kvalitativne parcialne odvode. Ker jih Quin ne, je to naslednja bistvena razlika med njima. V čem se ta razlika odraža in zakaj je pomembna, pokažemo na enostavni, globalno monotoni funkciji $f(x, y) = x^3 - y$. Njena parcialna odvoda sta $f_x = 3x^2$ in $f_y = -1$. Quin zgradi kvalitativno drevo na sliki 4.4.

Quin je pravilno opazil (drevo na sliki 4.4), da f povsod narašča z x. Drevo je za odvisnost f od x sicer pravilno, vendar ni dovolj porezano. Namesto osmih listov bi bil dovolj en sam. Odvisnosti f od y (f pada z y) Quin ni zaznal. Razlog temu je prevladujoč vpliv spremenljivke x nad y. Z drugimi besedami, že majhna sprememba x-a močno poveča f, medtem ko relativno velika sprememba y-a še vedno ne zadošča, da bi pretehtala vpliv x-a. Razloge za težavnost računanja kvalitativnih odvisnosti v takih primerih smo obravnavali v razdelku 2.5.2. Quin ne vsebuje nobenega mehanizma, ki bi mu omogočal odkriti, da f pada z y. Padé, kot smo pokazali v razdelku s poskusi, s to funkcijo nima večjih težav. Veliko večino odvodov izračuna pravilno in npr. klasifikacijskim pravilom ali drevesom ni težko zgraditi pravilnega kvalitativnega modela.

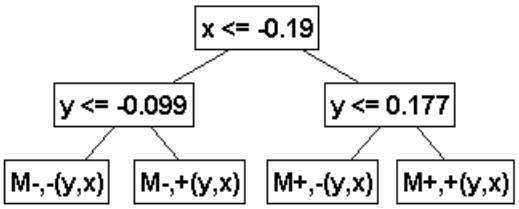


(a) Padéjevo kvalitativno drevo za odvisnost f od x.



(b) Padéjevo kvalitativno drevo za odvisnost f od y.

Slika 4.5: Kvalitativni drevesi za odvod $f(x, y) = xy$ po x in y. Kvalitativni parcialni odvodi so izračunani s τ-regresijo.



Slika 4.6: Kvalitativno drevo za $f(x, y) = xy$, zgrajeno s Quinom (leva veja ustreza pogoju iz vozlišča, desna ne).

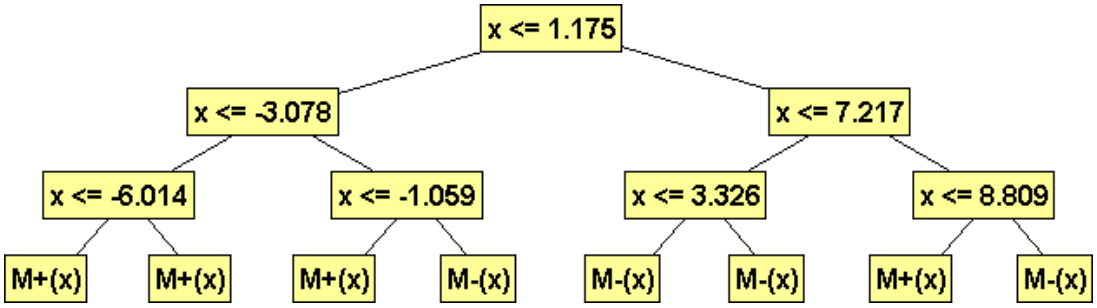
Funkcija xy

Funkcija $f(x, y) = xy$ je primer funkcije, pri kateri je parcialni odvod po eni spremenljivki odvisen od druge: $f_x = y$ in $f_y = x$. Modela za kvalitativni odvisnosti f od x in y, izračunani s Padéjem, sta prikazana na sliki (4.5). Quin zgradi kvalitativno drevo na sliki 4.6. Oba algoritma sta pravilno odkrila vse odvisnosti. Nekaj razlik je le v vrednostih vejitev, kjer je enkrat boljši Quin, drugič pa Padé.

Funkcija $\sin x \sin y$

Najbolj očitna razlika med algoritmoma je v tem, da Quin gradi kvalitativna drevesa, medtem ko Padé računa kvalitativne parcialne odvode, s čimer problem učenja kvalitativnih odvisnosti iz numeričnih podatkov prevede na gradnjo modelov iz podatkov z diskretnim razredom. Uporabnika ne omejuje na klasifikacijska drevesa, temveč mu prepušča izbiro ustreznega algoritma. Primer funkcije, kjer so klasifikacijska drevesa neustrezna, je funkcija $f(x, y) = \sin x \sin y$. Kvalitativno drevo, ki ga zgradi Quin, je na sliki 4.7.

Padé izračuna parcialne odvode in uporabniku prepusti izbiro primernega algoritma za gradnjo modela. V tem primeru je razsevni diagram najboljša



Slika 4.7: Kvalitativno drevo za $f(x, y) = \sin x \sin y$, zgrajeno s Quinom (leva veja ustreza pogoju iz vozlišča, desna ne).

izbira za predstavitev kvalitativnih odvisnosti. Za razliko od Quina lahko na podlagi Padéjevih rezultatov prikažemo razred v odvisnosti od vsakega atributa posebej (sliki 2.7a in 2.7b).

Biljard

Primerjavo algoritmov nadaljujemo na pol-realni domeni *biljard*, ki smo jo opisali v razdelku 2.7. Na tej domeni primerjamo tako model kot nekaj praktičnih lastnosti obeh algoritmov ter čas izvajanja. Podatki vsebujejo 5000 učnih primerov, 4 atribute in razred. Cilj je zgraditi kvalitativni model, ki razloži odvisnost razreda (*angle*) od atributa *azimuth*. Quin ne ponuja možnosti modeliranja kvalitativnih odvisnosti za posamezne atribute, ampak vedno opazuje vse. Računanje parcialnih odvodov Padéju praktično ne ponuja druge možnosti kot računanje kvalitativnih odvisnosti za posamezne atribute. Uporabnik lahko naknadno združi več kvalitativnih odvisnosti v en atribut, če oceni, da je to smiselno. Primerjava časov izvajanja pokaže, da je Padé hitrejši. Padéjeva τ-regresija porabi za izračun odvoda *angle* po *azimuth* 82.57 sekunde, Quin pa 14536.39 sekunde, kar je nekaj več kot 4 ure, pri čemer ponovno poudarimo, da Quin vedno izračuna kvalitativne odvisnosti za vse atribute. Model, ki smo ga zgradili s Padéjem smo komentirali že v razdelku 2.7. Quinovo kvalitativno drevo (slika 4.8) sta ocenila eksperta za biljard. Ugotovila sta, da je slabo razumljivo, ker se atributa *velocity* in *elevation* pojavljata relativno pogosto, čeprav sta popolnoma nepomembna. Še bolj so ju motile kvalitativne omejitve z večjim številom atributov, kjer je potrebno sklepati v večrazsežnem prostoru (pet dimenzij). Tako sklepanje človeku ni blizu.

```
azimuth <= 4.194
|  velocity <= 4.213
|  |  elevation <= 19.88
|  |  |  velocity <= 3.774      M+,+,-,-(azimuth,follow,elevation,velocity)
|  |  |  velocity > 3.774       M-,+,-,+(velocity,follow,elevation,azimuth)
|  |  elevation > 19.88
|  |  |  azimuth <= 0.817       M+,-,-,-(follow,azimuth,velocity,elevation)
|  |  |  azimuth > 0.817        M-,+(follow,azimuth)
|  velocity > 4.213
|  |  azimuth <= -11.95
|  |  |  follow <= -0.033       M-(azimuth)
|  |  |  follow > -0.033        M-,+,-(azimuth,follow,elevation)
|  |  azimuth > -11.95          M+,+,-,+(azimuth,follow,elevation,velocity)
azimuth > 4.194
|  velocity <= 3.982
|  |  elevation <= 16.13
|  |  |  azimuth <= 6.018       M-(follow)
|  |  |  azimuth > 6.018        M-,-(azimuth,follow)
|  |  elevation > 16.13
|  |  |  elevation <= 24.77     M-,-(azimuth,follow)
|  |  |  elevation > 24.77      M-,-(follow,azimuth)
|  velocity > 3.982
|  |  follow <= -0.088
|  |  |  follow <= -0.372       M-(azimuth)
|  |  |  follow > -0.372        M-(azimuth)
|  |  follow > -0.088
|  |  |  azimuth <= 6.624       M-,+(follow,azimuth)
|  |  |  azimuth > 6.624        M-,-(azimuth,follow)
```

Slika 4.8: Quinovo kvalitativno drevo za biljard.

Časovna primerjava

Končajmo razdelek s primerjavo časov izvajanja in kratko diskusijo rezultatov. Ker Quin vedno izračuna kvalitativne odvisnosti za vse atribute, za potrebe časovne primerjave isto naredimo tudi s Padéjem. Poudarimo pa, da je Quin implementiran v C++, Padé pa v jeziku Python, ki se interpretira, kar bistveno podaljša čas izvajanja.

Tabela 4.1 prikazuje čase izvajanja algoritmov Quin in Padé (τ-regresija) na podatkih, na katerih smo algoritma primerjali.

podatki	Quin	Padé
$f(x, y) = \sin x \sin y$	37.8	6.8
$f(x, y) = x^3 - y$	15.84	7.2
$f(x, y, r_0, r_1, r_2) = x^2 - y^2$	1049.7	18.04
biljard	14536.39	82.57

Tabela 4.1: Primerjava časov izvajanja (v sekundah).

Na kratko lahko povzamemo, da tako Padé kot Quin računata kvalitativne odvisnosti iz podatkov, pri čemer se razlikujeta tako po definicji kvalitativnih odvisnosti kot po načinu računanja. Zaradi razlike v definiciji kvalitativnih odvisnosti je algoritma praktično nemogoče nepristransko primerjati, saj Quin poleg točnosti uporablja tudi mero dvoumnosti. Quin s pomočjo kvalitativnih vektorjev sprememb ugotavlja trende funkcije v poljubni smeri, medtem ko Padé računa numerične in kvalitativne parcialne odvode le v smereh danih atributov. Bistvena razlika med algoritmoma je, da je Quin omejen na gradnjo kvalitativnih dreves, Padé pa je predprocesor in omogoča modeliranje odvodov s splošnimi klasifikacijskimi metodami strojnega učenja. Z vidika skalabilnosti je Padé po naših izkušnjah v precejšnji prednosti tako glede števila atributov kot števila primerov. Končno presojo prepuščamo uporabniku, ki bo za želeni namen izbral primernejšega od obeh algoritmov.

Poglavje 5

Druge razvite metode za učenje kvalitativnih odvisnosti

Med razvojem osnovnih algoritmov, opisanih v tej disertaciji, smo razvili še več drugih algoritmov za kvalitativno modeliranje v različnih kontekstih. Ker odstopajo od osrednje tematike disertacije, jih v strnjeni obliki opisujemo v tem poglavju. Našteti algoritmi so zanimivi predvsem s teoretičnega vidika, dva od njih (parametrični Padé in Strudel) pa sta se zelo dobro odrezala na realnih poskusih z roboti.

Parametrični Padé [Žabkar et al., 2007a] je primeren za modeliranje v dinamičnih sistemih, kjer čas obravnavamo kot parameter in ne kot ostale atribute. Posplošitev Padéja na relacijske domene imenujemo Strudel (Structural derivatives learning) [Košmerlj et al., 2009]. Osnovna ideja Strudla je posploševanje razlik, ki opisujejo spremembe struktur. Qing [Žabkar et al., 2007b] je algoritem za učenje kvalitativnih modelov na osnovi diskretne Morseove teorije. Za razliko od Padéja išče kritične točke (minimume, maksimume in sedla) v podatkih in deli prostor na t.i. diske, področja z istimi kvalitativnimi lastnostmi. Poglavje zaključimo z Edgar-jem (Equation Discovery with Grammars And Regression) [Žabkar et al., 2007c], algoritmom za odkrivanje kvalitativno pravilnih enačb iz podatkov. Algoritme smo praktično preizkusili na manjšem številu konkretnih primerov, podrobnejša analiza pa bo predmet prihodnjega dela.

5.1 Parametrični Padé

Pri izbiri najbližjih sosedov v metodah iz razdelka 2.2 smo privzeli evklidsko razdaljo. Pri modeliranju dinamičnih sistemov, denimo, si pogosto želimo izbrati drugačno metriko. Zaporedje učnih primerov narekuje čas, ki ga v dinamičnih sistemih obravnavamo kot parameter. Vzemimo parametrično funkcijo:

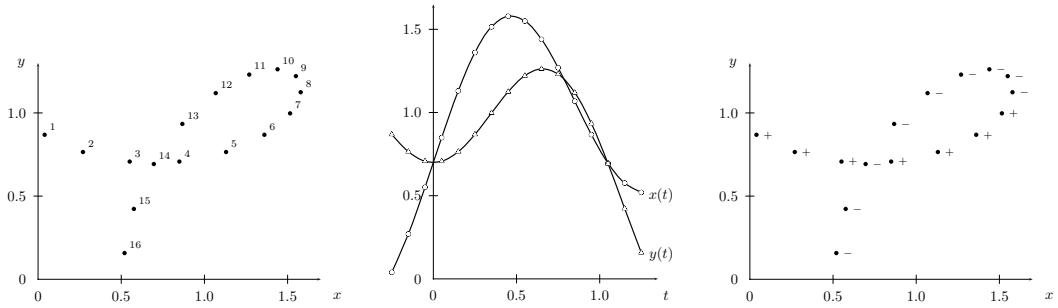
$$x(t) = \sin(3t)\cos t + 0.7,\ y(t) = \sin(3t)\sin t + 0.7, \tag{5.1}$$

za $t \in [0, 1.5]$ in odvajajmo y po x, pri čemer uporabimo verižno pravilo

$$\frac{dy}{dx} = \frac{dy}{dt}\frac{dt}{dx} = \frac{\dot{y}}{\dot{x}}.$$

Parametrični Padé, kot imenujemo Padéja s časovno metriko, naredi podobno. Na podatkih, v katerih vrstni red učnih primerov ni poljuben, temveč ga določa časovno zaporedje, računa parcialne odvode podobno kot Padé, le bližnje sosede izbira glede na čas. Izbira evklidske metrike bi Padéju povzročala težave v okolici točk 3, 4 in 14, kjer krivulja seka samo sebe. Ker sta točki 3 in 14 v evklidski metriki bliže kot 3 in 4, bi Padé izračunal, da funkcija najprej narašča, nato pa pada, kar bi bilo narobe.

Kot primer vzemimo prve štiri stolpce tabele 5.1 (slika 5.1a). Izračunajmo odvod razreda c po atributih x in y. Rezultat prikazuje tabela 5.1, odvod po x pa še slika 5.1c.



(a) Učni primeri za funkcijo 5.1. (b) Graf funkcij $x(t)$ in $y(t)$. (c) Kvalitativni odvod c_{Qx}.

Slika 5.1: Primer odvajanja vzorčene parametrične funkcije.

t	x	y	c	c_{Qx}	c_{Qy}
0	0.03	0.86	1	+	-
0.1	0.26	0.76	2	+	-
0.2	0.55	0.70	3	+	-
0.3	0.84	0.70	4	+	+
0.4	1.13	0.76	5	+	+
0.5	1.36	0.86	6	+	+
0.6	1.51	0.99	7	+	+
0.7	1.57	1.12	8	-	+
0.8	1.54	1.22	9	-	+
0.9	1.43	1.26	10	-	-
1.0	1.26	1.23	11	-	-
1.1	1.06	1.11	12	-	-
1.2	0.86	0.93	13	-	-
1.3	0.69	0.69	14	-	-
1.4	0.57	0.42	15	-	-
1.5	0.51	0.15	16	-	-

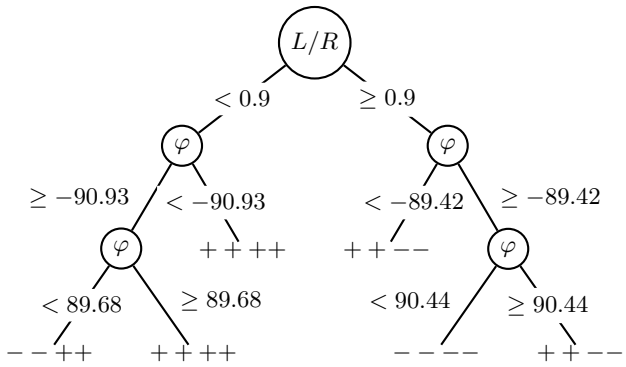
Tabela 5.1: Vhodni podatki (stolpci 1-4) in kvalitativna parcialna odvoda izračunana s parametričnim Padéjem.

5.1.1 Poskus: avtonomno učenje orientacije robota v prostoru

Scenarij za poskus učenja orientacije robota v prostoru je enak kot v primeru v razdelku 2.6. Poleg razdalje d in kota φ zdaj opazujemo še poti, ki ju opravita levo (S_L) in desno kolo (S_R). Robota omejimo na premikanje naprej, naprej in v desno ter naprej in v levo. Robot pozna svoji akciji s_L in s_R ter opazovanji, d in φ. Koordinatnega sistema ne pozna.

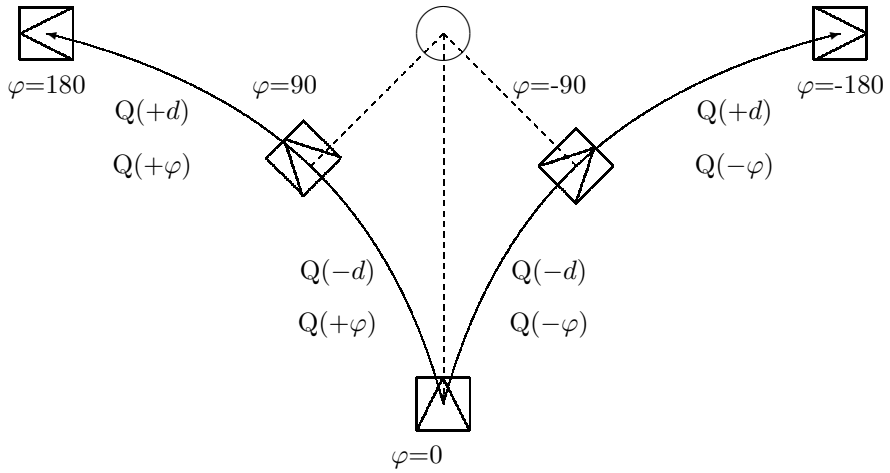
Robotov cilj je učenje kvalitativnega modela, ki opisuje odvisnost med njegovima akcijama in opazovanjema. Naučiti se mora torej: $d = Q(sS_L)$, $d = Q(sS_R)$, $\varphi = Q(sS_L)$ in $\varphi = Q(sS_R)$, kjer je s predznak odvoda po času (+ ali $-$). Iz poti koles z odvodom po času dobimo njuni hitrosti: $L = \dot{S}_L, R = \dot{S}_R$. Ker se učimo vse odvisnosti hkrati, uporabimo skrajšani zapis za zgornje kvalitativne odvisnosti, npr. "$++--$", ki pomeni: $d = Q(+S_L)$, $d = Q(+S_R)$, $\varphi = Q(-S_L)$ in $\varphi = Q(-S_R)$. Z besedami ta zapis razložimo takole: razdalja d narašča, ko naraščata S_L in S_R ($L, R > 0$), medtem ko kot φ pri istih pogojih pada. Za nalogo učenja kvalitativnega modela torej definiramo razred C kot

četverko predznakov, kot smo pravkar opisali. Končni model je kvalitativno drevo na sliki (5.2).

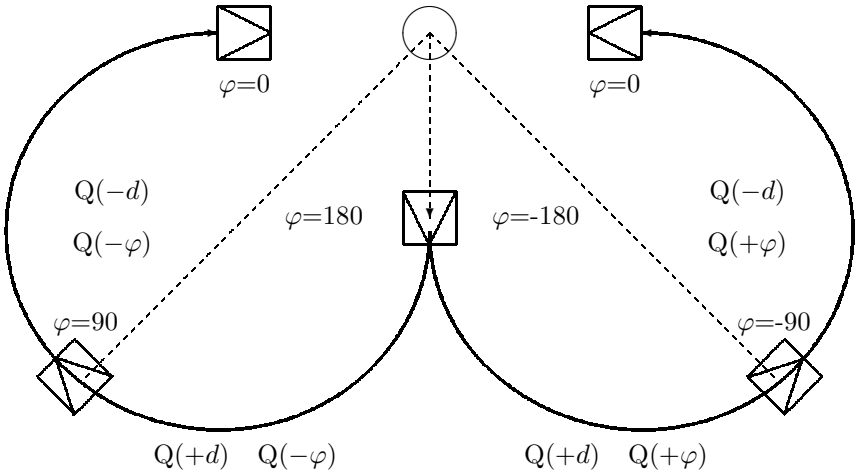


Slika 5.2: Kvalitativno drevo, ki opisuje odvisnosti med robotovima akcijama in opazovanjema.

Pri razlagi drevesa si pomagamo s slikama (5.3a) in (5.3b), ki prikazujeta območja z različnimi kvalitativnimi lastnostmi. Drevo razložimo takole: atribut v korenu, količnik hitrosti levega in desnega kolesa robota (L/R), razdeli primere na leve (levo poddrevo) in desne (desno poddrevo) zavoje robota. Nato se atributni prostor razdeli glede na kot φ. Kotom na intervalu $[-90°, 90°]$ ustreza skrajno levi list drevesa, ki pravi, da se razdalja zmanjšuje z opravljeno potjo, kot φ pa se pri tem povečuje. Pri levih zavojih ter kotih $-90° > \varphi > 90°$, razdalja in kot naraščata s prevoženo potjo. Razlaga desnega poddrevesa, ki opisuje desne zavoje, je podobna – kotom na intervalu $[-90°, 90°]$ ustreza peti list drevesa ($- - --$), ki pomeni, da se razdalja in kot zmanjšujeta, pri kotih izven tega intervala pa razdalja narašča, kot pa pada s prevoženo potjo.



(a) Koti pri obračanju robota v levo in desno pri začetnem kotu 0.



(b) Koti pri obračanju robota v levo in desno za začetni kot $\varphi = \pm 180$.

Slika 5.3: Razlaga domene in kvalitativni model za orientacijo robota v domeni *robot in žoga*.

5.2 QING

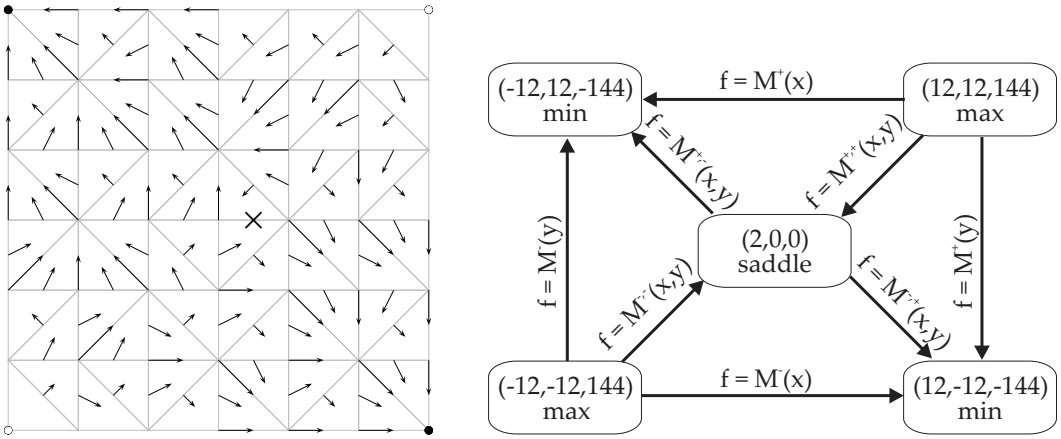
QING (Qualitative INduction Generalized) je algoritem za gradnjo kvalitativnih modelov iz numeričnih podatkov. Osnovan je na diskretni Morseovi teoriji (DMT) [Forman, 1995, 2001; Zomorodian, 2005], ki sodi na področje računske topologije. To daje algoritmu trdno teoretično osnovo.

Qing na podatkih, ki jih opisujejo zvezni atributi in zvezni razred naredi kvalitativno analizo razreda glede na atribute. Izhod algoritma je kvalitativno polje (QPolje) in njegova abstrakcija, ki ji rečemo kvalitativni graf (QGraf).

Od ostalih algoritmov za učenje kvalitativnih modelov se razlikuje v načinu deljenja prostora na podprostore z istimi lastnostmi. Za razliko od algoritmov, ki temeljijo na deljenju prostora glede na vrednosti atributov (drevesa, pravila), Qing triangulira atributni prostor in zgradi kvalitativno polje, ki za vsak učni primer pove smer spremembe razredne spremenljivke. S pomočjo tega odkrije kritične točke (maksimume, minimume, sedla). Omenjena predstavitev omogoča obravnavo šuma s t.i. metodo krajšanja (angl. canceling).

Opis algoritma pospremimo s preprostim primerom funkcije $f(x, y) = xy$. Naj bodo učni primeri podani kot točke v $\mathbb{R}^n$, kjer je n število atributov in naj bo f zvezni razred, za katerega privzamemo, da predstavlja vrednosti gladke Morseove funkcije v teh točkah. V primeru $n = 2$, je množica učnih primerov $\{(x_i, y_i, z_i), i = 1, \ldots, k\}$ vzorčena funkcija $z = f(x, y)$ na $D \subset \mathbb{R}^2$. Naš cilj je analizirati funkcijo f z uporabo DMT in ugotoviti kvalitativno obnašanje f.

Najprej trianguliramo D. Vrednosti razreda v danih točkah razširimo na diskretno Morseovo funkcijo, ki je definirana na trikotnikih naše triangulacije. Za slednjo uporabljamo Delaunayevo triangulacijo, implementirano v programu Qhull [Barber et al., 1996]. Za izračun kritičnih točk funkcije uporabimo diskretno Morseovo teorijo [Forman, 1995]. Kritične točke rekonstruiramo iz kvalitativnega polja (slika 5.4a) z algoritmom iz [King et al., 2005].



(a) (b)

Slika 5.4: Primer kvalitativnega polja (a) za funkcijo $f(x, y) = xy$. Puščice kažejo v smeri padanja funkcije. Legenda: minimum. . . •, maksimum. . . ∘, sedlo . . . ×. Kvalitativni graf (b) za $f(x, y) = xy$.

Definicija: *V-pot* v danem kvalitativnem polju je ali zaporedje točk in daljic, ki te točke povezujejo $v_1 s_1 v_2 s_2, \ldots v_n s_n$ tako, da za vsak i par (v_i, s_i) pripada QPolju, ali zaporedje daljic in trikotnikov $s_1 t_1 s_2 t_2 \ldots s_n t_n$ tako, da za vsak i, (s_i, t_i) pripada QPolju. V-pot določa pot po trikotnikih ali njihovih robovih po katerih vrednost funkcije monotono pada.

Šum obravnavamo tako, da pare kritičnih točk, ki so povezane z eno samo V-potjo in se v funkcijski vrednosti razlikujejo za manj kot vnaprej podan parameter *persistence*, okrajšamo. V-pot povezuje par kritičnih točk q_α in p_ω, če je p_1 krajišče ali rob q_α, in je p_ω krajišče ali rob q_n. Krajšanje dosežemo s prevezavo parov na V-poti v nova para $(q_\alpha, p_1), (q_1, p_2), \ldots, (q_n, p_\omega)$, s čimer odstranimokritični točki q_α in p_ω. Parameter *persistence* običajno nastavimo na vrednost merske napake pri zajemanju podatkov.

Kritične točke, skupaj s QPoljem, predstavljajo kvalitativni model ciljne funkcije oziroma razredne spremenljivke. Tak kvalitativni model lahko uporabimo npr. v Q^2 učenju [Šuc and Bratko, 2003; Šuc et al., 2004; Žabkar et al., 2006], medtem ko za razlago ni primeren, ker je običajno prezapleten. Zato QPolje povzamemo z novo strukturo, ki ji rečemo QGraf. QGraf je abstrakcija QPolja in služi kot grafična predstavitev podatkov ali kot orodje za kvalitativno simulacijo. Formalno QGraf definiramo takole:

Definicija *QGraf* je par (E, V), kjer je V množica vozlišč, $V = \{c \mid c$ *je kritična točka* $\}$, in E množica usmerjenih povezav, $E = \{(a, b) |\ a, b \in V$; *obstaja V-pot v QPolju, ki se začne v a in konča v b* $\}$.

Slaba stran algoritma Qing je njegova odvisnost od triangulacije in časovna zahtevnost. Po vzoru razvoja metod v Padéju se tu odpirajo možnosti za nadaljni razvoj algoritma.

5.3 EDGAR

Avtomatsko odkrivanje enačb [Todorovski and Džeroski, 1997; Langley et al., 2002; Lenat, 1976; Križman et al., 1995; Bradley et al., 2001] je zelo živa veja strojnega učenja že vse od začetkov umetne inteligence. Ukvarja se z učenjem matematičnih enačb iz danih podatkov. Eden od osnovnih pristopov k odkrivanju enačb je generiranje velike množice formul, med katerimi izberemo tisto, ki se najbolj prilega podatkom. V kontekstu Q^2 učenja smo temu pristopu dodali kvalitativne omejitve, ki smo jih pridobili s Padéjem in klasifikacijskimi pravili.

Razvili smo algoritem EDGAR [Žabkar et al., 2007c], ki išče enostavne, kvalitativno pravilne enačbe, ki se čim bolje prilegajo podatkom. Algoritem je zanimiv s teoretičnega vidika in deluje na enostavnih primerih, trenutna implementacija pa ni kos realnim problemom.

Algoritmi, kot so: Goldhorn [Križman et al., 1995], Lagramge [Todorovski and Džeroski, 1997] in PRET [Bradley et al., 2001] uporabljajo simbolični pristop za odkrivanje enačb iz podatkov. Na različne načine sestavljajo enačbe ter s prilagajanjem podatkom uravnavajo vrednosti prostih koeficientov v enačbah. Lagramge uporabniku omogoča celo definicijo gramatike za generiranje enačb [Todorovski and Džeroski, 1997; Langley et al., 2002].

Problem teh pristopov je, da ignorirajo kvalitativne omejitve, kar lahko vodi do nesmiselnih rezultatov. Vzemimo za primer naslednjo enačbo, ki opisuje spreminjanje težnega pospeška v odvisnosti od razdalje od površine Zemlje:

$$g = G\frac{M}{r^2} = \frac{3.99 \times 10^{14}}{r^2},$$

kjer je G gravitacijska konstanta, M masa Zemlje in r razdalja od središča Zemlje do točke, kjer merimo težni pospešek g.

S pomočjo te enačbe smo vzorčili podatke za funkcijo $g(r)$ s korakom 200 na intervalu $[6371, 39971]$ (t.j. od površine Zemlje do višine geostacionarnih satelitov) v 169 točkah. Dodali smo Gaussov šum ($N(0, 0.5)$) in poskušali rekonstruirati enačbo z linearno kombinacijo naslednjih elementarnih funkcij:

$$\{1, r, r^{-1}, r^2, r^{-2}, r^3, r^{-3}, \sin r, \log r, \cos r, \exp r\},$$

kar pomeni, da smo uravnavali koeficiente funkcij kot so npr. $a + br^2 + \sin r$ in $a \log r + br^{-2}$.

Funkcije smo uredili glede na kombinacijo korena povprečne kvadrirane napake (RMSE) in principa najkrajšega opisa (MDL), kot je predlagano v [Todorovski and Džeroski, 1997]. Glede na izbrano mero je bila optimalna funkcija kar konstanta, za njo pa (RMSE=0.5013):

$$g(r) = \frac{3.928 \cdot 10^{14}}{r^2} - 0.124\cos(r).$$

Prvi člen je smiseln, medtem ko drugi očitno poskrbi za prilagajanje šumu. Enačbo je težko razložiti, saj namiguje, da težni pospešek niha, ko povečujemo r. Členu s kosinusom se lahko izognemo na več načinov, npr. s popravkom MDL,

ki bi opis še skrajšal. To predpostavlja, da rešitev poznamo, kar v splošnem seveda ni res, poleg tega pa je tudi v tem primeru očitno, da je vpliv MDL v izbrani meri že zelo visok, kar je vzrok temu, da je bila najbolje ocenjena funkcija kar konstanta.

EDGAR (Equation Discovery with Grammars And Regression) je algoritem za odkrivanje kvalitativno pravilnih enačb iz podatkov. Podobno kot Lagramge uporablja gramatiko, s katero uporabnik določi nastavke enačb, ki naj se generirajo. Poleg tega omogoča podajanje kvalitativnih omejitev, kot npr.: "y narašča z x za vse pozitivne vrednosti x". Algoritem je sledeč:

1. Uporabi generator funkcij za generiranje splošnih oblik funkcij, npr., $a + bx + cx^2$ je splošna oblika za polinom druge stopnje za spremenljivko x.

2. Izračunaj simbolični odvod splošne oblike funkcije, npr.

$$\frac{\partial(a + bx + cx^2)}{\partial x} = b + 2cx.$$

3. Simbolično reši sistem, ki vsebuje dane kvalitativne omejitve in omejuje koeficiente funkcij, npr. če vemo, da funkcija narašča z x za $x > 0$, mora algoritem poiskati taki vrednosti koeficientov b in c, ki ustrezata

$$\forall x, x > 0 : b + 2cx > 0.$$

   Rešitev je:

$$(b = 0 \wedge c > 0) \vee (b > 0 \wedge c \geq 0).$$

4. Izračunaj koeficiente funkcije tako, da minimiziraš RMSE glede na omejitve iz prejšnjega koraka, ki zagotavljajo kvalitativno pravilnost, npr. algoritem poišče vrednosti a, b in c tako, da velja: $(b = 0 \wedge c > 0) \vee (b > 0 \wedge c \geq 0)$, pri čemer se $a + bx + cx^2$ kar najbolj prilega podatkom.

Algoritem je težko implementirati zaradi tretjega koraka, ki se prevede na problem eliminacije kvantifikatorjev iz logičnih formul. Problem v splošnem ni rešljiv in še takrat, ko je rešljiv, praktično ni uporaben zaradi velike računske kompleksnosti. V naši implementaciji smo uporabili funkcijo *Reduce* iz programa za simbolično računanje Mathematica [Wolfram Research, Inc., 2005]. Prav tako je netrivialen zadnji korak, ki v splošnem rešuje nelinearni problem z omejitvami. V pričujočem poskusu ga rešujemo z Nelder-Meadovo metodo [Luersen and Le Riche, 2002].

Za opisani primer podamo kvalitativno omejitev, da težni pospešek pada z razdaljo, $g = Q(-r)$. Najboljša izmed rešitev je (RMSE = 0.4968):

$$g(r) = -0.0259 + \frac{4.096 \cdot 10^{14}}{r^2}.$$

Rešitev je podobna kot prej, le brez člena s kosinusom. Le-tega odstrani podana kvalitativna omejitev, ker kosinus ni monotono naraščajoča funkcija, kot veleva omejitev.

5.4 STRUDEL

Poglavje o sorodnih algoritmih zaključujemo z razširitvijo Padéja na časovne relacijske podatke. Algoritem STRUDEL (*angl. structural derivative learning*) računa t.i. strukturne odvode, ki smo jih definirali na osnovi topološkega odvoda iz matematike. Strudel torej opazuje spremembe strukture podatkov skozi čas.

Kot primer vzemimo logične programe v programskem jeziku Prolog, kjer dodajamo/brišemo predikate, spreminjamo njihove definicije itd. Naj bo S program v Prologu in privzemimo, da se S spreminja v diskretnih časovnih korakih. Zaporedje $S_1, S_2, \ldots, S_n$ označuje stanja S med spremembami. Strukturni odvod S po času označimo z D_i. Definiramo ga kot simetrično razliko sosednjih stanj, $D_i = S_i \triangle S_{i+1}$. Podobno kot Padé je Strudel predprocesor, katerega cilj je izračunati vse strukturne odvode med sosednjimi stanji. Odvode skušamo nato posplošiti, zato se iz njih učimo dinamike S. Odvodi D_i so na nek način le učni primeri, iz katerih avtomatsko tvorimo pozitivne in negativne primere za učenje z induktivnim logičnim programiranjem (ILP), npr. s programom HYPER [Bratko, 2001].

Vhod za algoritem Strudel predstavljata množici opazovanj O in akcij A v diskretnih časovnih točkah, predstavljeni kot program v Prologu. Opazovanja O so predikati, ki opisujejo stanje sveta, npr. meritve senzorjev robota. Prehod med zaporednima stanjema omogoča akcija, npr. premik robota:

$$O_{t1} \wedge A_{t1} \longrightarrow O_{t2}.$$

Kvalitativno spremembo D_i sestavljajo dejstva, ki so se spremenila med časoma T_i in T_{i+1} ali niso obstajali v T_i in so se pojavila v T_{i+1} ali so obstajala v T_i in jih ne najdemo v T_{i+1}. D_i je torej nov predikat, ki opisuje razliko med T_i in

T_{i+1}. Argumenti D_i so kar argumenti akcije A_i. Predikat D_i takoj posplošimo tako, da vse argumente v njem zamenjamo s spremenljivkami:

$$\begin{array}{lcl} \text{Di}(a,\, b)\text{:-} & & \text{Di}(A,\, B)\text{:-} \\ \quad \text{o1}(a,\, 1), & \longrightarrow & \quad \text{o1}(A,\, C), \\ \quad \text{o2}(c,\, 1). & & \quad \text{o2}(D,\, C). \end{array}$$

Z uporabo mehanizma prilagajanja posplošimo odvode, s čimer ekvivalentne odvode uvrstimo v iste skupine, ekvivalenčne razrede. To nam omogoča, da v skupine razvrstimo tudi akcije z istimi učinki – enostavno damo v isto skupino tiste akcije A_i, ki imajo v isti skupini odvod D_i. Zdaj lahko začnemo z učenjem vzrokov za razlike med skupinami akcij. Naš cilj je učenje učinkov akcij. Za učenje uporabimo algoritem HYPER.

HYPER pričakuje množico pozitivnih in množico negativnih učnih primerov. Ker se učimo pogojev, ki ločujejo posamezne skupine akcij, pozitivne in negativne primere definiramo takole: vse akcije iz skupine A_i označimo kot pozitivne primere, vse akcije iz ostalih skupin ($A_j, j \neq i$) pa za negativne primere. Opazovanja O damo HYPER-ju kot predznanje. Za vsako skupino akcij se HYPER nauči predikata:

a(*argumenti*):-
    pogoj(*argumenti*),
    odvod(*argumenti*).

Z opisanim postopkom se je robot, opremljen s kamero, samostojno naučil modelov, ki jih zlahka razumemo kot koncepte stabilnosti enostavne strukture (npr. stolpa iz kock), premičnosti objektov in števila prostostnih stopenj.

Poglavje 6

Zaključek

Eden glavnih ciljev področja umetne inteligence je zgraditi računalnik (robota, stroj), ki bo kar se da dobro posnemal človeški način razmišljanja. To pomeni, da mora računalnik ta način tudi razumeti in se na podoben način tudi izražati (oboje spada v posnemanje človeškega razmišljanja). Zgledi iz vsakdanjega življenja nas prepričajo, da ljudje večinoma razmišljamo kvalitativno. Celo večina sklepanja v naravoslovnih znanostih, ki se od kvalitativnega načina razmišljanja morda še najbolj oddaljijo, je kvalitativnega – v običajnem jeziku, brez formul in enačb. Med različne vrste abstrakcij, ki nam lajšajo kvalitativno sklepanje, uvrščamo poleg naravnega jezika tudi skice, diagrame, grafe, kretnje, glasbo ipd.

Kvalitativni modeli predstavljajo del tega, kar računalniku omogoča kvalitativno sklepanje. V preteklosti so kvalitativni modeli kot del ekspertnih sistemov nastajali postopoma, ročno, kot zapisovanje ekspertnega znanja. V poplavi podatkov, ki se danes pojavljajo na vseh področjih, narašča potreba po avtomatskem učenju kvalitativnih modelov.

V tej disertaciji smo predlagali nove postopke strojnega učenja kvalitativnih modelov, ki temeljijo na kvalitativnih parcialnih odvodih, ki smo jih poprej definirali. Pojem kvalitativnega parcialnega odvoda smo razširili za uporabo na domenah z diskretnim razredom, kjer smo definirali verjetnostni kvalitativni parcialni odvod. Razvite metode smo preizkusili s številnimi poskusi. Na umetnih domenah smo proučevali točnost metod, njihovo odpornost na šum in obnašanje v prisotnosti naključnih atributov. Ugotovili smo, da so metode relativno zelo točne, točnost in odpornost na šum pa se še povečata, ko s pomočjo izračunanih odvodov zgradimo kvalitativni model. Možnost dvo-stopenjskega

učenja štejemo med glavne prednosti razvitih pristopov v primerjavi z drugimi, sorodnimi algoritmi.

Poskusi na podatkih iz simulatorja za biljard so pokazali, da se Padé dobro odreže tudi pri modeliranju zelo kompleksnih fizikalnih pojavov. Fizikalni modeli, ki so uporabljeni v simulatorju so zmožni simulirati praktično vso kompleksnost realne igre, kar sta potrdila tudi eksperta za biljard, ki sta komentirala naučeni model.

Uporaba Padéja in njegove parametrične različice pri obeh realnih poskusih z robotom dokazuje veliko več kot le odpornost na šum v realnih razmerah. Potrjuje namreč naša izhodišča iz uvoda, kjer trdimo, da so kvalitativni modeli pogosto bolj uporabni, točni, zanesljivi, enostavni in razumljivi, kot numerični modeli. Numerični model je izredno kompleksen že v tako enostavnem scenariju, kot je opazovanje površine žoge v sliki kamere. Žoga namreč lahko na sliki izginja na več načinov, npr. na eni, dveh ali treh stranicah hkrati. Če razmišljamo v širšem kontekstu avtonomnega aktivnega učenja robota, ko robot med učenjem preverja kvaliteto svojega modela, vidimo, da lahko to veliko bolj učinkovito počne z enostavnim in točnim kvalitativnim modelom kot s kompleksnim numeričnim modelom.

Do istega zaključka pridemo tudi pri analizi podatkov o bakterijskih okužbah pri starostnikih. Skupina starejših pacientov zaradi staranja populacije in drugih specifičnih lastnosti postaja pereč družbeni problem, zaradi česar je med drugim tudi raziskovalno zelo zanimiva. Rezultati algoritma Qube so po mnenju specialistov odlični. Nekateri modeli odpirajo celo nova raziskovalna vprašanja na področju infektologije.

Podobno velja tudi za področje kvalitativnega sklepanja. Ob tem, da so praktično vsi kvalitativni modeli, tudi na realnih podatkih, zelo enostavni, se postavlja vprašanje, zakaj je temu tako. To vprašanje se po eni strani navezuje na izbiro jezika hipotez – ali so za učenje kvalitativnih modelov iz kvalitativnih parcialnih odvodov dovolj klasifikacijska drevesa oziroma pravila ali je potrebno poseči po bogatejših jezikih, npr. ILP? Po drugi strani je zanimivo tudi vprašanje, zakaj so kvalitativni modeli enostavnejši od numeričnih – zaenkrat smo se zadovoljili zgolj z intuitivno razlago in primeri iz vsakdanjega življenja, znanstvenega odgovora na to vprašanje pa nimamo.

Padé in Qube se sicer zgledujeta po matematični definiciji parcialnega odvoda, ob predpostavki, da je atributni prostor ortogonalen. Algoritma sta dovolj

robustna, da dobro delujeta tudi v primerih, ko to ni res. Kakšne so alternativne možnosti, katere smeri odvajanja so koristne? Nadaljne delo v tej smeri smo začeli z algoritmom QING.

Nadaljne delo smo odprli tudi na drugih področjih. Parametrični Padé odpira nove možnosti v povezavi z aktivnim učenjem. Po našem mnenju je zanimiv predvsem na področju kognitivne robotike, predvsem zaradi primernosti kvalitativnih modelov pri avtonomnem učenju robotov. Zanimiva veja nadaljnih raziskav na istem področju se odpira tudi z algoritmom STRUDEL, ki se uči iz relacijskih podatkov.

Kljub očitni naklonjenosti kvalitativnim modelom na koncu nikakor nočemo pustiti vtisa, da so numerični modeli neuporabni. Razviti algoritem EDGAR uporablja kvalitativne modele kot omejitve pri gradnji numeričnih modelov.

Literatura

Alciatore, D. G. (2004). *The Illustrated Principles of Pool and Billiards*. Sterling; 1st edition.

Asuncion, A. and Newman, D. J. (2007). UCI machine learning repository.

Atkeson, C., Moore, A., and Schaal, S. (1997). Locally weighted learning. *Artificial Intelligence Review*, 11:11–73.

Barber, C. B., Dobkin, D. P., and Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4):469–483.

Ben-David, S., von Luxburg, U., and Pal, D. (2006). A sober look at clustering stability. In Lugosi, G. H.-U. S., editor, *19th Annual Conference on Learning Theory*, pages 5–19, Berlin, Germany. Springer.

Ben-Yehuda, A. and Weksler, M. (1992). Host resistance and the immune system. *Clin Geriatr Med*, 8(4):701–11.

Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999). When is "Nearest Neighbor" Meaningful? In *In Int. Conf. on Database Theory*, pages 217–235.

Bradley, E., Easley, M., and Stolle, R. (2001). Reasoning about nonlinear system identification. *Artificial Intelligence*, 133:139–188.

Bratko, I. (2001). *Prolog Programming for Artificial Intelligence*. Addison Wesley.

Bratko, I., Mozetič, I., and Lavrač, N. (1989). *KARDIO: a study in deep and qualitative knowledge for expert systems*. MIT Press, Cambridge, MA, USA.

Bratko, I., Muggleton, S., and Varšek, A. (1991). Learning qualitative models of dynamic systems. In *Proc. Inductive Logic Programming ILP-91*, pages 207–224.

Bratko, I. and Šuc, D. (2003). Learning qualitative models. *AI Magazine*, 24(4):107–119.

Castle, S., Norman, D., Yeh, M., Miller, D., and Yoshikawa, T. (1991). Fever response in elderly nursing home residents: are the older truly colder? *J Am Geriatr Soc*, 39(9):853–7.

Cestnik, B. (1990). Estimating probabilities: A crucial task in machine learning. In *ECAI*, pages 147–149.

Coghill, G. M., Srinivasan, A., and King, R. D. (2008). Qualitative system identification from imperfect data. *J. Artif. Int. Res.*, 32(1):825–877.

Coghill, G. M.; Garrett, S. M. and King, R. D. (2002). Learning qualitative models in the presence of noise. In *Proc. Qualitative Reasoning Workshop QR'02*.

Coiera, E. (1989). Generating qualitative models from example behaviours. Technical Report 8901, University of New South Wales.

Dasgupta, S. (2010). Strange effects in high dimension. *Commun. CACM*, 53(2):96.

de Kleer, J. (1977). Multiples representations of knowledge in a mechanics problem-solver. In *IJCAI'77: Proceedings of the 5th international joint conference on Artificial intelligence*, pages 299–304, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

de Kleer, J. (1979). *Causal and Telcological Reasoning in Circuit Recognition.* PhD thesis, MIT, MA, USA.

de Kleer, J. and Brown, J. (1984). A qualitative physics based on confluences. *Artificial Intelligence*, 24:7–83.

Džeroski, S. and Todorovski, L. (1993). Discovering dynamics. In *International Conference on Machine Learning*, pages 97–103.

Džeroski, S. and Todorovski, L. (1995). Discovering dynamics: from inductive logic programming to machine discovery. *Journal of Intelligent Information Systems*, 4:89–108.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874.

Fontanarosa, P. B., Kaeberlein, F. J., Gerson, L. W., and Thomson, R. B. (1992). Difficulty in predicting bacteremia in elderly emergency patients. *Annals of Emergency Medicine*, 21(7):842–8.

Forbus, K. (1984). Qualitative process theory. *Artificial Intelligence*, 24:85–168.

Forman, R. (1995). A discrete morse theory for cell complexes. In Yau, S. T., editor, *Geometry, Topology and Physics for Raoul Bott*. International Press.

Forman, R. (2001). A user's guide to discrete morse theory. In *Proceedings of the 2001 International Conference on Formal Power Series and Algebraic Combinatorics, A special volume of Advances in Applied Mathematics.*

Frank, E., Hall, M., and Pfahringer, B. (2003). Locally weighted naive Bayes. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI 2003).*

Friedman, N. and Goldszmidt, M. (1996). Building classifiers using Bayesian networks. In *Proceedings of the thirteenth national conference on artificial intelligence*, pages 1277–1284. AAAI Press.

Gerçeker, R. K. and Say, A. C. C. (2006). Using polynomial approximations to discover qualitative models. In *In Proceedings of the 20th International Workshop on Qualitative Reasoning*, Hanover, New Hampshire.

Gleckman, R. and Hibert, D. (1982). Afebrile bacteremia. a phenomenon in geriatric patients. *JAMA*, 248(12):1478–81.

Hau, D. and Coiera, E. (1997). Learning qualitative models of dynamic systems. *Machine Learning Journal*, 26:177–211.

Jeffries, C. (1974). Qualitative stability and digraphs in model ecosystems. *Ecology*, 55(6):1415–1419.

Kalagnanam, J. and Simon, H. A. (1992). Directions for qualitative reasoning. *Computational Intelligence*, 8(2):308–315.

Kalagnanam, J., Simon, H. A., and Iwasaki, Y. (1991). The mathematical bases for qualitative reasoning. *IEEE Intelligent Systems*, 6(2):11–19.

Kalagnanam, J. R. (1992). *Qualitative analysis of system behaviour*. PhD thesis, Pittsburgh, PA, USA.

King, H. C., Knudson, K., and Mramor Kosta, N. (2005). Generating discrete morse functions from point data. *Exp. math.*, 14(4):435–444.

Kononenko, I. (1991). Semi-naive bayesian classifier. In *EWSL*, pages 206–219.

Košmerlj, A., Leban, G., Žabkar, J., and Bratko, I. (2009). Xpero project Deliverable D4.3, Gaining insights about objects functions, properties and interactions.

Križman, V., Džeroski, S., and Kompare, B. (1995). Discovering dynamics from measured data. *Electrotechnical Review*, 62:191–198.

Kuipers, B. (1986). Qualitative simulation. *Artificial Intelligence*, 29:289–338.

Langley, P. and Sage, S. (1994). Induction of selective Bayesian classifiers. In *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, pages 399–406. Morgan Kaufmann.

Langley, P., Sanchez, J., Todorovski, L., and Džeroski, S. (2002). Inducing process models from continuous data. In *Proceedings of The Nineteenth International Conference on Machine Learning*.

Lenat, D. B. (1976). *Am: an artificial intelligence approach to discovery in mathematics as heuristic search*. PhD thesis, Stanford, CA, USA.

Luersen, M. A. and Le Riche, R. (2002). Globalized Nelder-Mead method for engineering optimization. In *ICECT'03: Proceedings of the third international conference on Engineering computational technology*, pages 165–166, Edinburgh, UK. Civil-Comp press.

Marco, C., Schoenfeld, C., Hansen, K., Hexter, D., Stearns, D., and Kelen, G. (1995). Fever in geriatric emergency patients: clinical features associated with serious illness. *Ann Emerg Med*, 26(1):18–24.

May, R. M. (1973). Qualitative stability in model ecosystems. *Ecology*, 54(3):638–641.

Mellors, J. W., Horwitz, R. I., Harvey, M. R., and Horwitz, S. M. (1987). A simple index to identify occult bacterial infection in adults with acute unexplained fever. *Arch Intern Med*, 147(4):666–71.

Moore, E. H. (1920). On the reciprocal of the general algebraic matrix. In *Bulletin of the American Mathematical Society*, number 26, pages 394–395.

Možina, M., Demšar, J., Kattan, M. W., and Zupan, B. (2004). Nomograms for visualization of naive bayesian classifier. In Boulicaut, J.-F., Esposito, F., Giannotti, F., and Pedreschi, D., editors, *PKDD*, volume 3202 of *Lecture Notes in Computer Science*, pages 337–348. Springer.

Mozetič, I. (1987a). Learning of qualitative models. In *EWSL*, pages 201–217.

Mozetič, I. (1987b). The role of abstractions in learning qualitative models. In *Proc. Fourth Int. Workshop on Machine Learning*. Morgan Kaufmann.

Papavasiliou, D. (2009). Billiards manual. Technical report.

Penrose, R. (1955). A generalized inverse for matrices. In *Proc. of the Cambridge Philosophical Society*, number 3 in 51, pages 406–413.

Pfitzenmeyer, P., Decrey, H., Auckenthaler, R., and Michel, J. (1995). Predicting bacteremia in older patients. *J Am Geriatr Soc*, 43(3):230–5.

Ramachandran, S., Mooney, R., and Kuipers, B. (1994). Learning qualitative models for systems with multiple operating regions. In *Working Papers of the 8th International Workshop on Qualitative Reasoning about Physical Systems*, Japan.

Richards, B., Kraan, I., and Kuipers, B. (1992). Automatic abduction of qualitative models. In *Proceedings of the National Conference on Artificial Intelligence*. AAAI/MIT Press.

Rockwood, K. (1989). Acute confusion in elderly medical patients. *J Am Geriatr Soc*, 37(2):150–4.

Samuelson, P. A. (1983). *Foundations of Economic Analysis*. Harvard University Press; Enlarged edition.

Say, A. and Kuru, S. (1996). Qualitative system identification: deriving structure from behavior. *Artificial Intelligence*, 83:75–141.

Sokal, R. R. and Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin*, 28:1409–1438.

Šuc, D. (2003). *Machine Reconstruction of Human Control Strategies*, volume 99 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, Amsterdam, The Netherlands.

Šuc, D. and Bratko, I. (2001). Induction of qualitative trees. In De Raedt, L. and Flach, P., editors, *Proceedings of the 12th European Conference on Machine Learning*, pages 442–453. Springer. Freiburg, Germany.

Šuc, D. and Bratko, I. (2003). Improving numerical accuracy with qualitative constraints. In Lavrač, N., Gamberger, D., Blockeel, H., and Todorovski, L., editors, *Proceedings of the 14th European Conference on Machine Learning*, pages 385–396. Springer. Dubrovnik, Croatia.

Šuc, D., Vladušič, D., and Bratko, I. (2004). Qualitatively faithful quantitative prediction. *Artificial Intelligence*, 158(2):189–214.

Todorovski, L. (2003). *Using Domain Knowledge for Automated Modeling of Dynamic Systems with Equation Discovery*. PhD thesis, University of Ljubljana, Ljubljana, Slovenia.

Todorovski, L. and Džeroski, S. (1997). Declarative bias in equation discovery. In *In Proceedings of the 14th International Conference on Machine Learning*.

Todorovski, L., Dzeroski, S., Srinivasan, A., Whiteley, J., and Gavaghan, D. (2000). Discovering the structure of partial differential equations from example behavior. In *In Proceedings of the Seventeenth International Conference on Machine Learning*.

urad RS, S. (2010). Demografsko socialno področje – prebivalstvo, http://www.stat.si/tema_demografsko_prebivalstvo.asp.

Wikipedia (2010). Nomogram — wikipedia, the free encyclopedia. [Online; accessed 13-April-2010].

Wolfram Research, Inc. (2005). *Mathematica, Version 5.2*. Champaign, IL, USA.

Yoshikawa, T. (2000). Epidemiology and unique aspects of aging and infectious diseases. *Clin Infect Dis*, 30(6):931–3.

Žabkar, J., Bratko, I., and Demšar, J. (2007a). Learning qualitative models through partial derivatives by Padé. In *Proc. of the 21th International Workshop on Qualitative Reasoning*, Aberystwyth, U.K.

Žabkar, J., Jerše, G., Mramor, N., and Bratko, I. (2007b). Induction of qualitative models using discrete morse theory. In *Proceedings of the 21st Workshop on Qualitative Reasoning*, Aberystwyth.

Žabkar, J., Možina, M., Bratko, I., and Demšar, J. (2009). Discovering monotone relations with padé. In *ECML 2009 workshop: Learning Monotone Models From Data*.

Žabkar, J., Sadikov, A., Bratko, I., and Demšar, J. (2007c). Qualitatively constrained equation discovery. In *Proceedings of the 21st Workshop on Qualitative Reasoning*, Aberystwyth.

Žabkar, J., Žabkar, R., Vladušič, D., Čemas, D., Šuc, D., and Bratko, I. (2006). Q^2 prediction of ozone concentrations. *Ecological Modelling*, 191:68–82.

Ženko, B. and Džeroski, S. (2008). Learning classification rules for multiple target attributes. In Washio, T., Suzuki, E., Ting, K., and Inokuchi, A., editors, *Advances in Knowledge Discovery and Data Mining*, volume 5012 of *Lecture Notes in Computer Science*, pages 454–465. Springer Berlin / Heidelberg.

Zheng, Z., Webb, G. I., and Ting, K. M. (1999). Lazy Bayesian rules: a lazy semi-naive Bayesian learning technique competitive to boosting decision trees. In *Proc. 16th International Conf. on Machine Learning*, pages 493–502. Morgan Kaufmann, San Francisco, CA.

Zomorodian, A. (2005). *Topology for Computing*. Cambridge University Press, New York, NY.

Zupan, B., Leban, G., and Demšar, J. (2004). Orange: Widgets and visual programming, a white paper.