

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Vesna Tanko

**Kovariančne funkcije v modelih na
podlagi Gaussovih procesov**

DIPLOMSKO DELO
NA UNIVERZITETNEM ŠTUDIJU

Mentor: doc. dr. Janez Demšar

Somentor: prof. dr. Juš Kocijan

Ljubljana, 2009



Št. naloge: 01558/2009

Datum: 05.04.2009

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Kandidat: **VESNA TANKO**

Naslov: **KOVARIANČNE FUNKCIJE V MODELIH NA PODLAGI GAUSSOVIH
PROCESOV**

COVARIANCE FUNCTIONS FOR GAUSSIAN PROCESS MODELS

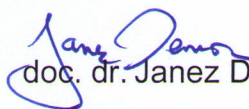
Vrsta naloge: Diplomsko delo univerzitetnega študija

Tematika naloge:

Gaussovi procesi so stohastični procesi, v katerih so vektorji neodvisnih spremenljivk porazdeljeni po Gaussovi normalni porazdelitvi, korelacije med njimi pa določa kovariančna funkcija. V zadnjem času se Gaussovi procesi vedno pogosteje uporabljajo tudi kot osnova za modeliranje v regresijskih (in tudi klasifikacijskih) problemih.

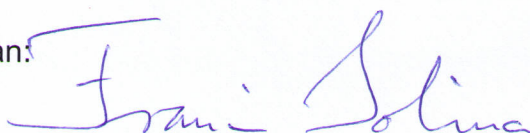
Ena pomembnih odločitev pri sestavljanju modela na podlagi Gaussovih procesov je izbor kovariančne funkcije. V okviru diplomske naloge opišite najpogosteje uporabljene kovariančne funkcije, analizirajte njihove teoretične lastnosti in preučite njihovo delovanje na praktičnih primerih.

Mentor:

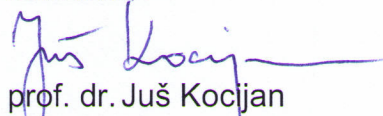

doc. dr. Janez Demšar



Dekan:


prof. dr. Franc Solina

Somentor:


prof. dr. Juš Kocijan

Rezultati diplomskega dela so intelektualna lastnina Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavlanje ali izkoriščanje rezultatov diplomskega dela je potrebno pisno soglasje Fakultete za računalništvo in informatiko ter mentorja.

Besedilo je oblikovano z urejevalnikom besedil $\text{\LaTeX}$.

Namesto te strani **vstavite** original izdane teme diplomskega dela s podpisom mentorja in dekana ter žigom fakultete, ki ga diplomant dvigne v študentskem referatu, preden odda izdelek v vezavo!

IZJAVA O AVTORSTVU

diplomskega dela

Spodaj podpisani/-a Vesna Tanko,

z vpisno številko 63020326,

sem avtorica diplomskega dela z naslovom:

Kovariančne funkcije v modelih na podlagi Gaussovih procesov

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal/-a samostojno pod mentorstvom
doc. dr. Janeza Demšarja
in somentorstvom
prof. dr. Juša Kocijana
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela
- soglašam z javno objavo elektronske oblike diplomskega dela v zbirki "Dela FRI".

V Ljubljani, dne 9. 9. 2009

Podpis avtorja/-ice:

Zahvala

Za pomoč pri izdelavi diplomskega dela se zahvaljujem mentorju doc. dr. Janezu Demšarju in delovnemu mentorju prof. dr. Jušu Kocijanu ter kolegom Odseka za sisteme in vodenje na Institutu Jožef Stefan.

Kazalo

Povzetek	1
Abstract	2
1 Uvod	3
2 Gaussovi procesi	6
2.1 Osnove verjetnostnega modeliranja	6
2.1.1 Osnovni pojmi	6
2.1.2 Bayesovo modeliranje	7
2.2 Modeliranje z Gaussovimi procesi	8
2.2.1 Kovariančna funkcija	8
2.2.2 Parametri kovariančne funkcije	9
2.2.3 Napovedovanje	10
2.2.4 Interpretacija	11
2.2.5 Vrednotenje	12
2.2.6 Programska podpora	13
3 Kovariančne funkcije	14
3.1 Stacionarne kovariančne funkcije	15
3.1.1 Konstantna kovariančna funkcija	15
3.1.2 Gaussova kovariančna funkcija	16
3.1.3 Eksponentna kovariančna funkcija	18
3.1.4 Racionalni kvadratični polinom	19
3.1.5 Matérnova kovariančna funkcija	19
3.1.6 Periodična kovariančna funkcija	22
3.2 Nestacionarne kovariančne funkcije	23
3.2.1 Linearna kovariančna funkcija	24
3.2.2 Polinomska kovariančna funkcija	24
3.2.3 Kovariančna funkcija nevronske mreže	26

3.2.4	Vhodno odvisen model šuma	28
3.2.5	Prostorsko spremenljiva dolžinska merila	28
3.3	Izgradnja novih kovariančnih funkcij iz obstoječih	29
3.4	Avtomatsko določanje ustreznosti	30
4	Primeri modeliranja časovnih vrst	32
4.1	Modeliranje globalnega segrevanja	32
4.1.1	Opis podatkov	32
4.1.2	Interpolacija	33
4.1.3	Ekstrapolacija	50
4.1.4	Sklep	60
4.2	Modeliranje življenjske dobe zobnika	61
4.2.1	Opis testnega okolja in zajema podatkov	61
4.2.2	Modeliranje in vrednotenje modela	62
4.2.3	Napovedovanje kritične točke	63
4.2.4	Sklep	66
5	Zaključek	68
A	Primer uporabe GP modela	70
	Seznam slik	72
	Seznam tabel	75
	Literatura	76

Seznam uporabljenih kratic in simbolov

GP - Gaussov proces

GP model - model na podlagi Gaussovih procesov

KF - kovariančna funkcija

Povzetek

V diplomskem delu je poudarek na uporabi Gaussovih procesov v strojnem učenju, natančneje izbiri kovariančnih funkcij, ki predstavljajo pomembnejši del modeliranja na podlagi Gaussovih procesov. Za izbiro ustreznih kovariančnih funkcij je potrebno poznati osnove modeliranja z Gaussovimi procesi, zato so v diplomsko delo vključene tudi te. Opisu osnov omenjenega načina modeliranja sledi opis najpogostejše uporabljenih kovariančnih funkcij.

Predstavljena sta še dva konkretna primera modeliranja na katerih smo demonstrirali izbiro in uporabo kovariančnih funkcij. Na podlagi rezultatov modeliranja in primerjave modelov na eni vrsti podatkov smo želeli dobiti smernice za modeliranje podobnih podatkov ter priporočila za določitev primernih kovariančnih funkcij. Pridobljene izkušnje smo kasneje uporabili za modeliranje drugih podatkov.

Ugotovili smo, da je izbira kovariančne funkcije odvisna od narave podatkov in predvsem namena modeliranja, s čimer mislimo posameznikovo predstavo oz. predznanje o izhodu modela. Drugače povedano: izbrana kovariančna funkcija za modeliranje poglajene krivulje ne bo enaka kovariančni funkciji, s pomočjo katere na izhodu GP modela želimo ravno premico.

Ključne besede:

Gaussov proces, model na podlagi Gaussovih procesov, kovariančna funkcija, stacionarnost, gladkost, hiperparametri, časovne vrste.

Abstract

The main topic of this thesis are Gaussian processes for machine learning, more precisely the selection of covariance function, which represents one of the most important issues of GP modeling. To understand covariance function, it is necessary to know GP modeling basics, thus they are included in the thesis. Besides GP modeling basics, the most popular covariance functions are described. Their selection and use are presented on two particular cases of GP modeling.

The first case was used to study covariance functions and provide us with guidelines for similar data modeling and recommendations to determine appropriate covariance function. Finally, we used the experiences gained through the experiments and made predictions for the second case of data.

We realized, that selection of covariance function depends on data structure and primarily purpose of modeling, by which we mean individuals expectations and background knowledge of GP model output. In other words: selection of covariance function for smooth function modeling differs from selection of covariance function for linear line.

Key words:

Gaussian process, Gaussian process model, covariance functions, stationarity, smoothness, hyperparameters, time series.

Poglavje 1

Uvod

Človek si je vedno želel bolje spoznati samega sebe in ustvariti sisteme, ki bi njegove zmožnosti presegle. Človeško učenje je zelo počasno. Potrebno je precej let delovanja, da si človek nabere zadosti izkušenj in tako postane strokovnjak na nekem ozkem področju. Poleg tega svojih izkušenj ne more preprosto prenesti na mlajše kolege, pri vsem tem pa veljajo zanj vse slabe lastnosti, kot za vsakega človeka: je pozabljiv in zmotljiv, se hitro utrudi, na njegovo delo vplivajo zunanji dejavniki, kot so vreme, bolezen, razpoloženje itd. Ekspertni sistemi so sposobni pomagati strokovnjakom pri njihovem delu, v izjemnih primerih pa jih lahko celo nadomestijo. Od ekspertnih sistemov se zahteva, da znajo uporabniku svoje rešitve problemov tudi obrazložiti in argumentirati. Le takemu sistemu bo človek lahko zaupal in mu prepustil pomembne odločitve. Ekspertni sistemi imajo vse prednosti, ki veljajo za računalnik: so zanesljivi, neutrudljivi, ponovljivi, zakodirano znanje je trivialno prenosljivo, pri odločitvah lahko upoštevajo velike količine podatkov, zbrane v preteklosti. Kljub prednostim pa ekspertni sistemi v bližnji prihodnosti ne morejo izpodrinuti strokovnjakov, saj kljub veliki hitrosti računalnikov, ki lahko obdelujejo ogromne količine podatkov, ne morejo preseči človekovega širokega znanja in spomina. Glavna prednost ljudi pa je, da so prilagodljivi in svoje znanje dinamično spreminjajo in izpopolnjujejo. Z ustreznimi algoritmi je potrebno narediti današnje računalnike manj toge in prilagodljive novim situacijam in problemom [12].

Veda, ki išče rešitve teh problemov je umetna inteligenca, področje pa imenujemo strojno učenje. Strojno učenje pomeni pridobivanje znanja na podlagi izkušenj. Gre torej za posploševanje na podlagi podatkov, s katerim lahko kasneje sklepamo o primerih, kakršnih v učnih primerih ni bilo. Strojno učenje se močno opira na statistiko, saj se tudi statistika ukvarja s podatki, a se v na-

sprotju z njo strojno učenje bolj osredotoča na algoritme, ki omogočajo učenje iz podatkov.

Strojno učenje ima širok spekter uporabnosti, saj se uporablja pri medicinskih diagnozah, računalniškem vidu, bioinformatiki, iskalnikih, analizi poslovanja z vrednostnimi papirji, detekciji ponarejanja dokumentov, strojnem vidu, računalniških igrah, robotiki, razpoznavanju DNA sekvenc, razpoznavanju govora in pisave itd.

Osnovni princip strojnega učenja je avtomatsko opisovanje (modeliranje) pojavov iz podatkov [12]. Rezultat učenja iz podatkov so lahko pravila, funkcije, relacije, sistemi enačb, verjetnostne porazdelitve ipd., ki so lahko predstavljene z različnimi formalizmi: odločitvenimi pravili, odločitvenimi drevesi, regresijskimi drevesi, Bayesovimi mrežami, nevronskimi mrežami itd. Naučeni modeli poskušajo razlagati podatke, iz katerih so bili modeli tvorjeni, in se lahko uporabljajo za določanje pri opazovanju modelirnega procesa v bodočnosti.

Metode strojnega učenja delimo glede na način uporabe naučenega znanja: klasifikacija, regresija, učenje asociacij in logičnih relacij, učenje sistemov enačb in razvrščanje. V tem diplomskem delu bo podrobneje obravnavana regresija z modeli na podlagi Gaussovih procesov, katere naloga je za problem, opisan z množico atributov, določiti vrednost odvisne (regresijske) spremenljivke, ki je zvezna. Regresijski prediktor mora imeti na nek način predstavljeno zvezno funkcijo, ki preslika prostor atributov v napovedano vrednost. Ta funkcija je lahko podana vnaprej ali pa je naučena iz podatkov. Naloga učnega algoritma je torej iz množice opisov primerov z znanimi vrednostmi odvisne spremenljivke izračunati zvezno funkcijo, ki jo lahko uporabimo za določanje vrednosti regresijske spremenljivke za nove primere. Regresijske prediktorje ločimo glede na način predstavitve regresijske funkcije. Najbolj pogosti regresorji so regresijska drevesa, linearna regresija, lokalna utežena regresija, regresija po metodi podpornih vektorjev, usmerjene umetne nevronske mreže, v tem diplomskem delu pa bo obravnavano še precej novo področje - modeliranje na podlagi Gaussovih procesov.

Namen tega diplomskega dela je narediti pregled trenutno najbolj uporabljanih kovariančnih funkcij, ki se uporabljajo pri strojnem učenju z Gaussovimi procesi. Poleg tega bi želeli v diplomskem delu prikazati uporabo različnih kovariančnih funkcij na praktičnih primerih modeliranja.

Diplomsko delo je razdeljeno na štiri glavne dele. Za uvodnim poglavjem je podan opis modeliranja na podlagi Gaussovih procesov. Sledi poglavje, kjer so predstavljene pogostejše uporabljene kovariančne funkcije ter načini, kako lahko s kombinacijo obstoječih funkcij dobimo nove. Za tem sta podana še primera

modeliranja z uporabo Gaussovih procesov in prikazom izbire kovariančnih funkcij.

Poglavje 2

Gaussovi procesi

2.1 Osnove verjetnostnega modeliranja

2.1.1 Osnovni pojmi

Naključna spremenljivka je spremenljivka, katere vrednost je odvisna od naključja. Naključni proces (ang. *stochastic process*) je večkratna realizacija naključne spremenljivke v odvisnosti od nekih neodvisnih spremenljivk (npr. časa, prostora). Porazdelitve vrednosti posameznih realizacij naključne spremenljivke so lahko poljubne, opišemo jih s porazdelitvenim zakonom in zalogo vrednosti, s čimer je naključna spremenljivka tudi popolnoma določena. Glede na zalogo vrednosti ločimo diskretne in zvezne naključne spremenljivke, glede na obliko porazdelitvenega zakona pa različne standardne in nestandardne porazdelitve (npr. Gaussova, Poissonova itd.).

Naključni vektor je vektor, katerega komponente so naključne spremenljivke. Če imamo dve ali več med seboj povezanih naključnih spremenljivk, njihovo večkratno realizacijo imenujemo naključni proces.

Če so vrednosti naključne spremenljivke porazdeljene po normalnem (Gaussovem) porazdelitvenem zakonu, takemu procesu pravimo Gaussov proces (GP). Drugače povedano: naključni proces je Gaussov proces, če je za vsak vektor neodvisnih spremenljivk $\mathbf{x}$, vrednost funkcije $f(\mathbf{x})$ porazdeljena po normalni (Gaussovi) porazdelitvi. GP je povsem določen z vektorjem srednjih vrednosti μ in kovariančno funkcijo $C(\mathbf{x}_i, \mathbf{x}_j)$, prek katere je definirana njegova kovariančna matrika $\mathbf{K}$ [13, 2, 11].

2.1.2 Bayesovo modeliranje

Pri modeliranju statičnih funkcij, kjer je učna množica sestavljena iz n vektorjev D -dimenzionalnih neodvisnih vhodnih spremenljivk $\mathbf{X}$ in pripadajočim vektorjem n izhodnih točk $\mathbf{t}$, podanih v obliki $\mathcal{D} = \{\mathbf{X}, \mathbf{t}\} = \{(\mathbf{x}_i, t_i), i=1 \dots n\}$, želimo najti funkcijo f , ki naj čimbolje opisuje relacijo med vhodno-izhodnimi pari $(\mathbf{x}_i, t_i)$. Z vidika Bayesovega modeliranja, lahko modeliranje statičnega procesa z nelinearno funkcijo $f = f(\mathbf{x}, \boldsymbol{\omega})$, katere parametre $\boldsymbol{\omega}$ določimo glede na podatke $\mathcal{D}$, predstavimo v obliki Bayesovega teorema:

$$p(\boldsymbol{\omega}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\omega})p(\boldsymbol{\omega})}{p(\mathcal{D})} \quad (2.1)$$

kjer je:

- $p(\boldsymbol{\omega})$ - apriorna verjetnostna porazdelitev, ki vsebuje predhodno znanje o parametrih funkcije (ang. *prior*), ki ponavadi predpostavljaajo zveznost, frekvenčno razporeditev moči itd.,
- $p(\mathcal{D}|\boldsymbol{\omega})$ - verjetnostna porazdelitev učne množice pri danih parametrih funkcije (ang. *likelihood*),
- $p(\mathcal{D})$ - verjetnostna porazdelitev učne množice (ang. *evidence*), ki v primeru Bayesovega modeliranja služi kot normalizacijska konstanta,
- $p(\boldsymbol{\omega}|\mathcal{D})$ - posteriorna verjetnostna porazdelitev parametrov $\boldsymbol{\omega}$ pri dani učni množici $\mathcal{D}$ (ang. *posterior*).

Bayesov teorem združuje predhodno znanje o parametrih $p(\boldsymbol{\omega})$ z znanjem, dobljenim v obliki učne množice $\mathcal{D}$. Z upoštevanjem zapisa $\mathcal{D} = \{\mathbf{X}, \mathbf{t}\}$ lahko Bayesov teorem zapišemo v obliki:

$$p(\boldsymbol{\omega}|\mathbf{X}, \mathbf{t}) = \frac{p(\mathbf{t}|\boldsymbol{\omega}, \mathbf{X})p(\boldsymbol{\omega})}{p(\mathbf{t}|\mathbf{X})} \quad (2.2)$$

Parametre $\boldsymbol{\omega}$ na podlagi verjetnostnega zapisa, pri predpostavkah o Gaussovi porazdelitvi $p(\mathbf{t}|\boldsymbol{\omega}, \mathbf{X})$ in $p(\boldsymbol{\omega})$, lahko določimo s kriterijsko funkcijo:

$$J(\boldsymbol{\omega}) = -\log(p(\mathbf{t}|\boldsymbol{\omega}, \mathbf{X})p(\boldsymbol{\omega})) \quad (2.3)$$

Z minimizacijo kriterijske funkcije določimo maksimum pogojne verjetnostne gostote $p(\boldsymbol{\omega}|\mathbf{X}, \mathbf{t})$ in s tem oceno najverjetnejše vrednosti parametrov $\boldsymbol{\omega}$. Za

predikcijo porazdelitve izhoda t^* pri vходу x^* pri poznavanju $p(\omega|\mathbf{X}, \mathbf{t})$ pa je potrebno izračunati integral (marginalizacijo) po parametrih [13]:

$$p(t^*|\mathbf{x}^*, \mathbf{X}, \mathbf{t}) = \int p(t^*|\mathbf{x}^*, \mathbf{X}, \mathbf{t}, \omega) p(\omega|\mathbf{X}, \mathbf{t}) d\omega \quad (2.4)$$

2.2 Modeliranje z Gaussovimi procesi

Zamislimo si regresijski problem. Imamo nabor vektorjev, združenih v matriko $\mathbf{X}$ in nabor izhodnih točk $\mathbf{y}$. Želimo najti funkcijo $f(\mathbf{x})$, ki naj čimbolje opisuje relacijo med N vhodno-izhodnimi pari $(\mathbf{x}_i, y_i)$.

Model na podlagi Gaussovih procesov, ali krajše GP model, je primer neparametričnega verjetnostnega modela in se ravna po Bayesovem principu modeliranja. Pri Bayesovem modeliranju, namesto parametrizacije funkcije, znanje uporabimo za določitev porazdelitvenega zakona nad neko družino funkcij, ki preslikavajo vhodne podatke $\mathbf{x}_i$ v izhodne y_i . Pri tem dopuščamo večjo verjetnost funkcij, za katere menimo, da se pri opisu sistema verjetneje pojavljajo. Predhodno znanje (ang. *prior*) je vključeno tako, da odraža mnenje o preslikavi med vhodi in izhodi in običajno predpostavlja gladkost preslikave (podobni vhodi naj bi se preslikali v podobne izhode). Ko vključimo v model še verjetnost učne množice (ang. *likelihood*), sestavljene iz N vhodno-izhodnih parov $(\mathbf{x}_i, y_i)$, dobimo posteriorno porazdelitev za predikcijo modela (ang. *posterior*) [3].

Vhod v GP model so posamezne vrednosti neodvisnih spremenljivk, zbrane v vhodnem vektorju $\mathbf{x}$, medtem ko je izhod iz GP modela verjetnostna porazdelitev izhodne vrednosti $f(\mathbf{x})$ pri danem vhodnem vektorju.

2.2.1 Kovariančna funkcija

Vrednost kovariančne funkcije $C(\mathbf{x}_i, \mathbf{x}_j)$ izraža korelacijo med posameznima izhodoma $f(\mathbf{x}_i)$ in $f(\mathbf{x}_j)$ modela, obravnavana kot dve medsebojno povezani naključni spremenljivki, glede na vhoda $\mathbf{x}_i$ in $\mathbf{x}_j$:

$$\text{cov}(y_i, y_j) = C(\mathbf{x}_i, \mathbf{x}_j) \quad (2.5)$$

V splošnem je kovariančna funkcija lahko poljubna funkcija, ki tvori nenegativno definitno kovariančno matriko $\mathbf{K}$ za poljuben nabor vhodnih vektorjev. S

stališča modeliranja je primernejša izbira take kovariančne funkcije, ki močnejše korelira izhodne točke, ki so si v vhodnem prostoru bližje [13]. Kovariančne funkcije so lahko stacionarne, nestacionarne, periodične itd., podrobneje pa jih bomo obravnavali v naslednjih poglavjih. Za zdaj naj povemo, da kovariančne funkcije, ki določa obliko neznane funkcije $f(\mathbf{x})$, navadno ne poznamo, lahko pa iz znanja o splošnih lastnostih funkcije $f(\mathbf{x})$ sklepamo o njeni obliki. Pogosto ugotavljamo, da se v podatkih nahaja tudi šum, zato želimo vpeljati v kovariančno funkcijo tudi model šuma.

2.2.2 Parametri kovariančne funkcije

Parametre kovariančne funkcije imenujemo hiperparametri; s tem poudarimo, da so to parametri sicer neparametričnega modela, ki določajo verjetnostno porazdelitev nad družino funkcij. Vsaka funkcija vsebuje parameter, ki govori o varianci izhoda, ter parametre, ki odražajo pomembnost posamezne komponente vhodnega vektorja. Najbolj verjetne vrednosti hiperparametrov pri določeni kovariančni funkciji določimo prek posteriorne verjetnostne porazdelitve parametrov:

$$p(\Theta|\mathbf{X}, \mathbf{t}) = \frac{p(\mathbf{t}|\Theta, \mathbf{X})p(\Theta)}{p(\mathbf{t}|\mathbf{X})} \quad (2.6)$$

Optimalne vrednosti hiperparametrov določimo z iskanjem največjega logaritma porazdelitve $p(\mathbf{t}|\Theta, \mathbf{X})$, ki je logaritem Gaussovega procesa:

$$\log(p(\mathbf{t}|\Theta, \mathbf{X})) = -\frac{1}{2}\log(|\mathbf{K}|) - \frac{1}{2}\mathbf{t}^\top \mathbf{K}^{-1}\mathbf{t} - \frac{n}{2}\log(2\pi) \quad (2.7)$$

Metoda se imenuje metoda največje podobnosti (ang. *maximum likelihood*). Za iskanje minimuma se lahko uporablja katerakoli metoda. Ena izmed možnih je uporaba metode konjugiranih gradientov zaradi enostavnih analitičnih izračunov parcialnih odvodov. Metoda je občutljiva na začetno izbiro hiperparametrov (padec v lokalni minimum), poleg tega pa je računsko zahtevna, saj vsak korak optimizacije zahteva izračun inverzne kovariančne matrike dimenzije $n \times n$, kjer je n število podatkov v učni množici.

Poleg omenjene metode je za aproksimacijo integrala (2.4) pogosta tudi numerična integracija nad celotno porazdelitvijo hiperparametrov (MCMC metode), dobljeno z optimizacijo verjetnosti učnih podatkov [11], obstajajo pa tudi druge metode, npr. rekurzivne [14].

2.2.3 Napovedovanje

Za predikcijo porazdelitve izhoda y^* pri novem vhodu $\mathbf{x}^*$ velja:

$$p(y^*|\mathbf{y}) = \frac{p(\mathbf{y}, y^*)}{p(\mathbf{y})} \quad (2.8)$$

Pri modeliranju z GP predpostavimo n -dimenzionalno porazdelitev izhoda $p(\mathbf{y})$:

$$p(\mathbf{y}) = \frac{1}{Z} \exp\left(-\frac{1}{2}(\mathbf{y} - \mu)^\top \mathbf{K}^{-1}(\mathbf{y} - \mu)\right) \quad (2.9)$$

kjer je $\mathbf{K}$ kovariančna matrika normalne porazdelitve in Z normalizacijska konstanta [13]. Ob upoštevanju zgornjih enačb (2.8) in (2.9) izpeljemo:

$$p(y^*|\mathbf{y}) = \frac{Z_n}{Z_{n+1}} \exp\left(-\frac{1}{2}(\mathbf{y}, y^*)^\top \mathbf{K}_{n+1}^{-1}(\mathbf{y}, y^*) - \mathbf{y}^\top \mathbf{K}_n^{-1} \mathbf{y}\right) \quad (2.10)$$

V zgornji enačbi je kovariančna matrika $\mathbf{K}_n$ (reda $n \cdot n$) določena na podlagi izbrane kovariančne funkcije $C(\mathbf{x}_i, \mathbf{x}_j)$ in n vhodnih vektorjev učne množice $\mathcal{D}$, kovariančna matrika $\mathbf{K}_{n+1}$ (reda $(n+1) \cdot (n+1)$) pa na podlagi iste učne množice, razširjene z novim vhodom $\mathbf{x}^*$. Med njima velja relacija:

$$\mathbf{K}_{n+1} = \begin{bmatrix} \begin{bmatrix} \mathbf{K}_n \\ [\mathbf{k}] \end{bmatrix} \\ \begin{bmatrix} [\mathbf{k}^\top] \\ [k] \end{bmatrix} \end{bmatrix} \quad (2.11)$$

Vektor $\mathbf{k} = (C(\mathbf{x}_1, \mathbf{x}^*) \dots C(\mathbf{x}_n, \mathbf{x}^*))$ vsebuje korelacije izhoda y^* z izhodi modela $\mathbf{y}$, skalar $k = C(\mathbf{x}^*, \mathbf{x}^*)$ pa varianco izhoda y^* . Enačbo (2.10) z upoštevanjem relacij v zapisu (2.11) lahko izrazimo v obliki:

$$p(y^*|\mathbf{y}) = \frac{1}{Z} \exp\left(-\frac{(y^* - \mu_{y^*})^2}{2\sigma_{y^*}^2}\right) \quad (2.12)$$

$$\mu_{y^*} = \mathbf{k}^\top \mathbf{K}^{-1} \mathbf{t} \quad (2.13)$$

$$\sigma_{y^*}^2 = k - \mathbf{k}^\top \mathbf{K}^{-1} \mathbf{t} \quad (2.14)$$

V zgornjih enačbah sta prikazana izraza za izračun srednje vrednosti μ_{y^*} in variance $\sigma_{y^*}^2$ izhoda y^* , ki je normalna porazdelitev za napovedovanje vrednosti izhoda procesa na podlagi vhodnega vektorja $\mathbf{x}^*$. Pričakovana vrednost izhoda y^* je tako enaka srednji vrednosti porazdelitve μ_{y^*} [13].

2.2.4 Interpretacija

GP model je torej sestavljen iz dveh delov:

- iz parov vhodno/izhodnih učnih podatkov $\mathcal{D}$, ki predstavljajo obnašanje neznanega sistema, in
- kovariančne funkcije $C(\mathbf{x}_i, \mathbf{x}_j)$ z znanimi oz. optimiranimi hiperparametri ω , ki pove, v kakšnem medsebojnem razmerju so podatki $\mathcal{D}$.

Ker GP model potrebuje informacijo o neznani funkciji v obliki učnih vhodov in izhodov tudi po učenju, je model neparametričen. Hiperparametri namreč prek kovariančne funkcije samo povedo, kako se učna informacija uporabi za napovedovanje, ni pa v njih spravljena informacija o opisovani funkciji/sistemu, kot v parametričnem modelu.

Na vektor $\mathbf{k}^\top \mathbf{K}_n^{-1}$ v izrazu za srednjo vrednost napovedanega izhoda lahko gledamo kot na vektor uteži, ki določa utežitev posameznih učnih izhodov y_i , v $\mathbf{y}$ glede na s hiperparametri uteženo evklidsko razdaljo med učnimi in testnim vhodnim vektorjem v vhodnem prostoru. Ta linearna kombinacija učnih izhodov se lahko razume kot glajenje v GP modelu vsebovane informacije o neznaniem sistemu (učni podatki). Še drugače si lahko srednjo vrednost napovedi μ_{y^*} predstavljamo kot linearno kombinacijo N jedernih (ang. *kernel*) funkcij, usrediščenih v učnih točkah $\mu_{y^*} = \sum_{i=1}^N \alpha_i C(\mathbf{x}^*, \mathbf{x}_i)$. Izhod iz sistema je eden izmed vzorcev iz dobljene normalne porazdelitve.

Majhna varianca $\sigma_{y^*}^2$ napovedane porazdelitve izhoda pomeni večje zaupanje v napovedi in obratno. Izraz za varianco je sestavljen iz dveh delov. Od prvega dela k , ki predstavlja apriorno varianco GP modela, je odštet izraz $\mathbf{k}^\top \mathbf{K}^{-1} \mathbf{k}$. Ta zmanjša apriorno varianco GP modela pri $\mathbf{x}^*$ zaradi učnih podatkov in se večja z večjo kovarianco med učnimi in testnim vhodom. Drugače

povedano: čim podobnejši je testni vhod že znanim (učnim) vhodom v vhodnem prostoru, večje je zaupanje GP modela v točnost napovedi. Prav variacija, odvisna tudi od lege testnega vhoda glede na učne, je ena izmed glavnih prednosti GP modela pred drugimi modeli [2]. Preprosta ilustracija uporabe GP modela se nahaja v dodatku A.

2.2.5 Vrednotenje

Čeprav je vrednotenje zelo pomemben korak v modelirnem postopku, ki pove, kako dober je dobljen model, mu dostikrat namenjamo premalo pozornosti. Z vrednotenjem preverimo ujemanje matematičnega modela in obravnavanega sistema. Kvaliteto predikcije modela lahko merimo na več načinov, najbolj pogoste mere pa so:

- povprečna absolutna napaka MAE (ang. *mean absolute error*)
- povprečna kvadratična napaka MSE (ang. *mean squared error*)
- negativni logaritem gostote napake LPD (ang. *minus log-predicted density error*)
- povprečna relativna kvadratična napaka MRSE (ang. *mean relative square error*)
- negativni logaritem verjetnostne porazdelitve učne množice (ang. *minus log-marginal likelihood*)

Kot je razvidno iz imena, mera MAE predstavlja povprečno odstopanje predvidenih vrednosti od ciljnih. MAE je definirana z izrazom:

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \quad (2.15)$$

kjer f_i predstavlja ciljno vrednost (target), y_i pa predvideno vrednost modela. Podobno mera MSE predstavlja povprečno kvadrirano odstopanje predvidenih vrednosti od ciljnih in je definirana z izrazom:

$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2 \quad (2.16)$$

V izračunu mere LPD, poleg razlike med odzivom procesa in srednjo vrednostjo

modela, nastopa tudi varianca predikcije. Tako mera LPD podaja informacijo o povprečni kvadratični napaki, normirani z vrednostjo variance predikcije. Uporablja se predvsem pri Bayesovem modeliranju, kakršno je tudi modeliranje na podlagi Gaussovih procesov. LPD je definirana z izrazom:

$$LPD = \frac{1}{2n} \sum_{i=1}^n \left(\log(2\pi) + \log(\sigma) + \frac{(f_i - y_i)^2}{\sigma} \right) \quad (2.17)$$

kjer f_i predstavlja ciljno vrednost (ang. *target*), y_i napovedano vrednost modela, σ pa varianco predikcije. MRSE je mera, katere vrednost je neodvisna od vrednosti podatkov in je definirana z izrazom:

$$MRSE = \sqrt{\frac{\sum_{i=1}^n (f_i - y_i)^2}{\sum_{i=1}^n (f_i)^2}} \quad (2.18)$$

Obstaja še ena možnost vrednotenja modela, tj. navzkrižno vrednotenje (ang. *cross-validation*). Navzkrižno vrednotenje izvedemo z vrednotenjem odziva modela na podatkih, ki niso bili uporabljeni za učenje modela. To pomeni, da moramo imeti na voljo vsaj dve različni množici podatkov. Ker te možnosti nimamo vedno na razpolago, si lahko pomagamo tudi tako, da obstoječo množico podatkov razdelimo na del za učenje in del za vrednotenje. Obstaja tudi n -kratno navzkrižno vrednotenje (ang. *n-fold cross validation*). Pri tej metodi vrednosti hiperparametrov iščemo kot običajno, le da učne podatke razbijemo na n delov. Za učenje uporabimo $n - 1$ delov, za vrednotenje pa tistega, ki ostane. Postopek ponovimo n -krat, vsakič z drugimi podatki za vrednotenje. Ekstremni primer je navzkrižno vrednotenje z izpuščanjem (ang. *leave one out - LOO cross-validation*). Največji problem tega postopka je računska zahtevnost, saj moramo naučiti n modelov in poiskati povprečne vrednosti hiperparametrov.

2.2.6 Programska podpora

Pri modeliranju smo uporabili programsko orodje v okolju Matlab [9]. Programsko orodje predstavlja dve temeljni funkciji: funkcija za optimizacijo hiperparametrov po metodi največje podobnosti in funkcija, ki na podlagi podanih hiperparametrov in učnih točk omogoča napovedovanje izhoda modela pri podanem vходу. Poleg omenjenih funkcij so na voljo še funkcije za izračun kovariančnih matrik, ki smo jih po potrebi dopolnili.

Poglavje 3

Kovariančne funkcije

Vloga kovariančne funkcije je pri modeliranju z Gaussovimi procesi zelo pomembna. Napovedane verjetnostne porazdelitve, ki nastopajo pri danih podatkih, so v glavnem odvisne od kovariančne funkcije in njenih hiperparametrov.

Kovariančna funkcija $C(\mathbf{x}_i, \mathbf{x}_j)$ izraža mero podobnosti med vhodoma $\mathbf{x}_i$ in $\mathbf{x}_j$. Za realne procese je navadno sestavljena iz dveh delov:

$$C(\mathbf{x}_i, \mathbf{x}_j) = C_f(\mathbf{x}_i, \mathbf{x}_j) + C_n(\mathbf{x}_i, \mathbf{x}_j) \quad (3.1)$$

Prvi, tj. *funkcijski* del $C_f(\mathbf{x}_i, \mathbf{x}_j)$, opisuje lastnosti neznanega sistema, ki ga želimo modelirati, drugi, *šumni* del $C_n(\mathbf{x}_i, \mathbf{x}_j)$ pa predstavlja varianco šuma. Pogosto predpostavljamo, da je šum naključen. Iz tega sledi, da ne pričakujemo nobene korelacije med šumom in določenimi izhodi, ter da vpliva samo na diagonalne elemente kovariančne matrike:

$$C_n(\mathbf{x}_i, \mathbf{x}_j) = \delta_{ij} \theta_n^2 \quad (3.2)$$

Z δ_{ij} je označena Kroneckerjeva delta funkcija:

$$\delta_{ij} = \begin{cases} 0; & i \neq j \\ 1; & i = j \end{cases} \quad (3.3)$$

Ker je vsota dveh nenegativno definitnih funkcij zopet nenegativno definitna funkcija, lahko funkciji $C_f(\mathbf{x}_i, \mathbf{x}_j)$ in $C_n(\mathbf{x}_i, \mathbf{x}_j)$ izbiramo ločeno. Kovariančne funkcije delimo na stacionarne in nestacionarne.

3.1 Stacionarne kovariančne funkcije

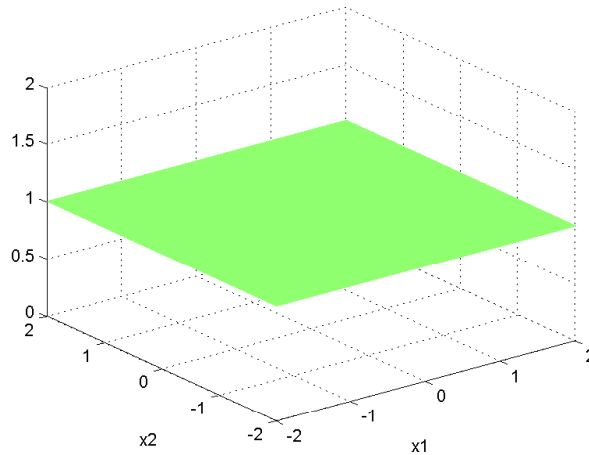
Stacionarne kovariančne funkcije $C_f(\mathbf{x}_i, \mathbf{x}_j)$ so tiste, pri katerih je vrednost funkcije odvisna samo od relativne lege vhodnih vektorjev $\mathbf{x}_i$ in $\mathbf{x}_j$ oz. od njune medsebojne razdalje: $r = |\mathbf{x}_i - \mathbf{x}_j|$.

3.1.1 Konstantna kovariančna funkcija

Najpreprostejša oblika kovariančne funkcije je funkcija, ki zavzame isto vrednost na celotnem območju. To je konstantna kovariančna funkcija, opisuje pa jo naslednji izraz:

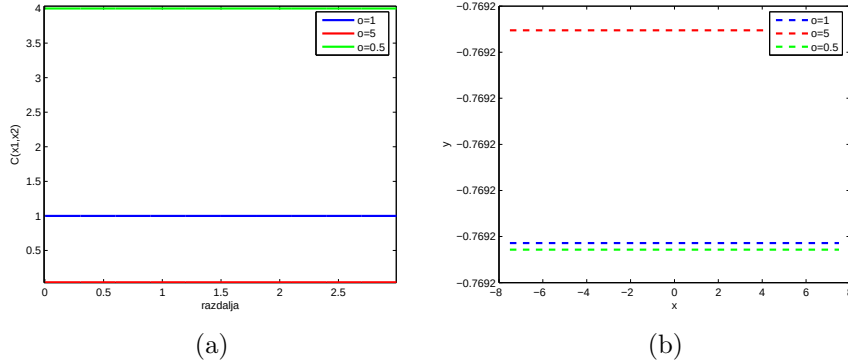
$$C_f(r) = \frac{1}{\theta_1^2} \quad (3.4)$$

Določena je z enim samim hiperparametrom θ_1 , ki predstavlja skalirni faktor variance učnih podatkov. Konstantno kovariančno funkcijo prikazuje slika 3.1. Vpliv hiperparametra θ_1 na poljubno funkcijo lahko prikažemo s preprostim



Slika 3.1: Konstantna kovariančna funkcija s hiperparametrom $\theta_1 = 1$

primerom. Na sliki 3.2 je prikazana konstantna kovariančna funkcija s podanimi različnimi vrednostmi hiperparametra θ_1 in vpliv le-teh na modeliranje preproste funkcije. Zaradi preprostosti konstantno kovariančno funkcijo, kot npr. funkcijo belega šuma (enačba 3.2), ponavadi uporabljamo v kombinaciji z drugimi kovariančnimi funkcijami. Opis kombiniranja različnih kovariančnih funkcij je opisan v poglavju 3.3.



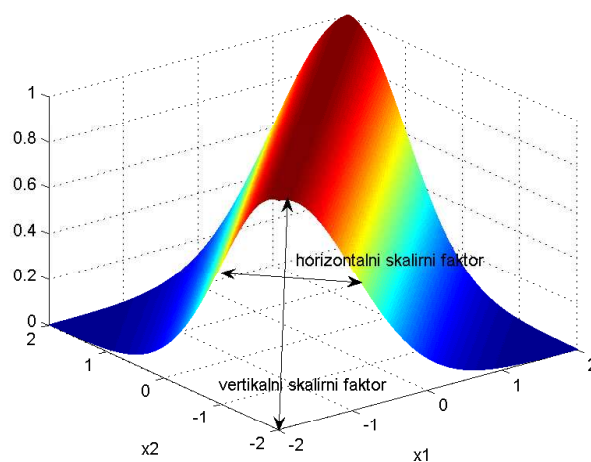
Slika 3.2: Konstantna kovariančna funkcija pri različnih vrednostih hiperparametrov (a) in z njo modelirana poljubno izbrana funkcija (b)

3.1.2 Gaussova kovariančna funkcija

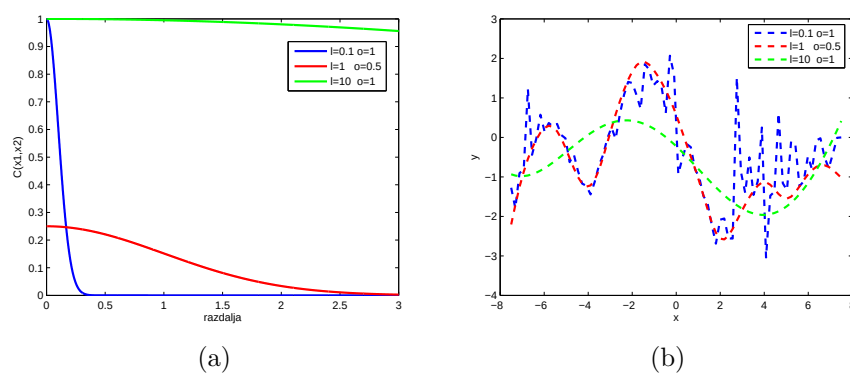
Pri predpostavki stacionarnosti (kovarianca med dvema točkama je odvisna samo od medsebojne razdalje in ne tudi od premika v prostoru) in gladkosti funkcije, je najpogosteje uporabljena funkcija. Gaussova kovariančna funkcija je neskončno mnogokrat odvedljiva. Rezultat modeliranja s to funkcijo je gladka krivulja. Gaussovo kovariančno funkcijo opisuje naslednji izraz:

$$C_f(r) = \theta_1^2 \exp\left(-\frac{r^2}{2l^2}\right) \quad (3.5)$$

Hiperparameter θ_1 je skalirni faktor variance učnih podatkov oz. vertikalni skalirni faktor, hiperparameter l (horizontalni skalirni faktor) pa določa relativno vlogo (utež) razdalje po vhodni spremenljivki $\mathbf{x}$ pri celotni vrednosti kovariance. Gaussovo kovariančno funkcijo prikazuje slika 3.3. Gladkost lahko demonstriramo z risanjem preprostih funkcij. Na sliki 3.4 je prikazana Gaussova kovariančna funkcija z različnimi vrednostmi hiperparametrov in vpliv le-teh na modeliranje preproste funkcije. Vidimo, da se z večanjem horizontalnega skalirnega faktorja več gladkost funkcije, medtem ko se z večanjem skalirnega faktorja variance več amplituda funkcije. Na splošno uporaba Gaussove kovariančne funkcije pomeni predpostavko, da funkcija, ki jo želimo identificirati, kaže gladko in neprekinjeno obnašanje z močno korelacijo med izhodi in vhodi, ki so si blizu.



Slika 3.3: Gaussova kovariančna funkcija s hiperparametroma $\theta_1 = 1$ in $l = 1$



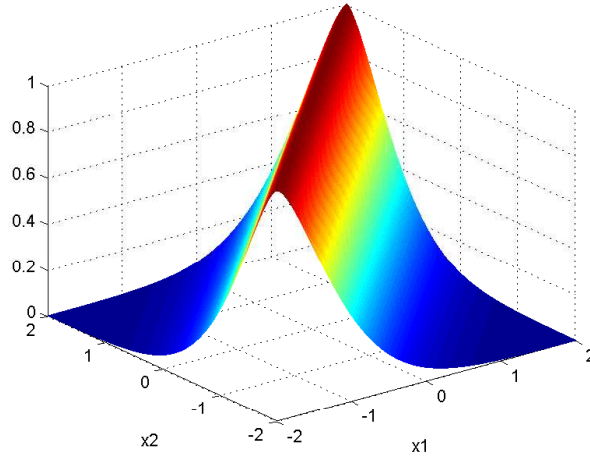
Slika 3.4: Gaussova kovariančna funkcija pri različnih vrednostih hiperparametrov (a) in z njo modelirana poljubno izbrana funkcija (b)

3.1.3 Eksponentna kovariančna funkcija

Eksponentno kovariančno funkcijo opisuje naslednji izraz:

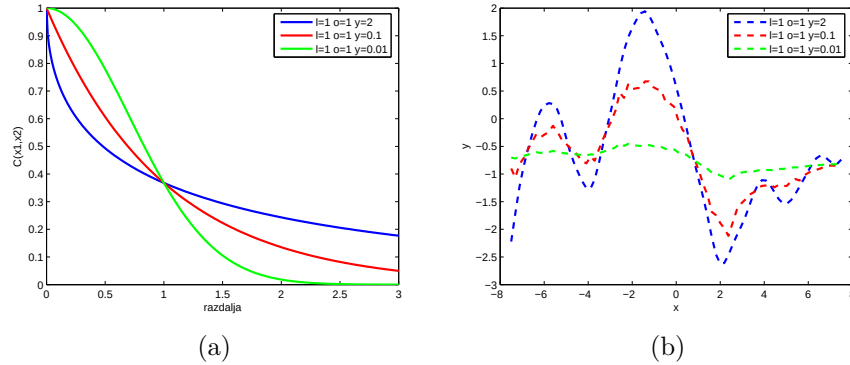
$$C_f(r) = \theta_1^2 \exp\left(-\left(\frac{r}{l}\right)^\gamma\right) \text{ za } 0 < \gamma \leq 2 \quad (3.6)$$

Hiperparameter θ_1 je skalirni faktor variance učnih podatkov, hiperparameter l (horizontalni skalirni faktor) določa relativno vlogo (utež) razdalje po vhodni spremenljivki $\mathbf{x}$ pri celotni vrednosti kovariance, hiperparameter γ pa določa vrednost eksponenta. Eksponentno kovariančno funkcijo prikazuje slika 3.5. Na sliki 3.6 je prikazana eksponentna kovariančna funkcija z različnimi



Slika 3.5: Eksponentna kovariančna funkcija s hiperparametri $\theta_1 = 1$, $l = 1$ in $\gamma = 1.5$

vrednostmi hiperparametrov in vpliv le-teh na modeliranje preproste funkcije. Ker je bil vpliv vertikalnega in horizontalnega skalirnega faktorja obravnavan že pri Gaussovi kovariančni funkciji, smo tukaj obravnavali le vpliv spreminjanja vrednosti eksponenta γ . Vidimo, da večanje vrednosti eksponenta γ poveča amplitudo modelirane funkcije. Eksponentna kovariančna funkcija je ekvivalentna predhodno opisani Gaussovi kovariančni funkciji, ko eksponent γ zavzame vrednost $\gamma = 2$. Kljub temu pa je neskončno mnogokrat odvedljiva le za $\gamma = 2$, zato je ta funkcija manj prilagodljiva in ne tako pogosto uporabljena [15], poleg tega je izhod iz GP modela manj gladka funkcija. Največkrat se uporabi pri vrednosti $\gamma = 1$ [1].



Slika 3.6: Eksponentna kovariančna funkcija pri različnih vrednostih hiperparametrov (a) in z njo modelirana poljubno izbrana funkcija (b)

3.1.4 Racionalni kvadratični polinom

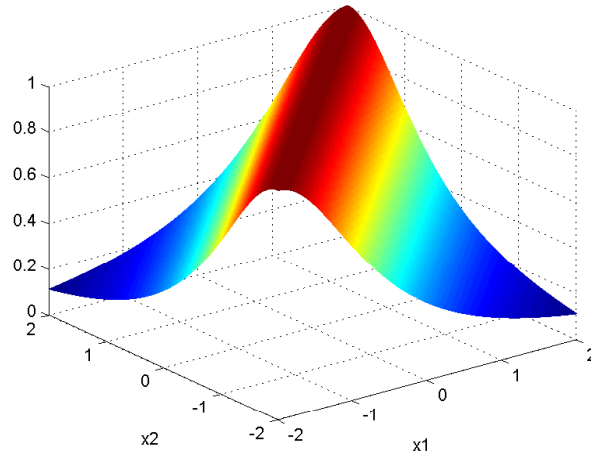
Racionalni kvadratični polinom lahko smatramo kot *skalirno mešanico* oz. neskončno vsoto Gaussovih kovariančnih funkcij z različnimi vrednostmi horizontalnega skalirnega faktorja [15]. Prav tako kot Gaussova, je tudi ta funkcija neskončno mnogokrat odvedljiva. Racionalni kvadratični polinom opisuje naslednji izraz:

$$C_f(r) = \theta_1^2 \left(1 + \frac{r^2}{2\alpha l^2} \right)^{-\alpha} \quad (3.7)$$

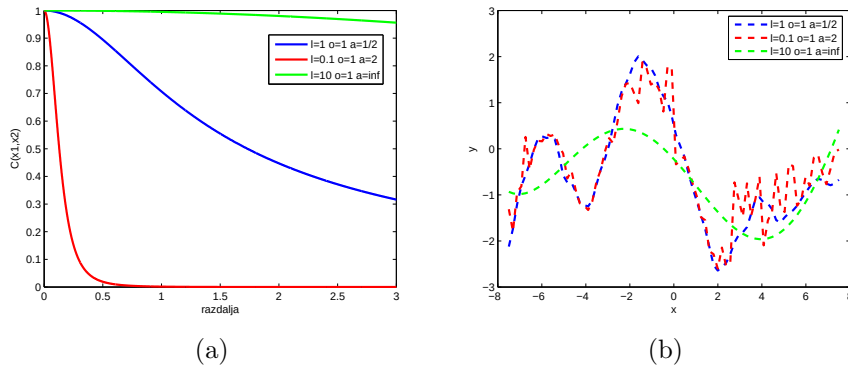
Hiperparameter θ_1 je skalirni faktor variance učnih podatkov, hiperparameter l (horizontalni skalirni faktor) določa relativno vlogo (utež) razdalje po vhodni spremenljivki $\mathbf{x}$ pri celotni vrednosti kovariance, α je pozitiven hiperparameter. Pri $\alpha \rightarrow \infty$ racionalni kvadratični polinom postane enak Gaussovi kovariančni funkciji. Racionalni kvadratični polinom prikazuje slika 3.7. Na sliki 3.8 je prikazan racionalni kvadratični polinom z različnimi vrednostmi hiperparametrov in vpliv le-teh na modeliranje preproste funkcije. Prednost racionalnega kvadratičnega polinoma v primerjavi z Gaussovo kovariančno funkcijo je, da lahko vsebuje več horizontalnih skalirnih faktorjev.

3.1.5 Matérnova kovariančna funkcija

Matérnova kovariančna funkcija omogoča nadzor nad relativno gladkostjo in odvedljivostjo modelirane funkcije. Opisuje jo naslednji izraz:



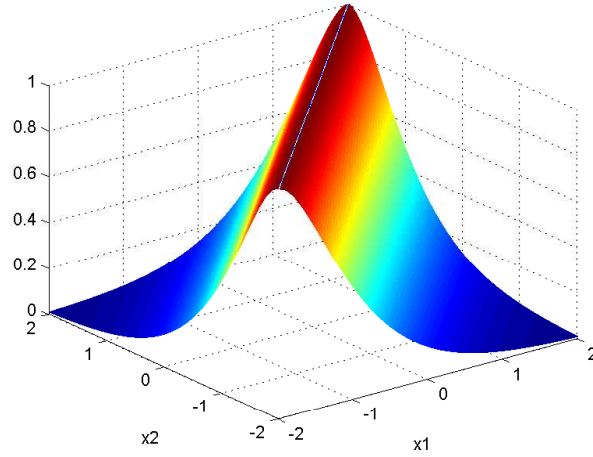
Slika 3.7: Racionalni kvadratični polinom s hiperparametri $\theta_1 = 1$, $l = 1$ in $\alpha = 1$



Slika 3.8: Racionalni kvadratični polinom pri različnih vrednostih hiperparametrov (a) in z njo modelirana poljubno izbrana funkcija (b)

$$C_f(r) = \theta_1^2 \left(\frac{2^{1-\nu}}{\Gamma(\nu)} \right) \left(\frac{\sqrt{2\nu}r}{l} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}r}{l} \right) \quad (3.8)$$

Hiperparameter θ_1 je skalirni faktor variance učnih podatkov, hiperparameter l (horizontalni skalirni faktor) določa relativno vlogo (utež) razdalje po vhodni spremenljivki $\mathbf{x}$ pri celotni vrednosti kovariance, K_ν predstavlja modificirano Besselovo funkcijo, hiperparameter ν pa služi nadzoru odvedljivosti poljubne funkcije in s tem vpliva na gladkost. Funkcija je k -krat odvedljiva, če in samo če $\nu > k$. Pri $\nu \rightarrow \infty$ Matérnova kovariančna funkcija postane enaka Gaussovi. Matérnovo kovariančno funkcijo prikazuje slika 3.9. Funkcija postane še pose-



Slika 3.9: Matérnova kovariančna funkcija s hiperparametri $\theta_1 = 1$, $l = 1$ in $\nu = 3/2$

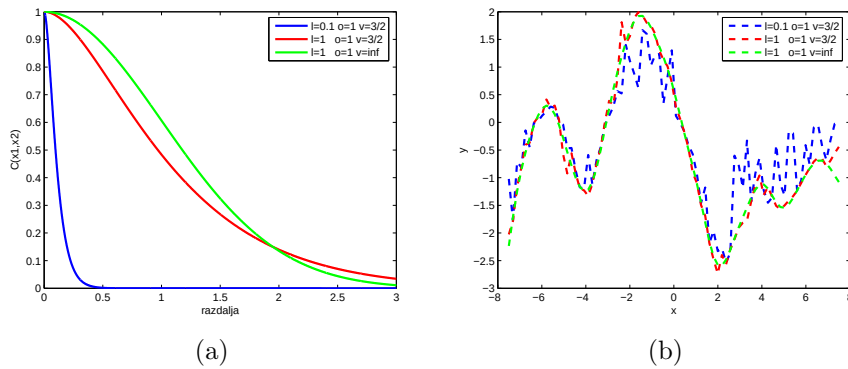
bej preprosta pri $\nu = p + 1/2$, kjer je p nenegativno celo število. V strojnem učenju se največkrat uporabljata $\nu = 3/2$ in $\nu = 5/2$:

$$C_{f_{\nu=3/2}}(r) = \left(1 + \frac{\sqrt{3}r}{l} \right) \exp \left(- \frac{\sqrt{3}r}{l} \right) \quad (3.9)$$

$$C_{f_{\nu=5/2}}(r) = \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2} \right) \exp \left(- \frac{\sqrt{5}r}{l} \right), \quad (3.10)$$

medtem ko proces pri $\nu = 1/2$ (v tem primeru eksponentna kovariančna funkcija za $\gamma = 1$) postane zelo grob [15], pri $\nu \geq 7/2$ pa je pri končni množici

učnih podatkov težko razlikovati med ν in $\nu \rightarrow \infty$ (v tem primeru Gaussova kovariančna funkcija). Vpliv na gladkost lahko demonstriramo z risanjem preprostih funkcij. Na sliki 3.10 je prikazana Matérnova kovariančna funkcija z različnimi hiperparametri in vpliv le-teh na modeliranje preproste funkcije. Vidimo, da manjšanje hiperparametra ν naredi funkcijo manj gladko. V primerjavi z Gaussovo kovariančno funkcijo, kovarianca lahko pada precej hitreje, kar povzroči znatno zmanjšanje gladkosti. Pri uporabi Gaussove kovariančne



Slika 3.10: Matérnova kovariančna funkcija pri različnih vrednostih hiperparametrov (a) in z njo modelirana poljubno izbrana funkcija (b)

funkcije predpostavimo gladkost modelirane funkcije. Če glede na predznake o sistemu ugotovimo, da funkcija ni gladka, raje uporabimo Matérnovo kovariančno funkcijo.

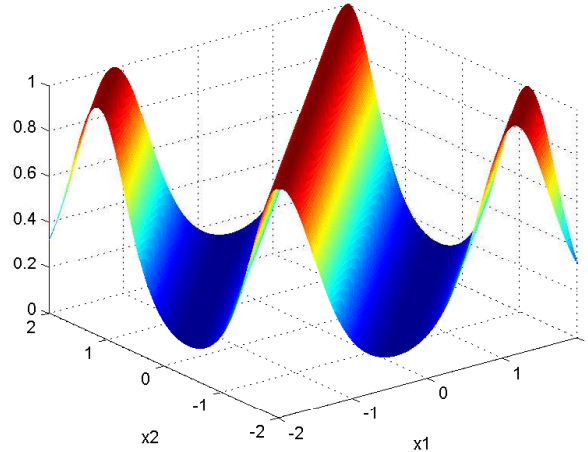
3.1.6 Periodična kovariančna funkcija

Funkcije, ki jih opisuje periodična kovariančna funkcija nimajo močne korelacije samo med točkami, ki so si blizu, ampak tudi med točkami, ki so med seboj oddaljene za neko dolžino (periodo). Primer periodične kovariančne funkcije opisuje naslednji izraz:

$$C_f(r) = \theta_1^2 \exp\left(-\frac{2(\sin(\frac{\pi}{\omega}r))^2}{l^2}\right), \quad (3.11)$$

vzet po [15], ki smo mu dodali periodo. Hiperparameter θ_1 je skalirni faktor variance učnih podatkov, hiperparameter l (horizontalni skalirni faktor) določa relativno vlogo (utež) razdalje po vhodni spremenljivki $\mathbf{x}$ pri celotni vrednosti kovariance, hiperparameter ω pa določa periodo. Periodično kovariančno

funkcijo prikazuje slika 3.11. Na sliki 3.12 je prikazana periodična kovariančna



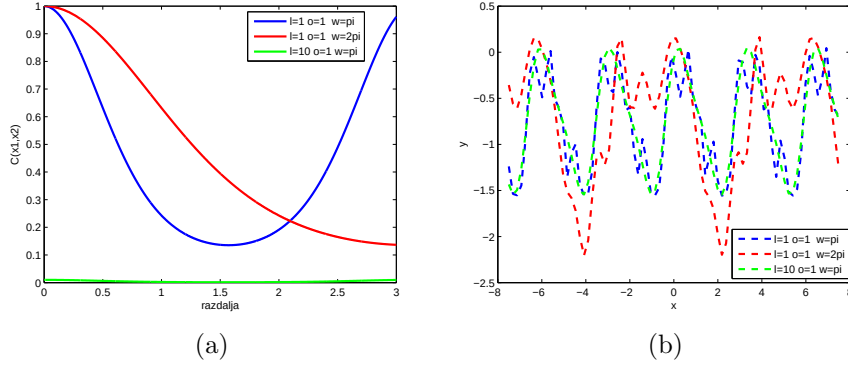
Slika 3.11: Periodična kovariančna funkcija s hiperparametri $\theta_1 = 1$, $l = 1$ in $\omega = \pi$

funkcija z različnimi vrednostmi hiperparametrov in vpliv le-teh na modeliranje preproste funkcije. Vidimo, da je perioda modelirane funkcije enaka vrednosti parametra ω , medtem ko horizontalni skalirni faktor določa stopnjo gladkosti znotraj periode; večji kot je, bolj je funkcija gladka. Pri večjem odmiku stran od znanih podatkov napovedi z uporabo Gaussove kovariančne funkcije težijo h konstantni vrednosti. Kadar pa v nasprotju s tem predvidevamo, da obstaja neka ponavljajoča težnja v množici podatkov, je primernejša uporaba periodične kovariančne funkcije. Ta omogoča periodično modeliranje tudi izven obsega učnih podatkov (ekstrapolacija) pri predpostavki, da se modelirana funkcija nadaljuje v neskončnost z enako periodo.

3.2 Nestacionarne kovariančne funkcije

Medtem ko je mogoče veliko večino množic podatkov učinkovito modelirati z uporabo stacionarnih kovariančnih funkcij, obstajajo določeni primeri, v katerih je potrebno uporabiti nestacionarne funkcije.

Nestacionarne funkcije so tiste, katerih vrednost $C_f(\mathbf{x}_i, \mathbf{x}_j)$ je odvisna od absolutne lege vhodnih vektorjev $\mathbf{x}_i$ in $\mathbf{x}_j$ v prostoru. Uporabimo jih kadar predpostavimo nestacionarnost procesa (kovarianca med dvema točkama je odvisna tako od medsebojne razdalje, kot tudi od premika v prostoru). V



Slika 3.12: Periodična kovariančna funkcija pri različnih vrednostih hiperparametrov (a) in z njo modelirana poljubno izbrana funkcija (b)

nadaljevanju so opisane najpogostejše uporabljene nestacionarne kovariančne funkcije.

3.2.1 Linearna kovariančna funkcija

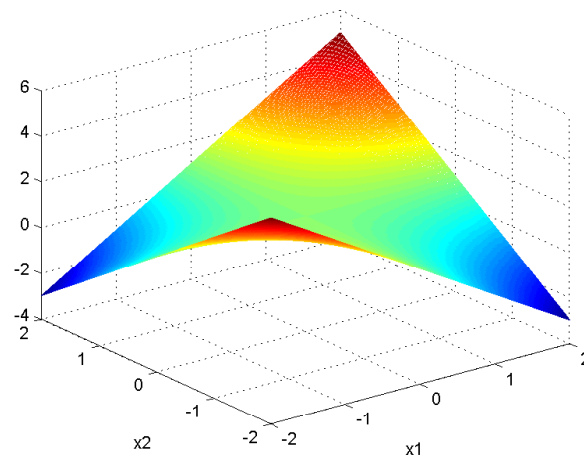
Linearna kovariančna funkcija je primer funkcije, sestavljene iz skalarnega produkta dveh vektorjev $\mathbf{x}_i \cdot \mathbf{x}_j$. Opisuje jo naslednji izraz:

$$C_f(\mathbf{x}_i, \mathbf{x}_j) = \theta_1^2 \mathbf{x}_i \cdot \mathbf{x}_j + \theta_2^2 \quad (3.12)$$

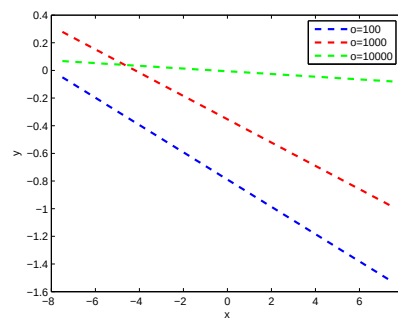
Hiperparameter θ_1 je skalirni faktor variance učnih podatkov, θ_2 pa določa vertikalno negotovost. V primeru ko je $\theta_1^2 = 0$, je funkcija homogena. Uporablja pa se tudi oblika linearne kovariančne funkcije, kjer je $\theta_1 = \theta_2$. Linearno kovariančno funkcijo prikazuje slika 3.13. Na sliki 3.14 je prikazana linearna kovariančna funkcija z različnimi vrednostmi hiperparametrov in vpliv le-teh na modeliranje preproste funkcije. Pri večjem odmiku stran od znanih podatkov, napovedi z uporabo Gaussove kovariančne funkcije težijo h konstantni vrednosti. Kadar predvidevamo, da obstaja neka linearna težnja v množici podatkov, lahko v model dodamo linearno komponento z uporabo linearne kovariančne funkcije.

3.2.2 Polinomska kovariančna funkcija

Polinomska kovariančna funkcija predstavlja razširitev linearne kovariančne funkcije in je definirana z izrazom:



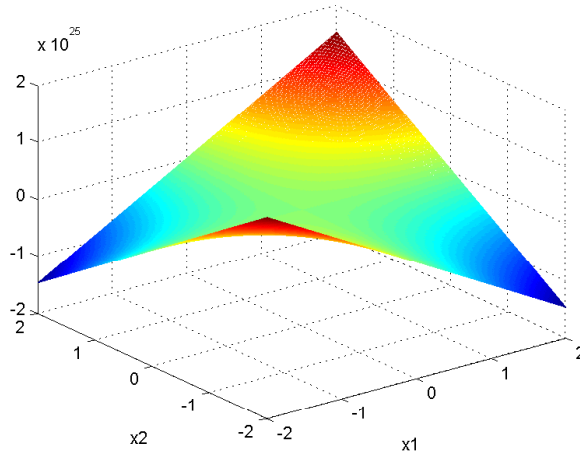
Slika 3.13: Linearna kovariančna funkcija s hiperparametri $\theta_1 = \theta_2 = 1$



Slika 3.14: Z linearno kovariančno funkcijo narisana poljubno izbrana funkcija pri različnih vrednostih hiperparametrov

$$C_f(\mathbf{x}_i, \mathbf{x}_j) = (\theta_1^2 \mathbf{x}_i \cdot \mathbf{x}_j + \theta_2^2)^\nu \quad (3.13)$$

Hiperparameter θ_1 je skalirni faktor variance učnih podatkov, θ_2 določa vertikalno negotovost, hiperparameter ν pa predstavlja stopnjo polinoma. Tudi v primeru uporabe te funkcije se lahko uporablja oblika, kjer je $\theta_1 = \theta_2$. Polinomska kovariančna funkcijo prikazuje slika 3.15. Na sliki 3.16 je prikazana

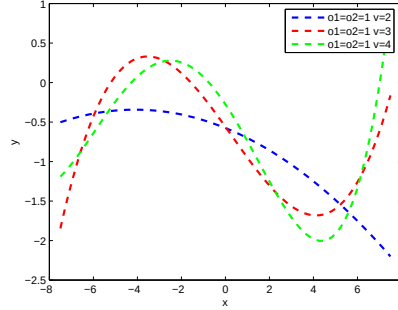


Slika 3.15: Polinomska kovariančna funkcija s hiperparametri $\theta_1 = \theta_2 = 1$ in $\nu = 10$

polinomska kovariančna funkcija z različnimi vrednostmi hiperparametrov in vpliv le-teh na modeliranje preproste funkcije. Vidimo, da spreminjanje hiperparametra ν vpliva na stopnjo modelirane funkcije. Izbira polinomske kovariančne funkcije je za regresijske probleme rahlo neprimerna, saj varianca hitro narašča z $|\mathbf{x}|$ za $|\mathbf{x}| > 1$, zato pa so toliko bolj uporabne pri klasifikaciji [15]. Kljub temu lahko polinomsko kovariančno funkcijo porabimo pri ekstrapolaciji, kadar predvidevamo, da obstaja neka polinomska funkcija podobna težnja v množici podatkov.

3.2.3 Kovariančna funkcija nevronske mreže

Do sedaj smo obravnavali le primere kovariančnih funkcij, sestavljenih iz skalarne produkta dveh vektorjev $\mathbf{x}_i$ in $\mathbf{x}_j$. Obstajajo tudi primerne kovariančne

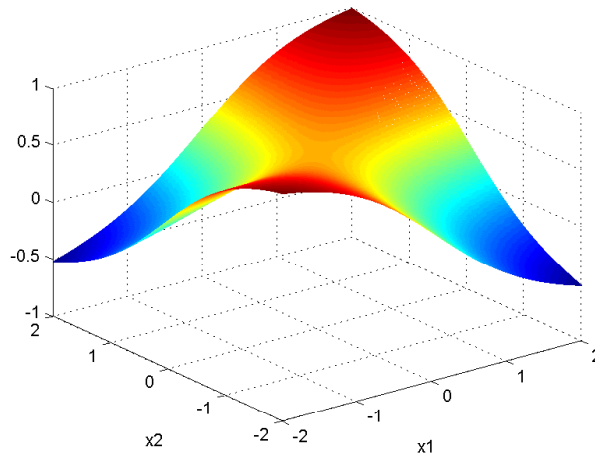


Slika 3.16: S polinomsko kovariančno funkcijo narisana poljubno izbrana funkcija pri različnih vrednostih hiperparametrov

funkcije drugih oblik. Ena izmed njih je kovariančna funkcija nevronske mreže [15], ki jo opisuje naslednji izraz:

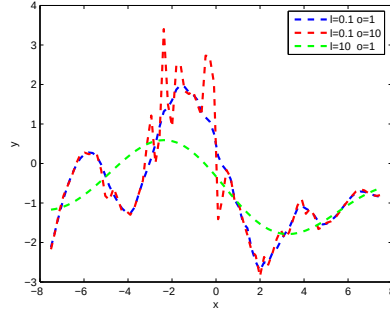
$$C_f(\mathbf{x}_i, \mathbf{x}_j) = \theta_1^2 \frac{2}{\pi} \sin^{-1} \left(\frac{2\tilde{\mathbf{x}}_i^\top \sum \tilde{\mathbf{x}}_j}{\sqrt{1 + 2\tilde{\mathbf{x}}_i^\top \sum \tilde{\mathbf{x}}_i} \sqrt{1 + 2\tilde{\mathbf{x}}_j^\top \sum \tilde{\mathbf{x}}_j}} \right), \quad (3.14)$$

kjer $\tilde{x}_i = (1, x_1, \dots, x_d)^\top$, $\sum$ pa predstavlja enotsko matriko pomnoženo z l^{-2} . Kovariančno funkcijo nevronske mreže prikazuje slika 3.17. Podobno kot pri



Slika 3.17: Kovariančna funkcija nevronske mreže s hiperparametrom $\theta_1 = 1$
ostalih kovariančnih funkcijah, tudi pri tej vključimo skalirni faktor variance

učnih podatkov θ_1 in horizontalni skalirni faktor l . Vpliv različnih vrednosti omenjenih hiperparametrov na preprosto funkcijo je prikazan na sliki 3.18. Vidimo, da povečanje horizontalnega skalirnega faktorja povzroči hitrejšo spremembo funkcije.



Slika 3.18: S kovariančno funkcijo nevronske mreže narisan poljubna funkcija pri različnih vrednostih hiperparametrov

3.2.4 Vhodno odvisen model šuma

Predhodno smo obravnavali samo model šuma, ki je bil neodvisen od vhodov. Povsem smiselno pa je pričakovati, da se lahko stopnja šuma spreminja tudi v odvisnosti od vhodnih podatkov. Vgradnja take odvisnosti v GP model je povsem naravna in je lahko definirana z naslednjim izrazom:

$$\theta_2(\mathbf{x}_m; \Theta) = e^{(\sum_{j=1}^J \beta_j \Phi_j(\mathbf{x}_m))} \quad (3.15)$$

kjer $\Phi_j(\mathbf{x})$ predstavlja množico baznih funkcij in $\beta_j \in \Theta$ primerne koeficiente. Tak model šuma ni stacionaren, je pa pozitivno definiten, ker prispeva samo k diagonalnim elementom kovariančne matrike. Primer uporabe se nahaja v [4].

3.2.5 Prostorsko spremenljiva dolžinska merila

V običajnih stacionarnih funkcijah predpostavljamo, da je v posamezni smeri vhodnega vektorja horizontalni skalirni faktor konstanten. Lahko si predstavljamo primer slabega modela glede na podatke. Če npr. določimo, da l_d postanejo funkcije $\mathbf{x}$, ne moremo pa preprosto nadomestiti parametrizirane

oblike za l_d , saj s tem ne dobimo splošne pozitivne definitne kovariančne funkcije. Primer take kovariančne funkcije je podrobneje obdelan v [15] in ga opisuje funkcija:

$$C_f(\mathbf{x}_i, \mathbf{x}_j) = \prod_{d=1}^D \left(\frac{2l_d(x_i)l_d(x_j)}{l_d^2(x_i) + l_d^2(x_j)} \right)^{1/2} e^{-\sum_{d=1}^D \left(\frac{(x_{i_d} - x_{j_d})^2}{l_d^2(x_i) + l_d^2(x_j)} \right)} \quad (3.16)$$

3.3 Izgradnja novih kovariančnih funkcij iz obstoječih

Glede na lastnosti Gaussovega procesa lahko kombiniramo različne kovariančne funkcije in tako dobimo nove stacionarne in nestacionarne kovariančne funkcije. Glede na predznanje o podatkih jih lahko med seboj poljubno seštevamo, množimo in spreminjamo, kar omogoča izgradnjo čim boljšega modela. V nadaljevanju bomo predstavili, kako lahko kombiniramo in spreminjamo različne kovariančne funkcije in tako dobimo nove.

Vsota

Ker je edini pogoj za kovariančno funkcijo to, da zagotovi pozitivno definitnost kovariančne matrike in ker je vsota dveh pozitivno definitnih kovariančnih funkcij $C_1(\mathbf{x}_i, \mathbf{x}_j)$ in $C_2(\mathbf{x}_i, \mathbf{x}_j)$ zopet pozitivno definitna kovariančna funkcija $C(\mathbf{x}_i, \mathbf{x}_j) = C_1(\mathbf{x}_i, \mathbf{x}_j) + C_2(\mathbf{x}_i, \mathbf{x}_j)$, lahko nove funkcije poljubno sestavljamo iz že obstoječih. Tak primer je že bil predstavljen pri obravnavi šuma. Ta lastnost omogoča združevanje funkcij, ki imajo različne horizontalne skalirne faktorje.

Produkt

Podobno kot vsota, je tudi produkt dveh pozitivno definitnih kovariančnih funkcij $C_1(\mathbf{x}_i, \mathbf{x}_j)$ in $C_2(\mathbf{x}_i, \mathbf{x}_j)$ zopet pozitivno definitna kovariančna funkcija $C(\mathbf{x}_i, \mathbf{x}_j) = C_1(\mathbf{x}_i, \mathbf{x}_j) \cdot C_2(\mathbf{x}_i, \mathbf{x}_j)$. Lastnost lahko posplošimo na produkt večih kovariančnih funkcij, iz česar sledi, da je $C^p(\mathbf{x}_i, \mathbf{x}_j)$ veljavna kovariančna funkcija za $p \in \mathbb{N}$.

Vertikalno skaliranje

Pri tej operaciji se stacionarna kovariančna funkcija pretvori v nestacionarno. Naj bo $g(\mathbf{x}_i) = a(\mathbf{x}_i)f(\mathbf{x}_i)$, kjer je $a(\mathbf{x}_i)$ deterministična funkcija in $f(\mathbf{x}_i)$ naključni proces. Novo kovariančno funkcijo predstavlja zapis $cov(g(\mathbf{x}_i), g(\mathbf{x}_j)) = a(\mathbf{x}_i)k(\mathbf{x}_i, \mathbf{x}_j)a(\mathbf{x}_j)$. Metodo lahko uporabimo za normalizacijo kovariančnih funkcij tako, da izberemo $a(\mathbf{x}_i) = k^{-1/2}(\mathbf{x}_i, \mathbf{x}_i)$, kjer predpostavimo $k(\mathbf{x}_i, \mathbf{x}_i) > 0 \forall \mathbf{x}_i$ in dobimo:

$$k'(\mathbf{x}_i, \mathbf{x}_j) = \frac{k(\mathbf{x}_i, \mathbf{x}_j)}{\sqrt{k(\mathbf{x}_i, \mathbf{x}_i)}\sqrt{k(\mathbf{x}_j, \mathbf{x}_j)}} \quad (3.17)$$

Izraz zagotovi $k'(\mathbf{x}_i, \mathbf{x}_i) = 1 \forall \mathbf{x}_i$ [15].

Konvolucija

Novo veljavno kovariančno funkcijo lahko dobimo tudi s konvolucijo. Tudi pri tej metodi se stacionarna kovariančna funkcija pretvori v nestacionarno. Če je $h(\mathbf{x}_i, \mathbf{z}_i)$ kovariančna funkcija in $g(x) = \int h(\mathbf{x}_i, \mathbf{z}_i)f(\mathbf{z}_i) d\mathbf{z}_i$ preslikava, potem zapis $cov(g(\mathbf{x}_i), g(\mathbf{x}_j)) = \int \int h(\mathbf{x}_i, \mathbf{z}_i)k(\mathbf{z}_i, \mathbf{z}_j)h(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{z}_i d\mathbf{z}_j$ predstavlja novo kovariančno funkcijo [15].

3.4 Avtomatsko določanje ustreznosti

Do sedaj smo obravnavali le kovariančne funkcije, ki dajo enako težo vsem komponentam vhodnega vektorja $\mathbf{x}$ oz. so primerne le za eno vhodno spremenljivko. Pristop, ki omogoča obravnavo več vhodnih spremenljivk, imenujemo avtomatsko določanje ustreznosti (ang. *automatic relevance determination - ARD*) in bo opisan v nadaljevanju.

Uporaba ARD pristopa pride v poštev pri vseh, tako stacionarnih kot tudi nestacionarnih, kovariančnih funkcijah. Pristop ARD izkoristi obliko kovariančne funkcije za ocenitev relativne pomembnosti prispevka posameznega vhoda glede na primerjavo skalirnih faktorjev. Pri optimizaciji GP modela, kjer se določijo optimalne vrednosti hiperparametrov, lahko uporabimo pristop ARD in tako pripomoremo k optimizaciji strukture modela. Pristop ARD ima posebej velik prispevek pri identifikaciji sistemov, kjer imamo pomanjkanje predznanja glede na naravo primernih regresorjev - vhodov v model. To je velika prednost modeliranja z GP, saj pripomore k razumevanju vpliva različnih

vhodov na model. S praktičnega vidika lahko pristop ARD izkoristimo za izbiro celotne strukture obravnavanega modela, kjer lahko nepomembne vhode eliminiramo in tako izboljšamo računsko učinkovitost modela ter navsezadnje tudi zmožnost interpretacije sistema. Umestitev ARD pristopa v modeliranje z Gaussovimi procesi bomo prikazali na primeru Gaussove kovariančne funkcije (3.5). Funkcijo lahko izvedemo v neizotropični obliki, kjer vsaki komponenti vhodnega vektorja, namesto enakega hiperparametra l , pripišemo drug hiperparameter l_d in s tem poskrbimo za nadzor nad horizontalnim skalirnim faktorjem. Novo obliko Gaussove kovariančne funkcije lahko zapišemo kot:

$$C_f(\mathbf{x}_i, \mathbf{x}_j) = \theta_1^2 e^{\left(-\frac{1}{2} \sum_{d=1}^D \frac{(x_{i_d} - x_{j_d})^2}{l_d^2}\right)} \quad (3.18)$$

kjer je x_{i_d} d -ta komponenta vektorja D -dimenzionalnega vektorja x_i in l_d hiperparameter pripadajoče komponente. Na l_d lahko gledamo kot razdalja v tej določeni smeri, na kateri pričakujemo, da se vrednost izhoda znatno spreminja. Povedano drugače: večji ko je l_d , manjši vpliv ima d -ta komponenta vhodnega vektorja na izhod. Več o pristopu ARD je napisano v [15].

Poglavje 4

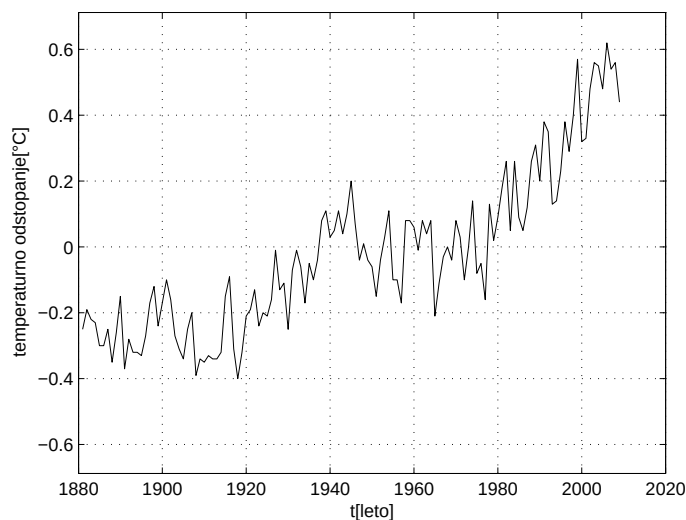
Primeri modeliranja časovnih vrst

4.1 Modeliranje globalnega segrevanja

V tem poglavju bo predstavljena uporaba različnih kovariančnih funkcij in njihovih kombinacij na primeru modeliranja globalnega segrevanja. Poleg same izbire kovariančne funkcije je pri postavitvi modela pomembna tudi določitev začetnih vrednosti za učenje hiperparametrov. Te sicer z metodo največje podobnosti postavimo na optimalno vrednost, a smo zaradi možnega padca v lokalni minimum (pri metodi konjugiranih gradientov) napravili več iteracij optimizacije pri naključnih vrednostih hiperparametrov in tako skušali dobiti optimalen minimum. Uporaba GP modelov je prikazana na primerih interpolacije in ekstrapolacije s poskusom napovedovanja prihodnosti, ki pa naj služi zgolj kot zanimivost, saj ga je težko primerno ovrednotiti.

4.1.1 Opis podatkov

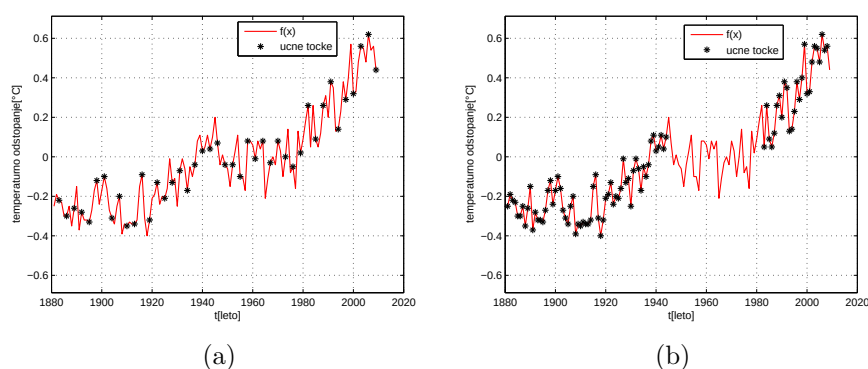
Podatki predstavljajo letna povprečja temperaturnega odstopanja od povprečnih vrednosti temperatur oceanov in zemeljskega površja [8]. Povprečja odstopanj so izmerjena za leta med 1880 in 2009 in so prikazana na sliki 4.1. Odstopanje je izračunano glede na povprečje temperature oceanov in zemeljskega površja v letih 1951-1980. Razlog za uporabo odstopanj namesto absolutnih vrednosti je ta, da se absolutna temperatura spreminja bolj izrazito v manjšem območju, medtem ko temperaturno odstopanje lahko ustrezno predstavi tudi večje območje. Temperaturne anomalije so namreč močno korelirane izven radija 1000 km, za absolutne temperature pa to ne velja.



Slika 4.1: Odstopanje temperature površja in oceanov v letih od 1880 do 2009

4.1.2 Interpolacija

Uporaba različnih kovariančnih funkcij za interpolacijo je prikazana na dveh primerih učnih podatkov. Pri prvem smo kot učno množico določili tretjino vseh podatkov, enakomerno porazdeljenih na celotnem intervalu (slika 4.2a). Pri drugem primeru pa smo učne podatke razdelili na dva intervala, kjer prvi zajema prvo polovico podatkov, drugi pa zadnjo petino podatkov (slika 4.2b). Preostale točke smo uporabili za vrednotenje.

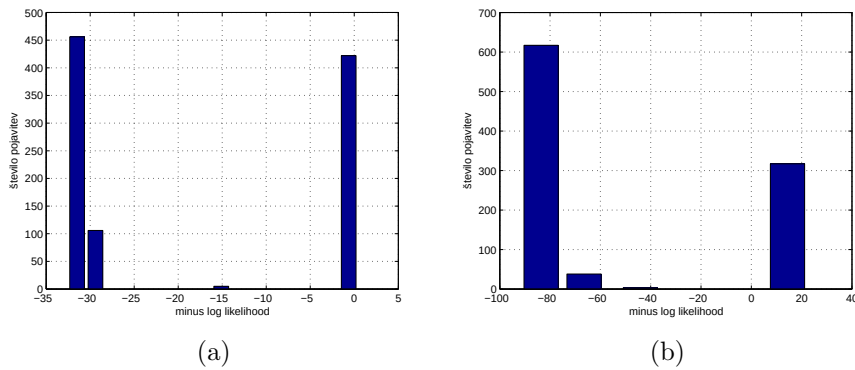


Slika 4.2: Učne točke pri interpolaciji

Prvi korak pri modeliranju na podlagi Gaussovih procesov je izbira primerne kovariančne funkcije, s katero je tvorjena kovariančna matrika. V nadaljevanju je prikazano modeliranje z različnimi kovariančnimi funkcijami, vrednotenje modelov in postopek določitve najprimernejše kovariančne funkcije ter optimalnih hiperparametrov.

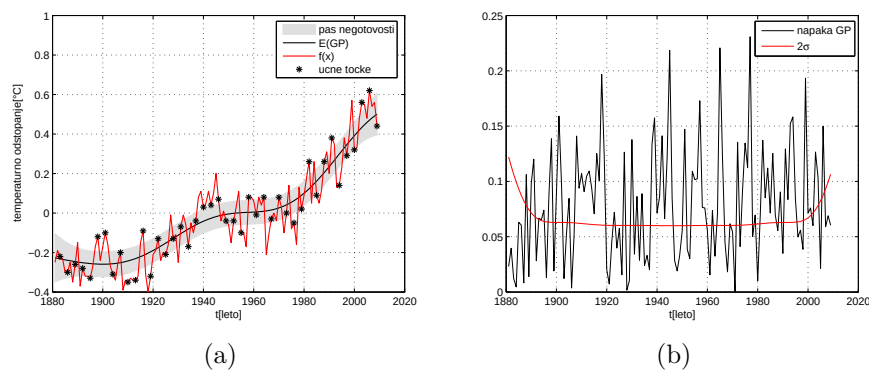
Zaradi narave podatkov, ki ne kažejo nikakršne povezanosti med seboj, smo sprva uporabili **Gaussovo kovariančno funkcijo** (3.5), saj je najprimernejša za uporabo, ko o podatkih ne vemo nič. Poleg tega podatke zgladi in omogoča boljšo predstavo o njih.

Sledila je določitev optimalnih vrednosti hiperparametrov θ_1 (skalirni faktor variance učnih podatkov), l (horizontalni skalirni faktor) in θ_n (varianco šuma). Začetne vrednosti le-teh smo nastavili z naključnim generatorjem, po optimizaciji z metodo največje podobnosti pa dobili vrednosti $l = 29.3209$, $\theta_1 = 0.2881$ in $\theta_n = 0.0916$ za prvi primer učne množice (slika 4.2a) in $l = 23.4530$, $\theta_1 = 0.2978$ in $\theta_n = 0.0835$ za drugi (slika 4.2b). Da bi se izognili padcu v lokalni minimum, smo naredili 1000 iteracij optimizacije. Pri tem smo dobili štiri lokalne minimume. Število pojavov posameznega minimuma prikazuje slika 4.3. Dejanska vrednost lokalnega minimuma se nahaja na sredini stolpca. Višina stolpca pove kakšna je verjetnost pojava posameznega lokalnega minimuma. Širina stolpca pa, kakšna je razdalja med najmanj oddaljenima minimumoma in razdalja med najbolj oddaljenima minimumoma in je posledica normiranja slike v Matlabu. Drugače povedano, ozek stolpec pomeni veliko oddaljenost med skrajnima minimumoma in majhno oddaljenost med najbližjima minimumoma.



Slika 4.3: Število pojavov minimumov pri interpolaciji z Gaussovo kovariančno funkcijo pri 43 (a) in 89 učnih točkah (b)

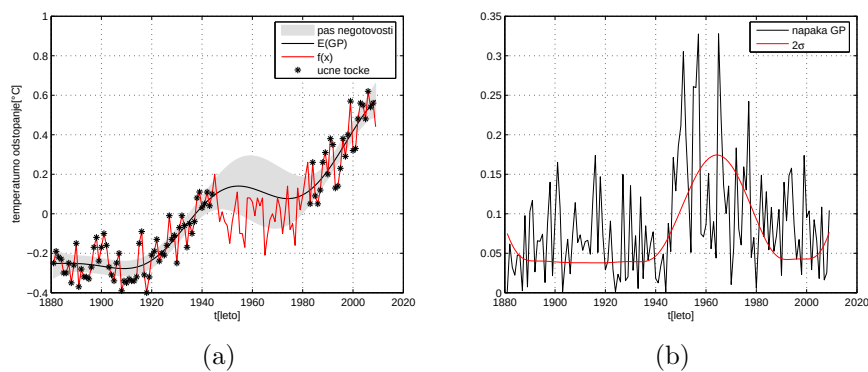
Na sliki 4.4a lahko vidimo predikcijo modela. Varianca modela je, ne glede na razredčenost učnih točk, relativno nizka, narašča pa na mestih, ki so oddaljena od učnih točk, kar lahko vidimo na sliki 4.7b. Na sliki je poleg variance prikazana tudi napaka modela. Napaka modela je enakomerno porazdeljena po celotnem intervalu testnih točk, saj Gaussova kovariančna funkcija pri optimalnih vrednostih hiperparametrov funkcijo le pogladi in ne teži k temu, da bi zadela natančne vrednosti.



Slika 4.4: Interpolacija z Gaussovo kovariančno funkcijo (a) ter napaka in tolerančni pas (2σ) izhoda (b) pri 43 učnih točkah

Poglejmo si še demonstracijo Gaussove kovariančne funkcije na drugem primeru učnih točk. Predikcija GP modela in njegova interpretacija sta prikazana na sliki 4.5. Pričakovano se varianca modela in s tem pas negotovosti v primerjavi s prejšnjim primerom na intervalih $1880 < t < 1945$ in $1983 < t < 2009$ zmanjša, saj so učne točke gostejše posejane. Podobno povedanemu, se varianca na območju stran od učnih točk precej poveča. Podobno velja tudi za napako GP modela, ki je na intervalu brez učnih točk precej večja kot v prejšnjem primeru. Ker pa je napaka v okolici učnih točk v primerjavi s prejšnjim primerom manjša, se tudi vrednost mer ne spremeni drastično (tabela 4.1). Iz povedanega sledi, da je uporaba Gaussove kovariančne funkcije primerna za obe vrsti učnih podatkov in jo uporabimo takrat, kadar kot rezultat GP modela želimo dobiti gladko krivuljo.

Poglejmo si še primer modeliranja z **eksponentno kovariančno funkcijo** (3.6), ki je sicer zelo podobna Gaussovi, le da ni neskončno mnogokrat odvedljiva in posledično naredi izhod iz GP modela manj gladek. Tudi tukaj smo za začetne vrednosti hiperparametrov uporabili naključni generator, po optimizaciji pa dobili vrednosti $l = 44.1238$, $\theta_1 = 0.2281$, $\theta_n = 0.0874$ in $\gamma = 1.5713$ za

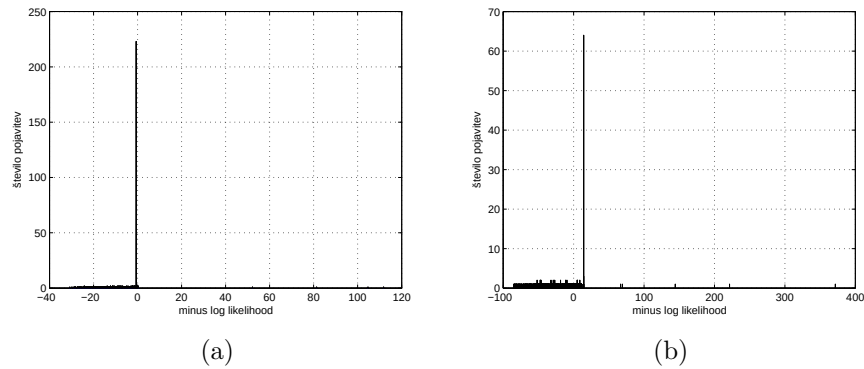


Slika 4.5: Interpolacija z Gaussovo kovariančno funkcijo (a) ter napaka in tolerančni pas (2σ) izhoda (b) pri 89 učnih točkah

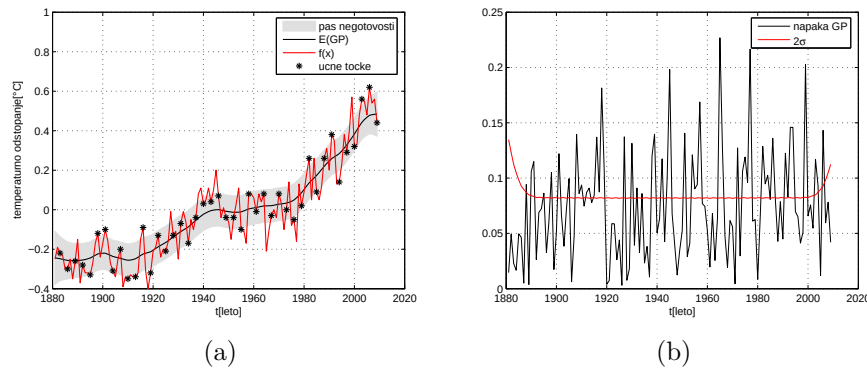
	MAE	MSE	LPD	MRSE	LL
43 učnih točk	0.0804	0.0094	-0.5049	0.3894	-31.4869
89 učnih točk	0.1334	0.0250	-0.2330	1.5119	-83.5593

Tabela 4.1: Vrednosti mer pri interpolaciji z Gaussovo kovariančno funkcijo. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

prvi primer učne množice (slika 4.2a) in $l = 53.4903$, $\theta_1 = 0.2667$, $\theta_n = 0.0720$ in $\gamma = 1.1792$ za drugi (slika 4.2b). Rezultat modeliranja je prikazan na slikah 4.7 in 4.8. Iz slik in vrednosti mer (tabela 4.2) lahko razberemo, da so vrednosti mer pri eksponentni kovariančni funkciji malenkost boljše. Kljub temu pa za modeliranje naših podatkov ni najbolj primerna, ker težko zadanemo globalni minimum pri iskanju optimalnih vrednosti hiperparametrov, kar potrjuje slika 4.6.

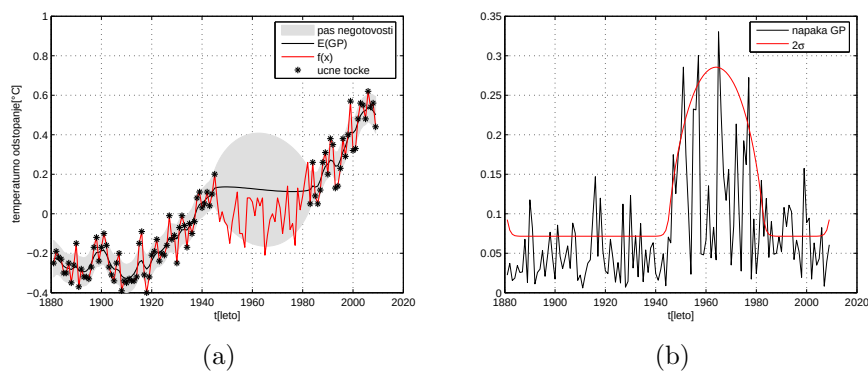


Slika 4.6: Število pojavov minimumov pri interpolaciji z eksponentno kovariančno funkcijo pri 43 (a) in 89 učnih točkah (b)



Slika 4.7: Interpolacija z eksponentno kovariančno funkcijo (a) ter napaka GP modela in tolerančni pas (2σ) izhoda (b) pri 43 učnih točkah

Podobno, le v manjši meri, se zgodi pri modeliranju z **racionalnim kvadratičnim polinomom** (3.7), ki predstavlja vsoto Gaussovih kovariančnih

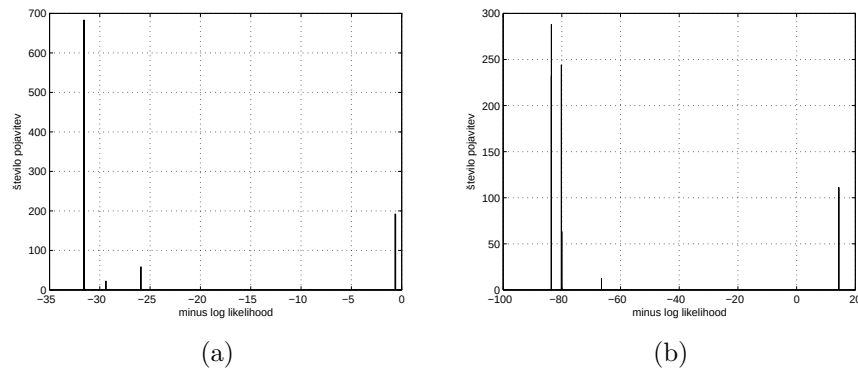


Slika 4.8: Interpolacija z eksponentno kovariančno funkcijo (a) ter napaka GP modela in tolerančni pas (2σ) izhoda (b) pri 89 učnih točkah

	MAE	MSE	LPD	MRSE	LL
43 učnih točk	0.0774	0.0086	-0.5226	0.3735	-30.7618
89 učnih točk	0.1313	0.0243	-0.4298	1.4917	-84.5311

Tabela 4.2: Vrednosti mer pri interpolaciji z eksponentno kovariančno funkcijo. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

funkcij. Rezultati modeliranja so zato zelo podobni rezultatom modeliranja z Gaussovo kovariančno funkcijo, kar potrjujeta sliki 4.10 in 4.11 ter vrednosti mer (tabela 4.3). Tudi tukaj smo za začetne vrednosti hiperparametrov uporabili naključni generator, po optimizaciji pa dobili vrednosti $l = 35.6412$, $\theta_1 = 0.3189$, $\theta_n = 0.0915$ in $\alpha = 2.1346$ za prvi primer učne množice (slika 4.2a) in $l = 21.0626$, $\theta_1 = 0.2877$, $\theta_n = 0.0814$ in $\alpha = 0.4974$ za drugi (slika 4.2b). Kljub dobrim rezultatom pa tudi pri primerjavi racionalnega kvadratičnega polinomoma in Gaussove kovariančne funkcije za modeliranje raje izberemo Gaussovo, ker z večjo verjetnostjo dobimo najmanjšega izmed lokalnih minimumov pri optimizaciji hiperparametrov (slika 4.9).

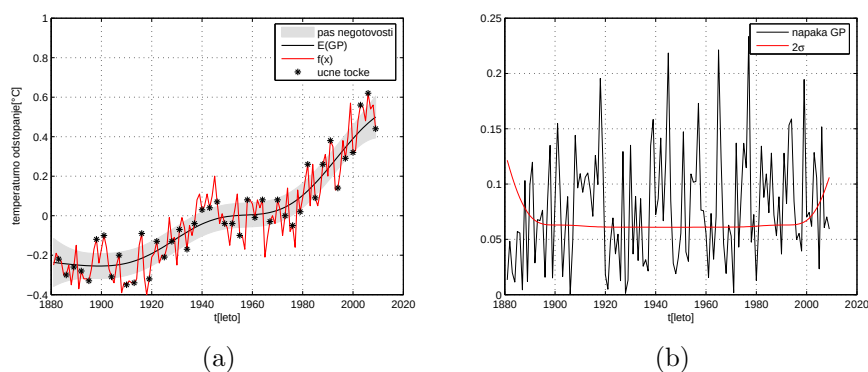


Slika 4.9: Število pojavov minimumov pri interpolaciji z racionalnim kvadratičnim polinomom pri 43 (a) in 89 učnih točkah (b)

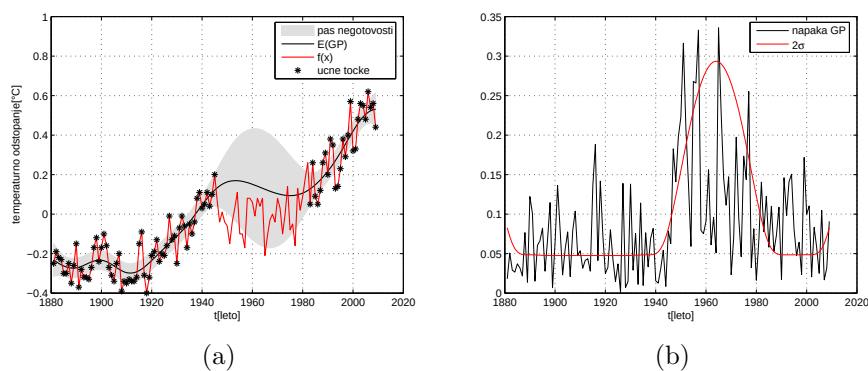
	MAE	MSE	LPD	MRSE	LL
43 učnih točk	0.0805	0.0094	-0.503	0.3905	-31.5706
89 učnih točk	0.1404	0.0273	-0.3193	1.5783	-83.5811

Tabela 4.3: Vrednosti mer pri interpolaciji z racionalnim kvadratičnim polinomom. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

Glede to, da je **Matérnova kovariančna funkcija** (3.8) podobna Gaussovi, le da omogoča nadzor nad relativno gladkostjo, je izgleda najbolj primerna za uporabo modeliranja naših podatkov. Slike 4.13 in 4.14 prikazujejo modeliranje z Matérnovo kovariančno funkcijo z enako pridobljenimi začetnimi

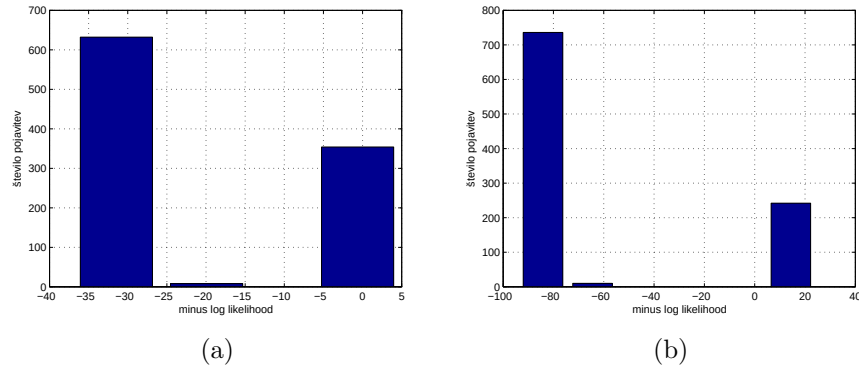


Slika 4.10: Interpolacija z racionalnim kvadratičnim polinomom (a) ter napaka GP modela in tolerančni pas (2σ) izhoda (b) pri 43 učnih točkah

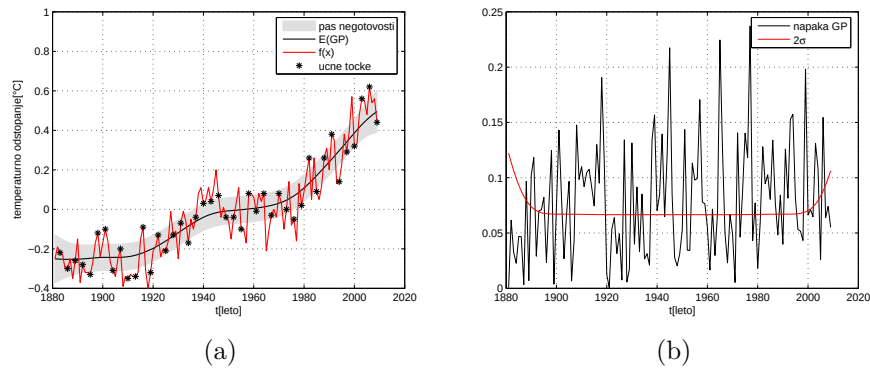


Slika 4.11: Interpolacija z racionalnim kvadratičnim polinomom (a) ter napaka GP modela in tolerančni pas (2σ) izhoda (b) pri 89 učnih točkah

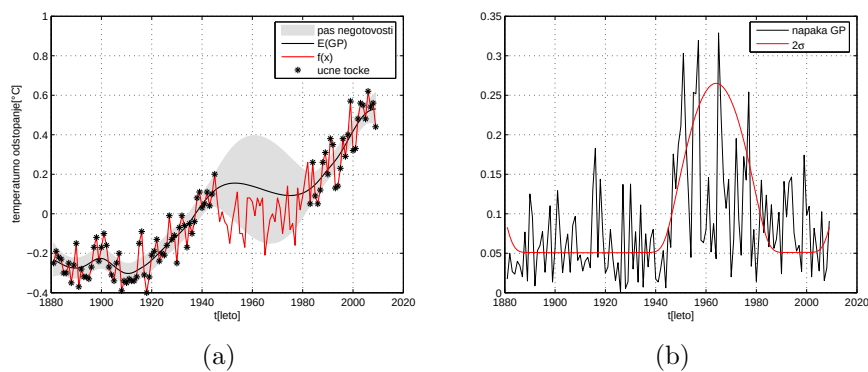
vrednostmi hiperparametrov kot pri predhodno opisanih primerih. Po optimizaciji smo dobili vrednosti $l = 77.7337$, $\theta_1 = 0.3846$, $\nu = 1.5$ in $\theta_n = 0.0913$ za prvi primer učne množice (slika 4.2a) in $l = 34.7716$, $\theta_1 = 0.2931$, $\nu = 1.5$ in $\theta_n = 0.0804$ za drugi (slika 4.2b). Vrednosti mer (tabela 4.4) so malenkost boljše kot pri predhodno uporabljenih kovariančnih funkcijah. Poleg tega je verjetnost, da najdemo optimalen lokalni minimum pri obeh primerih učnih podatkov zelo velika, kar vidimo na sliki 4.12. Slednje potrjuje zgoraj omenjeno predpostavko, da je Matérnova kovariančna funkcija primernejša za vhodne podatke, ki niso opisani z gladko krivuljo.



Slika 4.12: Število pojavov minimumov pri interpolaciji z Matérnovo kovariančno funkcijo pri 43 (a) in 89 učnih točkah (b)



Slika 4.13: Interpolacija z Matérnovo kovariančno funkcijo (a) ter napaka GP modela in tolerančni pas (2σ) izhoda (b) pri 43 učnih točkah

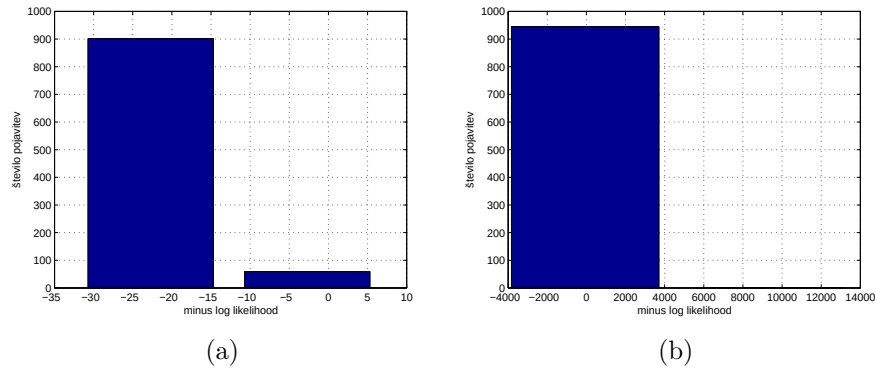


Slika 4.14: Interpolacija z Matérnovno kovariančno funkcijo (a) ter napaka GP modela in tolerančni pas (2σ) izhoda (b) pri 89 učnih točkah

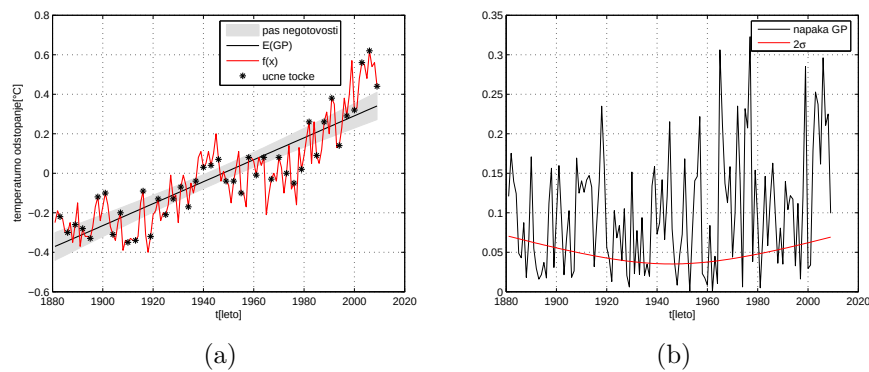
	MAE	MSE	LPD	MRSE	LL
43 učnih točk	0.0806	0.0094	-0.5019	0.3906	-31.4551
89 učnih točk	0.1341	0.0252	-0.3646	1.5160	-84.1464

Tabela 4.4: Vrednosti mer pri interpolaciji z Matérnovno kovariančno funkcijo. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

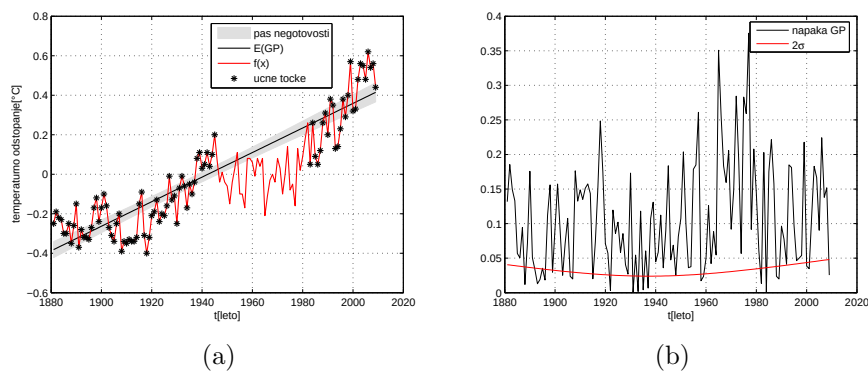
Pokažimo še primer modeliranja z nestacionarnimi kovariančnimi funkcijami. Najprej smo uporabili **linearno kovariančno funkcijo** (3.12), kjer velja $\theta_1 = \theta_2$. Po optimizaciji smo dobili vrednosti $l = 3.7408$ in $\theta_1 = 0.1155$ za prvi primer učne množice (slika 4.2a) in $l = 3.6408$ in $\theta_1 = 0.1150$ za drugi (slika 4.2b). Pozitivna lastnost linearne kovariančne funkcije je ta, da padec v optimalen lokalni minimum ni problematičen (slika 4.15), so pa zato vrednosti mer (tabela 4.5) precej slabše kot pri modeliranju s stacionarnimi funkcijami. To je razumljivo, saj je rezultat modelirana ravna premica čez podatke.



Slika 4.15: Število pojavov minimumov pri interpolaciji z linearno kovariančno funkcijo pri 43 (a) in 89 učnih točkah (b)



Slika 4.16: Interpolacija z linearno kovariančno funkcijo (a) ter napaka GP modela in tolerančni pas (2σ) izhoda (b) pri 43 učnih točkah

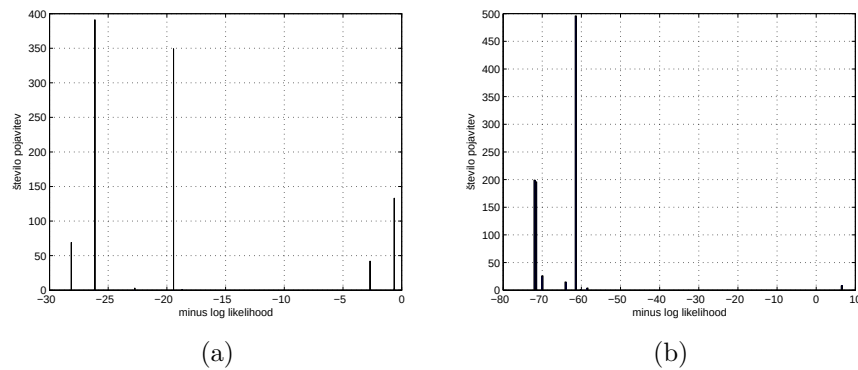


Slika 4.17: Interpolacija z linearno kovariančno funkcijo (a) ter napaka GP modela in tolerančni pas (2σ) izhoda (b) pri 89 učnih točkah

	MAE	MSE	LPD	MRSE	LL
43 učnih točk	0.1050	0.0169	-0.3476	0.5231	-22.7301
89 učnih točk	0.1406	0.0289	-0.1626	1.6248	-58.4516

Tabela 4.5: Vrednosti mer pri interpolaciji z linearno kovariančno funkcijo. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

Primernejša za naše podatke je **kovariančna funkcija nevronske mreže** (3.14). Vrednosti hiperparametrov po optimizaciji so $l = 566.7146$, $\theta_1 = 111.3055$ in $\theta_n = 0.0994$ za prvi primer učne množice (slika 4.2a) in $l = 52.8945$, $\theta_1 = 28.9393$ in $\theta_n = 0.0863$ za drugi (slika 4.2b). Vrednosti mer so za oba primera učnih podatkov boljše od linearne, zanimiva pa je primerjava s stacionarnimi funkcijami. Pri prvem primeru učnih podatkov so vrednosti mer slabše kot pri slednjih. Ko pa gre za interpolacijo podatkov med dvema intervaloma (drugi primer učnih podatkov), se kovariančna funkcija nevronske mreže obnese malenkost bolje. Problematičen pa je padec v optimalen lokalni minimum, čigar verjetnost je manj kot 20 odstotna. Poleg tega je vrednost najmanjšega lokalnega minimuma (*minus log-likelihood*) precej večja, kot pri modeliranju s stacionarnimi kovariančnimi funkcijami.

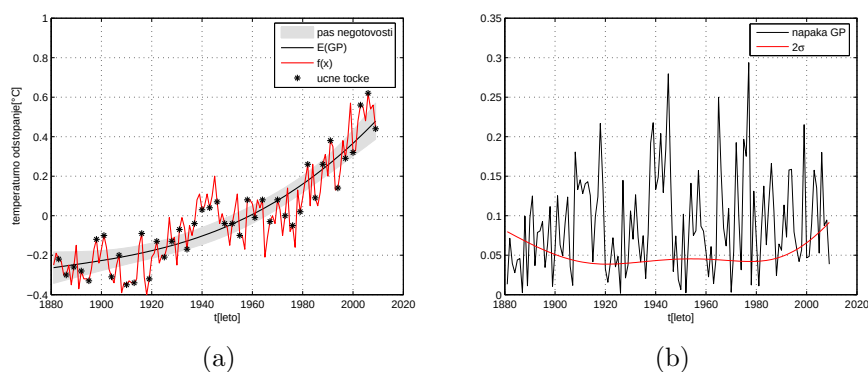


Slika 4.18: Število pojavov minimumov pri interpolaciji z kovariančno funkcijo nevronske mreže pri 43 (a) in 89 učnih točkah (b)

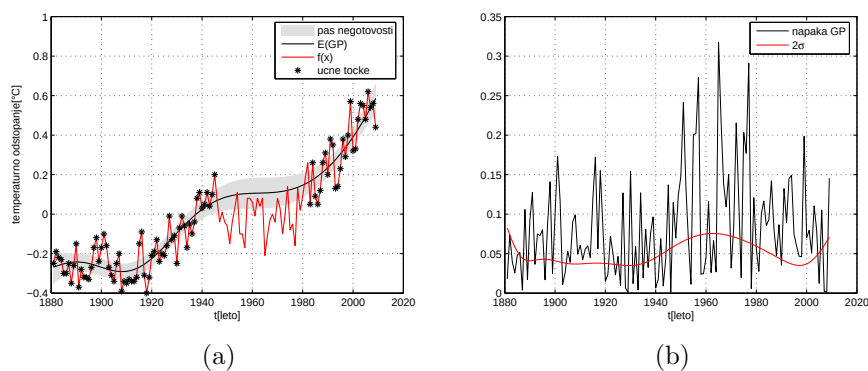
	MAE	MSE	LPD	MRSE	LL
43 učnih točk	0.106	0.0173	-0.4005	0.5291	-28.1469
89 učnih točk	0.1278	0.0259	-0.2673	1.5399	-71.9610

Tabela 4.6: Vrednosti mer pri interpolaciji s kovariančno funkcijo nevronske mreže. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

Za konec demonstrirajmo še uporabo kombinacije različnih kovariančnih funkcij. Glede na naravo podatkov lahko sklepamo, da se ne glede na to, da v splošnem naraščajo, temperaturna odstopanja periodično ponavljajo. Zato v



Slika 4.19: Interpolacija kovariančno funkcijo nevronske mreže (a) ter napaka GP modela in tolerančni pas (2σ) izhoda (b) pri 43 učnih točkah

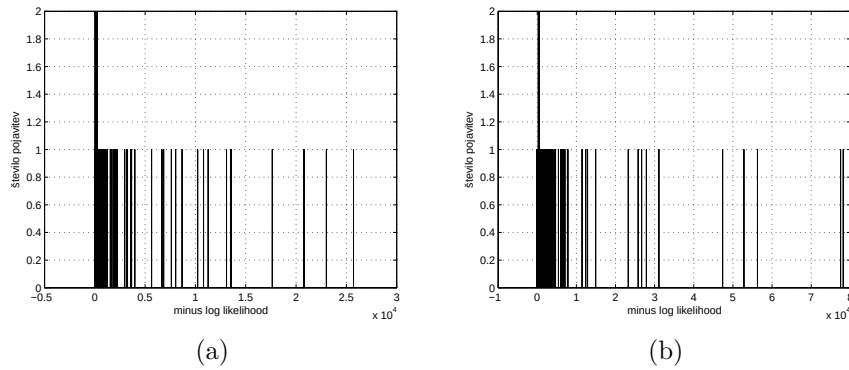


Slika 4.20: Interpolacija s kovariančno funkcijo nevronske mreže (a) ter napaka GP modela in tolerančni pas (2σ) izhoda (b) pri 89 učnih točkah

model vključimo kovariančno funkcijo, ki je **produkt periodične in Gaussove kovariančne funkcije**. Novo kovariančno funkcijo prikazuje izraz:

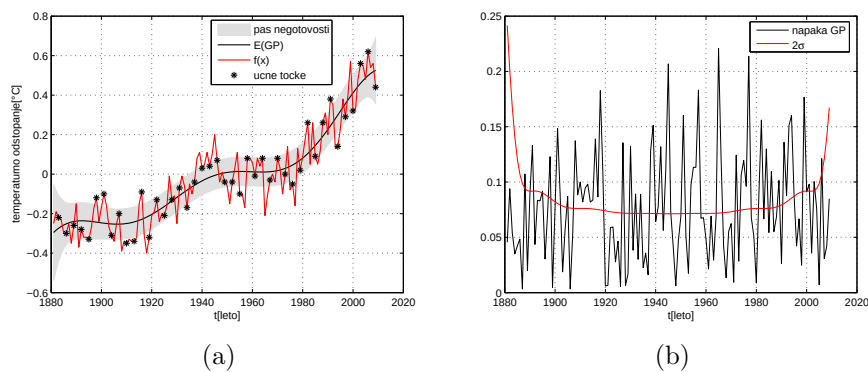
$$C(\mathbf{x}_i, \mathbf{x}_j) = \theta_1^2 \theta_2^2 e^{-\frac{2(\sin(\frac{\pi}{\omega}(x_i - x_j)))^2}{l_{PER}^2} - \frac{(x_i - x_j)^2}{2l_{SE}^2}} + \delta_{ij} \theta_n^2 \quad (4.1)$$

Po optimizaciji so vrednosti hiperparametrov $l_{SE} = 90.7196$, $\theta_1 = 0.7299$, $l_{PER} = 1.0000$, $\theta_2 = 0.7299$, $\omega = 1.0017$ in $\theta_n = 0.1050$ za prvi primer učne množice (slika 4.2a) in $l_{SE} = 24.9383$, $\theta_1 = 0.5714$, $l_{PER} = 1.000$, $\theta_2 = 0.5714$, $\omega = 1.000$ in $\theta_n = 0.0827$ za drugi (slika 4.2b). Rezultat modeliranja je prikazan na slikah 4.22 in 4.23. Podobno kot pri obravnavi Gaussove, sta tudi tukaj varianca in napaka modela večji na intervalih brez učnih točk. Iz tabele 4.7 je razvidno, da se obravnavana kombinacija še posebno dobro obnese pri učnih točkah, ki so enakomerno porazdeljene. Kljub temu pa je neprimerna za uporabo, ker je možnost padca v optimalen lokalni minimum le 0.1% (slika 4.21a).

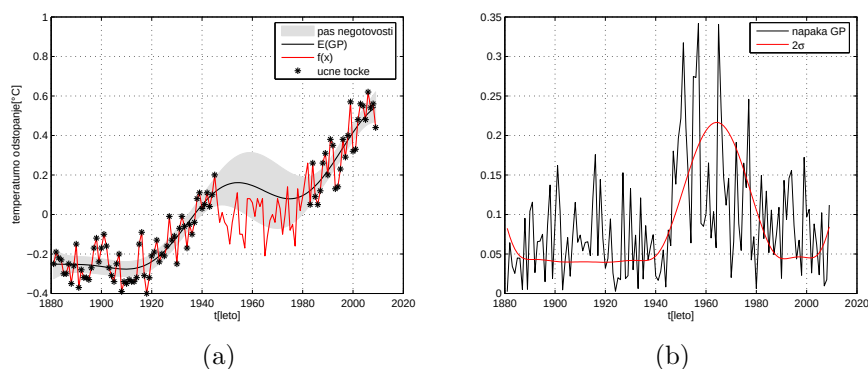


Slika 4.21: Število pojavov minimumov pri interpolaciji z zmnožkom Gaussove in periodične kovariančne funkcije pri 43 (a) in 89 učnih točkah (b)

V tabelah 4.8 in 4.9 so zbrane vse mere modeliranja pri interpolaciji. Vidimo, da se v primeru podatkov s 43 učnimi točkami najbolj obnese eksponentna kovariančna funkcija, saj ima najmanjšo vrednost vseh mer, razen LPD. Sledi ji Gaussova kovariančna funkcija. LPD je najmanjša pri zmnožku Gaussove in periodične kovariančne funkcije. Najslabše se obnese linearna kovariančna funkcija. Na neprimernost uporabe te funkcije nakazuje že vrednost LL, ki je izmerjena le na podlagi učnih podatkov, a vseeno uporabna za vrednotenje.



Slika 4.22: Interpolacija z zmnožkom Gaussove in periodične kovariančne funkcije (a) ter napaka GP modela in tolerančni pas (2σ) izhoda (b) pri 43 učnih točkah



Slika 4.23: Interpolacija z zmnožkom Gaussove in periodične kovariančne funkcije (a) ter napaka GP modela in tolerančni pas (2σ) izhoda (b) pri 89 učnih točkah

	MAE	MSE	LPD	MRSE	LL
43 učnih točk	0.0780	0.0089	-0.9113	0.3795	-31.4822
89 učnih točk	0.1419	0.0277	-0.2127	1.5923	-83.5575

Tabela 4.7: Vrednosti mer pri interpolaciji z zmnožkom Gaussove in periodične kovariančne funkcije. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

V primeru podatkov z 89 učnimi točkami se najbolje obnese kovariančna funkcija nevronske mreže, med stacionarnimi pa je zopet najboljša eksponentna kovariančna funkcija. Tudi tukaj sledi Gaussova kovariančna funkcija. Iz povedanega sledi, da se stacionarne funkcije pri interpolaciji obnašajo podobno, ne glede na obliko učne množice.

	MAE	MSE	LPD	MRSE	LL
Gaussova kovariančna funkcija	0.0804	0.0094	-0.5049	0.3894	-31.4869
Eksponentna kovariančna funkcija	0.0774	0.0086	-0.5226	0.3735	-30.7618
Racionalni kvadratični polinom	0.0805	0.0094	-0.503	0.3905	-31.5706
Matérnova kovariančna funkcija	0.0806	0.0094	-0.5019	0.3906	-31.4551
Linearna kovariančna funkcija	0.1050	0.0169	-0.3476	0.5231	-22.7301
Kovariančna funkcija nevronske mreže	0.0951	0.0134	-0.4005	0.4652	-28.1469
Zmnožek Gaussove in periodične KF	0.0780	0.0089	-0.9113	0.3795	-31.4822

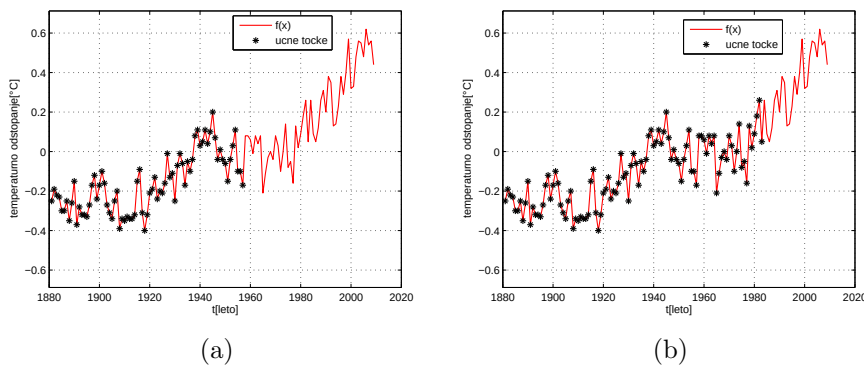
Tabela 4.8: Vrednosti mer pri interpolaciji za 43 učnih točk za vse kovariančne funkcije. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

	MAE	MSE	LPD	MRSE	LL
Gaussova kovariančna funkcija	0.1334	0.0250	-0.2330	1.5119	-83.5593
Eksponentna kovariančna funkcija	0.1313	0.0243	-0.4298	1.4917	-84.5311
Racionalni kvadratični polinom	0.1404	0.0273	-0.3193	1.5783	-83.5811
Matérnova kovariančna funkcija	0.1341	0.0252	-0.3646	1.5160	-84.1464
Linearna kovariančna funkcija	0.1406	0.0289	-0.1626	1.6248	-58.4516
Kovariančna funkcija nevronske mreže	0.1167	0.0206	-0.2673	1.3716	-71.9610
Zmnožek Gaussove in periodične KF	0.1419	0.0277	-0.2127	1.5923	-83.5575

Tabela 4.9: Vrednosti mer pri interpolaciji za 89 učnih točk za vse kovariančne funkcije. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

4.1.3 Ekstrapolacija

Pri ekstrapolaciji smo množico učnih točk določili nekoliko drugače kot pri interpolaciji. Za večino primerov smo izbrali interval 103 učnih točk (slika 4.24b), ostale pa uporabili za vrednotenje rezultatov. Le za Gaussovo smo poleg omenjene množice učnih točk, uporabili še interval 77 učnih točk (slika 4.24a) in tako prikazali pomen velikosti učne množice.

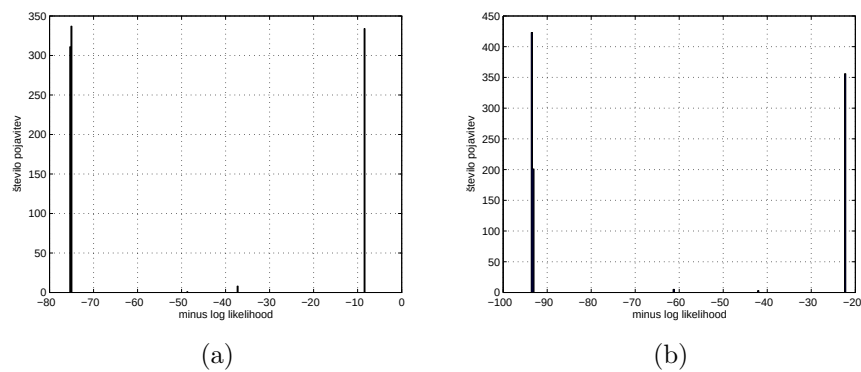


Slika 4.24: Učne točke pri ekstrapolaciji

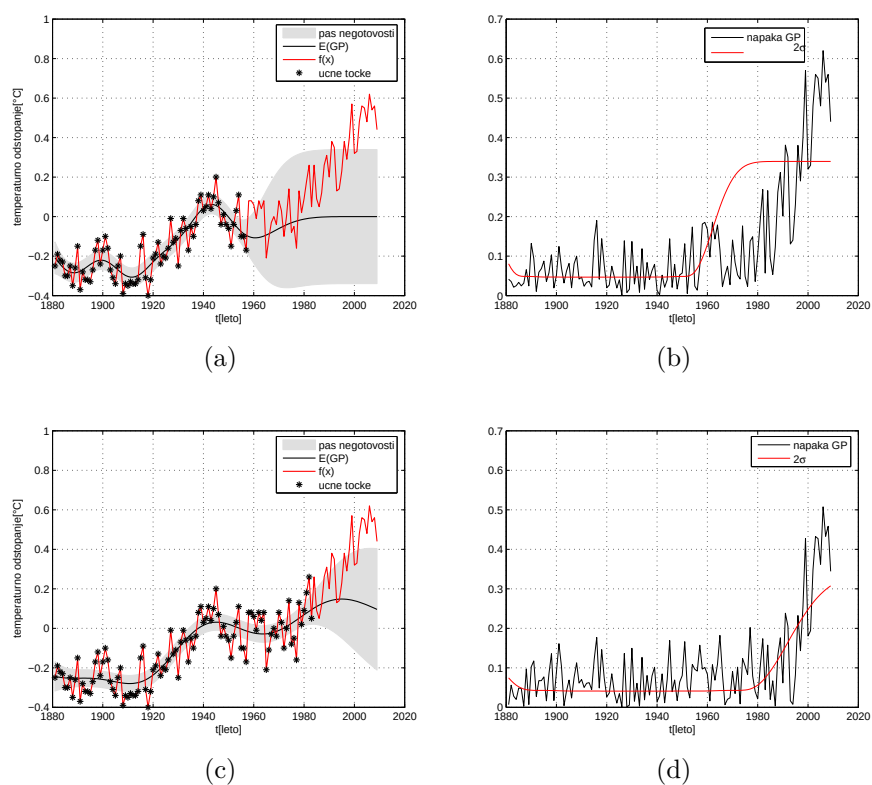
Pri izbiri začetne vrednosti hiperparametrov vseh kovariančnih funkcij smo se tudi pri ekstrapolaciji odločili za uporabo naključnega generatorja. Po optimizaciji smo za **Gaussovo kovariančno funkcijo** (3.5) dobili naslednje vrednosti: $l = 10.3667$, $\theta_1 = 0.1698$ in $\theta_n = 0.0762$ za prvi primer učne množice (slika 4.24a) in $l = 9.4712$, $\theta_1 = 0.1554$ in $\theta_n = 0.0830$ za drugi (slika 4.24b). Iz tabele 4.10 in slik 4.25 ter 4.26 je vidno, da večja učna množica pripomore k boljšim rezultatom modeliranja. Pri večji učni množici lažje zadanemo optimalen lokalni minimum pri iskanju optimalnih vrednosti hiperparametrov (slika 4.25b). Poleg tega sta tudi napaka modela in varianca manjši pri večji učni množici (sliki 4.26b in 4.26d). Prikazan vpliv velikosti učne množice pri ekstrapolaciji bomo posplošili na vse kovariančne funkcije.

Pri rezultatih modeliranja z Gaussovo kovariančno funkcijo opazimo tudi, da pri večjem odmiku od učnih podatkov napovedi težijo h konstantni vrednosti. Enak problem se pojavi pri vseh ostalih stacionarnih kovariančnih funkcijah, zato je za ekstrapolacijo bolj smiselno uporabiti nestacionarne ali pa kombinacijo obeh.

Pri modeliranju z **eksponentno kovariančno funkcijo** (3.6) smo po optimizaciji dobili naslednje vrednosti hiperparametrov: $l = 77.6828$, $\theta_1 = 0.1883$,



Slika 4.25: Število pojavov minimumov pri ekstrapolaciji z Gaussovo kovariančno funkcijo pri 77 (a) in 103 učnih točkah (b)

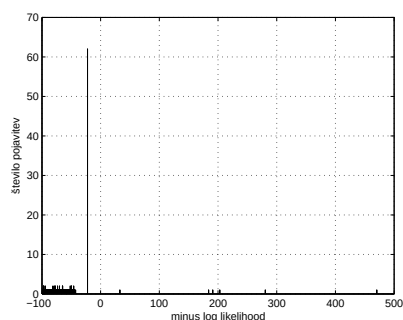


Slika 4.26: Ekstrapolacija z Gaussovo kovariančno funkcijo (a) in (c) ter napaka in tolerančni pas (2σ) izhoda (b) in (d)

	MAE	MSE	LPD	MRSE	LL
77 učnih točk	0.2279	0.0811	0.4173	1.0169	-75.2861
103 učne točke	0.2206	0.0732	0.4377	0.7174	-93.4693

Tabela 4.10: Vrednosti mer pri ekstrapolaciji z Gaussovo kovariančno funkcijo. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

$\theta_n = 0.0295$ in $\gamma = 0.4540$. Zanimiva je opazka, da je logaritem verjetnostne porazdelitve učne množice najmanjši, kljub temu pa je model eden izmed slabših (tabela 4.16). To je sicer razumljivo, saj je vrednost logaritma verjetnostne porazdelitve učne množice neodvisna od vrednosti testnih podatkov, pa vendar je relativna nizkost pogoj za dober model.

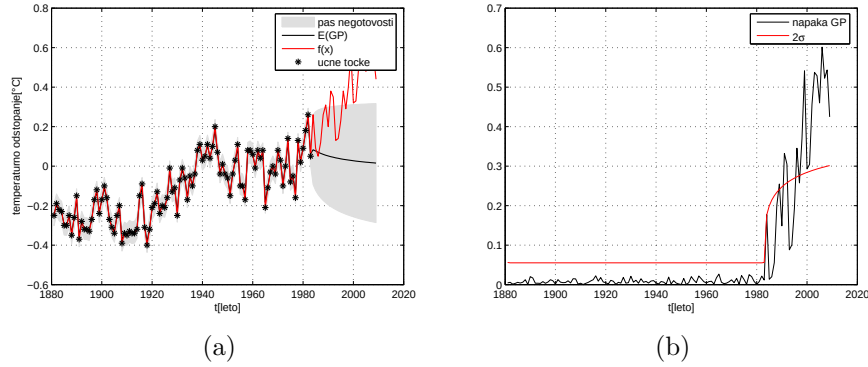


Slika 4.27: Število pojavov minimumov pri ekstrapolaciji z eksponentno kovariančno funkcijo

	MAE	MSE	LPD	MRSE	LL
103 učne točke	0.2995	0.1227	1.8046	0.9288	-97.5966

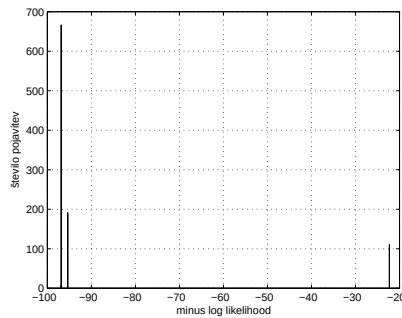
Tabela 4.11: Vrednosti mer pri ekstrapolaciji z eksponentno kovariančno funkcijo. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

Podobno se zgodi pri modeliranju z **racionalnim kvadratičnim polinomom** (3.7), čigar vrednosti mer (tabela 4.12) so malenkost boljše kot pri modeliranju z eksponentno kovariančno funkcijo, logaritem verjetnostne porazdelitve učne množice pa je malenkost večji. Glede na povedano bi med



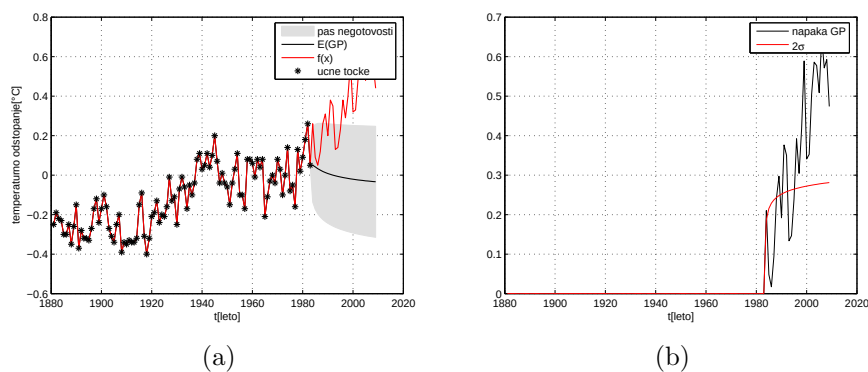
Slika 4.28: Ekstrapolacija z eksponentno kovariančno funkcijo (a) ter napaka in tolerančni pas (2σ) izhoda (b)

slednjima funkcijama raje izbrali racionalni kvadratični polinom, ker ima optimizacija vrednosti hiperparametrov zelo majhno verjetnost za padec v lokalni minimum, vseeno pa tako eksponentna kovariančna funkcija, kot tudi racionalni kvadratični polinom, nista primerna na modeliranje take vrste podatkov. Po optimizaciji so bile vrednosti hiperparametrov pri modeliranju z racionalnim kvadratičnim polinomom naslednje: $l = 0.8848$, $\theta_1 = 0.2156$, $\theta_n = 0.0000$ in $\alpha = 0.0444$.



Slika 4.29: Število pojavov minimumov pri ekstrapolaciji z racionalnim kvadratičnim polinomom

Tako kot pri interpolaciji se tudi pri ekstrapolaciji najbolje obnese uporaba **Matérnovе kovariančne funkcije** (3.8). Poleg dobrih vrednosti mer (tabela 4.13) je tudi verjetnost za padec v neoptimalen lokalni minimum pri iskanju optimalnih vrednosti hiperparametrov najmanjša izmed obravnavanih funkcij

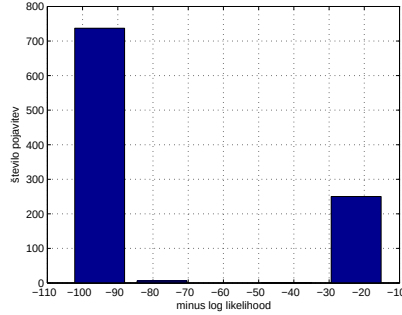


Slika 4.30: Ekstrapolacija z racionalnim kvadratičnim polinomom (a) ter napaka in tolerančni pas (2σ) izhoda (b)

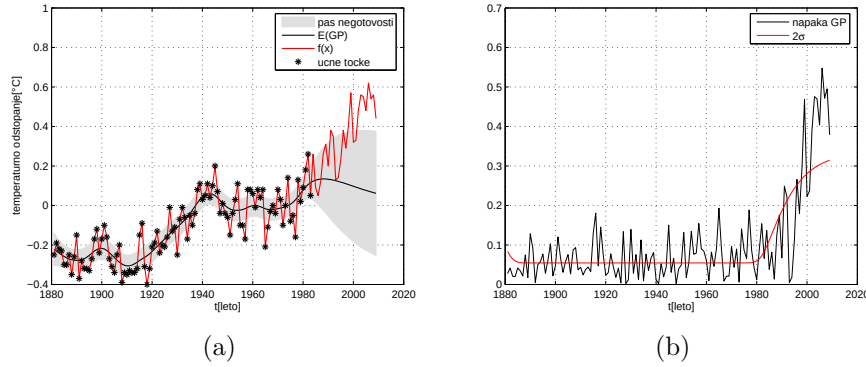
	MAE	MSE	LPD	MRSE	LL
103 učne točke	0.3410	0.1527	2.3829	1.0361	-96.8599

Tabela 4.12: Vrednosti mer pri ekstrapolaciji z racionalnim kvadratičnim polinomom. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

(slika 4.31). Po optimizaciji so bile vrednosti hiperparametrov $l = 19.8349$, $\theta_1 = 0.1662$, $\nu = 1.5$ in $\theta_n = 0.0819$. Izhod GP modela pa je, podobno kot pri Gaussovi kovariančni funkciji (slika 4.26c), poglajena krivulja z varianco, ki se večja z oddaljenostjo od učnih točk (slika 4.32).



Slika 4.31: Število pojavov minimumov pri ekstrapolaciji z Matérnovo kovariančno funkcijo



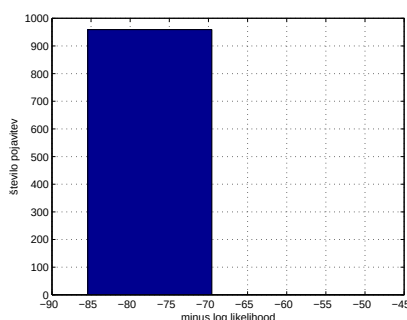
Slika 4.32: Ekstrapolacija z Matérnovo kovariančno funkcijo (a) ter napaka in tolerančni pas (2σ) izhoda (b)

Zaradi neustreznosti stacionarnih kovariančnih funkcij, smo preverili še modeliranje z nestacionarnimi; točneje z **linearno kovariančno funkcijo** (3.12), kjer velja $\theta_1 = \theta_2$, in **kovariančno funkcijo nevronske mreže** (3.14). Izbira ene izmed omenjenih kovariančnih funkcij je odvisna od namena modeliranja. Če na izhodu GP modela želimo krivuljo, ki se ucinim podatkom kar najbolje prilaga (slika 4.36a), izberemo kovariančno funkcijo nevronske mreže. V

	MAE	MSE	LPD	MRSE	LL
103 učne točke	0.2449	0.0887	0.6386	0.7897	-95.1539

Tabela 4.13: Vrednosti mer pri ekstrapolaciji z Matérnovno kovariančno funkcijo. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

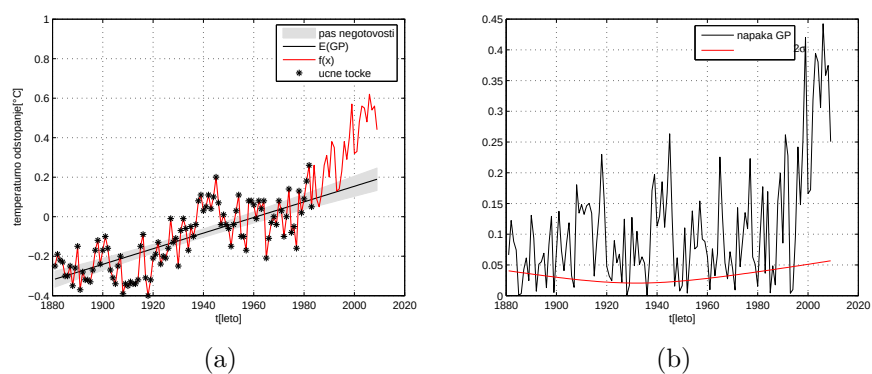
nasprotju s tem, nam izbira linearne kovariančne funkcije omogoča boljše uje-manje modelirane funkcije s testnimi podatki tudi daleč od učnih točk (slika 4.34a). Povedano potrjujejo tudi vrednosti mer v tabeli 4.14. Te so povsod boljše pri linearni kovariančni funkciji, razen pri LPD, ki poleg napake modela upošteva še varianco. Dobra lastnost obeh kovariančnih funkcij je ta, da optimizacija vrednosti hiperparametrov ni problematična (sliki 4.33 in 4.35). Po optimizaciji smo za linearno kovariančno funkcijo dobili vrednosti $\theta_1 = 4.4073$ in $\theta_n = 0.1039$, za kovariančno funkcijo nevronske mreže pa $l = 0.0628$, $\theta_1 = 0.6448$ in $\theta_n = 0.0878$.



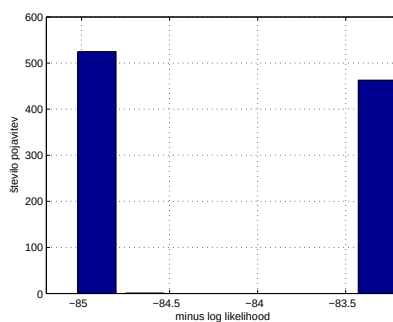
Slika 4.33: Število pojavov minimumov pri ekstrapolaciji z linearno kovariančno funkcijo

	MAE	MSE	LPD	MRSE	LL
Linearna kovariančna funkcija	0.2053	0.0606	1.3200	0.6525	-77.4915
Kovariančna funkcija nevronske mreže	0.2922	0.1136	4.6635	0.8937	-84.9126

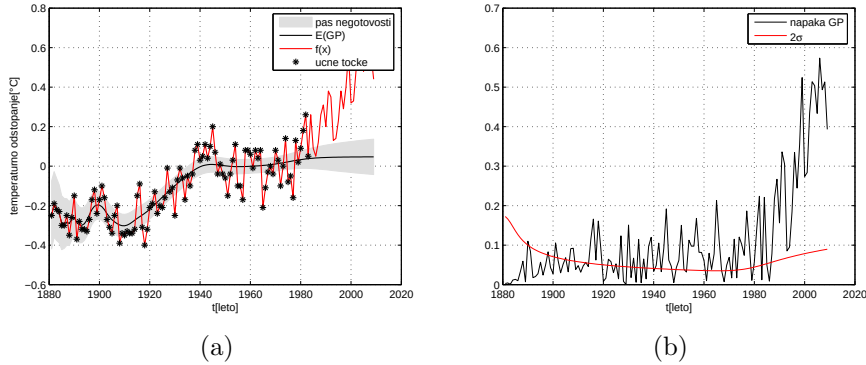
Tabela 4.14: Vrednosti mer pri ekstrapolaciji z nestacionarnimi kovariančnimi funkcijami. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.



Slika 4.34: Ekstrapolacija z linearno kovariančno funkcijo (a) ter napaka in tolerančni pas (2σ) izhoda (b)



Slika 4.35: Število pojavov minimumov pri ekstrapolaciji s kovariančno funkcijo nevronske mreže



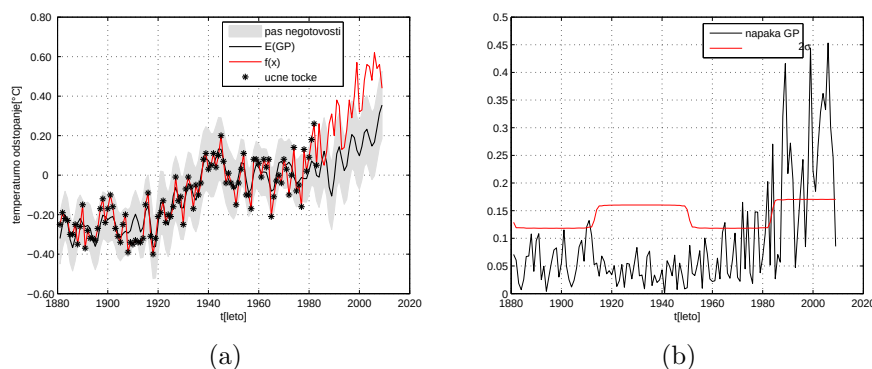
Slika 4.36: Ekstrapolacija s kovariančno funkcijo nevronske mreže (a) ter napaka in tolerančni pas (2σ) izhoda (b)

Za izboljšavo modeliranja s stacionarnimi kovariančnimi funkcijami smo uporabili še pristop kombiniranja več kovariančnih funkcij. Da bi pri modeliranju z Gaussovo kovariančno funkcijo preprečili težnjo k vrednosti nič daleč od učnih točk, smo ji prišteli še linearno kovariančno funkcijo. Kljub temu pa je vsota obeh vrnila enak rezultat, kot samo Gaussova, kar lahko vidimo iz vrednosti mer v tabeli 4.15. Modeliranje smo opravili še z vsoto periodične in linearne kovariančne funkcije. Model s slednjo kombinacijo se je, sodeč po vrednostih mer (tabela 4.15), izkazal za enega boljših. Po optimizaciji hiperparametrov smo za **vsoto Gaussove in linearne** dobili vrednosti $l_{SE} = 1.2931$, $\theta_{SE} = 0.0856$, $\theta_{LIN} = 4.4853$ in $\theta_n = 0.0598$ ter $l_{PER} = 1.0000$, $\theta_{PER} = 0.4000$, $\omega = 1.0000$. $\theta_{LIN} = 1258.8$, in $\theta_n = 0.1000$ za **vsoto periodične in linearne**.

	MAE	MSE	LPD	MRSE	LL
Vsota Gaussove in linearne KF	0.2206	0.0732	0.4377	0.7174	-93.4693
Vsota periodične in linearne KF	0.2228	0.0647	0.5175	0.6746	574.2940

Tabela 4.15: Vrednosti mer pri ekstrapolaciji s kombinacijami kovariančnih funkcij. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

V tabeli 4.16 so zbrane vse mere modeliranja pri ekstrapolaciji. Vidimo, da se najbolje obnese Gaussova kovariančna funkcija in vsota periodične in linearne kovariančne funkcije. Dober je tudi model z linearno kovariančno funkcijo, a je mera LPD prevelika, da bi to funkcijo uvrstili med najboljše. Zanimivo



Slika 4.37: Ekstrapolacija z vsoto periodične in linearne kovariančne funkcije (a) ter napaka in tolerančni pas (2σ) izhoda (b)

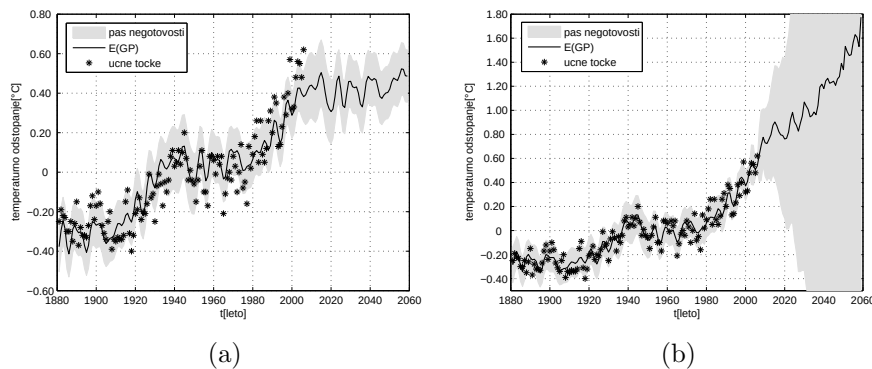
je, da LL pri vsoti periodične in linearne kovariančne funkcije zavzame zelo veliko vrednost glede na ostale. Do odstopanja pride, ker kriterijska funkcija ne zadosti našim željam, hiperparametre pa smo nastavili s poskušanjem, ne pa glede na minimalno vrednost LL.

	MAE	MSE	LPD	MRSE	LL
Gaussova KF za 77 učnih točk	0.2279	0.0811	0.4173	1.0169	-75.2861
Gaussova KF za 103 učne točke	0.2206	0.0732	0.4377	0.7174	-93.4693
EkspONENTNA kovariančna funkcija	0.2995	0.1227	1.8046	0.9288	-97.5966
Racionalni kvadratični polinom	0.3410	0.1527	2.3829	1.0361	-96.8599
Matérnova kovariančna funkcija	0.2449	0.0887	0.6386	0.7897	-95.1539
Linearna kovariančna funkcija	0.2053	0.0606	1.3200	0.6525	-77.4915
Kovariančna funkcija nevronske mreže	0.2922	0.1136	4.6635	0.8937	-84.9126
Vsota Gaussove in linearne KF	0.2206	0.0732	0.4377	0.7174	-93.4693
Vsota periodične in linearne KF	0.2228	0.0647	0.5175	0.6746	574.2940

Tabela 4.16: Vrednosti mer pri ekstrapolaciji za vse kovariančne funkcije. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

Za konec je prikazan še primer napovedovanja globalnega segrevanja za 50 let naprej. Pri tem primeru smo uporabili nekoliko drugačen pristop kot v predhodno opisanih primerih. Vrednosti hiperparametrov smo nastavili ročno, saj pri optimalnih vrednostih hiperparametrov kovariančna matrika ni bila

pozitivno definitna. Poleg tega smo v množico učnih točk vključili vse točke, razen zadnjih treh. Pri teh je namreč funkcija vhodnih podatkov padajoča, kar bi lahko privedlo do neprimernih rezultatov. Za kovariančno funkcijo pri modeliranju smo glede na predznanje sprva uporabili **vsoto linearne in periodične kovariančne funkcije** ter naslednje vrednosti hiperparametrov: $\theta_{LIN} = 1.5290$, $l_{PER} = 0.4538$, $\theta_{PER} = 59874$, $\omega = 0.0080$ in $\theta_n = 11.8224$. Ker pa podatki ne naraščajo linearno, ampak eksponentno z manjšimi nihanji, smo model izboljšali tako, da smo uporabili **vsoto polinomske in periodične kovariančne funkcije**, z vrednostjo hiperparametra $\nu = 4.0552$. Rezultati modeliranja z obema kombinacijama kovariančnih funkcij so vidni na sliki 4.38. Če privzamemo, da je boljši model na sliki 4.38b, ker se modelirana funkcija bolj prilega učnim točkam, lahko predvidevamo, kakšno bo temperaturno odstopanje v obsegu naslednjih 50 let, vendar s precejšnjo negotovostjo.



Slika 4.38: Napovedovanje globalnega ogrevanja z vsoto periodične in linearne kovariančne funkcije (a) ter vsoto periodične in polinomske kovariančne funkcije (b)

4.1.4 Sklep

Na podlagi poskusov smo ugotovili, da se večina stacionarnih kovariančnih funkcij pri interpolaciji obnaša podobno. Ker pomembno vlogo pri modeliranju igra določitev začetnih vrednosti hiperparametrov, lahko rečemo, da je za naš primer podatkov najboljša Matérnova kovariančna funkcija. S stališča iskanja globalnega minimuma je primernejša oblika učne množice, kjer so učne točke razdeljene na dva ločena intervala. Uporaba kombinacij kovariančnih funkcij v našem primeru pri interpolaciji ni bila smiselna, saj vpliva na kom-

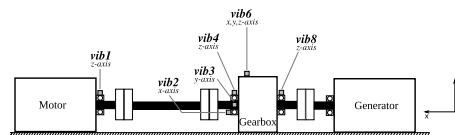
pleksnost, rezultati pa niso bistveno boljši. Za uporabo kombinacije smo se raje odločili pri ekstrapolaciji, vendar pa moramo v tem primeru paziti na vrednosti hiperparametrov, saj je možno, da optimalne vrednosti (tj. kadar je logaritem verjetnostne porazdelitve učne množice najmanjši) ne zagotavljajo želenega rezultata, saj so naše želje drugačne, kot pa to določa uporabljena kriterijska funkcija. Z uporabo kombinacije smo preprečili pomanjkljivost stacionarnih kovariančnih funkcij, da se daleč od učnih točk vrednosti modelirane funkcije približujejo vrednosti nič.

4.2 Modeliranje življenjske dobe zobnika

V tem poglavju je predstavljen način napovedovanja stopnje obrabljenosti strojnega elementa, glede na katero bi lahko sprožili varnostni alarm in tako preprečili daljšo prekinitev proizvodnje.

4.2.1 Opis testnega okolja in zajema podatkov

Podatki, iz katerih lahko sklepamo na obrabo zobnika, so bili merjeni na testni postaji. Testna postaja je sestavljena iz motorja in generatorja z enofaznim prestavnim razmerjem (ang. *gearbox*), povezanimi preko zobnikov. Motor je standardni motor na enosmerno napetost, ki ga napaja Simoreg DC pogon. Generator se uporablja kot breme. Proizvedena moč generatorja se vrača nazaj v sistem in tako proizvede bremensko silo. Testna postaja je opremljena z osmimi senzorji (slika 4.39) za merjenje vibracij, s katerimi je določena vrednost značilke, ki posredno govori o obrabljenosti zobnika. Signali vseh osmih senzorjev so bili pridobljeni hkrati in analizirani z analizo ovojnice (ang. *envelope analysis*). Za modeliranje smo izbrali podatke osmega senzorja, ki ima največjo občutljivost [5].

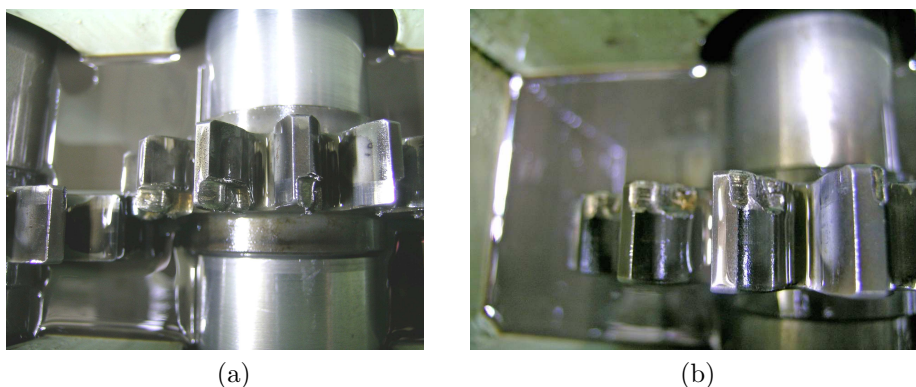


Slika 4.39: Postavitev senzorjev

Test je bil narejen pri konstantnem navoru 82.5 Nm in konstantni hitrosti 990 rpm, ki je proizvedla zobniško frekvenco (ang. *gear mesh frequency*) $f_{gm} = 396$ Hz, vhodno vrtilno hitrost gredi $f_i = 16.5$ Hz in izhodno vrtilno hitrost gredi $f_o = 24.75$ Hz. Signal je bil vzorčen z vzorčno frekvenco $f_o = 80$ kHz.

Posamezno zajemanje podatkov je trajalo 5 sekund in bilo ponovljeno vsakih 10 min. Celoten poskus je trajal 65 ur.

Delovanje testne postaje je bilo ustavljeno preden je zobnik odpovedal; tj. pri vrednosti značilke 70358. Na koncu so bili poškodovani vsi zobniki, tako vhodne kot tudi izhodne gredi (slika 4.40), kar je povzročilo veliko frekvenco vibracij, ki je bila razlog zaustavitve pred popolno odpovedjo.

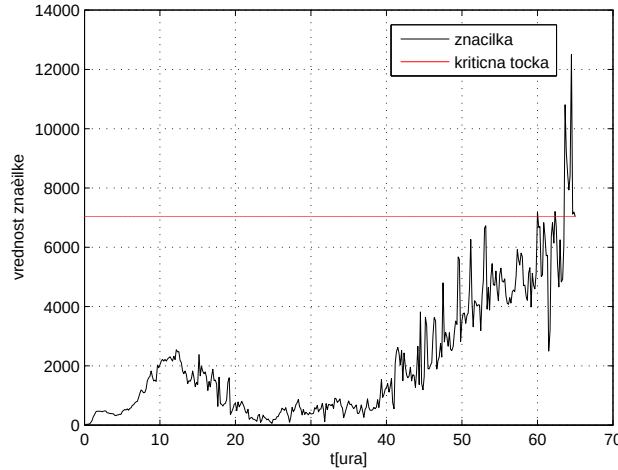


Slika 4.40: Poškodovani vhodna (a) in izhodna (b) gred

4.2.2 Modeliranje in vrednotenje modela

Napovedovanja življenjske dobe zobnika smo se lotili tako, da smo sprva določili kovariančne funkcije, ki bi najbolje opisale potek življenjske dobe zobnika. Ker je problem ekstrapolacijski, smo izločili uporabo le stacionarnih funkcij.

Za modeliranje smo uporabili dve različni kovariančni funkciji, tj. **vsoto Matérnove in polinomske kovariančne funkcije ter kovariančno funkcijo nevronske mreže**. Omenjene kovariančne funkcije smo izbrali na podlagi izkušenj modeliranja globalnega segrevanja, opisanega v preteklih poglavjih. Za Matérnovo funkcijo smo se odločili zaradi modeliranja podatkov, kjer na izhodu ne predpostavljamo gladkosti. Poleg tega je verjetnost, da pri optimizaciji hiperparametrov za ekstrapolacijo z metodo največje podobnosti najdemo optimalen lokalni minimum, velika. Ker pa je Matérnova kovariančna funkcija stacionarna in posledično neprimerna za ekstrapolacijo, smo ji prišteli še polinomsko kovariančno funkcijo. Hiperparametri so po optimizaciji zasedli



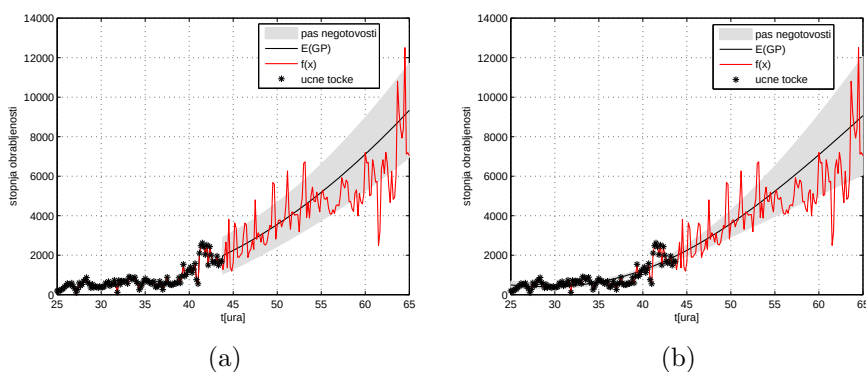
Slika 4.41: Meritev obrabe zobnika na osmem senzorju

vrednosti $l_{MAT} = 0.0077$, $\theta_{MAT} = 448.6755$, $\theta_{POL} = 0.0007$, $\nu = 1.8497$ in $\theta_n = 0.0163$. Iz podobnih razlogov smo izbrali še kovariančno funkcijo nevronske mreže. Pri slednji so bile optimalne vrednosti hiperparametrov naslednje: $l = 42.7599$, $\theta_1 = 8252$ in $\theta_n = 391.3491$. Rezultate modeliranja za različne velikosti učnih množic lahko vidimo na slikah 4.42 (za 112 učnih točk), 4.43 (za 144 učnih točk) in 4.44 (za 192 učnih točk). Poglavitna razlika med modeloma je, da kovariančna funkcija nevronske mreže opiše podatke z gladko krivuljo, medtem ko se model, opisan s kombinacijo, popolnoma prilaga učnim točkam. Posledično je varianca modela pri kombinaciji v območju učnih točk minimalna, se pa poveča v oddaljenosti od učnih točk. Obratno velja za modeliranje s kovariančno funkcijo nevronske mreže, kjer varianca modela postopoma narašča. Poleg tega so vrednosti mer boljše pri modeliranju s kovariančno funkcijo nevronske mreže.

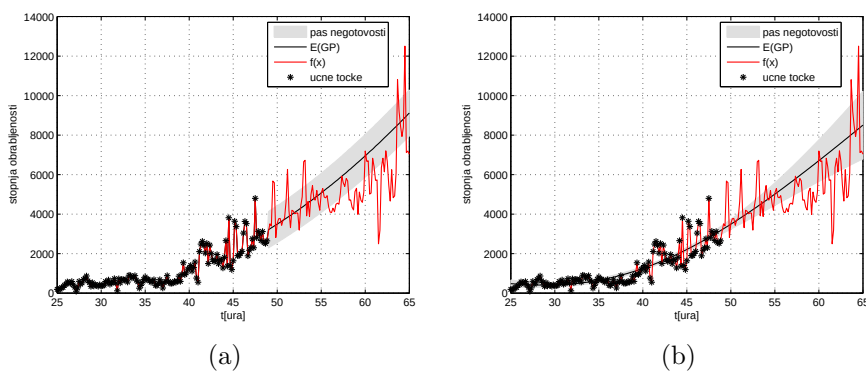
Poleg omenjenih kovariančnih funkcij smo poskusili še vsoto Gaussove in linearne kovariančne funkcije. Ta se ni izkazala za primerno, saj je modelirana funkcija naraščala prepočasi.

4.2.3 Napovedovanje kritične točke

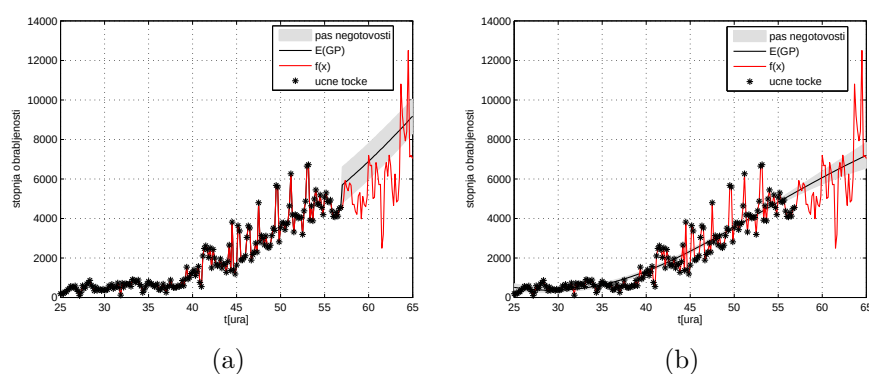
Kritična točka predstavlja vrednost značilke, pri kateri bi bilo smiselno delovanje orodja ustaviti. V našem primeru so upravljalci testne postaje za kritično vrednost značilke določili vrednost 70358. Z modeliranjem želimo napovedati čas pojava kritične vrednosti značilke, da bi lahko avtomatsko sprožili varno-



Slika 4.42: Modeliranje zobnika z vsoto Matérnove in polinomske kovariančne funkcije (a) kovariančne funkcije nevronske mreže (b) za 112 učnih točk



Slika 4.43: Modeliranje zobnika z vsoto Matérnove in polinomske kovariančne funkcije (a) kovariančne funkcije nevronske mreže (b) za 144 učnih točk



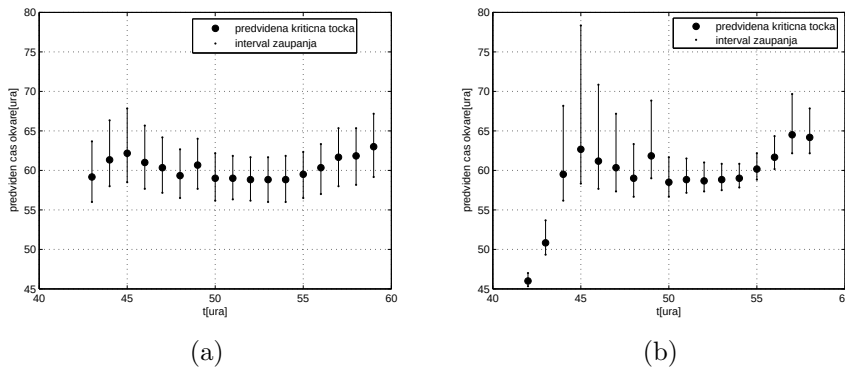
Slika 4.44: Modeliranje zobnika z vsoto Matérnove in polinomske kovariančne funkcije (a) kovariančne funkcije nevronske mreže (b) za 192 učnih točk

	MAE	MSE	LPD	MRSE
Matérnova in polinomska KF 112	1074.8017	2173852.995	9.0171	0.3009
KF nevronske mreže 112	1065.9104	2104769.3764	11.472	0.29606
Matérnova in polinomska KF 144	1160.0941	2467277.0031	11.3916	0.2874
KF nevronske mreže 144	1038.399	2075211.9588	9.8599	0.26354
Matérnova in polinomska KF 192	1623.3136	3901929.5609	9.0171	0.31502
KF nevronske mreže 192	1095.1353	2400581.1221	8.9025	0.24709

Tabela 4.17: Vrednosti mer pri modeliranju orodja za različne velikosti učnih množic. Vse mere razen LL so izračunane na podlagi testne množice, medtem ko je slednja izračunana na podlagi učne množice.

stni alarm.

Pri napovedovanju točke odpovedi smo uporabili kovariančni funkciji in vrednosti hiperparametrov, opisanih v predhodnem podpoglavju. Napovedovanje smo pričeli pri 35. uri delovanja. Točko odpovedi smo napovedovali s korakom ene ure delovanja orodja in sicer tako, da smo vedno znova naredili predikcijo na podlagi vseh podatkov, ki smo jih imeli na voljo do tistega trenutka. Pri vsaki predikciji smo znova optimizirali vrednosti hiperparametrov, saj se optimalne vrednosti spreminjajo v odvisnosti od učne množice. Rezultati napovedovanja za vsoto Matérnovne in polinomske kovariančne funkcije so prikazani na sliki 4.45a, za kovariančno funkcijo nevronske mreže pa na sliki 4.45b. Interval zaupanja prikazuje 95% odstopanje od predvidenega časa, določen pa je na podlagi vrednosti variance napovedi GP modela, ki prva doseže kritično vrednost značilke.



Slika 4.45: Napovedovanje kritične točke z vsoto Matérnovne in polinomske KF (a) ter KF nevronske mreže (b)

Na sliki 4.45a vidimo, da v prvem primeru (slika 4.45a) kritično točko prvič dosežemo po 43. uri delovanja zobnika, odpovedalo pa naj bi po 59. uri. V drugem primeru (slika 4.45b) kritično točko prvič dosežemo po 42. uri delovanja orodja, odpovedalo pa naj bi po 46. uri. V resnici je bila kritična vrednost prvič dosežena pri 60. uri delovanja.

4.2.4 Sklep

Rezultati kažejo, da lahko GP model, pridobljen iz merjenih podatkov, učinkovito napoveduje prihodnje obnašanje značilke in ga zato lahko uporabimo za napovedovanje časa, ki bi bil smislen za sproženje varnostnega alarma. Ocenju-

jemo, da je z uporabljenim načinom modeliranja mogoče napovedati kritično točko obrabe za 15 do 20 ur vnaprej. Ta čas naj bi operaterjem in vzdrževalcem omogočal dovolj časa za zamenjavo orodja, brez povzročanja nepotrebnega zastoja proizvodnje.

Poglavje 5

Zaključek

V diplomskem delu smo opisali modeliranje časovnih vrst z modeli na podlagi Gaussovih procesov. Posebej smo opisali kovariančne funkcije, ki se pri omenjenem načinu modeliranja najpogosteje uporabljajo in jih delimo na stacionarne in nestacionarne. Prednost modeliranja z GP modeli pred podobnimi metodami je, da poleg napovedi da tudi zaupanje v napovedi modela. Kot negativno lastnost lahko navedemo veliko računsko zahtevnost pri optimizaciji hiperparametrov, ki raste s tretjo potenco velikosti učne množice.

Ker je kovariančna funkcija lahko katerakoli funkcija, ki tvori pozitivno definitno matriko, smo v diplomskem delu napisali nekaj smernic za izbiro primerne kovariančne funkcije. Smernice smo ovrednotili na podlagi primera podatkov, kjer smo za modeliranje uporabili več različnih kovariančnih funkcij in primerjali rezultate posameznih modelov. Izkušnje in ugotovitve smo uporabili za določanje kovariančne funkcije za modeliranje v drugem primeru podatkov.

Ugotovili smo, da je izbira kovariančne funkcije ključnega pomena za delovanje GP modela. Z njo izrazimo znanje o obnašanju oz. splošnih lastnostih neznanega sistema, naj je to apriorno ali pa pridobljeno s poskušanjem. Največkrat, zlasti kadar ne poznamo ozadja povezave med podatki, za funkcijski del kovariančne funkcije uporabimo kar Gaussovo kovariančno funkcijo, saj sta gladkost izhoda in stacionarnost mnogokrat predpostavljene lastnosti. To se je izkazalo za ustrezno le, kadar smo napovedi modela uporabili za interpolacijo. Pri uporabi GP modela za ekstrapolacijo moramo upoštevati, da stacionarne kovariančne funkcije daleč od učnih točk težijo k vrednosti nič. Problem lahko rešimo z uporabo nestacionarnih kovariančnih funkcij ali pa s kombinacijo obeh. Kljub temu pa moramo biti pazljivi pri izbiri učnih točk, saj vsaka nova učna točka lahko vpliva na GP model.

Zavedati se moramo, da vse našete ugotovitve temeljijo le na modeliranju dveh primerov in moramo biti previdni pri njihovem posploševanju. Lahko pa te ugotovitve predstavljajo smernice za nadaljnje raziskovanje na področju modeliranja z modeli na podlagi Gaussovih procesov.

Dodatek A

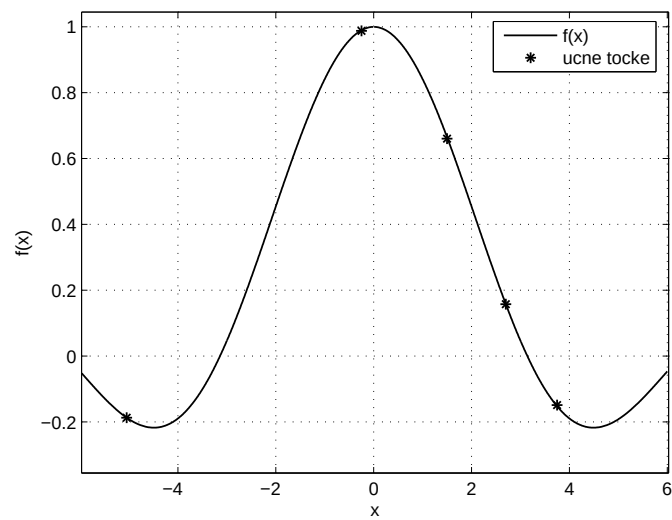
Primer uporabe GP modela

Želimo modelirati nelinearno funkcijo $f(x)$, odvisno od neodvisne spremenljivke x :

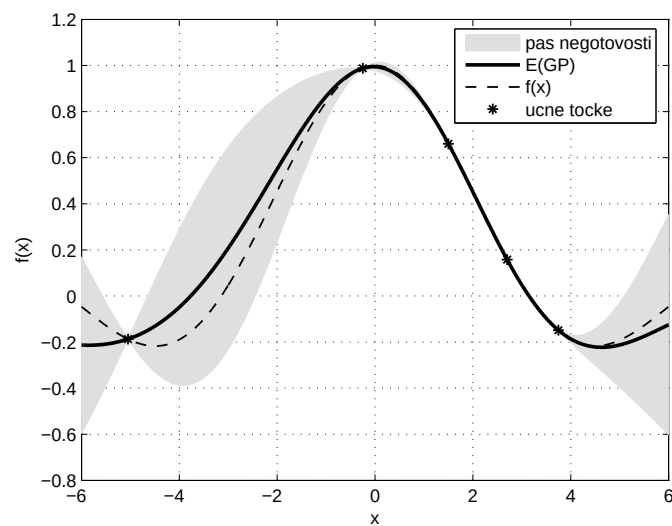
$$f(x) = \sin(x)/x + \nu \quad (\text{A.1})$$

na intervalu $x \in [-6, 6]$. Varianca Gaussovega šuma ν na izhodu je $\sigma^2 = 0.01$. Funkcija je predstavljena s petimi neenakomerno porazdeljenimi učnimi pari (točkami), ki predstavljajo vhodno/izhodno relacijo $x/f(x)$. Funkcija in učne točke so prikazani na sliki A.1.

Za modeliranje izberemo Gaussovo kovariančno funkcijo (3.5). Z optimizacijo določimo tri hiperparametre. Rezultati modeliranja so prikazani na sliki A.2. Lahko opazimo, da model slabo opisuje neznano funkcijo na področju, ki ni opisano z učnimi točkami, prav tako je predikcija slabša na redkeje opisanih področjih. Prednost GP modela je, da nas na slabše opisano področje opozori povečana varianca (negotovost).



Slika A.1: Nelinearna funkcija, ki jo želimo modelirati



Slika A.2: Izhod GP modela

Slike

3.1	Konstantna kovariančna funkcija	15
3.2	Konstantna kovariančna funkcija pri različnih vrednostih hiperparametrov	16
3.3	Gaussova kovariančna funkcija	17
3.4	Gaussova kovariančna funkcija pri različnih hiperparametrih . .	17
3.5	Eksponentna kovariančna funkcija	18
3.6	Eksponentna kovariančna funkcija pri različnih vrednostih hiperparametrov	19
3.7	Racionalni kvadratični polinom	20
3.8	Racionalni kvadratični polinom pri različnih vrednostih hiperparametrov	20
3.9	Matérnova kovariančna funkcija	21
3.10	Matérnova kovariančna funkcija pri različnih vrednostih hiperparametrov	22
3.11	Periodična kovariančna funkcija	23
3.12	Periodična kovariančna funkcija pri različnih vrednostih hiperparametrov	24
3.13	Linearna kovariančna funkcija	25
3.14	Linearna kovariančna funkcija pri različnih vrednostih hiperparametrov	25
3.15	Polinomska kovariančna funkcija	26
3.16	Polinomska kovariančna funkcija pri različnih vrednostih hiperparametrov	27
3.17	Kovariančna funkcija nevronske mreže	27
3.18	Kovariančna funkcija nevronske mreže pri različnih vrednostih hiperparametrov	28
4.1	Temperaturno odstopanje	33
4.2	Učne točke pri interpolaciji	33

4.3	Število pojavov minimumov pri interpolaciji z Gaussovo kovariančno funkcijo pri 43 učnih točkah	34
4.4	Interpolacija z Gaussovo kovariančno funkcijo pri 43 učnih točkah	35
4.5	Interpolacija z Gaussovo kovariančno funkcijo pri 89 učnih točkah	36
4.6	Število pojavov minimumov pri interpolaciji z eksponentno kovariančno funkcijo pri 43 učnih točkah	37
4.7	Interpolacija z eksponentno kovariančno funkcijo pri 43 učnih točkah	37
4.8	Interpolacija z eksponentno kovariančno funkcijo	38
4.9	Število pojavov minimumov pri interpolaciji z racionalnim kvadratičnim polinomom pri 43 učnih točkah	39
4.10	Interpolacija z racionalnim kvadratičnim polinomom pri 43 učnih točkah	40
4.11	Interpolacija z racionalnim kvadratičnim polinomom pri 89 učnih točkah	40
4.12	Število pojavov minimumov pri interpolaciji z Matérnovo kovariančno funkcijo pri 43 učnih točkah	41
4.13	Interpolacija z Matérnovo kovariančno funkcijo pri 43 učnih točkah	41
4.14	Interpolacija z Matérnovo kovariančno funkcijo pri 89 učnih točkah	42
4.15	Število pojavov minimumov pri interpolaciji z linearno kovariančno funkcijo pri 43 učnih točkah	43
4.16	Interpolacija z linearno kovariančno funkcijo pri 43 učnih točkah	43
4.17	Interpolacija z linearno kovariančno funkcijo pri 89 učnih točkah	44
4.18	Število pojavov minimumov pri interpolaciji s kovariančno funkcijo nevronske mreže pri 43 učnih točkah	45
4.19	Interpolacija s kovariančno funkcijo nevronske mreže pri 43 učnih točkah	46
4.20	Interpolacija s kovariančno funkcijo nevronske mreže pri 89 učnih točkah	46
4.21	Število pojavov minimumov pri interpolaciji z zmnožkom Gaussove in periodične kovariančne funkcije pri 43 učnih točkah . . .	47
4.22	Interpolacija z zmnožkom Gaussove in periodične kovariančne funkcije pri 43 učnih točkah	48
4.23	Interpolacija z zmnožkom Gaussove in periodične kovariančne funkcije pri 89 učnih točkah	48
4.24	Učne točke pri ekstrapolaciji	50
4.25	Število pojavov minimumov pri ekstrapolaciji z Gaussovo kovariančno funkcijo pri 43 učnih točkah	51
4.26	Ekstrapolacija z Gaussovo kovariančno funkcijo	51

4.27	Število pojavov minimumov pri ekstrapolaciji z eksponentno kovariančno funkcijo	52
4.28	Ekstrapolacija z eksponentno kovariančno funkcijo	53
4.29	Število pojavov minimumov pri ekstrapolaciji z racionalnim kvadratičnim polinomom	53
4.30	Ekstrapolacija z racionalnim kvadratičnim polinomom	54
4.31	Število pojavov minimumov pri ekstrapolaciji z Matérnovo kovariančno funkcijo	55
4.32	Ekstrapolacija z Matérnovo kovariančno funkcijo	55
4.33	Število pojavov minimumov pri ekstrapolaciji z linearno kovariančno funkcijo	56
4.34	Ekstrapolacija z linearno kovariančno funkcijo	57
4.35	Število pojavov minimumov pri ekstrapolaciji s kovariančno funkcijo nevronske mreže	57
4.36	Ekstrapolacija s kovariančno funkcijo nevronske mreže	58
4.37	Ekstrapolacija z vsoto periodične in linearne kovariančne funkcije	59
4.38	Napovedovanje globalnega ogrevanja	60
4.39	Postavitev senzorjev	61
4.40	Poškodovani vhodna in izhodna gred	62
4.41	Meritev obrabe zobnika	63
4.42	Modeliranje zobnika za 112 učnih točk	64
4.43	Modeliranje zobnika za 144 učnih točk	64
4.44	Modeliranje zobnika za 192 učnih točk	65
4.45	Napovedovanje kritične točke	66
A.1	Nelinearna funkcija	71
A.2	Izhod GP modela	71

Tabele

4.1	Vrednosti mer pri interpolaciji z Gaussovo kovariančno funkcijo	36
4.2	Vrednosti mer pri interpolaciji z eksponentno kovariančno funkcijo	38
4.3	Vrednosti mer pri interpolaciji z racionalnim kvadratičnim polinomom	39
4.4	Vrednosti mer pri interpolaciji z Matérnovο kovariančno funkcijo	42
4.5	Vrednosti mer pri interpolaciji z linearno kovariančno funkcijo .	44
4.6	Vrednosti mer pri interpolaciji s kovariančno funkcijo nevronske mreže	45
4.7	Vrednosti mer pri interpolaciji z zmnožkom Gaussove in periodične kovariančne funkcije	48
4.8	Vrednosti mer pri interpolaciji za 43 učnih točk	49
4.9	Vrednosti mer pri interpolaciji za 89 učnih točk	49
4.10	Vrednosti mer pri ekstrapolaciji z Gaussovo kovariančno funkcijo	52
4.11	Vrednosti mer pri ekstrapolaciji z eksponentno kovariančno funkcijo	52
4.12	Vrednosti mer pri ekstrapolaciji z racionalnim kvadratičnim polinomom	54
4.13	Vrednosti mer pri ekstrapolaciji z Matérnovο kovariančno funkcijo	56
4.14	Vrednosti mer pri ekstrapolaciji z nestacionarnimi kovariančnimi funkcijami	56
4.15	Vrednosti mer pri ekstrapolaciji s kombinacijami kovariančnih funkcij	58
4.16	Vrednosti mer pri ekstrapolaciji	59
4.17	Vrednosti mer pri modeliranju orodja	65

Literatura

- [1] P. Abrahamsen, *A Review of Gaussian Random fields and Correlation Functions*, April 1997
- [2] K. Ažman, *Identifikacija dinamičnih sistemov z Gaussovimi procesi*, Doktorska disertacija, Univerza v Ljubljani, Ljubljana, 2007
- [3] K. Ažman, *Identifikacija dinamičnih sistemov z Gaussovimi procesi z vključenimi lokalnimi modeli*, Magistrsko delo, Univerza v Ljubljani, Ljubljana, 2004
- [4] K. Ažman, J. Kocijan, *Identifikacija dinamičnega sistema z znanim modelom šuma z modelom na osnovi Gaussovih procesov*, Zbornik petnajste elektrotehniške in računalniške konference (ERK 2006), Portorož, 2006.
- [5] M. Gašperin, P. Boškoski, Dj. Juričić *Prognosis of gear health using stochastic dynamical models with online parameter estimation*, IJS, Ljubljana, 2009
- [6] A. Geiger, *Gaussian Processes for Machine Learning*, Doktorska disertacija, Universität Karlsruhe, June 2007
- [7] M. N. Gibbs, *Bayesian Gaussian Processes for Regression and Classification*, Doktorska disertacija, University of Cambridge, Cambridge
- [8] (2009) Global-mean monthly, annual and seasonal land-ocean temperature index. Dostopno na:
<http://data.giss.nasa.gov/gistemp/>
- [9] (2006) GPML toolbox for Matlab. Dostopno na:
<http://www.gaussianprocess.org/gpml/code/matlab/doc/>
- [10] B. Grašič, *Napovedovanje povišanih koncentracij ozona z uporabo umetnih nevronske mreže, Gaussovih procesov in mehke logike*, Magistrsko delo, Univerza v Ljubljani, Ljubljana, 2005

- [11] J. Kocijan, *Modeliranje dinamičnih sistemov z umetnimi nevronskimi mrežami in sorodnimi metodami*, založba Univerze v Novi Gorici, Nova Gorica, 2007
- [12] I. Kononenko, *Strojno učenje*, Založba FE in FRI, Ljubljana, 2005
- [13] B. Likar, *Prediktivno vodenje nelinearnih sistemov na osnovi Gaussovih procesov*, Magistrsko delo, Univerza v Ljubljani, Ljubljana, 2004
- [14] D. Petelin, *Sprotno učenje modelov na podlagi Gaussovih procesov*, Diplomsko delo, Univerza v Ljubljani, Ljubljana, 2009
- [15] C. E. Rasmussen, C. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2006