

Univerza v Ljubljani
Fakulteta za računalništvo in informatiko

Ocenjevanje zanesljivosti posameznih napovedi z analizo občutljivosti regresijskih modelov

doktorska disertacija

Zoran Bosnić

Mentor: prof. dr. Igor Kononenko

Ljubljana, avgust 2007

Radmili, Milanu in Jeleni



Numquam se minus solum quam cum solus esset.

(Cicero)

Povzetek

Disertacija obravnava področje ocenjevanja zanesljivosti posameznih napovedi regresijskih modelov, zgrajenih z uporabo nadzorovanega učenja. V nasprotju s povprečnimi ocenami točnosti regresijskih modelov kot celote nam ocene zanesljivosti posameznih napovedi lahko zagotavljajo dodatno koristno informacijo, ki je lahko v pomoč pri presoji o uporabnosti in posledičnih ukrepih na podlagi teh napovedi. Pogosta področja aplikacije takšne uporabe ocenjevanja zanesljivosti napovedi so medicinska diagnostika, finančne aplikacije in kontrolni sistemi.

V disertaciji je kot novost zasnovana metoda za ocenjevanje zanesljivosti regresijskih napovedi, ki temelji na analizi občutljivosti regresijskih modelov. Razvita metoda meri občutljivost modelov na podlagi sprememb njihovih napovedi, ki nastopijo kot posledica razširjanja učne množice primerov z dodatnim primerom, katerega napoved nas zanima. Analiza zatem primerja spremembo napovedi tega primera s tako dobljenim modelom. V disertaciji predlagamo uporabo novih ocen zanesljivosti na osnovi izmerjenih sprememb, katerih uspešnost primerjamo s klasičnimi ali prilagojenimi ocenami zanesljivosti z drugih področij. Ker je analiza občutljivosti pristop, ki obravnava regresijski model kot parameter postopka, je cela predstavljena metoda splošna in neodvisna od regresijskega modela.

V okviru razvoja nove metodologije disertacija obravnava tudi problem samodejne izbire ocene zanesljivosti, ki bo za dani regresijski model in problemsko domeno najbolj korelirala z napako pri napovedi. Reševanja tega problema se lotimo z dvema pristopoma: z metaučenjem, v okviru katerega predlagamo tudi način atributnega

opisa meta-problemskega prostora, in s pristopom notranjega prečnega preverjanja.

Testiranje uspešnosti ocen zanesljivosti je opravljeno s koreliranjem z napako napovedi posameznih testnih primerov, uspešnost pa je analizirana na podlagi statističnega ovrednotenja izračunanih korelacijskih koeficientov. Poskusi so bili izvedeni na 28 problemskih domenah in z uporabo 8 regresijskih modelov. Med rezultati je izstopala uspešnost novorazvite ocene pri uporabi z regresijskimi drevesi, kjer je ocena značilno korelirala z napako pri 82% testnih domen. V okviru testiranja uporabnosti je bila metodologija preizkušena tudi v realni domeni s področja medicinske prognostike. Prevladujoče število korelacij ocen zanesljivosti z napako napovedi in rezultati testiranja v realni domeni so pokazali potencial uporabe predlagane metodologije v praksi.

Abstract

The dissertation discusses the reliability estimation of individual regression predictions in the field of supervised learning. In contrast with average measures for the evaluation of model accuracy, the reliability estimates for individual predictions can provide additional information which could be beneficial for evaluating the usefulness of the prediction and possible consequential actions. The most common application areas for this challenge are medical diagnosis, financial applications and control systems.

As a novelty, the dissertation proposes a method for reliability estimation of regression predictions, which is based on the sensitivity analysis approach. It defines a framework for measuring the regression model sensitivity based on the change of their predictions, which are influenced by expanding the learning set with the additional example. We propose new reliability estimates which are based on the observed prediction changes and compare them with traditional or adapted reliability estimates from related research fields. Since sensitivity analysis is an approach which considers the regression model as its parameter, the developed method itself is general and independent of the regression model.

The dissertation additionally studies the problem of optimal reliability estimate selection based on the given problem domain and the regression model. We discuss and define two possible solutions of this problem, based on metalearning and internal cross-validation approach. In the context of metalearning we propose possible attribute description of the metalearning problem.

The testing of reliability estimates was performed by correlating the estimates with the prediction error. The performance of estimates was statistically evaluated. For testing, 28 standard benchmark domains from publicly accessible repositories and with 8 regression models were used. The testing results showed usefulness of the proposed reliability estimates for use with the regression trees, where the proposed estimate correlated with the prediction error in 82% of the testing domains. The proposed methodology was tested in a real domain from the area of medical prognostics. The significant number of correlations between the reliability estimates and the prediction error in the majority of tests has showed the potential of the proposed methodology in practice.

Zahvala

Iskreno se zahvaljujem mentorju, prof. dr. Igorju Kononenku, za vso pomoč pri izdelavi doktorske disertacije. Hvaležen sem za vse strokovne nasvete, ideje, razpoložljivost in za mnoge konstruktivne nasvete, ki sem jih bil deležen tudi ob vikendih in preko SMS-sporočil.

Zahvaljujem se članoma Laboratorija za kognitivno modeliranje, doc. dr. Marku Robniku-Šikonji in doc. dr. Matjažu Kukarju, ki sta bila del okolja, v katerem je disertacija nastajala. Obema se zahvaljujem za kvalitetne ideje in strokovne nasvete.

Največja podpora pri nastanku disertacije je prišla s strani ožje družine. Zahvaljujem se mami Radmili in očetu Milanu za vso skrb in motivacijo pri študiju. Zahvaljujem se sestrici Jeleni, ki me je s svojo ambicioznostjo velikokrat spomnila na to, kakšen je bil njen starejši brat pred leti, in mi s tem dala energijo za nadaljevanje dela. Hvala tudi ostalim članom družine, Draganu, babicama, dedku, tetam in stricu.

Prijatelju Martinu Vrtačniku se najlepše zahvaljujem za lektoriranje naloge.

Za podporo nenazadnje hvala tudi prijateljem in prijateljicam Andreji, Damjanu, Gregorju, Jasni, Juretu, Katji, Maji, Mateju, Miši, Mojci, Nastji, Nataši, Nini, Petji, Petru, Špeli, Tatjani in Urošu ter vsem ostalim, ki jim je ta zahvala naredila krivico in jih pomotoma ni omenila.

Kazalo

1	Uvod	1
1.1	Motivacija in cilji	3
1.2	Prispevki k znanosti	5
1.3	Pregled vsebine	6
2	Pregled področja	9
2.1	Regresijsko napovedovanje	10
2.1.1	Regresijska drevesa	11
2.1.2	Linearna regresija	12
2.1.3	Umetne nevronske mreže	13
2.1.4	Lokalno utežena regresija	14
2.1.5	Posplošeni aditivni model	15
2.1.6	Metoda podpornih vektorjev	16
2.1.7	Bagging z regresijskimi drevesi	16
2.1.8	Naključni gozdovi regresijskih dreves	17
2.2	Analiza občutljivosti	18
2.3	Ocenjevanje zanesljivosti napovedi	20
2.3.1	Razširitve napovednih metod	20
2.3.2	Metode, neodvisne od napovednih modelov	25
2.4	Učenje z vzorčenjem primerov iz učne množice	26
2.5	Uporaba neoznačenih primerov pri nadzorovanem učenju	27

2.6	Terminologija področja in njeno poenotenje	30
3	Razvoj ocen zanesljivosti	33
3.1	Razvoj ocen z uporabo analize občutljivosti	33
3.1.1	Vhodne vrednosti analize občutljivosti	34
3.1.2	Izhodne vrednosti analize občutljivosti	38
3.1.3	Ocene zanesljivosti	42
3.2	Druge ocene zanesljivosti	46
3.2.1	Varianca modela bagging	46
3.2.2	Ocena z uporabo lokalnega prečnega preverjanja	47
3.2.3	Zanesljivost na podlagi gostote učnih primerov	48
3.2.4	Ocena lokalne napake	49
4	Testiranje in kombiniranje ocen zanesljivosti	51
4.1	Testiranje ocen zanesljivosti	51
4.2	Kombiniranje ocen zanesljivosti	55
5	Samodejna izbira najbolj ustrezne ocene zanesljivosti	57
5.1	Metaučenje v prostoru domen	58
5.1.1	Testiranje dinamične izbire ocene	58
5.1.2	Razlaga izbire in uspešnosti ocen	59
5.2	Izbira ocene z notranjim prečnim preverjanjem	68
6	Aplikacija ocen zanesljivosti na realni domeni	71
6.1	Množica podatkov	72
6.2	Izbira napovednega modela	72
6.3	Testiranje ocen zanesljivosti v praksi	75
7	Diskusija in zaključek	79
7.1	Pregled metodologije in rezultatov	79
7.2	Časovne zahtevnosti algoritmov	83
7.3	Nadaljnje delo	87
A	Podrobni rezultati testiranja	99

B Knjižnica algoritmov v R	113
B.1 Branje testnih podatkov	114
B.2 Izračun ocen zanesljivosti	115
B.2.1 Prva faza	115
B.2.2 Druga faza izračuna	116
B.3 Metaučenje	117
B.3.1 Tvorjenje učne množice	117
B.3.2 Napovedovanje ocene	118
B.4 Dinamična izbira ocene z notranjim prečnim preverjanjem	118
B.5 Dostopnost algoritmov	119

Poglavje *1*

Uvod

Področje strojnega učenja se ukvarja z gradnjo programskih sistemov, ki se z izkušnjami samodejno izpopolnjujejo [Kononenko in Kukar, 2007; Mitchell, 1997]. Ukvarja se z avtomatskim generiranjem znanja iz zbranih že rešenih učnih primerov in s predstavitvijo generiranega znanja v simbolni ali numerični obliki. Rezultat učenja je znanje, ki ga sistem uporabi za reševanje novih nalog.

Pri obravnavi sistemov za strojno učenje ločimo med učnim algoritmom, ki iz množice podatkov in predznanja tvori novo znanje, in med izvajalnim algoritmom, ki naučeno znanje uporablja za reševanje novih problemov [Kononenko, 2005]. Za generirano znanje ponavadi zahtevamo, da čim bolj ustreza vhodnim podatkom in eksplicitno ali implicitno podanemu predznanju. To znanje lahko uporabimo bodisi

za boljše razumevanje problema, ki ga obdelujemo, bodisi za napovedovanje neznanih primerov.

Naloga napovedovanja se lotevamo z uporabo različnih formalizmov (odločitvena pravila, odločitvena drevesa, regresijska drevesa, nevronske mreže ...), ki jih generiramo na končni množici znanih podatkov. Ti modeli se nato lahko uporabljajo za napovedovanje vrednosti podatkov, ki niso bili vključeni v učni proces gradnje modela. Z nekaterimi modeli pa si lahko tudi pomagamo pri razlagi podatkov, na katerih so bili generirani. Učni proces, ki ga izvajajo algoritmi strojnega učenja, modelira odvisnost med množico neodvisnih spremenljivk in vrednostjo ene ali več odvisnih (ciljnih) spremenljivk. Glede na naravo ciljne spremenljivke ločimo probleme na klasifikacijske in regresijske. Pri klasifikacijskih problemih je ciljna spremenljivka diskretna, njenemu omejenemu številu vrednosti pa pravimo tudi razredi; pri regresijskih problemih je ciljna (regresijska) spremenljivka zvezna in ima s tem (teoretično) neskončno zalogo vrednosti. Temu ustrezno imenujemo naučeno znanje iz prvega primera *klasifikator* (angl. *classifier*) in iz drugega *regresijski prediktor* (angl. *regression predictor*) ali *regresor* (angl. *regressor*).

S področjema strojnega učenja in napovedovanja je tesno povezan tudi pojem *zanesljivosti* (angl. *reliability*), ki ga je možno glede na področje rabe opredeliti na različne načine. V splošnem je pojem zanesljivosti definiran kot *zmožnost osebe ali sistema opravljati svoje funkcije tako v običajnih kot tudi v neugodnih ali nepričakovanih okoliščinah* [Wikipedia, 2007b]. Na področju *programskega inženirstva* (angl. *software engineering*) se pojem zanesljivosti uporablja predvsem v kontekstu robustnosti in odpornosti na napake, v statistiki ta pojem ponazarja konsistentnost množice meritev ali podatkov, praktične znanosti pa zanesljivost definirajo kot mero konsistentnosti ponovitev pri večkratni izvedbi merjenj pod enakimi pogoji. Na področju strojnega učenja se izraz *zanesljivost* nanaša predvsem na pravilnost generiranega znanja (*točnost* napovednega sistema). Ker pa točnost sistema na neoznačenih podatkih, ki niso bili vključeni v učni postopek, ni znana, je primerneje govoriti o *oceni točnosti* in s tem posredno o *oceni zanesljivosti* sistema, kakor pa o njegovi *točnosti*¹.

Poleg optimizacije razumljivosti in kompleksnosti je izboljšanje točnosti predmet

¹*Zanesljivost* je bolj splošen pojem od *točnosti*, ki je le eden možnih aspektov vrednotenja zanesljivosti sistema. Natančen povzetek terminologije področja je podan v razdelku 2.6.

številnih že opravljenih raziskav [Drucker, 1997; Breiman, 1997; Ridgeway in sod., 1999; Schapire, 1999]. Ocena zanesljivosti je lahko pomembna odločitvena informacija pri različnih napovednih sistemih, kjer uporabniki potrebujejo informacijo o stopnji zaupanja v napoved sistema. Ta informacija je lahko kritičnega pomena pri odločitvah, katerih posledica je lahko finančna, strokovna ali drugačna večja škoda (npr. odločitev pri nakupu delnic, napoved medicinske diagnoze itd.).

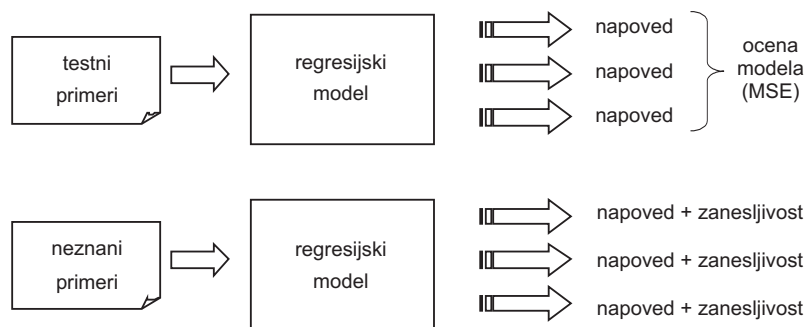
Ker je ocenjevanje zanesljivosti na področju strojnega učenja še pretežno neraziskano in ni metodološko standardizirano, je s tem tudi raziskovalno zanimivo. Disertacija se ukvarja z različnimi pristopi ocenjevanja zanesljivosti posameznih napovedi na področju regresijskega napovedovanja.

1.1 Motivacija in cilji

Metode za ocenjevanje točnosti regresijskih modelov najpogosteje ocenjujejo točnost modela kot celote. To pomeni, da z uporabo primerov iz neodvisne testne množice povprečijo njihove posamezne doprinose napak (npr. srednja kvadratna napaka, relativna srednja kvadratna napaka). Čeprav je takšen pristop primeren za medsebojno primerjavo točnosti različnih modelov, nam istočasno ne zagotavlja informacije o zanesljivosti *posameznih* regresijskih napovedi [Crowder in sod., 1991].

Na področju kritičnih napovednih sistemov (npr. medicinska diagnostika, finančne aplikacije, kontrolni sistemi) nam lokalna informacija o zaupanju v posamezne napovedi sistema lahko omogoča dodaten aspekt presoje dobljenih napovedi. Npr. za uporabnike sistemov v medicinski diagnostiki je večjega pomena ocena pravilnosti posamezne diagnoze pacienta kot pa splošna ocena točnosti napovednega sistema, saj je ocena zanesljivosti posamezne diagnoze tista, na podlagi katere bodo prinesli strokovno odločitev za danega pacienta.

Razliko med obema pristopoma ocenjevanja točnosti prikazuje slika 1.1. Na sliki lahko opazimo večjo splošnost pristopa ocenjevanja posameznih napovedi. Namreč, za razliko od agregirane povprečne ocene regresijskega modela (v našem primeru je to srednja kvadratna napaka – MSE), ki za izračun potrebuje množico označenih testnih primerov, lahko pristop ocenjevanja zanesljivosti posameznih napovedi uporabimo s



Slika 1.1: Ocena zanesljivosti celotnega regresijskega modela s srednjo kvadratno napako – MSE (zgoraj) v primerjavi z oceno zanesljivosti posameznih regresijskih napovedi (spodaj).

poljubnimi neoznačenimi učnimi primeri.

Za ocenjevanje zanesljivosti se v praksi uporabljajo statistični intervali zaupanja. Čeprav ti intervali omogočajo verjetnostno interpretacijo točnosti napovedi, so vezani na specifično verjetnostno porazdelitev in niso izračunljivi pri večini regresijskih modelov. Pri modelih, ki omogočajo izračun intervalov zaupanja, se njihova definicija od modela do modela razlikuje. Za večjo uporabnost takšnega pristopa ocenjevanja zanesljivosti bi torej potrebovali bolj splošno metodo, ki ni vezana na uporabljen regresijski model.

Ker viri nezanesljivosti napovedi velikokrat niso kvantitativno merljivi (poglaviti viri so: šum v podatkih, način posplošitve, pristranost učnega algoritma, mehanizmi izogibanja pretiranemu prilaganju itd.), jih s statističnim opisovanjem problemskega prostora ni možno analizirati. Tudi različne definicije pojma *zanesljivost* so za praktično rabo nekoristne, ker jo opisujejo kot kvalitativno lastnost. Za potrebe praktične implementacije ocenjevanja zanesljivosti potrebujemo torej kvantitavno definicijo tega pojma oziroma metodo za njeno merjenje.

Za razliko od navedenih pristopov je zanesljivost napovedi možno ocenjevati z uporabo tehnike *analize občutljivosti* (angl. *sensitivity analysis*) [Saltelli in sod., 2003]. Ker je to tehnika, ki opazuje delovanje sistema kot celoto in se ne spušča v podrobnosti njegove implementacije, je možno z njo kvantitativno zajeti in izmeriti širši spekter *posebnosti* v obnašanju učnega modela, posredno tudi tisti del, ki je opisljiv kvalitativno.

V doktorski disertaciji se ukvarjamo z zasnovo, razvojem in testiranjem ocen za-

nesljivosti, ki so izpeljane z uporabo tehnike analize občutljivosti in na osnovi lokalnih značilnosti problemskega prostora. Razvite ocene primerjamo z uveljavljenimi metodami za ocenjevanje regresijskih napovedi in preizkušamo različne načine združevanja ocen napovedi z namenom pridobitve bolj optimalne ocene. Metodologijo, predstavljeno v okviru disertacije, testiramo tudi na realnem regresijskem problemu napovedi ponovitve raka na dojki in na podlagi povratne informacije strokovnjakov iz Onkološkega inštituta v Ljubljani vrednotimo njeno delovanje.

1.2 Prispevki k znanosti

Deli disertacije so bili objavljeni v naslednjih pomembnejših člankih: [Bosnić in Kononenko, 2007b,a; Bosnić in Kononenko, 2005; Bosnić in sod., 2003]. Disertacija vsebuje naslednje prispevke k znanosti:

1. **Analiza dosedanjih raziskav na področju disertacije in predlagana poenotena terminologija.** Opravljena sta pregled in analiza raziskav s področij, ki so sorodna področju disertacije. Izdelan je povzetek pogostih pojmov, ki se na tem raziskovalnem podpodročju uporabljajo, podana je tudi razlaga njihove rabe.
2. **Razvoj formalnega ogrodja za uporabo analize občutljivosti v regresijskem napovedovanju.** Predstavljena sta način kontroliranega spreminjanja vhodnih podatkov za analizo občutljivosti in opazovanje izhodnih podatkov – napovedi z namenom njihovega kombiniranja v ocene zanesljivosti.
3. **Ocena uporabnosti pristopa analize občutljivosti pri večih regresijskih modelih.** Analizirana je smiselnost uporabe predlaganega pristopa za ocenjevanje zanesljivosti napovedi pri 8 razširjenih regresijskih modelih. Rezultati analize kažejo na uporabnost ocen zanesljivosti pri kompleksnejših modelih, ki pred lokalnim modeliranjem tudi delijo problemski prostor.
4. **Razvoj in ovrednotenje različnih mer zanesljivosti regresijskih napovedi.** Poleg dveh ocen, zasnovanih z analizo občutljivosti, je zasnovana tudi

nova, neodvisna lokalna ocena zanesljivosti. Dve že obstoječi oceni zanesljivosti sta ravno tako prilagojeni in posplošeni za uporabo s poljubnim regresijskim modelom.

5. **Primerjalna analiza razvitih mer zanesljivosti s standardnimi metodami ocenjevanja zanesljivosti posameznih napovedi.** Ocene so ovrednotene s testiranjem na 28 testnih domenah in z uporabo 8 regresijskih modelov. V okviru testiranja so novo razvite ocene primerjane s standardnimi.
6. **Predlog postopkov za dinamično izbiro optimalne ocene zanesljivosti.** Predlagana je uporaba metaučnja in notranjega prečnega preverjanja v kontekstu samodejne izbire optimalne ocene zanesljivosti za dano kombinacijo domene in regresijskega modela.
7. **Dostopnost algoritmov za ocenjevanje zanesljivosti napovedi za uporabo v statističnem paketu R.** Algoritmi, organizirani v 10 modulov, so dokumentirani in na razpolago za javno uporabo (glej priloženo zgočenko in spletno stran <http://lkm.fri.uni-lj.si/zoranb>).

1.3 Pregled vsebine

Disertacija obsega 7 poglavij. V drugem poglavju je podrobno predstavljeno področje regresijskega napovedovanja s pregledom razširjeno uporabljenih regresijskih modelov. Poglavje predstavlja tudi podroben pregled raziskovalnih področij, ki so povezana s temo disertacije, to so: analiza občutljivosti, ocenjevanje zanesljivosti napovedi, metode perturbiranja učne množice in uporaba neoznačenih primerov pri strojnem učenju. Podan je tudi predlog poenotenja terminologije področja.

Tretje poglavje opisuje razvoj ocen zanesljivosti z uporabo tehnike analize občutljivosti in definira tudi druge ocene, neodvisne od tega postopka, ki jih v nadaljevanju uporabljamo za primerjavo s predlagano metodologijo.

Četrto poglavje opisuje različne pristope testiranja uspešnosti ocen zanesljivosti. Poglavje vsebuje tudi rezultate testiranj ocen zanesljivosti, pridobljenih z linearnim kombiniranjem posameznih ocen.

Poglavje 5 analizira dva pristopa k dinamični izbiri optimalne ocene zanesljivosti glede na izbrano domeno in regresijski model. Poglavje se ukvarja tudi z razlago o poglavitnih kriterijih za izbiro posamezne ocene in analizira težavnost testnih domen v smislu naloge ocenjevanja zanesljivosti napovedi.

Poglavje 6 obravnava uporabnost razvite metode v realni domeni s področja zdravstva. Povzetek in ovrednotenje skupnih rezultatov podaja zaključno poglavje 7, ki navaja tudi pregled idej za nadaljnje delo.

Poglavje 2

Pregled področja

Predlagana disertacija spada na širše področje računalniških znanosti, podrobneje na področje umetne inteligence in strojnega učenja. V okviru tega področja se ukvarja s podpodročjem ocenjevanja zanesljivosti posameznih napovedi v okviru regresijskega napovedovanja in z izboljšanjem njihove točnosti. V nadaljevanju podajamo pregled področij, ki so vezana na temo disertacije, to so: regresijsko napovedovanje, analiza občutljivosti, ocenjevanje zanesljivosti napovedi, učenje z vzorčenjem primerov iz učne množice in uporaba neoznačenih primerov. Raziskave z vsakega izmed teh področij predstavljajo temelj ali motivacijo za zasnovo metode ocenjevanja zanesljivosti z uporabo analize občutljivosti.

2.1 Regresijsko napovedovanje

Cilj regresijskega problema je izpeljati model, ki povezuje vrednost ciljne zvezne spremenljivke y z neznanim primerom x_i , ki je vektor vrednosti – atributov $(x_{i1}, x_{i2}, \dots, x_{in})$. Vhod v algoritem za iskanje neznane regresijske funkcije (gradnjo modela) je množica primerov, ki jo imenujemo *učna množica*. Pravimo, da so primeri v učni množici označeni, kar pomeni, da imajo znano vrednost iskane spremenljivke y . Učni primer lahko torej zapišemo kot vektor

$$\vec{x}_i = (x_{i1}, x_{i2}, \dots, x_{in}, y_i) \quad (2.1)$$

kjer je i indeks učnega primera v učni množici \mathcal{L} , $|\mathcal{L}| = L$, $i = 1 \dots L$. Vrednosti atributov x_{i1}, \dots, x_{in} podanih učnih primerov predstavljajo vhodne vrednosti za gradnjo napovednega modela. V primeru napovedovanja težavnosti obolenja pacientov so ti atributi lahko npr. množica opazovanih simptomov pacienta, pri napovedovanju velikosti odobrenih kreditov pa so to lahko podatki o dosedanjem poslovanju bančnega komitenta s svojimi računi. Vrednost y_i v izrazu (2.1) je tista, ki jo pri neoznačenih primerih napovedujemo, in je v zgornjih primerih enaka iskani težavnosti obolenja pacienta oziroma znesku odobrenega kredita.

Pri tako definiranim primeru \vec{x} in iskani zvezni vrednosti y je regresijski problem definiran [Ridgeway, 2000] kot učenje funkcije $f(\vec{x})$, ki minimizira pričakovano izgubo:

$$J(f) = E_{\vec{x}, y}[C(y, f(\vec{x}))] \quad (2.2)$$

Izraz (2.2) predstavlja *posplošeno napako funkcije* $f(\vec{x})$. V njem nastopa C kot funkcija, ki meri ceno napake, ko za določeni primer napovemo vrednost $f(\vec{x})$, medtem ko je prava vrednost iskane označbe primera enaka y . Za regresijo po metodi najmanjših kvadratov na ta način torej velja:

$$C(y, f(\vec{x})) = (y - f(\vec{x}))^2 \quad (2.3)$$

Funkcija E opredeljuje, na kakšen način se napake napovedi posameznih primerov združujejo v skupno napako napovednega modela. Za E se ponavadi uporablja vsota

vseh vrednosti funkcije C po posameznih primerih.

V nadaljevanju so podrobneje opisani regresijski prediktorji, ki so bili uporabljeni za testiranja v raziskovalnem delu naloge.

2.1.1 Regresijska drevesa

Zasnova regresijskih dreves [Breiman in sod., 1984] je podobna binarnim odločitvenim drevesom, ki jih uporabljamo za klasificiranje primerov v razred, ki je element vnaprej podane diskretne množice. Ker pri regresijskih drevesih napovedujemo vrednost realne spremenljivke, pravimo, da imajo primeri pripisano *označbo* (angl. *label*). Regresijsko drevo torej modelira zvezo med atributi in zvezno spremenljivko.

Vsako notranje vozlišče regresijskih dreves predstavlja odločitev glede na določeno mejno vrednost izbranega atributa (za zvezne attribute) ali podmnožico zaloge vrednosti (za diskretne attribute). Glede na primerjavo atributa s tem mejnim kriterijem se pri napovedovanju določi, ali bo postopek v drevesu nadaljeval z levim ali desnim naslednikom vozlišča. Postopek se ustavi v listih drevesa, ki predpisujejo funkcijo, ki primerom določi napovedano označbo [Karalič, 1992].

Jedro algoritma za gradnjo regresijskega drevesa je zasnovano na rekurzivni delitvi začetne množice primerov v predvidoma dve podmnožici. Algoritem na vsakem koraku upošteva najboljšo možno delitev množice primerov po izbranem kriteriju, ki je običajno minimizacija pričakovane nečistosti množice [Karalič, 1992; Bratko, 2001]:

$$I_{exp} = p_l \cdot I_l + p_d \cdot I_d \quad (2.4)$$

V izrazu (2.4) predstavljata p_l in p_d verjetnosti, da bo primer uvrščen v levo oziroma desno poddrevo, I_l in I_r pa sta meri nečistosti le-teh. Za izračun verjetnosti p_l in p_d ponavadi uporabljamo relativno frekvenco primerov glede na posamezno vejo (pot, ki vodi iz vozlišča k naslednikom). Ker je napovedovani razred pri regresijskih drevesih zvezna spremenljivka, se nam kot mera nečistosti intuitivno ponuja varianca označb primerov v učni množici. Z gradnjo drevesa rekurzivno nadaljujemo, dokler ni izpolnjen izbrani ustavitveni pogoj. Ta je običajno npr. največja dovoljena nečistost primerov v listih drevesa, največje število primerov v listih drevesa ali pa kombinacija

pogojev. V listu priredimo novemu primeru napoved v skladu z modelom, ki je v listu definiran (npr. srednja vrednost označb primerov v listu, mediana, regresijska premica itd.).

Pri gradnji regresijskih dreves se srečujemo s težavo pretiranega prilaganja drevesa učnim podatkom, ki ga rešujemo z uporabo različnih pristopov. Najbolj uporabljena pristopa sta omejevanje rasti drevesa med gradnjo in rezanje drevesa po gradnji. Čeprav je na prvi pogled pristop omejevanja rasti drevesa bolj neposredno vezan na izbiro stopnje splošnosti grajenega modela, se v praksi izkaže, da je izbira ustreznega ustavitvenega kriterija težak problem. Boljše rezultate dosegamo torej s pristopom, kjer pretirano prilagojeno drevo porežemo.

2.1.2 Linearna regresija

Linearna regresija [Chambers in Hastie, 1992] je metoda regresijskega napovedovanja, ki modelira odvisnost med odvisno spremenljivko y_i in neodvisnimi spremenljivkami $x_{i1}, x_{i2}, \dots, x_{in}$:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in} + \varepsilon \quad (2.5)$$

kjer β_0 nastopa kot konstanten člen, členi $\beta_i, i = 1, \dots, n$ nastopajo kot koeficienti neodvisnih spremenljivk, ε pa predstavlja nepričakovano ali nepojasnjeno varianco odvisne spremenljivke. Metodo imenujemo *linearno* zato, ker je odvisna spremenljivka modelirana kot linearna funkcija neodvisnih spremenljivk in parametrov.

Za izračun regresijskega modela se najpogosteje uporablja metoda najmanjših kvadratov (Gauss, 1820). Če na podani učni množici minimiziramo vsoto kvadratov odstopanj po odvisni koordinati in s parcialnim odvajanjem poiščemo ekstreme funkcije

$$f(\beta_0, \dots, \beta_n) = \sum_{k=1}^n (y_k - \beta_0 - \beta_1 x_{k1} + \dots + \beta_n x_{kn} - \varepsilon)^2 \quad (2.6)$$

dobimo z rešitvijo sistema enačb vrednosti nastopajočih koeficientov. Za dvorazsežni prostor, torej za model $y = \beta_0 + \beta_1 x$, na ta način izpeljemo [Jamnik, 1995]:

$$\beta_1 = \frac{n \sum_k x_k y_k - \sum_k x_k \sum_k y_k}{n \sum_k x_k^2 - (\sum_k x_k)^2} = \frac{\sum_k (x_k - \bar{x})(y_k - \bar{y})}{\sum_k (x_k - \bar{x})^2} \quad (2.7)$$

$$\beta_0 = \bar{y} - b\bar{x} \quad (2.8)$$

Omenjeni sistem enačb se v praksi običajno rešuje v matrični obliki. Za reševanje tega sistema se je v praksi najbolj uveljavila uporaba metode *SVD* (angl. *singular value decomposition*) (Press, 1992), ki odstrani vpliv manj pomembnih spremenljivk in s tem zagotovi izračun razumljivejšega linearnega modela.

2.1.3 Umetne nevronske mreže

McCulloch in Pitts [McCulloch in Pitts, 1943; Kononenko, 1990; Dobnikar, 1990] sta leta 1943 matematično formulirala model biološkega živčnega sistema z uporabo modela umetne živčne celice ali nevrona. Ta je definiran kot celica z več vhodnimi povezavami in eno izhodno povezavo. Vsaki vhodni povezavi pripada utež w_k , $k = 1, \dots, n$, s katero se pri računanju izhodne vrednosti o utežijo vrednosti vhodnih povezav i_k , $k = 1, \dots, n$, torej:

$$o = \sigma \left(\sum_{k=1}^n w_k i_k - P \right) \quad (2.9)$$

V izrazu (2.9) vidimo, da je poleg vsote produktov vhodnih vrednosti z utežmi izhod odvisen tudi od posebne razmejitvene vrednosti P , ki jo imenujemo prag. Uteženo vsoto vhodov, upoštevajoč tudi vrednost praga, nam v izhodno vrednost preslika funkcija σ , ki jo imenujemo tudi izhodna ali aktivacijska funkcija. Za to funkcijo se najpogosteje izbere funkcija `arctan` ali druga iz družine sigmoidnih funkcij.

Z medsebojnim povezovanjem vhodov in izhodov več umetnih nevronov lahko le-te povežemo v strukturo, ki jo imenujemo umetna nevronska mreža. Nevroni v mrežah so najpogosteje organizirani v več kaskadnih plasti, med katerimi velja pravilo usmerjenosti. To pravilo določa, da izhod izbranega nevrona lahko vstopa kot vhod nevrona, ki se nahaja v izključno naslednji kaskadni plasti. Umetno nevronske mrežo ponavadi sestavlja *vhodni nivo*, eden ali več *skritih nivojev* in *izhodni nivo*, ki je pri regresijskem napovedovanju običajno sestavljen iz enega samega nevrona.

Za dosego pravilnega preslikanja vhodnih vrednosti v izhodne je potrebno izvesti učenje nevronske mreže. Kot pomembno prelomnico na tem področju štejemo delo avtorjev Rummelharta, Hintona in Williamsa leta 1986, ki so predlagali t. i. posplošeno delta učno pravilo – algoritem za *učenje nevronskih mrež z vzratnim razširjanjem napake* (angl. *back-propagation of error, BPG*).

Učenje po omenjeni metodi poteka v več zaporednih iteracijah ali *epochah* [Tveter, 2003; Brierley in Batty, 1999; Michie in sod., 1994]. V vsaki iteraciji predstavimo nevronske mreži nov primer iz učne množice. Glede na razliko med izhodno in želeno (ciljno) vrednostjo na izhodnem nivoju nato z uporabo delta pravila sklepamo na potrebne popravke uteži ter izvedemo popravke v smeri od zadnjega (izhodnega) nivoja mreže proti vhodnemu.

2.1.4 Lokalno utežena regresija

Metoda napovedovanja z lokalno uteženo regresijo je sorodna metodi k najbližjih sosedov. Gre za neparometrično statistično metodo s področja *lokalnega modeliranja* oziroma *lenega učenja* (angl. *lazy learner*) [Hastie in sod., 2001; Wettschereck in sod., 1997].

Označbo novega neoznačenega primera izračunamo s pomočjo vnaprej izbranega števila učnih primerov, ki so v atributnem prostoru novemu primeru najmanj oddaljeni; kot alternativno rešitev pa lahko ob ustrezni jedrni funkciji uporabimo tudi vse učne primere. Napoved izračunamo z uteženim povprečenjem označb najbližjih sosedov, tako da bližji primeri na označbo vplivajo močnejše kakor tisti, ki so bolj oddaljeni. Utež sosedov kot funkcijo razdalje določimo s poljubno *jedrno funkcijo* (angl. *kernel function*), ki je najpogosteje zvončaste oblike (npr. izpeljanka funkcije gostote verjetnosti zvezne normalne porazdelitve).

Napoved označbe y neoznačenega primera \vec{x} izračunamo po naslednjem obrazcu:

$$y = \frac{\sum_i^N f\left(\frac{D(\vec{x}_i, \vec{x})}{h}\right) \cdot y_i}{\sum_i^N f\left(\frac{D(\vec{x}_i, \vec{x})}{h}\right)} \quad (2.10)$$

V izrazu (2.10) je D izbrana metrika razdalje med primeroma, f jedrna funkcija, N število učnih primerov in parameter h t. i. *pasovna širina* [Cleveland in Loader, 1995]. Slednji parameter igra vlogo normalizacijske konstante glede na velikost atributnega prostora učnih primerov. Kadar v izrazu (2.10) nastopa le k učnih primerov, se pogosto določi, da je vrednost parametra h enaka razdalji med neoznačenim primerom in najbolj oddaljenih od podanih k primerov [Torgo, 1997]. Najpogosteje pa se uporablja vrednost $h = 1$.

Metoda za lokalno modeliranje se je uveljavila kot uspešna in robustna, ima pa to pomanjkljivost, da ne omogoča vpogleda v naučeno znanje. Zato v primerih, ko je interpretabilnost ključnega pomena, raje uporabljamo druge metode.

2.1.5 Posplošeni aditivni model

Posplošeni aditivni model (angl. *generalized additive model*) [Wood, 2006; Hastie in Tibshirani, 1990] je regresijska metoda, ki je posplošen naslednik linearne regresije. Namesto uteževanja neodvisnih spremenljivk s konstantnimi faktorji, kot v enačbi (2.5), je posplošeni aditivni model definiran kot

$$y_i = \beta_0 + f(x_{i1}) + f(x_{i2}) + \dots + f(x_{in}) \quad (2.11)$$

Funkcija $f(x)$ je pri tem lahko izbrana kot poljubna, najpogosteje neparametrična funkcija. Na ta način posplošeni aditivni model zagotavlja večji potencial natančnejšega modeliranja učnih podatkov, zaradi istega razloga pa je istočasno naklonjen tudi pretiranemu prileganju učnim podatkom. V izogib tega se najpogosteje uporablja glajenje izhodne funkcije in notranje prečno preverjanje modela.

Učenje modela se izvaja s postopkom *vzratnega prileganja* (angl. *backfitting*) (metoda po Gauss-Seidlu), ki glede na napako modela v iteracijah izmenično prilagaja in gladi njegove parametre. Kot kriterij prileganja se najpogosteje uporablja kriterij napake najmanjših kvadratov, za glajenje pa lokalne regresijske metode ali gladilne funkcije.

2.1.6 Metoda podpornih vektorjev

Metoda podpornih vektorjev (angl. *support vector machines*) [Vapnik, 1995] v regresiji je prilagoditev klasifikacijske metode podpornih vektorjev [Christiannini in Shawe-Taylor, 2000; Chang in Lin, 2001]. Osnovna ideja klasifikacijske metode je v atributnem prostoru poiskati hiperravnino, ki optimalno loči različne razrede primerov. Kot *optimalna* ravnina je definirana tista, ki je enako in najbolj oddaljena od najbližjih primerov iz obeh razredov. Najbližjim primerom optimalne hiperravnine pravimo *podporni vektorji*, razdalji hiperravnine od podpornih vektorjev pa *rob* (angl. *margin*).

Problem iskanja optimalne hiperravnine se rešuje kot minimizacijski problem s podanimi omejitvami (tvorimo jih iz množice točk – učnih primerov). Ker gre za kvadratični optimizacijski problem, lahko za njegovo reševanje uporabimo hitre algoritme [Burges, 1998; Kononenko, 2005].

Za uporabo v regresiji je metoda podpornih vektorjev prilagojena tako, da definira rob okoli regresijske hiperravnine na drugačen način [Smola in Schölkopf, 1998; Schölkopf in sod., 1998]. Znotraj roba predpostavlja, da je odvisna spremenljivka za vse primere eksaktno napovedana (odstopanje vrednosti odvisnih spremenljivk od napovednih vrednosti modela za te primere torej ne prispeva h kriterijski funkciji za optimizacijo modela), primeri na robu (imenujemo jih podporni vektorji) pa so tisti, ki dejansko določajo potek hiperravnine in s tem strukturo modela. Primeri zunaj roba so torej tisti, ki prispevajo k napaki, ki jo optimizira kriterijska funkcija.

Nasprotno kot pri klasifikacijski metodi podpornih vektorjev želimo tu rob minimizirati, hkrati pa želimo minimizirati tudi napako na učnih primerih. Problem iskanja regresijske optimalne hiperravnine je definiran tudi kot optimizacijski problem kriterijske funkcije, ki upošteva napake napovedi regresijske spremenljivke na učnih primerih.

2.1.7 Bagging z regresijskimi drevesi

Bagging [Breiman, 1996] je metoda, ki temelji na pristopu večkratne gradnje prediktorja na podlagi permutacij s ponavljanjem začetne učne množice in kombiniranju

rezultatov posameznih napovedi.

V prvem koraku metode *bagging* (ime izvira iz izraza *bootstrap aggregating*) se generira t. i. *bootstrap* učna množica B , ki ima enako število primerov kot prvotna učna množica E . Množico B pridobimo iz množice E z naključnim izbiranjem primerov s ponavljanjem. Zaradi večkratne izbire istih primerov je tako generirana množica B drugačna od E in predstavlja šibek povzetek porazdelitve učnih primerov v problemskem prostoru. Kombinatorične analize [Ridgeway, 2000] kažejo, da pri generiranju množice B na opisan način v povprečju v B ne vključimo približno 37% učnih primerov iz E .

Če na zgoraj opisani način konstruiramo več *bootstrap* učnih množic B_1, B_2, \dots, B_n , lahko na njihovi osnovi konstruiramo enako število regresijskih prediktorjev, od katerih vsak prispeva svojo napoved za nov, neoznačen primer. S povprečenjem teh posameznih napovedi ali njihovim uteženim kombiniranjem lahko izračunamo kombinirano napoved modela *bagging* kot samostojnega regresijskega prediktorja. Empirični rezultati so pokazali, da lahko na ta način dosežemo izboljšanje točnosti napovedi v primerjavi z napovedjo samostojno zgrajenega prediktorja.

Dodatna prednost metod, ki gradijo prediktorje na podmnožicah prvotne učne množice E , je ta, da lahko neuporabljene primere $E \setminus B$ uporabimo za oceno napake zgrajenega prediktorja (angl. *out-of-bag estimate*). Čeprav lahko z uporabo tega pristopa kombiniramo napovedi za poljuben regresijski model, se v praksi tehnika *bagging* najpogosteje uporablja s klasifikacijskimi ali regresijskimi drevesi.

2.1.8 Naključni gozdovi regresijskih dreves

V kombinaciji z regresijskimi drevesi je metoda naključnih gozdov [Breiman, 2001] regresijski prediktor, ki temelji na izboljšavi metode *bagging*. Prediktor je ravno tako sestavljen iz množice neodvisnih regresijskih dreves, ki so na enak način kakor pri metodi *bagging* zgrajeni na permutacijah s ponavljanjem prvotne učne množice in tudi njegova napoved je enaka kombinaciji napovedi posameznih dreves v gozdu.

Glavna razlika med obema metodama je v zgradbi regresijskih dreves, ki so v gozdu naključnih dreves zgrajena z vpeljavo elementa naključnosti (odtod ime *naključna drevesa*). Postopek gradnje naključnega drevesa poteka na naslednji način:

1. Naj bo N število učnih primerov in M število atributov učnih primerov.
2. Izberimo parameter algoritma m , $m \ll M$, ki je konstanten za ves postopek gradnje drevesa.
3. Z vzorčenjem prvotne učne množice izberemo permutacijo s ponavljanjem množice z N učnimi primeri. Ostale primere shranimo za ocenjevanje napake regresijskega drevesa.
4. Vsako notranje vozlišče drevesa določimo z obravnavo mejnih vrednosti le m naključno izbranih atributov (od skupno M) za to vozlišče.
5. S takim postopkom deljenja učne množice rekurzivno nadaljujemo, dokler drevesa ne zgradimo do konca. Drevesa ne porežemo.

Ker algoritem za gradnjo naključnega drevesa omejuje število obravnavanih atributov v vsakem vozlišču, je postopek gradnje hitrejši v primerjavi z gradnjo regresijskega drevesa. Ker jih obenem gradimo do največje možne globine in ne uporabljamo rezanja, dobimo s tem množico dreves, ki je vsako zase zelo specifično za problemski prostor, ki ga pokriva. Zaradi prednosti v hitrosti algoritma, se v praksi uporablja gradnja večjega števila dreves kakor pri metodi bagging, s čimer si model obenem zagotovi večjo pojasnjeno varianco problemskega prostora. Posledično daje model tudi na kombiniranih napovedih dobre rezultate in postaja v praksi vse bolj razširjen.

2.2 Analiza občutljivosti

Analiza občutljivosti (angl. *sensitivity analysis*) je tehnika, ki se uporablja za preučevanje vpliva sprememb parametrov in strukturnih lastnosti modela na njegovo strukturo in izhode [Breierova in Choudhari, 1996]. Analiza občutljivosti se običajno izvaja kot zaporedje poskusov, pri katerih uporabnik kontrolirano spreminja omenjene parametre modela in opazuje dinamiko sprememb.

Običajno so modeli odvisni od parametrov, ki so lahko težko ali celo nemogoče merljivi. V takih okoliščinah nam lahko omenjena tehnika pomaga oceniti zaupanje

v izbiro vrednosti parametrov modela. Z opazovanjem izhodov modela lahko uporabnik ugotovi, kateri parametri so za model pomembni in kakšna stopnja natančnosti je potrebna pri izbiri njihovih vrednosti. Ta tehnika je primerno orodje za izbiro modelov in ovrednotenje le-teh, istočasno pa ne zahteva znanja o matematičnih lastnostih modela, saj ga obravnava kot črno škatlo z vhodi in izhodi [Kleijnen, 2001].

Analiza občutljivosti je pristop, ki se najpogosteje uporablja na področju statistike, matematičnega programiranja [Saltelli in sod., 2003] in naravoslovnih znanosti [Saltelli in sod., 2005]. Čeprav je bila na področju strojnega učenja uporabljena za analizo posameznih regresijskih modelov (nevronskih mrež [Hashem, 1992] in Bayesovih mrež [Kjaerulff in van der Gaag, 2000]), je za področje disertacije bolj zanimivo področje splošnejšega ocenjevanja učnih algoritmov. To temo sta podrobneje obravnavala Bousquet in Elisseff [2002]. Z uporabo analize občutljivosti definirata pojem β -stabilnosti učnega algoritma kot algoritma, katerega pričakovana funkcija izgube se ne spremeni za več kot β ob majhnih spremembah v učni množici. Avtorja na podlagi postavljene definicije v nadaljevanju izpeljeta meje napake posplošitve z uporabo empirične napake in napake, dobljene s tehniko *izloči-enega*. V nadaljnjih raziskavah so bili dobljeni teoretični rezultati in meje napake uporabljeni in testirani na različnih učnih modelih [Bousquet in Elisseff, 2000; Bousquet in Pontil, 2003].

Na podoben način sta tudi Kearns in Ron [1997] definirala pojem *stabilnosti hipoteze*, ki jo testirata z odstranjevanjem posameznih primerov iz učne množice. Čeprav so opisane raziskave večinoma teoretične narave in temeljijo na teoriji VC (Vapnik-Chervonenkis) [Vapnik, 1995], nam ponujajo idejo spreminjanja učne množice in merjenja posledičnih sprememb modela [Bousquet in Elisseff, 2002]. S pozitivnimi teoretičnimi rezultati nas ravno tako motivirajo, da uporabimo to idejo na drugih podpodročjih strojnega učenja.

V okviru doktorske disertacije predlagamo aplikacijo analize občutljivosti tako, da na podlagi spreminjanja učne množice opazujemo spremembo v napovedi posameznih primerov. Glede na spremembe v opazovanih izhodih nato v nadaljevanju definiramo ocene zanesljivosti posameznih napovedi.

2.3 Ocenjevanje zanesljivosti napovedi

Dosedanje raziskave so se na problem ocenjevanja zanesljivosti posameznih napovedi sklicevale z različnim izrazoslovjem in v različnih kontekstih strojnega učenja, najpogosteje pa na področju klasifikacijskega napovedovanja.

S principom ocenjevanja zanesljivosti napovedi je pogosto tesno povezan princip *transdukcije* ali *transduktivnega sklepanja*. Konteksti rabe tega pojma so v strojnem učenju zelo različni, v osnovi pa transdukcija predstavlja *sklepanje s posameznega na posamezno* [Vapnik, 1995]. Za razliko od induktivnega pristopa, ki se ukvarja z modeliranjem splošnega znanja za napovedovanje neoznačenih primerov, se transduktivni pristop ukvarja s problemom, kako poiskati napoved neoznačenih primerov, ne da bi pri tem konstruiral splošno pravilo za njihovo napovedovanje. Zaradi opisane zasnove je pojem transduktivnega sklepanja povezan z algoritmi za *učenje na osnovi posameznih primerov* (angl. *instance-based, case-based learning*), med katere spada tudi algoritem *k* najbližjih sosedov. Istočasno pa je bolj splošen, saj je za razliko od omenjenega algoritma lahko transduktivno napovedovanje zasnovano tudi na drugih kriterijih (npr. na verjetnostni porazdelitvi) in ne nujno na razdaljah v problemskem prostoru in označbah najbližjih sosedov.

V nadaljevanju so povzete metode s področja klasifikacije in regresije z namenom izdelave pregleda nad obstoječo metodologijo s tega področja.

2.3.1 Razširitve napovednih metod

Najpogostejši pristop k ocenjevanju zanesljivosti posameznih napovedi regresijskega modela je razširitev specifičnega modela, da poleg napovedi poda tudi informacijo o njeni zanesljivosti. Takšen pristop ima to prednost, da je lahko dobro matematično argumentiran na osnovi formalizacije modela, istočasno pa ima to slabost, da je vezan na specifičen model in ga ni možno posplošiti.

Zanesljivost napovedi metode podpornih vektorjev. S kombinacijo iskanja napovedi s pridruženimi ocenami zanesljivosti in transduktivnega sklepanja se srečamo pri razširitvi klasifikacijske metode podpornih vektorjev avtorjev Gammerman, Vovk

in Vapnik [1998]. Predlagana metoda odpravlja pomankljivost osnovne metode podpornih vektorjev, ki ne izkorišča dodatne informacije pri postopku učenja modela, kot je porazdelitev primerov, koeficienti optimizacijske kriterijske funkcije ter delitev učnih primerov med podporne in ne-podporne vektorje. Ogradje ocenjevanja zanesljivosti klasifikacij s to metodo avtorji zastavijo s teoremom, ki pravi, da je verjetnost napačne klasifikacije neoznačenega primera x_{l+1} (podane so klasifikacije točk x_1, \dots, x_l) manjša oziroma največ enaka izrazu

$$\frac{E(\text{število podpornih vektorjev med } x_1, \dots, x_{l+1})}{l+1} \quad (2.12)$$

pri čemer je E matematično upanje, točke x_1, \dots, x_{l+1} pa so neodvisno porazdeljene na osnovi verjetnostne porazdelitve P . Za uporabo tega teorema v praksi bi morali torej poznati tudi P , poznamo pa le koordinate točk $(x_1, y_1), \dots, (x_l, y_l), x_{l+1}$. Ker ti podatki za izračun izraza (2.12) ne zadoščajo, je treba poiskati drugačen način aproksimacije te porazdelitve. Zato avtorji predlagajo uporabo t. i. *permutacijske mere neverjetnosti* (angl. *permutation measure of impossibility*) $p : Z^m \rightarrow \mathbb{R}$, s katero merijo kakovost *umestitve* izbranega primera s pripisanim izbranim razredom v porazdelitev P . Po izbiri funkcije p , za katero avtorji pokažejo, da mora zadoščati določenim formalnim zahtevam, lahko klasifikacijo¹ primera x_{l+1} in njeno zanesljivost izračunamo s pomočjo količin:

$$\mu_{-1} = \frac{1}{p((x_1, y_1), \dots, (x_l, y_l), (x_{l+1}, -1))} \quad (2.13)$$

$$\mu_1 = \frac{1}{p((x_1, y_1), \dots, (x_l, y_l), (x_{l+1}, 1))} \quad (2.14)$$

Z uporabo (2.13) in (2.14) lahko napovemo:

$$\text{razred} : \begin{cases} 1 & \text{če } \mu_1 > \mu_{-1} \\ -1 & \text{če } \mu_1 < \mu_{-1} \\ \text{poljuben} & \text{če } \mu_1 = \mu_{-1} \end{cases} \quad (2.15)$$

¹metodologija obravnava le dvorazredne klasifikacijske probleme z razredoma -1 in 1

in *nezanesljivost klasifikacije* (angl. *incertitude*):

$$\mu = \min(\mu_{-1}, \mu_1) \quad (2.16)$$

Iz izračunane nezanesljivosti klasifikacije μ pa lahko izrazimo *zanesljivost klasifikacije*, ki je enaka $1 - \mu$. Avtorji metode opozarjajo, da metoda ne dosega tako dobrih rezultatov v smislu točnosti napovedi kot osnovni klasifikacijski algoritem z metodo podpornih vektorjev, daje pa dobre rezultate pri ocenjevanju zanesljivosti napovedi. Ta vidik je lahko v nekaterih primerih celo pomembnejši kot uspešnost metode glede na klasifikacijsko točnost.

Zaupanje in verodostojnost napovedi metode podpornih vektorjev. Kot nadaljevanje zgornjega raziskovalnega dela avtorji Saunders, Gammerman in Vovk [1999] predstavijo razširitev opisane metode, ki dosega boljšo klasifikacijsko točnost. Poleg mere *zaupanja* v napoved (angl. *confidence*, definicija podana v nadaljevanju), avtorji predlagajo tudi dodatno mero *verodostojnosti* (angl. *credibility*) kot drug vidik na zanesljivost napovedi. Jedro metode temelji na postopku dodajanja k neoznačenih primerov iz testne množice v učno množico in gradnji ločenega prediktorja SVM za vsako možno kombinacijo klasifikacij teh k neoznačenih primerov. V nadaljevanju se avtorji omejujejo na obravnavo primera dvorazrednih klasifikacijskih problemov in na dodajanje le enega neoznačenega primera v učno množico, torej na $k = 1$. Predstavljena metoda izkorišča Lagrangeeve multiplikatorje – koeficiente α_i , ki so del reševanja optimizacijskega problema pri učenju modela. Iz teorije sledi, da visoke vrednosti koeficientov α_i pripadajo nenavadnim učnim primerom, ki jih redko srečamo. Čim manjša je torej vrednost Lagrangeevega koeficienta, tem bolj smo lahko prepričani v točnost napovedi. Na osnovi tega avtorji definirajo mero *verodostojnosti* (angl. *credibility*), katere velikost govori o meri zaupanja v napoved glede na porazdelitev preostalih in pogostost pojavljanja podobnih učnih primerov. Ocenjujejo jo z verjetnostno mero *p-value*, ki je definirana kot:

$$p - value = \frac{\#\{i : \alpha_i \geq \alpha_{new}\}}{l + 1} \quad (2.17)$$

V (2.17) predstavlja α_i Lagrangeev koeficient dodanega testnega primera, $l + 1$ pa število vseh primerov v učni množici. Algoritem klasificira neoznačen primer (x_{new}, y_{new}) v tisti razred y_{new} , za katerega je količina (2.17) največja. Na podoben način je definirana tudi mera *zaupanja* (angl. *confidence*), ki predstavlja stopnjo zavračanja klasifikacije nasprotnega razreda (pri dvorazrednem problemu):

$$Confidence = 1 - p_2 \quad (2.18)$$

kjer je p_2 vrednost verodostojnosti klasifikacije v nasprotni razred.

Preprosti pristopi k ocenjevanju zanesljivosti napovedi regresijskih modelov. Na področju regresijskega napovedovanja pristopajo avtorji k ocenjevanju zanesljivosti napovedi z uvedbo preprostih sprememb regresijskih modelov. Weigend in Nix [1994] predlagata razširitev modela večplastnega perceptrona z dodatnim izhodnim nevrom, ki za napoved učnega primera na vhodnih povezavah napoveduje zanesljivost napovedi. Za te namene se vsakemu učnemu primeru pridruži še lokalna informacija o srednji vrednosti in varianci označb v tisti točki problemskega prostora. Ocena zanesljivosti primera je tedaj enaka pripisani varianci, ki se med učenjem uporabi kot ciljna vrednost na izhodnem nevronu za napovedovanje zanesljivosti. Metoda na ta način tudi za neoznačene primere aproksimira pričakovano varianco v delu problemskega prostora, kamor primer spada, in z njo oceni zanesljivost napovedi. Rezultati so pokazali, da metoda uspešno ocenjuje lokalno zanesljivost. Ima pa to pomanjkljivost, da je zaradi izkoriščanja lokalnosti njen uspeh občutljiv na spremembe v problemskem prostoru in se zato pojavijo z njo težave že pri postopku prečnega preverjanja, ko se zaradi vzorčenja porazdelitev v učni množici spremeni. Ravno tako je pri učenju nevronske mreže treba skrbeti za to, da ne pride do pretiranega prileganja nevrona, ki napoveduje zanesljivost, lokalni varianci, tako da lahko njegova napoved ostane dovolj splošna tudi za neznane primere.

Nouretdinov, Melluish in Vovk [2001] uporabijo informacijo Lagrangeevih koeficientov pri metodi podpornih vektorjev [Saunders in sod., 1999] in definirajo verjetnostno mero p -vrednost, ki je primerna za uporabo pri metodi *ridge regression*. S predlaganim pristopom uspešno razširijo osnovni model z intervali zaupanja in ga

poimenujejo *ridge regression confidence machine*.

Heskes predlaga način ocenjevanja zanesljivosti agregiranega prediktorja regresijskih nevronskih mrež [Heskes, 1997] (bagging z uporabo nevronskih mrež). V svojem prispevku izpostavi razliko med *intervalom zaupanja* (angl. *confidence interval*) in *intervalom napovedi* (angl. *prediction interval*). Širino prvega definira kot interval, s katerim ponazarjamo stopnjo ujemanja med napovedano vrednostjo in pravo vrednostjo učnega ali testnega primera; širino drugega pa kot ponazoritev stopnje ujemanja napovedane vrednosti z označbo učnega primera (ki se od prave vrednosti lahko razlikuje zaradi šuma v podatkih). Poudarja, da je za aplikacijo na realnih problemih bolj praktična uporaba intervalov zaupanja in jih za neznani učni primer definira z uporabo variance napovedi nevronskih mrež v kombiniranem regresorju. Za izračun intervala napovedi avtor predlaga uporabo ločene nevronske mreže za modeliranje šuma posameznih označb primerov. Z uporabo predpostavke, da je šum aditiven pravim vrednostim primerov, avtor izpelje oceno intervala napovedi.

V nadaljnjih raziskavah Carney in Cunningham [1999] izboljšata delo Heskesa z uvedbo stabilnejših intervalov zaupanja in napovedi. Te dosežejo tako, da namesto izračuna variance vseh nevronskih mrež v agregatu le-tega razdelijo na manjše, enako velike dele in izračunajo povprečje varianc vsakega izmed njih. Na ta način dosežejo bolj zglajeno in stabilno oceno intervala zaupanja, ki skupaj z izpeljanim intervalom napovedi dajeta boljše rezultate.

Iz pregleda razširitev posameznih napovednih modelov vidimo, da uspešno dosegajo zastavljen cilj ocenjevanja zanesljivosti napovedi. Vendar pa imajo to pomanjkljivost, da so te razširitve specifične za svoj regresijski model in njihovega pristopa ni možno analogno uporabiti na drugih modelih. Ta ugotovitev nas motivira k razvoju pristopa, ki bo neodvisen od regresijskega modela in ga bo obravnaval le kot parameter metode. Ta idejna zasnova pa sovпада s pristopom analize občutljivosti, predstavljenim v razdelku 2.2.

2.3.2 Metode, neodvisne od napovednih modelov

Ocenjevanje zanesljivosti klasifikacij. Težavo splošnosti je odpravila transduktivna metoda za ocenjevanje zanesljivosti klasifikacij Kukarja in Kononenka [2001; 2002]. V splošnem lahko na osnovi deleža klasifikacij primerov v posamezne razrede izrazimo posteriorno verjetnost, s katero bo primer pripadal posameznemu razredu. Verjetnosti pripadnosti posameznim razredom oblikujejo diskretno verjetnostno porazdelitev, ki je funkcija učnega algoritma in učne množice podatkov. Ocenjevanje zanesljivosti klasifikacij temelji na opazovanju sprememb v verjetnostni porazdelitvi primerov med *začetnim* (induktivnim) in t.i. *transduktivnim* klasifikatorjem.

Začetni klasifikator je tisti, ki je induciran na celotni množici učnih primerov in ga uporabljamo za klasificiranje še neoznačenih primerov. Posamezen primer, ki smo ga označili s klasifikacijo začetnega klasifikatorja ali pa tudi katero drugo, lahko tedaj ponovno vključimo v učno množico in na njej ponovno zgradimo klasifikator. Ker smo s tem v učno množico vključili transducirano znanje, imenujemo ta klasifikator *transduktivni*. Transduktivni korak dodajanja primerov v učno množico ponovimo za vsak primer iz začetne učne množice. Na osnovi klasifikacij transduktivnega klasifikatorja lahko tako nato ponovno izrazimo verjetnostno porazdelitev po razredih, ki se zaradi drugačne učne množice razlikuje od prvotne.

Za kvantitativno izražanje spremembe verjetnostne porazdelitve primerov med začetnim in transduktivnim klasifikatorjem so v okviru metode definirane in testirane številne metrike, ki argumente – diskretne verjetnostne porazdelitve preslikujejo v realni interval $[0, 1]$ (variacijska razdalja, Bhattacharyyeva razdalja, harmonična srednja vrednost, normirana evklidska razdalja, kosinus kota med porazdelitvama itd.). Te metrike in njihove produktne kombinacije so uporabljene kot ocene zanesljivosti, ki jih, po izbiri pragovne vrednosti, uporabljajo za ločevanje množic pravih in napačnih klasifikacij. Definirane metrike so neodvisne od učnega algoritma, rezultati pa so pokazali, da uspešno ločijo pravilno in nepravilno klasificirane primere.

Ocenjevanje zanesljivosti regresijskih napovedi. V okviru nadaljnjega razvoja opisane metode sta Bosnić in Kononenko [2003; 2001] metodo za ocenjevanje zanesljivosti klasifikacij prilagodila regresijskemu napovedovanju. Uvedla sta pojem *trans-*

duktivnih napovedi kot napovedi regresijskih prediktorjev, zgrajenih na različno spremenjenih učnih množicah, ki sta jih z linearnim modelom in modelom lokalno utežene regresije uporabila za modeliranje ocene napake posameznih napovedi. Uspešni rezultati nadaljnjega dela [Bosnić, 2003; Bosnić in Kononenko, 2005] so še pokazali, da je z modeliranjem ocene napake z uporabo izbranih transduktivnih napovedi [Robnik-Šikonja in Kononenko, 1997; Robnik-Šikonja, 2001] možno dosegati statistično značilne popravke prvotnih napovedi regresorja in s tem tudi večjo točnost napovedi. Kljub uspešnim rezultatom pa je opisani pristop imel to pomanjkljivost, da je ocene zanesljivosti modeliral z ločenim regresijskim modelom, kar je predstavljalo največjo oviro pri časovni zahtevnosti sistema.

Modeliranje napake z lokalnim prečnim preverjanjem. Številne raziskave [Birattari in sod., 1998; Schaal in Atkeson, 1998, 1994; Giacinto in Roli, 2001; Woods in sod., 1997] prilagajajo metodo prečnega preverjanja lokalnim delom problemskega prostora in jo uporabljajo za ocenjevanje napake in s tem zanesljivosti napovedi posameznih primerov. V tem postopku za dani testni primer, ki mu želimo napovedati oceno zanesljivosti, generiramo množico modelov z uporabo k najbližjih sosedov. S postopkom izloči-enega (leave-one-out) generiramo za vsakega izmed sosedov svoj model, s katerim nato za soseda i izračunamo vrednost lokalne napovedi K_i . Ker so za primere v učni množici znane prave vrednosti označb C_i , nato najbližjim sosedom izračunamo vrednost absolutnih napak na lokalnih napovedih $|C_i - K_i|$. Oceno zanesljivosti napovedi primera tedaj izrazimo kot povprečje ali kot uteženo vsoto napak njegovih najbližjih sosedov.

2.4 Učenje z vzorčenjem primerov iz učne množice

Področje, ki je posredno povezano z ocenjevanjem učenja, je tudi področje metod, ki z iterativnim vzorčenjem učne množice generirajo serijo modelov z namenom izbire najboljšega modela ali kombiniranja napovedi posameznih modelov in s tem izračuna bolj stabilne skupne napovedi. Najbolj znani pristopi na tem področju sta *bagging* [Breiman, 1996] in *boosting* [Schapire, 1999; Drucker, 1997; Ridgeway in sod., 1999].

Znan je tudi pristop *pasting* [Breiman, 1997], ki poleg izboljšanja napovedne točnosti rešuje tudi težave z napovedovanjem v učnih množicah, ki so prevelike za hranjenje v delovnem spominu.

Tibshirani in Knight [1999] sta vpeljala *kriterij napihnjene variance* (angl. *covariance inflation criterion, CIC*), ki ga uporabljata za ocenjevanje modelov, generiranih na podvzorcih prvotne učne množice. V vsaki iteraciji algoritma se z merjenjem kovariance med vhom in izhodom prediktorja izvede izbira modela, ki naj bi po kriteriju CIC imel najmanjšo učno napako. Nadaljnje študije [Rosipal in sod., 2000] so pokazale, da je CIC primerna mera tudi za medsebojne primerjave modelov različnih domen in za te namene zadošča tudi brez uporabe časovno bolj zahtevnega prečnega preverjanja.

Elidan in sod. [2002] so na osnovi perturbiranja učne množice reševali problem izogibanja lokalnih optimumov pri učenju. Z analizo napovedi na perturbacijah učne množice predlagajo način uteževanja učnih primerov, s katerim se uspešno izognejo problemu lokalnih maksimumov.

Iz omenjenih pristopov vidimo, da z vzorčenjem in uteževanjem učnih primerov rešujejo problem minimizacije napake na celotnem učnem modelu in ne na posameznih napovedih. Vendar pa nas že sama ideja, da je z dodajanjem/odvzemanjem primerov iz učne množice možno priti do informacije o optimalni izbiri modela, motivira, da poskušamo raziskati, ali lahko s podobnim pristopom pridobimo tudi informacijo o optimalni izbiri posamezne napovedi oziroma o njeni zanesljivosti.

2.5 Uporaba neoznačenih primerov pri nadzorovanem učenju

Rezultati raziskav o uporabi neoznačenih učnih primerov pri nadzorovanem učenju kažejo, da lahko z vključitvijo teh primerov v postopek učenja značilno izboljšamo točnost napovednega modela [Seeger, 2000]. Čeprav točne označbe teh primerov niso znane, prispevajo dodatni učni primeri natančnejšo informacijo o porazdelitvi primerov v problemskem prostoru, kar omogoča natančnejše učenje.

Pri postopku dodajanja primerov v učno množico je treba poskrbeti za označitev

primerov. Najbolj preprost pristop reševanja tega problema je algoritem EM (Expectation - Maximization) [Dempster in sod., 1977], ki ga lahko povzamemo na naslednji način:

1. Konstruiraj model z uporabo le označenih učnih primerov.
2. Uporabi zgrajen model za izračun verjetnostne porazdelitve označitev neoznačenih primerov.
3. Neoznačene primere označi in jih v skladu z verjetnostno porazdelitvijo (število dodanih primerov) dodaj v učno množico.
4. Konstruiraj model na razširjeni učni množici.
5. Pojdi na korak 2 in ponavljaj postopek, dokler se model še značilno spreminja.

Raziskave na področju klasifikacije slik [Baluja, 1998] kažejo, da lahko s takšnim pristopom dosežemo nekaj odstotkov večjo klasifikacijsko točnost modela v primerjavi z modelom, ki uporablja samo označene učne primere.

Nadaljnje raziskave se bolj intenzivno osredotočajo na uporabo neoznačenih primerov v kontekstu *komplementarnega učenja* (angl. *co-learning*). Blum in Mitchell [Blum in Mitchell, 1998; Mitchell, 1999] pokažeta, da lahko z neoznačenimi primeri dopolnimo znanje označenih primerov v primeru, ko lahko zagotovimo dva različna pogleda na problemski prostor. To pomeni, da imamo za vsak učni primer podani dve množici neodvisnih in s tem redundantnih atributov (npr. problem klasifikacije spletnih strani lahko rešujemo na podlagi dveh zornih kotov: z analizo povezav, ki vodijo na neko spletno stran, ali pa na podlagi povezav, ki vodijo od obravnavane strani k drugim spletnim stranem). Avtorja predpostavita, da je vsak od teh pogledov na problemski prostor zadosten za reševanje celotnega napovednega problema ob predpostavki, da imamo zadostno število označenih učnih primerov. Na osnovi delitve atributov na dve podmnožici dobimo dve učni množici primerov, vsako s svojo neodvisno množico atributov. Z uporabo klasifikatorjev, zgrajenih na vsaki od teh množic, lahko tedaj napovemo vrednosti razredov neoznačenih primerov in primere nato dodamo v prvotno učno množico. Rezultati učenja na razširjeni učni množici so pokazali obetavno zmanjšanje klasifikacijske napake.

Poleg reševanja problema klasificiranja spletnih strani uporabljajo isti pristop tudi raziskave na drugih področjih. De Sa [1993] obravnava problem učenja iz popolnoma neoznačenih podatkov v domeni klasificiranja računalniških posnetkov. Težava, ki jo v tej domeni srečujemo, je ta, da je izračun razredov/označb primerov v učni množici lahko časovno drago in si ga ne moremo privoščiti. Kot rešitev avtorica predlaga obliko *samo-nadzorovanega učenja* (angl. *self-supervised learning*), ki problem, ravno tako kakor [Blum in Mitchell, 1998], razdeli na dva neodvisna problema (na področju obdelave posnetkov sta to analiza slikovnega signala in analiza zvočnega signala). Na vsaki izmed dveh podmnožic podatkov ločenega problema lahko tedaj zgradimo svoj klasifikator–nevronske mreže, ki primerom napoveduje njihove kodirne vektorje, določene na podlagi lokalne segmentacije problemskega prostora. Na začetku postopka učenja so kodirni vektorji naključno inicializirani v problemskem prostoru, vsak neoznačen primer pa je klasificiran v implicitni razred, ki ga določa najbližji kodirni vektor. Namesto izvajanja klasičnega nadzorovanega učenja vsakega klasifikatorja, je s postopkom samo-nadzorovanega učenja cilj doseči usklajenost napovedi ločenih klasifikatorjev za vsakega izmed podproblemov. Za reševanje tega problema avtorica predlaga združitev obeh nevronske mreže s skupnim izhodnim nivojem nevronov in učenje problema enakosti izhodov obeh mrež. Z minimizacijo razlik med izhodnimi kodirnimi vektorji obeh nevronske mreže ti vektorji skonvergirajo v točke, ki predstavljajo nova središča implicitnih razredov. Rezultati so pokazali, da daje predlagana metoda slabše rezultate kot pristopi nadzorovanega učenja, vendar pa uspešno rešuje problem napovedovanja z uporabo popolnoma neoznačenih podatkov v primerih, kadar lahko problem razčlenimo na dva neodvisna podproblema.

Omejitev delitve množice atributov v dve podmnožici odpravlja metoda Goldmanove [2000]. Predlagana metoda je uporabna v primerih, ko imamo opravka z dvema različnima prediktorjema, ki delita problemski prostor v ekvivalenčne razrede (ekvivalenčni razred definira kot npr. en list regresijskega drevesa). Po gradnji obeh neodvisnih modelov si modela v ekvivalenčne razrede navzkrižno označita in dodata prej neoznačene primere, ki potrjujejo zaupanje v lokalni model v ekvivalenčnem razredu (ne odstopajo od njega več, kot je vnaprej določena stopnja zaupanja). Postopek se ponovi v več iteracijah in s širitvijo učne množice značilno

zmanjša napovedno napako vsakega izmed prediktorjev.

Uspešnost raziskav uporabe neoznačenih primerov pri nadzorovanem učenju nam kaže, da so dodatni primeri, ki so generirani iz iste verjetnostne porazdelitve kot učni primeri, za postopek učenja koristni. Ta ideja pa nas motivira, da poskusimo ob razširjanju učne množice opazovati vpliv tega postopka na napovedi posameznih napovedi in na podlagi njihovih sprememb poskusiti oceniti njihovo zanesljivost.

2.6 Terminologija področja in njeno poenotenje

V nadaljevanju podajamo poenoten povzetek terminologije s področja disertacije in podajamo predlog uvedbe novih izrazov.

zanesljivost (angl. reliability)	Splošen pojem s področja vrednotenja sistemov, ki ponazarja lastnost entitete, da uspešno opravlja svoje funkcije [Wikipedia, 2007b]. V kontekstu strojnega učenja lahko kot zanesljivost sistema opredelimo njegove lastnosti, za katere menimo, da kritično vplivajo na uspešno izvajanje aspekta delovanja opazovane entitete. To so lahko: <i>točnost</i> , <i>netočnost</i> , <i>razpoložljivost</i> , <i>stopnja izpadov</i> , <i>odzivnost</i> in druge. V okviru našega dela se osredotočamo na pojem točnosti regresijskih modelov in njihovih napovedi, zato tudi pojem zanesljivosti uporabljamo v kontekstu točnosti in netočnosti regresijskih napovedi.
ocena zanesljivosti (angl. reliability estimate)	Ker se v praksi soočamo s problemom, da mere zanesljivosti niso kvantitativno definirane, zanesljivost ocenjujemo z <i>ocenami zanesljivosti</i> . Glede na definicijo pojma zanesljivosti v izbranem kontekstu z ocenjevanjem zanesljivosti torej ocenjujemo točnost ali napako regresijskih napovedi.
stabilnost (angl. stability)	Lastnost entitete, da se ji kritičen aspekt delovanja ne spremeni več kot za določen prag ob spremembi izbranih lastnosti entitete [Bousquet in Elisseeff, 2002].
občutljivost (angl. sensitivity)	Lastnost, sorodna <i>stabilnosti</i> . Kvantitativno izražena odvisnost spremembe delovanja aspekta entitete ob spremembi izbranih parametrov entitete [Breierova in Choudhari, 1996].

transdukcija	Širok pojem s splošnim pomenom <i>sklepanja s posameznega na posamezno</i> . V kontekstu ocenjevanja zanesljivosti je pri mnogih metodah [Gammerman in sod., 1998; Saunders in sod., 1999; Li in Wechsler, 2005] osnova za zasnovo ocen zanesljivosti, ki temeljijo na verjetnostnem merjenju kakovosti označitve novega testnega primera glede na porazdelitev preostalih učnih primerov. Pojem transdukcije se uporablja tudi v primerih, ko induciramo znanje z drugim namenom, kot je napovedovanje, npr. za ocenjevanje zanesljivosti napovedi [Kukar in Kononenko, 2002; Bosnić in sod., 2003].
zaupanje (angl. confidence)	Zaupanje je verjetnostno izražena ocena točnosti napovedi, vrednost zaupanja v napoved torej predstavlja verjetnost njene točnosti. Zaupanje napovedi ponavadi temelji na podlagi verjetnostne porazdelitve, definira se pa lahko tudi kot $1 - p_2$, kjer je p_2 verjetnost druge najbolj verjetne napovedi [Saunders in sod., 1999].
verodostojnost (angl. credibility)	Verjetnostna mera za ocenjevanje točnosti napovedi, sorodna zaupanju. Točnosti ne ocenjuje na podlagi verjetnostne porazdelitve, temveč z lastnim verjetnostnim modelom (npr. opazovanjem deleža podpornih vektorjev ali Lagrangeevih koeficientov [Saunders in sod., 1999]).
interval zaupanja (angl. confidence interval)	V statistiki interval števil, v katerem lahko parameter zavzame vrednost z določeno stopnjo zaupanja [Wikipedia, 2007a]. Tudi stopnja ujemanja med napovedano vrednostjo in pravo vrednostjo učnega primera (ki za razliko od njegove označbe ne vsebuje šuma) [Heskes, 1997].
interval napovedi (angl. prediction interval)	Stopnja ujemanja napovedi primera z njegovo vrednostjo označbe (ki se od prave vrednosti lahko razlikuje za šum).

V skladu z opisanimi pojmi predlagamo uporabo naslednjih terminov, ki so vezani na metodologijo, predstavljeno v disertaciji:

ocena točnosti (angl. accuracy estimate)	Aspekt ocenjevanja zanesljivosti napovedi; ocena, ki pozitivno korelira s točnostjo napovedi oziroma negativno korelira z velikostjo napake posamezne napovedi; sorodna zaupanju, vendar bolj splošna – nima verjetnostne interpretacije. Medtem, ko ima zaupanje zalogo vrednosti $[0, 1]$, ima lahko ocena točnosti poljubno zalogo vrednosti z intervala realnih števil.
---	--

2. PREGLED PODROČJA

ocena napake (angl. error estimate)	Tudi možen aspekt ocenjevanja zanesljivosti napovedi; ocena, obratna oceni točnosti napovedi, pozitivno korelira z velikostjo napake posamezne napovedi; nima verjetnostne interpretacije.
verjetnostna ocena napake (angl. probabilistic error estimate)	Ocena napake z verjetnostno interpretacijo, 1 – <i>zaupanje</i> .

Poglavje 3

Razvoj ocen zanesljivosti

3.1 Razvoj ocen z uporabo analize občutljivosti

Iz pregleda področja v poglavju 2 smo videli, da nam:

- tehnika analize občutljivosti omogoča obravnavati regresijski model kot črno škatlo oziroma parameter algoritma in nam zato omogoča razvoj metode, ki je neodvisen od uporabljenega modela,
- pristopi s permutiranjem učne množice ponujajo idejo, da je z različnim vzorčenjem učnih primerov s ponovitvami možno pridobiti regresijske modele, ki imajo različno napovedno točnost,

- področje uporabe neoznačenih primerov ponuja idejo, da lahko v učno množico vključimo nove učne primere tako, da jih označimo z uporabo regresijskega modela, zgrajenega na označenih podatkih.

V tem poglavju smo te ideje strnili skupaj in jih uporabili za razvoj ocen zanesljivosti z uporabo analize občutljivosti. Ideja naše metode sloni na lastnosti učnih algoritmov, da je njihova naloga doseči čim bolj natančno modeliranje učne množice podatkov. Na osnovi tega lahko pričakujemo, da se naučeno znanje ne bo bistveno spremenilo, če v učnih podatkih povzročimo zelo majhne spremembe. Če bi v nasprotju s tem razmišljanjem zaradi majhnih sprememb v učnih podatkih prišlo do velikih sprememb v izhodnih napovedih, bi to lahko razlagali, da je generiran regresijski model nestabilen. Ideja naše metode je torej v tem, da se osredotočamo na merjenje sprememb v izhodnih napovedih ob povzročanju sprememb v učni množici. Ob tem pa uporabljamo velikost spremembe kot mero nestabilnosti modela v točki – učnem primeru, ki ga preizkušamo.

Z opisanim pristopom se problema ocenjevanja zanesljivosti napovedi lotevamo z ocenjevanjem napake napovedi (glej terminologijo v razdelku 2.6). Razlika med tem pristopom in pristopom ocenjevanja točnosti je, da pri prvem najmanjša vrednost pomeni največjo zanesljivost (torej najmanjšo napako), pri drugem pa največjo napako (torej najmanjšo zanesljivost). Zato lahko predpostavimo, da sta si problema dualna in da obstaja preprosta preslikava z intervala posplošenih ocen napake na interval ocen zanesljivosti.

V nadaljevanju definiramo parametre, ki so potrebni za obravnavo regresijskih modelov z opisanim načinom analize občutljivosti.

3.1.1 Vhodne vrednosti analize občutljivosti

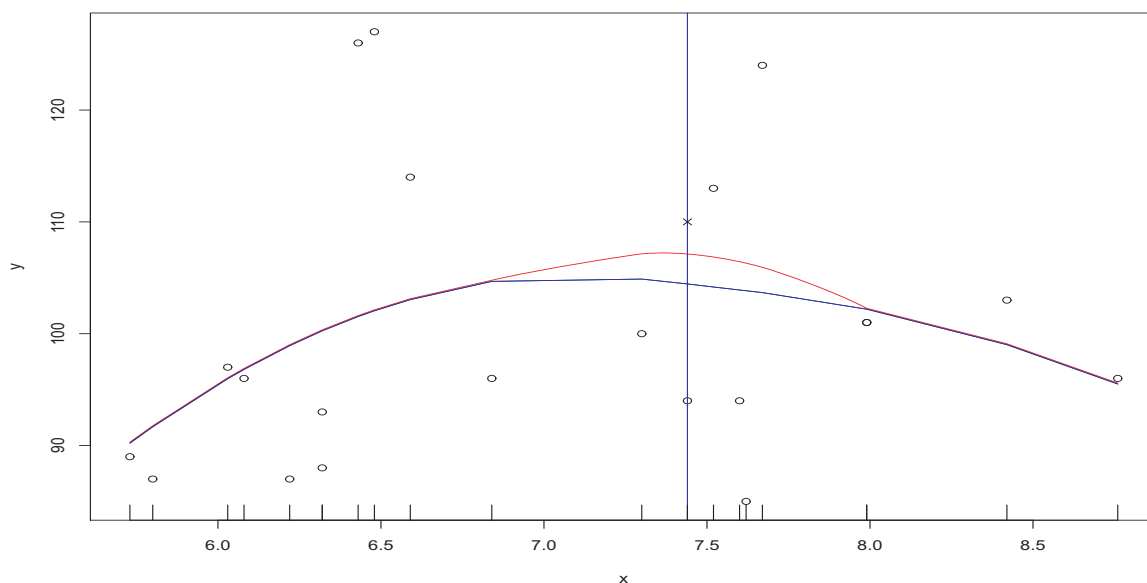
Za izbiro načina spreminjanja učne množice imamo veliko možnosti. Kot najbolj preprosta možnost se nam ponuja ideja razširiti učno množico ravno s primerom, za katerega napovedujemo vrednost odvisne spremenljivke in za katerega želimo oceniti zanesljivost te napovedi. Izbiro takšnega načina razširitve učne množice lahko argumentiramo tudi z naslednjimi ugotovitvami:

1. Raziskave s področja uporabe neoznačenih primerov kažejo, da obstaja odvisnost med točnostjo modela in dopolnjevanjem učne množice s primeri, ki izvirajo iz iste porazdelitve kakor preostali primeri v učni množici.
2. Teoretični rezultati formalizacije opisanega problema z uporabo principa MDL [Li in Vitányi, 1993] kažejo, da lahko najbolj informativno spremembo modela dosežemo, če dodamo tak učni primer, ki ga znanje prvotnega (nespremenjenega) modela slabo pokriva [Bosnić in Kononenko, 2007a]. V skladu s 1. točko to pomeni, da je v učno množico smiselno dodati učni primer s tako označbo, ki ni v skladu z naučenim znanjem regresijskega modela, torej ima drugačno označbo od prvotne napovedi. Na podobnih ugotovitvah temelji tudi področje aktivnega učenja [Cohn in sod., 1995], ki izbira primere za razširitev učne množice na podlagi različnih kriterijev, kot so področja problemskega prostora, v katerih še nimamo podatkov [Whitehead, 1991], področja prostora, v katerih je uspešnost napovedovanja slaba [Linden in Weber, 1992], področja prostora z veliko šuma v podatkih [Thrun in Möller, 1992] ali področja prostora, kjer bo dodajanje novega primera povzročilo največjo spremembo na napovednem modelu.
3. S spremembo učne množice želimo povzročiti spremembo v napovedi primera, za katerega napovedujemo zanesljivost, saj bomo zanesljivost izrazili kot funkcijo spremembe napovedi. Da to dosežemo, moramo v problemskem prostoru povzročiti lokalno spremembo v okolici točke, za katero računamo napoved. Ker ne vemo, kako daleč se bo vpliv nove točke na novem modelu propagiral, je najbolj smiselno, da nov primer dodamo v točko opazovanja in ne drugam (glej primer na sliki 3.1).

Z x označimo atributni vektor primera in z y njegovo označbo. Učni primer z znano označbo lahko torej predstavimo s parom (x, y) , z $(x, _)$ pa predstavimo neznani, neoznačeni primer.

Naj bo v nadaljevanju $(x, _)$ tisti neznani primer, za katerega želimo oceniti zanesljivost napovedi. Napoved primera K , katere zanesljivost želimo oceniti, izračunamo z uporabo modela M , torej $f_M(x) = K$. Ker je bila napoved K izračunana na

3. RAZVOJ OCEN ZANESLJIVOSTI



Slika 3.1: Povzročitev lokalne spremembe v regresijskem modelu z dodajanjem novega primera v učno množico (označen s križcem). Modra krivulja ponazarja začetni regresijski model, oranžna pa spremenjen model ob dodajanju novega učnega primera. Slika prikazuje primer dvorazsežnega regresijskega problema z enim atributom (x) in označbo y .

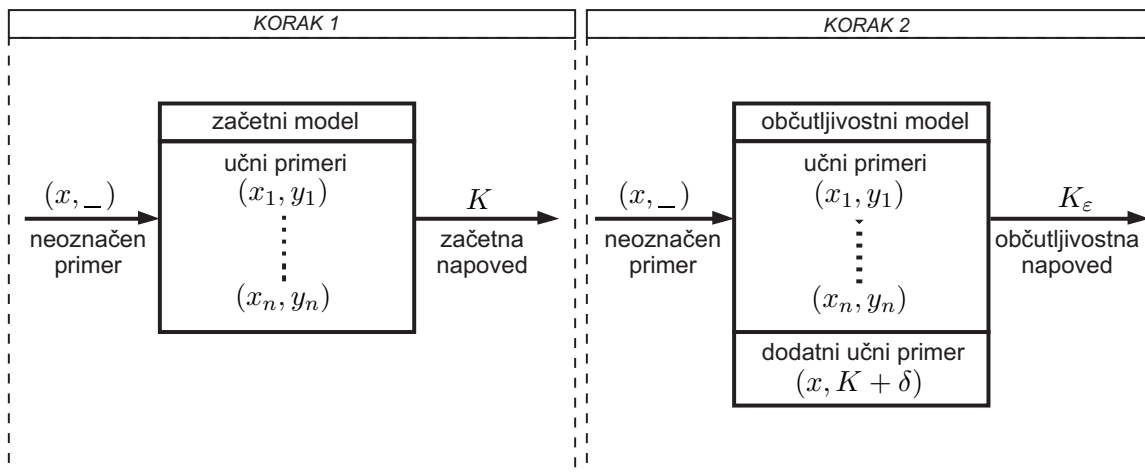
začetku postopka ocenjevanja zanesljivosti z modelom, konstruiranem na podlagi originalne (nespremenjene) učne množice, poimenujmo to napoved *začetna napoved*. V naslednjem koraku se odločimo, da bomo originalno učno množico razširili z obravnavanim neznanim primerom. Pred dodajanjem novega primera pa mu je treba določiti označbo. Odločimo se, da bomo primeru priredili označbo po vzorcu:

$$y = K + \delta \tag{3.1}$$

kjer predstavlja δ majhno aditivno spremembo (pozitivno ali negativno) glede na prvotno izračunano napoved K , ki jo torej uporabljamo kot centralno vrednost označbe y . V nadaljevanju izberemo tak δ , ki je sorazmeren z znanimi mejami označb, ki jih pridobimo iz primerov iz učne množice. Torej, če se označbe učnih primerov nahajajo na intervalu $[a, b]$ in če z ε predstavimo relativni delež tega intervala, izrazimo $\delta = \varepsilon(b - a)$.

Po izbiri vrednosti ε (in s tem posredno inkrementa δ) lahko razširimo učno množico z dodatnim učnim primerom (x, y) . Na podlagi razširjene učne množice

z $n + 1$ primeri nato zgradimo nov prediktor, ki ga imenujemo *občutljivostni model* M' . Občutljivostni model M' v nadaljevanju ponovno uporabimo za izračun napovedi primera $(x, _)$, s čimer zaradi spremembe učne množice dobimo spremenjeno napoved $f_{M'}(x) = K_\varepsilon$. Napoved K_ε poimenujemo z nazivom *občutljivostna napoved*. Za lažjo predstavo je opisan postopek ilustriran na sliki 3.2.



Slika 3.2: Shematski prikaz postopka analize občutljivosti. Slika prikazuje izračun začetne napovedi (korak 1) in izračun občutljivostne napovedi K_ε z uporabo občutljivostnega modela (korak 2).

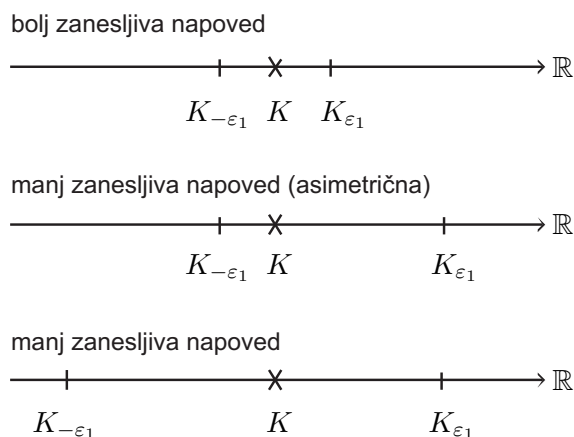
V postopku prirejanja označbe y neoznačenemu primeru lahko z izbiro različnih vrednosti $\varepsilon_k \in \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m\}$ pridobimo celo množico občutljivostnih napovedi

$$K_{\varepsilon_1}, K_{-\varepsilon_1}, K_{\varepsilon_2}, K_{-\varepsilon_2}, \dots, K_{\varepsilon_m}, K_{-\varepsilon_m} \quad (3.2)$$

Pri zasnovi ocen bomo tudi dejansko uporabili celo množico občutljivostnih napovedi za izračun razlik $K_\varepsilon - K$ med napovedmi občutljivostnega in začetnega prediktorja. Po uvedbi zgornje terminologije lahko (zaenkrat) na kvalitativen način ilustriramo predpostavke o medsebojni porazdelitvi začetne in občutljivostnih napovedi na realni osi, na podlagi katerih bomo sklepali, ali je bila napoved K *manj* ali *bolj zanesljiva* (slika 3.3). V razdelku 3.1.3 bomo informacijo o prikazani porazdelitvi opisali na kvantitativen način in jo uporabili kot oceno zanesljivosti.

Pred zasnovo ocen zanesljivosti, ki bodo temeljile na navedenih predpostavkah, si pogledjmo še predpostavke v zvezi z obnašanjem občutljivostnega modela (izhod-

3. RAZVOJ OCEN ZANESLJIVOSTI



Slika 3.3: Trije primeri predpostavke o točnosti napovedi na podlagi porazdelitve začetne in občutljivostnih napovedi na realni osi. Slika ponazarja kvalitativne primerjave zanesljivosti, pri čemer predpostavljamo, da večji intervali med napovedmi kažejo na večjo nezanesljivost prvotne napovedi.

nih vrednosti analize občutljivosti) in lastnosti obnašanja regresijskih modelov, ki so predmet obravnave.

3.1.2 Izhodne vrednosti analize občutljivosti

Na podlagi opisanega postopka pridobivanja občutljivostne napovedi lahko intuitivno predpostavimo, da velja naslednja relacija:

$$\varepsilon_1 < \varepsilon_2 \quad \Rightarrow \quad K_{\varepsilon_1} < K_{\varepsilon_2} \quad (3.3)$$

in posledično torej tudi

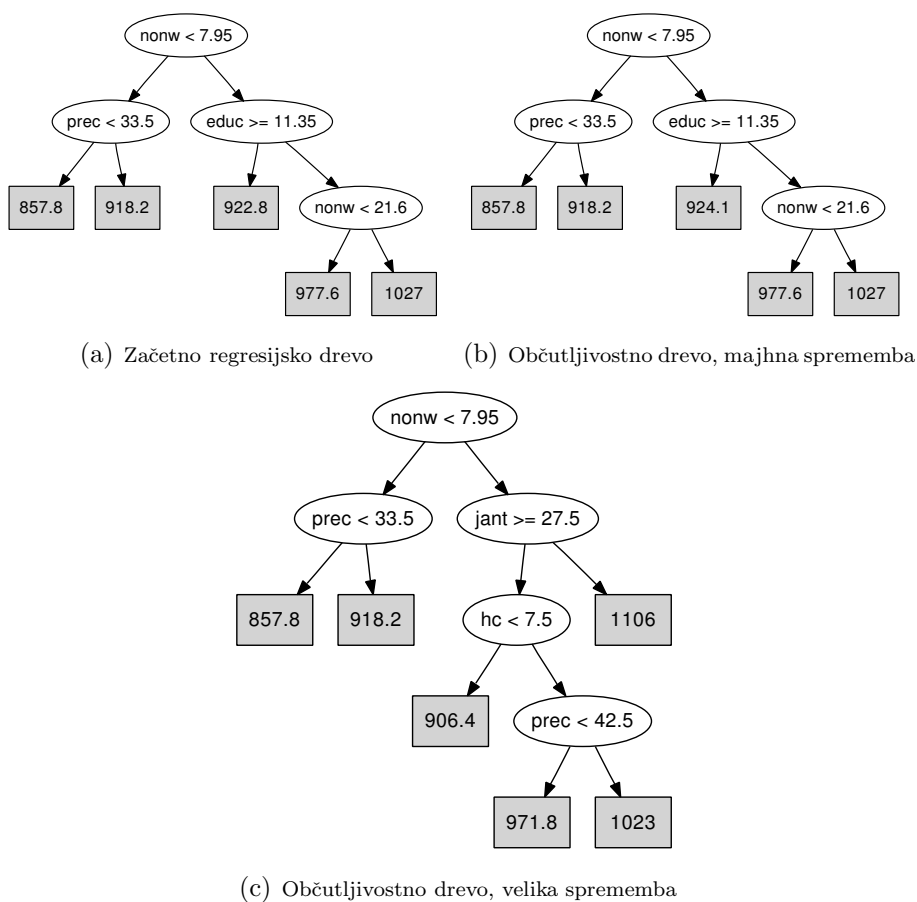
$$K_{-\varepsilon_m} < \dots < K_{-\varepsilon_1} < K < K_{\varepsilon_1} < K_{\varepsilon_2} < \dots < K_{\varepsilon_m}$$

za $\varepsilon_k \in \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m\}$, $\varepsilon_1 < \varepsilon_2 < \dots < \varepsilon_m$, $k = 1, \dots, m$. Čeprav ta predpostavka tudi empirično velja za večino primerov, pa so možne tudi izjeme, ki so posledica specifičnega obnašanja posameznih regresijskih modelov.

Enostavni in kompleksni regresijski modeli. Metodologijo, opisano v disertaciji, smo testirali z uporabo osmih regresijskih modelov (glej poglavje 4): regresijskimi drevesi, lokalno uteženo regresijo, linearno regresijo, nevronskimi mrežami, metodo podpornih vektorjev, posplošenim aditivnim modelom, metodo bagging na regresijskih drevesih in naključnimi gozdovi. Množico teh učnih algoritmov lahko razdelimo na *enostavne* in *kompleksne* po kriteriju, ali pred modeliranjem podatkov delijo problemski prostor na manjše enote ali ne. Po tem kriteriju lahko kot enostavne modele torej obravnavamo linearno regresijo, posplošeni aditivni model in lokalno uteženo regresijo, saj naenkrat modelirajo podatke z enim samim modelom. V nasprotju s tem pa ostalih pet modelov izvajajo bodisi posredno ali neposredno delitev problemskega prostora: regresijska drevesa glede na mejne vrednosti atributov v notranjih vozliščih, metoda podpornih vektorjev glede na primere znotraj/zunaj roba, nevronska mreža igra vlogo večplastne diskriminantne funkcije, bagging in naključni gozdovi pa so sestavljeni iz posameznih modelov, ki so vsak zase specifični.

Za obravnavo z uporabo analize občutljivosti so bolj zanimivi *kompleksni modeli*. Namesto le majhne spremembe v izhodni napovedi lahko pri njih majhna sprememba v učni množici povzroči drugačno delitev problemskega prostora, česar posledica je lahko občutna sprememba hipoteze. Posledica radikalne spremembe hipoteze je lahko večja razlika med začetno in občutljivostno napovedjo, kar bo pa v skladu z našimi pričakovanji pomenilo, da je bila že prvotna hipoteza nestabilna in s tem nezanesljiva. Takšno obnašanje modela regresijskih dreves je prikazano na sliki 3.4 za podatke iz domene *pollution* [Asuncion in Newman, 2007]. Slika 3.4(a) prikazuje začetno regresijsko drevo, zgrajeno na prvotni učni množici. Ob gradnji modela na spremenjeni učni množici lahko v nadaljevanju nastopita dva možna scenarija. Prvi je prikazan na sliki 3.4(b), na kateri vidimo, da se je zaradi novega učnega primera napoved v tretjem listu drevesa spremenila z vrednosti 922.8 na vrednost 924.1. Drugi možen scenarij pa je ta, da zaradi novega učnega primera pride do večje spremembe modela, kot je to prikazano na sliki 3.4(c). Vidimo lahko, da so se v desnem poddrevesu značilno spremenili kriteriji v notranjih vozliščih skupaj z vrednosti napovedi v listih. Naj v zvezi z ilustracijo primera opomnimo, da se je vrednost napovedi, čeprav smo uporabili pozitivno vrednost δ (glej izraz (3.1)), znižala z vrednosti 922.8 na 906.4.

3. RAZVOJ OCEN ZANESLJIVOSTI



Slika 3.4: Vpliv spreminjanja učne množice na gradnjo regresijskega drevesa.

Na ta način je to obenem ilustracija primera, ko pravilo (3.3) ne drži v vseh primerih, zlasti ne v tistih, ko pride do strukturne spremembe modela. Tega pojava ne obravnavamo kot oviro pri zasnovi naše metode, ampak zgolj kot še dodatni dobrodošel pokazatelj nestabilnosti hipoteze, kar je ravno tisto, kar želimo oceniti.

Formulacija občutljivostne napovedi pri enostavnih modelih. Enostavni in kompleksni modeli se po našem pojmovanju razlikujejo tudi po tem, da lahko pri enostavnih modelih občutljivostno napoved izrazimo kot funkcijsko odvisnost od začetne napovedi in novega učnega primera. Poglejmo si dva primera takšne formulacije občutljivostne napovedi, in sicer za lokalno uteženo regresijo in linearno regresijo.

Primer 1 Napoved primera x pri lokalno uteženi regresiji izračunamo v skladu s pravilom:

$$K = \frac{\sum_j^N \kappa(D(\mathbf{x}, \mathbf{x}_j)) \cdot C_j}{\sum_j^N \kappa(D(\mathbf{x}, \mathbf{x}_j))}$$

kjer D nastopa kot funkcija razdalje, κ kot jedrna funkcija, C_j prava vrednost primera j in \mathbf{x} kot atributni vektor primera. Občutljivostno napoved K_ε lahko tedaj izrazimo kot:

$$\begin{aligned} K_\varepsilon &= \frac{\sum_j^N [\kappa(D(\mathbf{x}, \mathbf{x}_j)) \cdot C_j] + [\kappa(D(\mathbf{x}, \mathbf{x})) \cdot (K + \delta)]}{\sum_j^N [\kappa(D(\mathbf{x}, \mathbf{x}_j))] + [\kappa(D(\mathbf{x}, \mathbf{x}))]} = \\ &= \frac{\sum_j^N [\kappa(D(\mathbf{x}, \mathbf{x}_j)) \cdot C_j] + [\kappa(0) \cdot (K + \delta)]}{\sum_j^N [\kappa(D(\mathbf{x}, \mathbf{x}_j))] + [\kappa(0)]} \end{aligned}$$

Primer 2 Omejimo se na primer obravnave linearne regresije v dvorazsežnem problemskem prostoru, kar pomeni, da modeliramo funkcijo $y = kx + n$, kjer k nastopa kot strmina regresijske premice in n kot presečišče z ordinatno osjo. V primeru linearne regresije tedaj velja:

$$\begin{aligned} K &= kx + n \\ k &= \frac{\sum_k (x_k - \bar{x})(y_k - \bar{y})}{\sum_k (x_k - \bar{x})^2} = \frac{n \sum_k x_k y_k - \sum_k x_k \sum_k y_k}{n \sum_k x_k^2 - (\sum_k x_k)^2} \\ n &= \bar{y} - k\bar{x} \end{aligned}$$

kjer n predstavlja število učnih primerov in \bar{x} povprečno vrednost spremenljivke x . Občutljivostno napoved K_ε lahko tedaj izrazimo kot:

$$\begin{aligned} K_{i,\varepsilon} &= k_\varepsilon x_i + n_\varepsilon \\ k_\varepsilon &= \frac{(n+1)[\sum_{k=1}^n x_k y_k + x(K+\delta)] - [\sum_{k=1}^n x_k + x][\sum_{k=1}^n y_k + K + \delta]}{(n+1)[\sum_{k=1}^n x_k^2 + x^2] - (\sum_{k=1}^n x_k + x)^2} \\ n_\varepsilon &= \frac{\bar{y}n + K + \delta}{n+1} - k_\varepsilon \frac{\bar{x}n + x}{n+1} \end{aligned}$$

V zgornji enačbi x_k , $k = 1, \dots, n$ predstavljajo učne primere, x pa novi primer.

Zaradi dodatne funkcije kompleksnih modelov, to je deljenje problemskega prostora, za njih občutljivostne napovedi ni možno izraziti na tako preprost način.

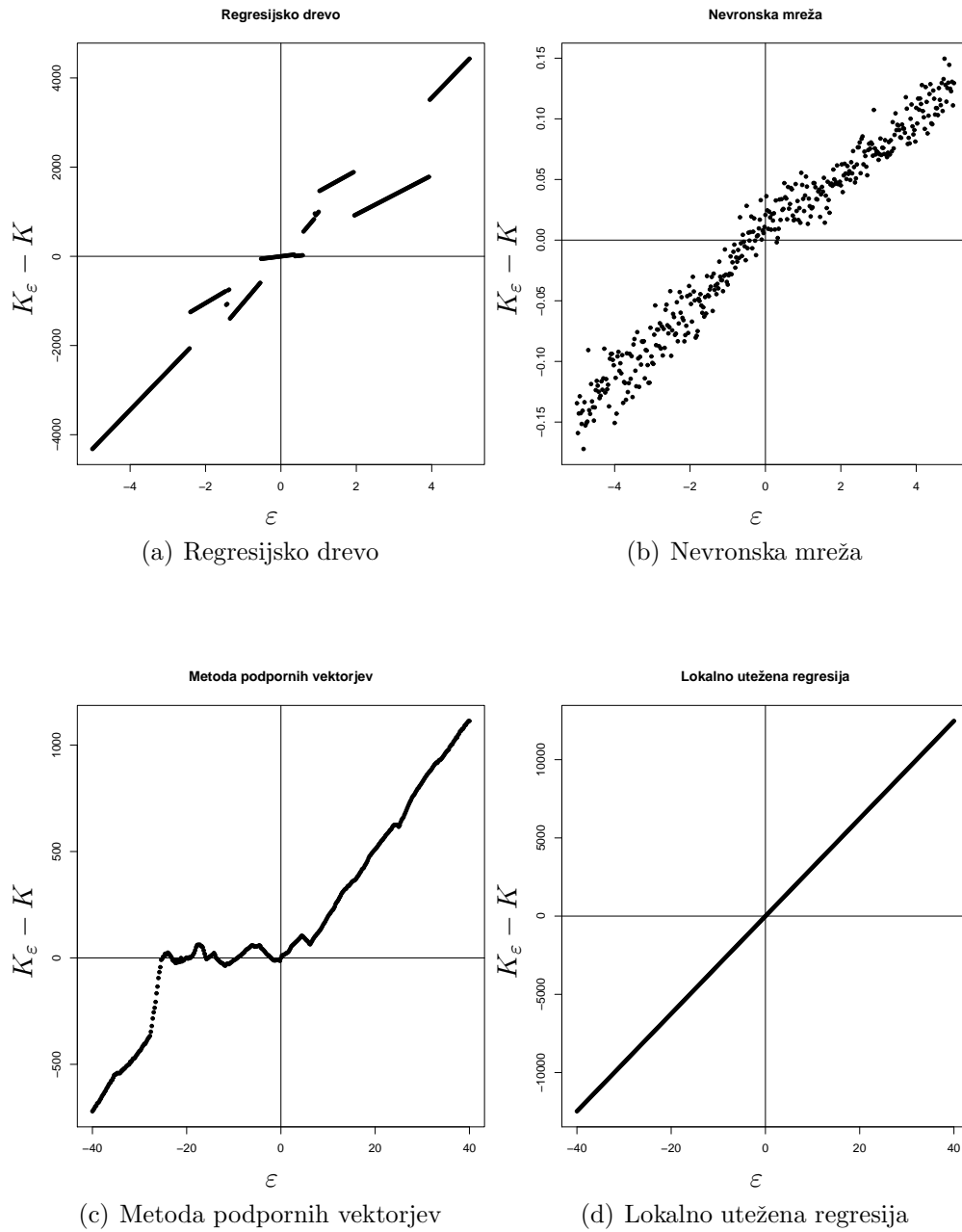
Izhodne karakteristike regresijskih modelov. Velikost razlike med začetno in občutljivostno napovedjo kot tudi obseg strukturne spremembe regresijskega modela je odvisen od vrednosti δ (ali ε) iz izraza (3.1). Primer te odvisnosti je grafično prikazan na sliki 3.5, kjer grafi za izbrani učni primer iz domene *strikes* prikazujejo odvisnost razlike $K_\varepsilon - K$ (ordinata) od vrednosti ε (abscisa). Iz teh prikazov, ki jih poimenujemo *karakteristični grafi regresijskih modelov*, lahko vidimo, da modeli lokalno utežene regresije (slika 3.5(d)), linearne regresije (slika 3.5(e)) in posplošenega aditivnega modela (slika 3.5(f)) ne prikazujejo lokalnih anomalij, ki bi jih lahko zajeli z oceno zanesljivosti.

Kritična lokalna področja, ki ponazarjajo lokalno občutljivost modelov, pa lahko zasledimo pri karakterističnih grafi kompleksnih modelov. Pri regresijskih drevesih (slika 3.5(a)) so te kritične regije definirane s posameznimi listi drevesa; pri metodi podpornih vektorjev (slika 3.5(c)) je kritično področje definirano z robom hiperravnine; pri nevronskih mrežah (slika 3.5(b)) so kritična območja posredno določena s pragovnimi elementi; pri metodi bagging (slika 3.5(g)) in naključnih gozdovih (slika 3.5(h)) pa posredno določena z listi vseh dreves v gozdu.

Na podlagi navedenih primerov in karakterističnih grafov na sliki 3.5 lahko zaključimo, da nam neodvisnost napovedi od delitve problemskega prostora pri enostavnih modelih ne omogoča zajeti *nestabilnega* obnašanja modela, kot smo ga definirali na začetku tega poglavja. Testiranja uporabe ocen zanesljivosti, definiranih z uporabo analize občutljivosti, na regresijskih modelih linearne regresije, posplošenega aditivnega modela in lokalno utežene regresije so ravno tako potrdila te zaključke.

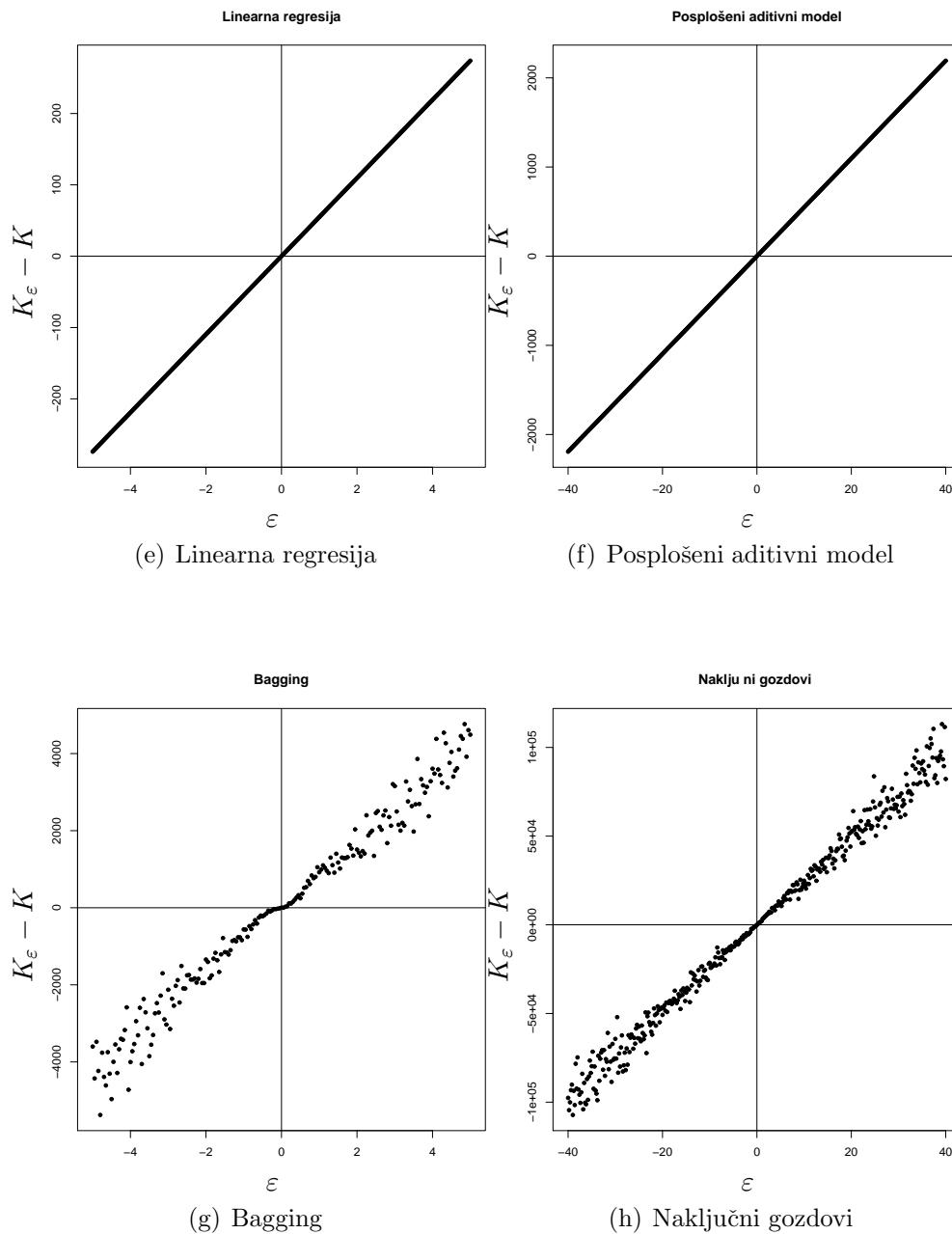
3.1.3 Ocene zanesljivosti

Za zasnovo ocen zanesljivosti, ki temeljijo na uporabi analize občutljivosti, uporabljamo razlike med začetno in občutljivostnimi napovedmi. Na podlagi teh razlik smo zasnovali dve oceni zanesljivosti, ki ju predstavljamo v nadaljevanju.



Slika 3.5: Občutljivost regresijske napovedi glede na spremembe v označbi enega učnega primera. Grafi prikazujejo odvisnost razlike $K_\epsilon - K$ (navpična os) od parametra ϵ (vodoravna os) za izbrani učni primer iz testne domene *strikes*.

3. RAZVOJ OCEN ZANESLJIVOSTI



Slika 3.5: (nadaljevanje) Občutljivost regresijske napovedi glede na spremembe v označbi enega učnega primera. Grafi prikazujejo odvisnost razlike $K_\epsilon - K$ (navpična os) od parametra ϵ (vodoravna os) za izbrani učni primer iz testne domene *strikes*.

Izračun vsake občutljivostne napovedi je odvisen od izbire konkretne vrednosti parametra ε (enačba (3.1)). Da bi se izognili izbiri enega samega ε in se s tem omejili na eno samo iteracijo postopka analize občutljivosti, smo ocene zanesljivosti definirali tako, da uporabljajo poljubno število občutljivostnih napovedi, ki so pridobljene z različnimi vrednostmi parametra ε . Na ta način razširimo okno, s katerim opazujemo nestabilnost modela pri različnih spremembah učne množice in zasnujemo ocene tako, da so manj občutljive na lokalne anomalije v problemskem prostoru.

Število uporabljenih parametrov ε torej predstavlja izbiro med nasprotujočima si kriterijema pridobivanja bolj stabilne ocene in časovno zahtevnostjo algoritma (za vsak ε je potreben izračun občutljivostne napovedi). Ker smo predpostavili, da enakost med začetno in občutljivostno napovedjo predstavlja največjo možno zanesljivost in ker se problema lotevamo s pristopa posplošenega ocenjevanja napake napovedi, definiramo tudi ocene zanesljivosti v skladu s tem, da vrednost 0 predstavlja oceno zanesljivosti najbolj zanesljive napovedi.

Predpostavimo, da imamo podano množico nenegativnih vrednosti parametrov ε , torej $E = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{|E|}\}$. Oceni¹ definiramo na naslednji način:

1. Ocena $SAvar$ (lokalna varianca):

$$SAvar = \frac{\sum_{\varepsilon \in E} (K_{\varepsilon} - K_{-\varepsilon})}{|E|} \quad (3.4)$$

Ocena je zasnovana tako, da povpreči vrednosti preko podanih parametrov ε . Temelji na ideji, da je pri zanesljivih napovedih sprememba v občutljivostnem modelu minimalna. Z upoštevanjem pozitivnega in negativnega ε (torej napovedi K_{ε} in $K_{-\varepsilon}$) želimo z oceno $SAvar$ zajeti širino celega intervala, prikazanega na sliki 3.3. Na ta način ta mera ustreza opisovanju lokalne variance občutljivostnih napovedi.

2. Ocena $SAbias$ (lokalna pristranost):

$$SAbias = \frac{\sum_{\varepsilon \in E} (K_{\varepsilon} - K) + (K_{-\varepsilon} - K)}{2|E|} \quad (3.5)$$

¹predpona SA v imenih ocen pomeni, da sta bili definirani z uporabo analize občutljivosti (Sensitivity Analysis)

V skladu s prikazom na sliki 3.3 lahko vidimo, da ocena $SAbias$ konstruira geometrijsko razliko med začetno napovedjo in napovedima K_ε in $K_{-\varepsilon}$ in s tem meri, ali je le-ta poudarjena v pozitivno ali negativno smer. Zaradi tega ima lahko tudi ocena $SAbias$ pozitivno ali negativno vrednost, ki nosi tudi informacijo o smeri, v katero je model bolj nestabilen. Ker ocena $SAbias$ meri asimetrijo med levim in desnim podintervalom, ustreza opisovanju lokalne pristranosti napovedi.

Oceni $SAvar$ in $SAbias$ smo v okviru testiranja primerjali z velikostmi napak (absolutnimi vrednostmi) testnih primerov. Ker pa nosi ocena $SAbias$ tudi informacijo o smeri občutljivosti modela, smo jo primerjali tudi s predznačeno vrednostjo napake testnih primerov.

Naj v tej točki omenimo, da sta obe oceni zanesljivosti podobni simetrični obliki formule za numerično odvajanje $K'(\varepsilon)$ [Press, 2002; Orel, 1997] kot odvisne spremenljivke modela in učnih podatkov. Odvod funkcije, torej strmino funkcije v izbrani točki, pa lahko obravnamo kot indikator občutljivosti modela, saj govori o tem, kako hitro se z majhnimi spremembami napoved spremeni, kar je v skladu z definicijo naših ocen zanesljivosti.

3.2 Druge ocene zanesljivosti

Poleg ocen, razvitih z uporabo analize občutljivosti, definirajmo še dodatne ocene zanesljivosti, ki jih bomo v nadaljevanju uporabili za medsebojno primerjavo uspešnosti.

3.2.1 Varianca modela bagging

V poglavju 2.3.1 smo omenili pristop za ocenjevanja zanesljivosti napovedi, ki se uporablja za agregate nevronske mreže [Heskes, 1997; Carney in Cunningham, 1999]. Ta pristop ocenjevanja zanesljivosti lahko posplošimo in ga uporabimo v kombinaciji s poljubnim regresijskim modelom v okviru pristopa bagging.

Predpostavimo, da uporabljamo agregiran prediktor, ki je sestavljen iz m regresijskih modelov, od katerih vsak prispeva napoved K_m . Označbo neoznačenega primera

tedaj napovemo v skladu s povprečno vrednostjo:

$$K = \frac{\sum_{i=1}^m K_i}{m} \quad (3.6)$$

oceno zanesljivosti napovedi *BAGV* pa definiramo kot varianco napovedi regresijskih modelov

$$BAGV = \frac{1}{N} \sum_{i=1}^m (K_i - K)^2 \quad (3.7)$$

Pri testiranju smo uporabili oceno *BAGV* v kombinaciji z vsemi preučevanimi regresijskimi modeli, varianco napovedi pa smo ocenili z generiranjem 50 modelov.

3.2.2 Ocena z uporabo lokalnega prečnega preverjanja

V skladu s postopkom lokalnega prečnega preverjanja, opisanega v razdelku 2.3.2, definirajmo oceno zanesljivosti *LCV* (local cross-validation), ki jo opišemo z algoritmom na sliki 3.6.

```

1 PROGRAM LCV
2   določi množico  $k$  najbližjih sosedov  $N = \{(x_1, C_1), \dots, (x_k, C_k)\}$ 
3   ZA VSAK  $(x_i, C_i) \in N$ 
4     konstruiraj regresijski model  $M$  na  $N \setminus (x_i, C_i)$ 
5     za  $(x_i, C_i)$  izračunaj napoved  $K_i$ 
5     za  $(x_i, C_i)$  izračunaj napako  $E_i = \text{abs}(C_i - K_i)$ 
6   END ZA VSAK
7    $LCV = \frac{\sum_i E_i}{k}$ 
8 END LCV
```

Slika 3.6: Pseudokoda za izračun ocene zanesljivosti *LCV*.

Pri testiranju smo uporabili oceno *LCV*, ki glede na velikost domene dinamično prilagaja število najbližjih sosedov tako, da je njihovo število enako $1/20$ učnih primerov. Namesto povprečenja lokalnih napak najbližjih sosedov smo uporabili tudi izboljšavo, da se prispevki posameznih napak povprečijo glede na razdaljo med testnim primerom in sosedom.

3.2.3 Zanesljivost na podlagi gostote učnih primerov

Za ocenjevanje zanesljivosti primerov lahko uporabimo tudi verjetnostni pristop, ki temelji na podlagi porazdelitve učnih primerov v problemskem prostoru [Wand in Jones, 1995]. Pri tem predpostavimo, da je zanesljivost napovedi za primere, ki se nahajajo v bolj gostem delu prostora, večja kakor zanesljivost napovedi za primere, ki se nahajajo v manj gostih predelih problemskega podprostora. Predpostavka temelji na tem, da je učni algoritem v gostem delu problemskega prostora imel na razpolago več lokalne informacije, ima pa to slabost, da upošteva samo razdaljo do primerov v atributnem prostoru in ne njihovih vrednosti označb. Zaradi tega je metoda zelo občutljiva na šum v podatkih in na situacije, kjer so bistveno različni primeri težko ločljivi med seboj.

Gostoto prostora v izbrani točki ocenimo kot aproksimacijo funkcije gostote verjetnosti pojavitve primera. Za aproksimiranje verjetnostne porazdelitve uporabimo metodo Parzenovih oken [Silverman, 1986; Jeon in Landgrebe, 1994]. Ob tem metodo računanja Gaussovih jeder v večdimenzijskem prostoru poenostavimo in število dimenzij zmanjšamo s kombinacijo uporabe funkcije razdalje [Pavešić, 2000; Alpaydin, 2004] in dvorazsežnih Gaussovih jeder [Härdle in sod., 2004].

Če imamo podano učno množico $L = ((x_1, y_1), \dots, (x_n, y_n))$, gostoto verjetnosti za primer $(x, _)$ izračunamo torej kot:

$$p(x) = \frac{\sum_{i=1}^n \kappa(D(x, x_i))}{n} \quad (3.8)$$

pri čemer je D izbrana funkcija razdalje, κ pa jedrna funkcija (v našem primeru funkcija normalne porazdelitve).

Če bi tako zasnovano aproksimacijo verjetnosti $p(x)$ uporabili kot oceno zanesljivosti in če bi bile predpostavke, na katerih temelji ocena, izpolnjene, bi imela ta ocena lastnost, da bi z napako korelirala negativno. To pomeni, da bi ocena imela višjo vrednost tam, kjer bi bila napaka manjša, in manjšo vrednost tam, kjer je napaka večja, kar je v nasprotju z definicijo ostalih ocen. Zato za naše potrebe in lažjo

primerjavo z drugimi ocenami oceno *DENS* obrnemo:

$$DENS(x) = \max_{i=1..k}(p(x_i)) - p(x) \quad (3.9)$$

da bi s tem dosegli pozitivno korelacijo in poenotenje s pričakovanimi smermi korelacij ostalih ocen z napakami.

3.2.4 Ocena lokalne napake

Lokalno napako napovedi primera smo poskušali oceniti tudi z oceno *CNK*, ki napako napovedi oceni kot razliko med povprečno vrednostjo označbe najbližjih sosedov v učni množici in napovedjo primera K .

Če ocenjujemo zanesljivost napovedi K za primer $(x, _)$ in je množica najbližjih sosedov $(x, _)$ enaka $N = \{(x_1, C_1), \dots, (x_k, C_k)\}$, je ocena CNK^2 enaka:

$$CNK = \frac{\sum_{i=1}^k C_i}{k} - K. \quad (3.10)$$

Pri testiranju smo uporabili oceno *CNK*, ki za izračun povprečne lokalne označbe uporablja 5 najbližjih sosedov. Ker z izračunom ocene *CNK* na podoben način kot pri oceni *SAbias* pridobimo informacijo o smeri ocenjene napake, smo pri testiranju poleg primerjave z absolutno vrednostjo napake testnih primerov to oceno primerjali tudi s predznačenimi vrednostmi napak.

²CNK je okrajšava za $C_{Neighbors} - K$.

3. RAZVOJ OCEN ZANESLJIVOSTI

Poglavje 4

Testiranje in kombiniranje ocen zanesljivosti

4.1 Testiranje ocen zanesljivosti

Ocene zanesljivosti, opisane v prejšnjem poglavju, so bile testirane na 28 standardnih testnih domenah iz repozitorijev UCI ML Repository [Asuncion in Newman, 2007] in StatLib DataSets Archive [StatLib, 2005]. Uporabljene so bile domene s področij zdravstva, ekologije, tehničnih aplikacij in matematike. Pregled lastnosti uporabljenih domen je prikazan v tabeli A.1 v dodatku A.

Ocene zanesljivosti smo testirali z osmimi regresijskimi modeli, implementiranimi

v statističnem paketu R [R, 2006; Venables in sod., 2006]. Podajamo okrajšave imen modelov, uporabljene v prikazih rezultatov in opise uporabljenih parametrov:

RT – regresijska drevesa: model *rpart* iz knjižnice *rpart*, kot kriterij nečistosti pri notranjih vozliščih uporablja varianco označb, v listih modelira povprečno vrednost označbe primerov,

LR – linearna regresija: model *lm* iz knjižnice *stats*, brez eksplicitnih parametrov,

NN – nevronska mreža: model *newff* iz knjižnice *AMORE*, tronivojski perceptron s 5 skritimi nevroni in aktivacijsko funkcijo *tansig(x)* (hiperbolični tangens), učenje z adaptivnim gradientnim spustom,

BAG – bagging: bagging s 50 regresijskimi drevesi,

SVM – metoda podpornih vektorjev: model *svm* iz knjižnice *e1071*, implementacija LIBSVM algoritma [Chang in Lin, 2001], RBF jedro tretje stopnje, parameter $\varepsilon = 0.1$,

LWR – lokalno utežena regresija: lokalno utežena regresija z različico funkcije Gaussove porazdelitve za uteževanje pomembnosti primerov glede na razdaljo,

RF – naključni gozdovi: model iz knjižnice *randomForest*, uporablja 100 naključnih dreves,

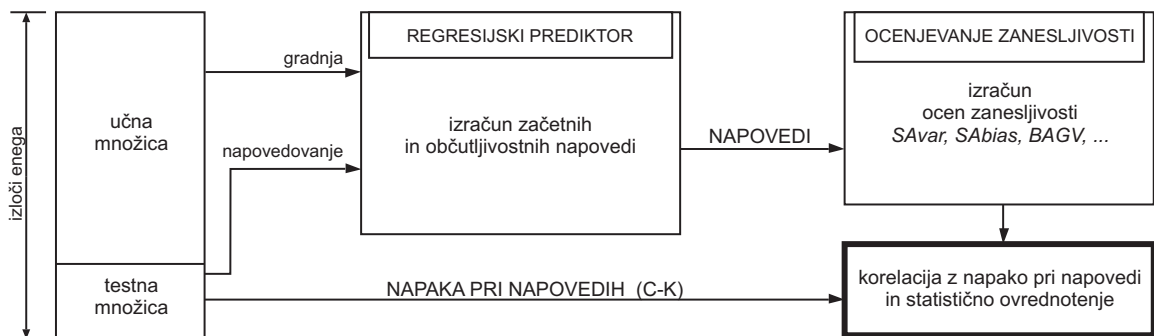
GAM – posplošeni aditivni model: model *gam* iz knjižnice *gam*, brez posebnih parametrov.

Testiranje ocen zanesljivosti je potekalo v obliki prečnega preverjanja na način *izloči-enega*. Vsakemu učnemu primeru, ki je bil izključen iz učne množice, je bila izračunana napoved na danem modelu, nato pa še 10 občutljivostnih napovedi ter posledično tudi oceni *SAvar*, *SAbias*. Primerom so bile istočasno izračunane tudi dodatne ocene *BAGV*, *LCV*, *DENS* in *CNK*. Z zaključkom prečnega preverjanja so bile na ta način izračunane ocene zanesljivosti za vse primere iz množice podatkov. Pred prehodom na naslednji korak prečnega preverjanja je bila učna množica vedno povrnjena v prvotno stanje, tako da se z vsakim korakom postopka ni iterativno razširjala, vsakič z dodatnim učnim primerom. Za izračun občutljivostnih

napovedi je bila uporabljena množica z naslednjimi vrednostmi parametra ε : $E = \{0.01, 0.1, 0.5, 1.0, 2.0\}$. Takšno množico vrednosti parametra ε smo izbrali z namenom, da preizkusimo različne stopnje spremembe označbe učnega primera - tako majhne kot tudi tiste, ki definirajo novo označbo izven intervala označb primerov učne množice (vrednosti 1.0 in 2.0).

Testiranje uspešnosti ocene smo izvedli z izračunom Pearsonovega korelacijskega koeficienta z napako napovedi učnih primerov (prave vrednosti označb primerov so pri prečnem preverjanju bile za izločene primere skrite pred učnim algoritmom). Kot smo omenili v poglavju 3, smo absolutne vrednosti vseh ocen korelirali z absolutnimi vrednostmi napak. Značilnost korelacijskega koeficienta smo ocenili z uporabo t -testa za ocenjevanje značilnosti korelacijskih koeficientov.

Ker pa oceni *SABias* in *CNK* vsebujeta tudi informacijo o smeri nestabilnosti modela, smo ti dve oceni dodatno korelirali še s predznačenimi vrednostmi napake. V nadaljevanju smo zaradi večje nazornosti prikazov tema ocenama dodali pripono $-p$ (*SABias-p*, *CNK-p*) ali $-a$ (*SABias-a*, *CNK-a*), glede na to ali so bile pri testiranju korelirane predznačene ali absolutne vrednosti ocen s predznačenimi ali absolutnimi vrednostmi napak. Pri vseh ostalih ocenah, ki pripone nimajo, to implicitno pomeni, da smo jih korelirali z absolutnimi vrednostmi napak napovedi. Postopek testiranja je shematsko prikazan na sliki 4.1.



Slika 4.1: Shematski prikaz postopka testiranja

Povzetek rezultatov, ki prikazuje uspešnost posameznih ocen glede na odstotek domen, na katerih je ocena dosegla značilno korelacijo, je prikazan v tabeli 4.1. Podrobnejši rezultati se nahajajo v tabelah A.2–A.9 v dodatku A.

4. TESTIRANJE IN KOMBINIRANJE OCEN ZANESLJIVOSTI

Tabela 4.1: Odstotek značilnih pozitivnih in negativnih korelacij med *ocenami zanesljivosti* in *napako pri napovedi*.

model	<i>SAvar</i> +/-	<i>SAbias-p</i> +/-	<i>SAbias-a</i> +/-	<i>BAGV</i> +/-	<i>LCV</i> +/-	<i>DENS</i> +/-	<i>CNK-p</i> +/-	<i>CNK-a</i> +/-
RT	46/0	82/0	50/0	64/0	36/0	36/4	86/0	68/0
LR	54/0	7/0	7/4	54/0	32/0	32/4	50/0	57/0
NN	39/4	18/4	29/4	50/0	36/0	25/4	36/4	39/4
BAG	46/4	21/0	11/0	57/0	50/0	36/7	25/0	46/0
SVM	46/4	36/7	25/0	46/0	61/0	39/4	29/11	36/0
LWR	39/7	4/7	11/7	46/0	46/0	43/7	25/11	32/0
RF	25/7	14/0	11/0	57/0	61/0	32/11	11/25	46/4
GAM	54/0	7/0	7/4	50/0	32/0	32/4	50/0	57/0
povprečno	44/3	24/2	19/2	53/0	44/0	34/6	39/6	48/1

Rezultati potrjujejo naša pričakovanja o smeri korelacij razvitih ocen zanesljivosti z napako. Vidimo namreč, da pri veliki večini ocen zanesljivosti in regresijskih modelov prevladuje število značilnih pozitivnih korelacij.

Najboljše povprečne rezultate smo dosegli pri uporabi ocen (v tem vrstnem redu) *BAGV*, *CNK-a* *LCV* in *SAvar*, najslabše pa pri korelaciji ocene *SAbias-a*. Iz rezultatov najbolj izstopa statistično značilna korelacija ocene *CNK-p* pri regresijskih drevesih, ki je značilno pozitivno korelirala z napako v 86% domen in pri nobeni domeni negativno. Na podoben način pri regresijskih drevesih izstopa tudi ocena *SAbias-p*, ki je dosegla značilno pozitivno korelacijo pri 82% domen. Vidimo lahko tudi, da je pri linearnih modelih (linearna regresija, posplošeni aditivni model) od ocene *BAGV* za odtenek bolj uspešna ocena *CNK-a*, ki značilno pozitivno korelira z napako v 57% domen in v 0% domen negativno.

Rezultati ocen *SAbias-a* in *SAbias-p* potrjujejo ugotovitve o neprimernosti uporabe analize občutljivosti za ocenjevanje zanesljivosti napovedi pri linearni regresiji, lokalno uteženi regresiji in posplošenem aditivnem modelu. Ocena zanesljivosti je dosegla slabše rezultate tudi pri ostalih modelih (vsi razen RT), katerih karakteristični grafi (slika 3.5) ne prikazujejo izrazite odsekoma zvezne funkcije.

Rezultati kažejo, da imata novozasnovani oceni *SAbias* in *CNK* tako kakor obstoječi oceni *BAGV* in *LCV* dober potencial za ocenjevanje napake napovedi. Ker je iz navedenih rezultatov razvidno, da so ocene različno uspešne pri različnih regre-

sijskih modelih, smo poskusili še kombinirati ocene z namenom pridobitve skupne ocene, ki bi bila dobra na vseh regresijskih modelih.

4.2 Kombiniranje ocen zanesljivosti

Iz rezultatov posameznih ocen smo opazili različno uspešnost ocen glede na uporabljen regresijski model. Z željo pridobiti kombinirano oceno, ki bo uspešnejša kakor vsaka samostojno, smo poskusili kombinirati pare ocen. Ocene smo kombinirali z modelom linearne kombinacije obeh ocen:

$$Ocena_{nova} = \gamma \cdot Ocena_1 + (1 - \gamma) \cdot Ocena_2 \quad (4.1)$$

pri $\gamma = 0.5$, torej s povprečenjem dveh ocen. Pri izbiri parov ocen za kombiniranje pa smo se omejili samo na podmnožico vseh možnih parov, ker vse kombinacije ocen niso smiselne. Namreč, s kombiniranjem ocen, ki uspešno korelirajo z absolutno napako, in ocen, ki korelirajo s predznačeno napako, bi pridobili oceno, za katero bi izgubili interpretacijo, ali še nosi informacijo o smeri napake ali ne. Zaradi tega se osredotočimo na kombinacije ocen, katerih uspešnost je prikazana v tabeli 4.2.

Iz rezultatov kombiniranja ocen vidimo, da je največjo uspešnost dosegla kombinacija ocen *BAGV* in *CNK-a*. Iz primerjave tabel 4.1 in 4.2 je razvidno, da je kombinirana ocena dosegla boljše rezultate (večje število značilnih pozitivnih korelacij z napako) pri nevronskih mrežah in metodi bagging kot vsaka izmed posameznih ocen. Kljub temu pa kombinacija teh ocen v povprečju značilno pozitivno korelira pri 54% domenah in negativno v 1%, kar pomeni, da izrazite prednosti pred dosedaj najboljšo oceno *BAGV* s kombiniranjem nismo pridobili.

Ker je kombinacija ocen dala dobre rezultate, se vendarle odločimo, da jo bomo obdržali za potrebe nadaljnjih poskusov. Za te potrebe uvedimo torej oceno *BVCK*, ki je za i -ti primer definirana kot:

$$BVCK_i = \frac{BAGV_i + CNK_i}{2} \quad (4.2)$$

Podrobnejši rezultati ocene *BVCK* so v tabeli A.10 v dodatku A.

4. TESTIRANJE IN KOMBINIRANJE OCEN ZANESLJIVOSTI

Tabela 4.2: Odstotek značilnih pozitivnih in negativnih korelacij med *linearnimi kombinacijami ocen zanesljivosti* in *napako pri napovedi*.

model	<i>SAvar</i>	<i>SAvar</i>	<i>SAvar</i>	<i>SAvar</i>	<i>SAvar</i>	<i>SAbias-p</i>	<i>SAbias-a</i>	<i>SAbias-a</i>
	<i>SAbias-a</i>	<i>BAGV</i>	<i>LCV</i>	<i>DENS</i>	<i>CNK-a</i>	<i>CNK-p</i>	<i>BAGV</i>	<i>LCV</i>
	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
RT	50/0	61/0	43/0	46/0	61/0	89/0	61/0	50/0
LR	54/0	54/0	39/4	54/0	61/0	50/0	54/0	32/0
NN	39/0	54/0	32/0	36/4	50/0	36/4	46/0	32/0
BAG	46/4	54/0	39/0	43/4	54/0	25/0	54/0	39/0
SVM	46/4	46/0	46/0	50/4	46/4	32/11	46/0	61/0
LWR	39/7	43/4	46/4	39/7	43/7	29/7	43/0	46/0
RF	25/7	39/4	25/4	25/7	36/4	14/18	50/0	50/0
GAM	54/0	57/0	39/4	54/0	61/0	50/0	50/0	32/0
povprečno	44/3	51/1	39/2	43/3	52/2	41/5	51/0	43/0

model	<i>SAbias-a</i>	<i>SAbias-a</i>	<i>BAGV</i>	<i>BAGV</i>	<i>BAGV</i>	<i>LCV</i>	<i>LCV</i>	<i>DENS</i>
	<i>DENS</i>	<i>CNK-a</i>	<i>LCV</i>	<i>DENS</i>	<i>CNK-a</i>	<i>DENS</i>	<i>CNK-a</i>	<i>CNK-a</i>
	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
RT	54/0	68/4	61/0	57/0	71/4	36/0	57/0	64/0
LR	32/4	57/0	43/0	50/0	50/0	32/0	39/0	57/0
NN	29/4	43/4	39/0	54/0	54/0	32/0	39/0	39/4
BAG	18/0	46/0	54/0	54/0	61/0	50/0	50/0	46/0
SVM	29/0	36/0	54/0	54/0	46/0	61/0	50/0	36/0
LWR	7/7	32/0	46/0	46/0	43/0	46/0	36/0	36/0
RF	11/0	46/4	61/0	50/0	50/0	61/0	61/0	46/4
GAM	32/4	57/0	46/0	54/0	54/0	32/0	39/0	57/0
povprečno	27/1	48/2	51/0	52/0	54/1	44/0	46/0	48/1

Poglavje 5

Samodejna izbira najbolj ustrezne ocene zanesljivosti

Rezultati testiranja ocen so pokazali dober potencial uporabe vseh ocen za namene ocenjevanja zanesljivosti napovedi. Istočasno smo pa iz rezultatov videli, da dosegajo ocene različno povprečno uspešnost pri različnih regresijskih modelih. Boljšo korelacijo ocen pri vseh modelih smo poskusili doseči v razdelku 4.2, zanimajo pa nas še pristopi k izbiri ocene, ki bo specifična za vsako posamezno domeno, ter morebitna razlaga, kdaj je uporaba posamezne ocene najbolj primerna. Ta problem v nadaljevanju rešujemo z uporabo dveh pristopov, metaučenja in notranjega prečnega preverjanja.

5.1 Metaučenje v prostoru domen

5.1.1 Testiranje dinamične izbire ocene

Pri pristopu z metaučanjem smo izdelali metaklasifikator, katerega cilj je bil za vsako kombinacijo domena/regresijski model predlagati optimalno oceno za ocenjevanje zanesljivosti primerov. Ocena zanesljivosti je torej v metaklasifikacijskem problemu predstavljala razred, pari domena/regresijski model so predstavljali učne primere, atributi učnih primerov pa so opisovali lastnosti domen in uporabljenega regresijskega modela. Definirali smo naslednjih 7 atributov:

- **regresijski model**, ki ga uporabljamo na domeni (v nadaljevanju *model*),
- **število učnih primerov** v domeni (*no.examples*),
- **število atributov** v domeni (*no.attr*),
- **relativna srednja kvadratna napaka**, ki jo regresijski model doseže na domeni pri desetkratnem prečnem preverjanju (*cv.rmse*),
- **povprečna gostota** problemskega prostora, vzorčena v učnih primerih domene (*avg.dens*),
- **povprečna razdalja do 5 najbližjih sosedov** v atributnem prostoru, povprečena preko vseh primerov domene (*avg.DA*),
- **povprečna razlika med napovedjo primera in napovedmi 5 najbližjih sosedov**, povprečena preko vseh primerov domene (*avg.DK*).

Z izračunom gornjih atributov smo vse kombinacije domena/model ($28 \times 8 = 224$ parov) združili v učno množico za metaklasifikator. Pri tem smo kot razred določili tisto oceno zanesljivosti, ki je na tej domeni dosegla največjo pozitivno korelacijo z napako, ne glede na to, ali je ta korelacija bila statistično značilna ali ne. Množica razredov je vsebovala oznake 9 ocen: *SAvar*, *SAbias-p*, *SAbias-a*, *CNK-p*, *CNK-a*, *LCV*, *BAGV*, *DENS in BVCK*. Da bi zagotovili nepristranost testiranja, smo pri napovedovanju ocene za testno domeno iz učne množice izločili primer, ki je povezan s poskusom.

Rezultati testiranja sistema z dinamično izbiro ocene zanesljivosti z metaklasifikatorjem so prikazani v tabeli 5.1. Iz rezultatov in iz primerjave z dosedaj v povprečju najbolj uspešno oceno *BVCK* vidimo, da smo z metaučenjem uspeli v povprečju doseči boljše rezultate, kar pomeni, da smo za posamezno domeno uspešno napovedali ustrezno oceno zanesljivosti. Podrobni rezultati so prikazani v tabeli A.11 v dodatku A.

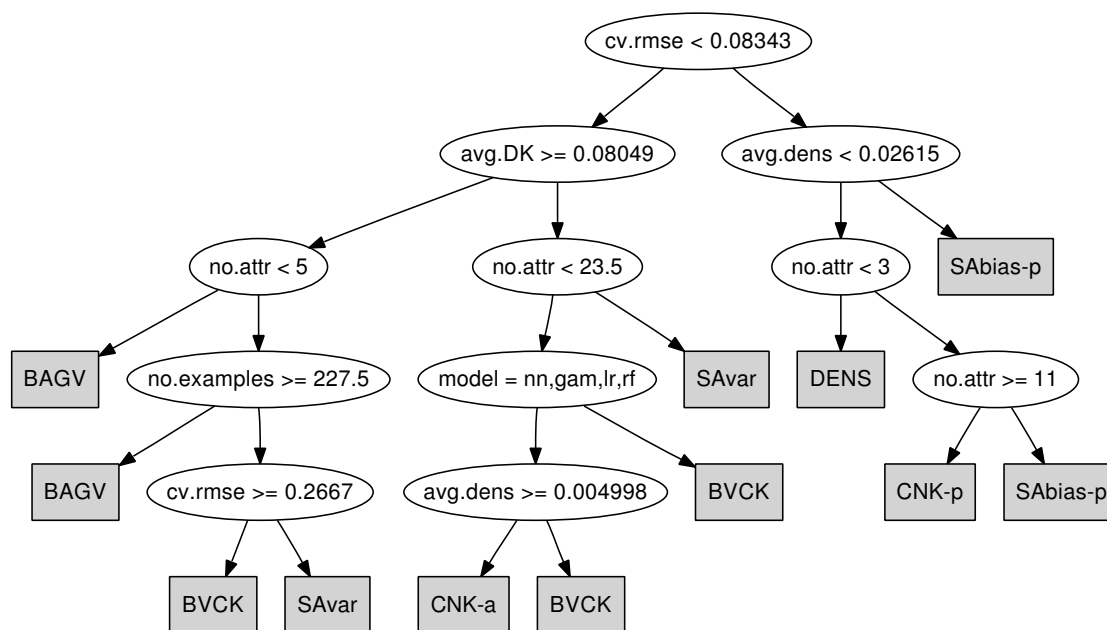
Tabela 5.1: Primerjava odstotka značilnih pozitivnih in negativnih korelacij med napako pri napovedi in oceno zanesljivosti *BVCK* ter oceno, ki je bila napovedana z uporabo metaučenja.

model	BVCK	Metaučenje
	+/-	+/-
RT	71/4	79/0
LR	50/0	54/0
NN	54/0	46/0
BAG	61/0	61/0
SVM	46/0	54/0
LWR	43/0	43/4
RF	50/0	64/0
GAM	54/0	54/0
povprečno	54/1	57/1

5.1.2 Razlaga izbire in uspešnosti ocen

Izbira optimalne ocene. Z uporabo učne množice za metaprediktor lahko poskusimo poiskati razlago, kdaj je uporaba katere izmed ocen najbolj primerna. Za te namene smo na učni množici zgradili klasifikacijsko drevo, ki je prikazano na sliki 5.1 (drevo je porezano z uporabo 1-SE pravila in cost-complexity rezanja [Breiman in sod., 1984; Torgo, 2003]).

Iz odločitvenega drevesa vidimo, da je pri napovedi ocene najbolj pomembna relativna srednja kvadratna napaka izbranega regresijskega modela na domeni, izračunana z desetkratnim prečnim preverjanjem (*cv.rmse*). Drevo prikazuje, da so za bolj uspešne regresorje bolj primerne ocene *DENS*, *CNK-p* in *SAbias-p*, za slabše regresorje pa ocene, ki imajo dober povprečni uspeh na večini domen (*BAGV*, *BVCK*, *SAvar* in *CNK-a*).

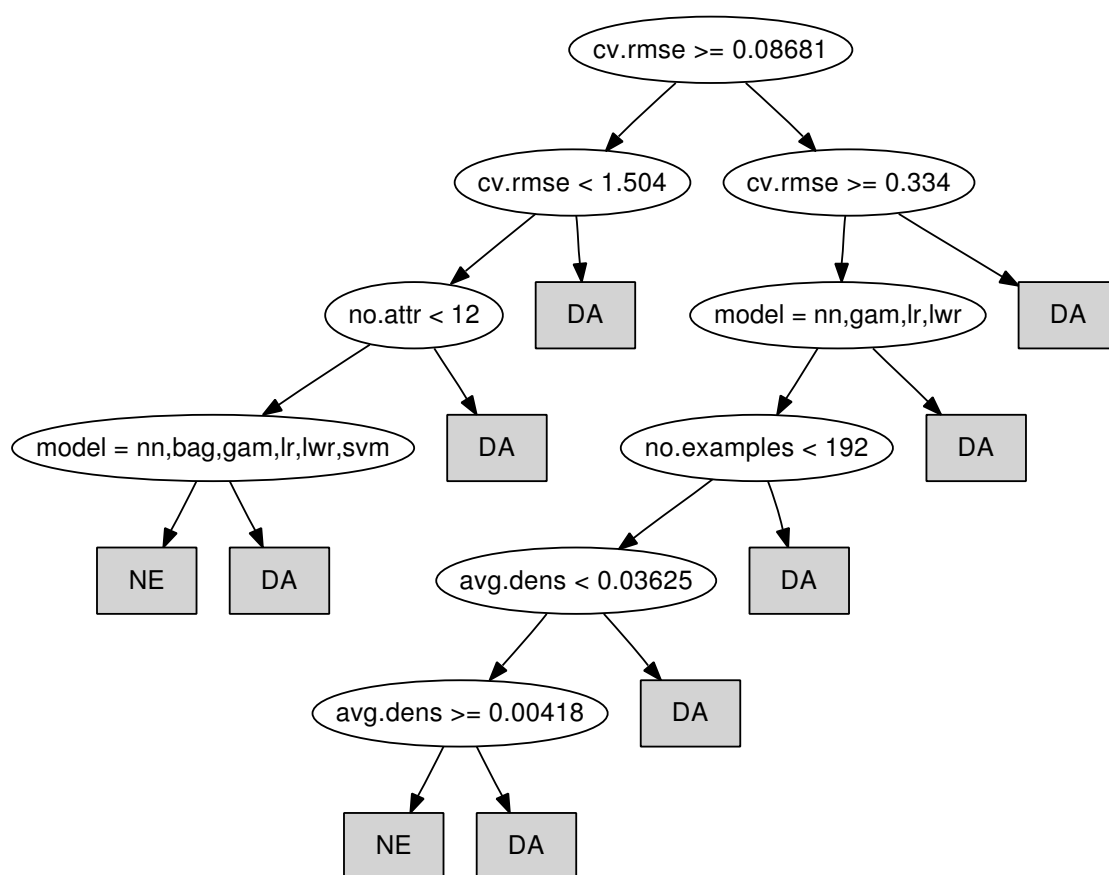


Slika 5.1: Klasifikacijsko drevo za napovedovanje optimalne ocene zanesljivosti

Značilnost pozitivne korelacije. V zgornjem primeru smo s prikazom klasifikacijskega drevesa na učni množici metaprediktorja poiskali razlago, kdaj je najbolj primerno uporabiti katero izmed ocen. Pri tem smo učenim primerom za metaučenje določili najboljšo oceno (razred) kot tisto, ki je na domeni dosegla najvišjo korelacijo z napako. Ker se pri tem nismo ozirali, ali je bila ta korelacija statistično značilna ali ne, smo po eni strani metaklasifikatorju omogočili učenje iz večjega števila učnih primerov, po drugi strani pa smo izgubili informacijo o tem, ali katera izmed ocen sploh dosega statistično značilen uspeh na tej domeni. Za analizo tega aspekta smo v nadaljevanju zgradili klasifikacijsko drevo na spremenjeni učni množici, ki ima namesto oznak ocen zanesljivosti kot možni vrednosti razredov le vrednosti DA in NE , ki ponazarjata, ali je katera izmed 9 ocen dosegla na kombinaciji domena/model značilno pozitivno korelacijo ali ne.

Opisano klasifikacijsko drevo, ki napoveduje uspešnost korelacije ocen z napako, je prikazano na sliki 5.2. Iz klasifikacijskega drevesa vidimo, da drevo v večini listov napoveduje uspešno korelacijo z vsaj eno izmed ocen zanesljivosti. Iz drevesa

vidimo tudi, da imajo ocene zanesljivosti večjo težavo z doseganjem ocen pri nevronskih mrežah, posplošenem aditivnem modelu, linearni regresiji in lokalno uteženi regresiji, saj odločitve v notranjih vozliščih o izbiri teh modelov vodijo do listov z oznako razreda *NE*. Podrobnejši izpis drevesa se nahaja na sliki A.1 v dodatku A. S podrobnejšo analizo drevesa smo ugotovili, da v lista, v katerih drevo napoveduje neuspešno korelacijo z razvitimi ocenami zanesljivosti, pade $29 + 11 = 40$ kombinacij domena/model, kar predstavlja 18% vseh uporabljenih kombinacij domen in modelov. Ker smo z dosedanjimi najbolj uspešnimi poskusi dosegli uspešno korelacijo ocen v (samo) 56% kombinacij domena/model, nas to motivira k poskusu izboljšanja tega rezultata.



Slika 5.2: Klasifikacijsko drevo za napovedovanje uspešnosti korelacije 9 ocen

Težavnost domen. Klasifikacijsko drevo na sliki 5.2 nam daje slutiti, da v določenih domenah ne moremo doseči korelacij napake napovedi z nobeno izmed razvitih ocen. Na osnovi skupnega števila ocen (1–9), s katerimi smo dosegli značilne pozitivne korelacije, lahko domene razvrstimo po njihovi težavnosti ocenjevanja zanesljivosti. Takšno razvrstitev prikazuje tabela 5.2.

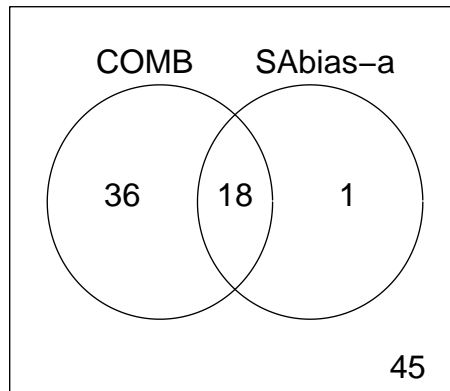
Za analizo odvisnosti uspešnosti ocen zanesljivosti pri posameznih testnih domenah smo uspešnost korelirali z lastnostmi domen: s številom primerov, številom atributov in napako RMSE, povprečeno preko vseh 8 obravnavanih regresijskih modelov. Korelacijske koeficiente in njihove stopnje značilnosti prikazuje tabela 5.3. Iz rezultatov lahko vidimo, da uspešnost značilno negativno korelira s povprečno napako RMSE. To pomeni, da imajo ocene zanesljivosti v povprečju boljšo uspešnost na domenah, na katerih regresijski modeli v povprečju naredijo manjšo napako. Ta ugotovitev se sklada s klasifikacijskim drevesom na sliki 5.2, saj iz podrobnejšega izpisa drevesa na sliki A.1 v dodatku A vidimo, da desno poddrevo (modeli z manjšo napako) napoveduje uspešno korelacijo ocen zanesljivosti pri 134 (60%) kombinacijah domena/model, levo poddrevo (modeli z večjo napako) pa 50 (22%) kombinacijam (spomnimo se, da v preostalih 18% kombinacij domena/model do značilnih korelacij ne pride).

Izboljšava metaučenja. Metaučenje smo zasnovali tako, da smo kot možne vrednosti razreda metaklasifikatorja uporabili vseh 9 razpoložljivih ocen, ne glede na to, ali so posamezne ocene dale dobre rezultate ali ne. Iz tabele 5.2 smo ravno tako videli, da je v večini kombinacij domena/model značilno pozitivno korelacijo z napako doseglo več ocen. Na tej podlagi se lahko vprašamo, ali bi se lahko pri metaučenju omejili na manjše število ocen zanesljivosti, ki bi še vedno uspešno korelirale z napako pri enakem številu kombinacij domena/model. Z izbiro manjšega števila bolj uspešnih ocen bi istočasno metaklasifikatorju onemogočili, da za izbrano kombinacijo domena/model napove oceno, ki daje na domeni manj optimalen rezultat kakor kakšna druga ocena.

Reševanja tega problema smo se lotili z analizo pokritosti kombinacij domena/model in povečevanjem števila razpoložljivih ocen. Če se omejimo samo na uporabo

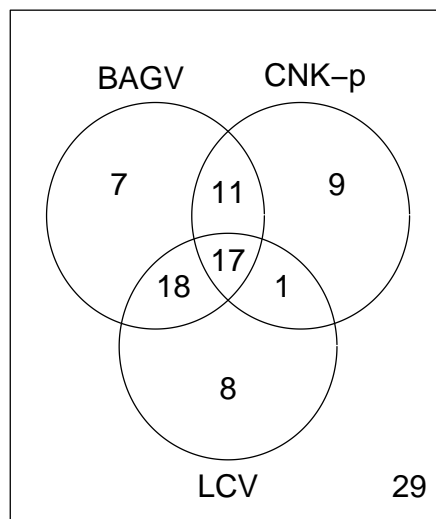
dveh ocen, lahko uspešnost poljubne kombinacije dveh ocen prikažemo z Vennovim diagramom [Ruskey in Weston, 2007], kot ga prikazuje vzorec na sliki 5.3. Kot najbolj optimalen par dveh ocen se odločimo izbrati tistega, ki je uspešen na največjem številu kombinacij domena/model (torej minimizira odstotek, prikazan zunaj krogov). Za naš primer smo ugotovili, da sta to oceni *CNK-p* in *LCV*, ki sta uspešni v 66% kombinacij domena/model oziroma neuspešni v 34% kombinacij domena/model.

Na enak način nadaljujemo pri izbiri treh ocen, ki jih tudi lahko prikažemo z Vennovim diagramom, kot ga prikazuje vzorec na sliki 5.4. Iz analize vseh kombinacij treh ocen smo ugotovili, da je optimalna uporaba ocen *CNK-p*, *LCV* in *SAvar*, ki so uspešne v 73% kombinacij domena/model. Pri analizi kombinacij štirih ali več ocen izgubimo možnost nazorne grafične predstavitve, optimalne kombinacije pa vendarle lahko predstavimo v tabelarični obliki. Rezultati analize optimalnih kombinacij različnega števila ocen so prikazani v tabeli 5.4.



Slika 5.3: Odstotek kombinacij domena/model, pri katerih oceni zanesljivosti uspešno korelirata z napako, predstavljen z Vennovim diagramom.

5. SAMODEJNA IZBIRA NAJBOLJ USTREZNE OCENE ZANESLJIVOSTI



Slika 5.4: Odstotek kombinacij domena/model, pri katerih ocene zanesljivosti uspešno korelirajo z napako, predstavljen z Vennovim diagramom treh ocen.

Tabela 5.2: Razvrstitev domen glede na uspešnost ocenjevanja zanesljivosti z 9 ocenami zanesljivosti. Stolpci *RT-GAM* kažejo število ocen, ki so v posamezni kombinaciji domena/model značilno pozitivno korelirale z napako. Stolpec *Uspeh* prikazuje povprečen odstotek ocen, ki so v domeni značilno pozitivno korelirale z napako. Zadnji trije stolpci podajajo lastnosti domen: število učnih primerov, število atributov in povprečno napako RMSE preko vseh regresijskih modelov.

										povpr.		
	RT	LR	NN	BAG	SVM	LWR	RF	GAM	Uspeh	#prim.	#atr.	cv.rmse
cpu	9	7	8	6	9	6	6	7	81%	209	6	0,318
fishcatch	8	6	7	8	8	7	6	6	78%	158	7	0,112
auto_price	9	6	9	7	6	6	6	6	76%	159	15	0,238
transplant	6	8	8	7	9	2	7	8	76%	131	2	0,818
hungarian	8	5	5	8	7	6	8	5	72%	294	13	0,664
cloud	9	5	5	6	9	5	6	5	69%	108	6	0,334
autohorse	8	6	6	5	7	2	5	6	63%	203	25	0,145
triazines	7	6	4	7	2	3	5	6	56%	186	60	4,735
auto93	8	3	6	5	3	6	5	3	54%	93	22	0,484
bodyfat	6	4	6	4	9	2	4	4	54%	252	14	0,064
pharynx	9	6	0	7	0	5	5	6	53%	195	11	0,660
servo	8	4	4	5	2	5	5	4	51%	167	4	0,662
pwlinear	8	5	3	1	4	6	0	5	44%	200	10	0,355
elusage	5	1	3	5	6	6	4	1	43%	55	2	0,578
pyrim	4	3	7	2	4	5	1	5	43%	74	27	1889,38
grv	8	4	2	6	1	0	1	3	35%	123	3	0,490
sleep	2	5	4	2	2	0	1	5	29%	58	7	1,255
echomonths	6	0	0	2	4	3	5	0	28%	130	9	0,698
pollution	5	6	0	0	1	0	0	6	25%	60	15	1,060
diabetes	3	3	2	1	1	3	0	2	21%	43	2	2,650
tumor	3	0	1	1	4	0	1	0	14%	86	4	1,315
wpbc	3	2	0	1	1	0	1	2	14%	198	32	1,119
basketball	4	1	0	0	0	0	1	1	10%	96	4	0,977
lowbwt	1	0	1	3	1	1	0	0	10%	189	9	0,459
breasttumor	2	0	0	0	2	0	0	0	6%	286	9	1,105
mbagrade	1	0	0	0	0	1	2	0	6%	61	2	1,276
fruitfly	1	0	0	0	0	0	1	0	3%	125	4	1,459
brainsize	0	0	0	0	0	1	0	0	1%	20	8	80,731

5. SAMODEJNA IZBIRA NAJBOLJ USTREZNE OCENE ZANESLJIVOSTI

Tabela 5.3: Korelacija uspešnosti ocen zanesljivosti z lastnostmi testnih domen.

korelacija uspešnosti ocen z	korelacijski koeficient	stopnja značilnosti
#primerov	0,357	0,062
#atributov	0,170	0,387
povprečen cv.rmse	-0,638	<0,001

Tabela 5.4: Optimalne kombinacije ocen glede na maksimiziranje pokritosti čim večjega števila kombinacij domena/model. Drugi stolpec prikazuje število uporabljenih ocen, tretji pa odstotek kombinacij domena/model, pri katerih nobena izmed ocen ni dosegla značilne korelacije z napako.

#	kombinacija ocen	število ocen	% neuspeha
1	CNK-p, LCV	2	34
2	CNK-p, LCV, SAvar	3	27
3	CNK-p, LCV, SAbias-p, SAvar	4	24
4	BAGV, CNK-p, DENS, LCV, SAbias-p	5	23
5	BAGV, CNK-p, LCV, SAbias-p, SAvar	5	23
6	CNK-p, BVCK, DENS, LCV, SAbias-p	5	23
7	CNK-p, DENS, LCV, SAbias-p, SAvar	5	23
8	BAGV, CNK-p, BVCK, DENS, LCV, SAbias-p	6	22
9	BAGV, CNK-p, DENS, LCV, SAbias-p, SAvar	6	22
10	BAGV, CNK-a, CNK-p, BVCK, DENS, LCV, SAbias-p	7	21
11	BAGV, CNK-a, CNK-p, DENS, LCV, SAbias-p, SAvar	7	21
12	BAGV, CNK-p, BVCK, DENS, LCV, SAbias-p, SAvar	7	21
13	BAGV, CNK-a, CNK-p, BVCK, DENS, LCV, SAbias-p, SAvar	8	21
14	BAGV, CNK-a, CNK-p, BVCK, DENS, LCV, SAbias-a, SAbias-p, SAvar	9	21

Tabela 5.5: Število pozitivnih in negativnih korelacij z napako pri spreminjanju števila uporabljenih ocen pri metaučanju.

kombinacija ocen iz tabele 5.4	14	13	12	11	10	9	8	7	6	5	4	3	2	1	
število ocen	9	8	7			6		5					4	3	2
% neuspeha	21	21	21			22		23					24	27	37
RT	79/0	79/0	82/0	79/0	86/0	82/0	82/0	79/0	89/0	79/0	75/0	82/0	71/0	86/0	
LR	54/0	46/0	71/0	61/0	54/0	61/0	61/0	54/0	61/0	64/0	61/0	61/0	61/0	54/0	
NN	46/0	50/0	61/4	57/0	50/0	54/0	61/0	54/0	54/0	43/0	46/0	57/0	46/4	39/4	
BAG	61/0	68/0	57/0	54/0	54/0	61/0	46/0	43/0	57/0	50/0	57/0	46/0	43/0	36/0	
SVM	54/0	61/0	57/0	57/0	61/0	43/0	54/0	43/4	54/7	43/7	39/0	43/0	43/4	46/7	
LWR	43/4	39/0	43/0	43/0	36/4	36/0	36/7	39/0	46/0	32/4	39/4	29/7	39/4	25/7	
RF	64/0	54/0	50/0	54/0	43/4	57/0	54/0	36/7	54/0	57/0	46/4	29/4	29/11	43/7	
GAM	54/0	50/4	61/0	61/0	57/0	61/0	61/0	57/0	57/0	61/0	57/0	61/0	61/0	54/0	
povprečno	57/1	56/1	60/1	58/0	55/1	57/0	57/1	51/1	59/1	54/1	53/1	51/1	49/3	48/3	

Iz prikazanih rezultatov metaučanja na podmnožicah ocen v tabeli 5.5 vidimo, da smo v skladu s predpostavkami, tudi z manjšim številom uporabljenih ocen uspeli doseči boljše rezultate korelacij ocen z napako. Pri uporabljeni kombinaciji 7 ocen pod oznako 12 (izločeni oceni *SAbias-p* in *CNK-p*) namreč vidimo, da je metaprediktor uspešno napovedal oceno, ki je značilno pozitivno korelirala z napako v povprečno 60% poskusov in značilno negativno korelirala z napako v povprečno 1% poskusov. Dober rezultat lahko zasledimo tudi pri kombinacijah 11 (izločeni *SAbias-a* in *BVCK*) in kombinaciji 6 (izločene ocene *SAvar*, *SAbias-a*, *CNK-a* in *BAGV*). Dober rezultat kombinacije 6 izstopa tudi po tem, ker je v primerjavi s kombinacijo 11 dosežen z uporabo učne množice, ki kombinacijam domena/model kot razred pripisuje manjši odstotek uspešnih ocen (z napako ne korelira 23% ocen zanesljivosti, pri kombinaciji 11 pa jih ne korelira 21%).

Najboljši rezultat (60% značilnih pozitivnih korelacij) lahko interpretiramo tudi na ta način, da smo z njim dosegli približno 76% možnega cilja, da bi za vsako kombinacijo domena/model napovedali oceno, ki bi zanjo uspešno korelirala z napako. Z uporabo 9 ocen smo namreč pokazali, da lahko pri 79% kombinacij domena/model dosežemo značilno pozitivno korelacijo ocen z napako (odstotek neuspeha je 21%), kar predstavlja torej največ, kar lahko z izbranim naborom ocen zanesljivosti na izbranih testnih podatkih dosežemo. Vendarle, zaključimo lahko, da smo z uporabo metaučanja uspešno izvedli izbiro ocen glede na vsako obravnavano problemsko

domeno in s kombiniranjem različnih ocen dosegli boljše rezultate.

5.2 Izbira ocene z notranjim prečnim preverjanjem

Problema dinamične izbire ocene glede na domeno smo se lotili tudi z notranjim prečnim preverjanjem. Jedro poteka postopka smo zasnovali kot dvojni, vgnezden postopek prečnega preverjanja:

- Zunanja zanka postopka 10-krat razdeli množico primerov na učne in testne primere. Za vse primere iz testne množice izračuna občutljivostne napovedi in lokalne parametre, ki so potrebni za izračun ocen zanesljivosti. Zunanja zanka uporabi notranjo zanko za predlog ocene zanesljivosti, ki ga shrani in počaka na zaključek celotnega postopka.
- Notranja zanka izvede prečno preverjanje na način izloči-enege. Vsem notranjim testnim primerom z uporabo notranje učne množice izračuna ocene zanesljivosti in jih po koncu notranje zanke korelira z napako. Zunanji zanki nato posreduje predlog ocene, ki najbolje korelira z napako.
- Konec postopka in testiranje: Po zaključku zunanje zanke iz 10 predlogov ocen zanesljivosti izberemo tisto, ki je preko vseh 10 ponovitev zunanje zanke dosegla najvišjo povprečno pozitivno korelacijo z napako. Uspešnost izbire ocene zanesljivosti ocenimo v skladu s statistično značilnostjo korelacije te ocene na testnih množicah zunanje zanke.

Povzetek rezultatov samodejne izbire ocene z notranjim prečnim preverjanjem je prikazan v tabeli 5.6. Podrobni rezultati se nahajajo v tabeli A.12 v dodatku A. Zaradi velike časovne zahtevnosti algoritma je bil ta pristop testiran na podmnožici 15 domen (brainsize, diabetes, elusage, sleep, pollution, mbagrade, pyrim, tumor, auto93, basketball, cloud, grv, fishcatch, autoprice, servo), ki so bile izbrane naključno in približno enakomerno vzorčene po razvrstitvi zahtevnosti domen, prikazanih v tabeli 5.2. Iz rezultatov vidimo, da smo s tem pristopom dosegli boljše rezultate kakor pri uporabi samostojnih ocen zanesljivosti, njihovih kombinacij ali metaučenja.

S tem smo pri opisanem pristopu obenem dosegli tudi dosedaj najvišje povprečno število značilnih pozitivnih korelacij (73%), medtem ko je število značilnih negativnih korelacij enako 0%.

Tabela 5.6: Odstotek značilnih pozitivnih in negativnih korelacij med napako pri napovedi in oceno zanesljivosti, ki je bila napovedana z notranjim prečnim preverjanjem. Preglednica podaja primerjavo uspešnosti med najbolj uspešno samostojno oceno zanesljivosti *BVCK*, metaučenjem in notranjim prečnim preverjanjem.

model	BVCK	Metaučenje	<i>Interno</i> <i>prečno preverjanje</i>
	+/-	+/-	+/-
RT	71/4	79/0	87/0
LR	50/0	54/0	73/0
NN	54/0	46/0	73/0
BAG	61/0	61/0	67/0
SVM	46/0	54/0	67/0
LWR	43/0	43/4	73/0
RF	50/0	64/0	60/0
GAM	54/0	54/0	80/0
povprečno	54/1	57/1	73/0

Primerjava frekvenc izbranih ocen. V tabeli 5.7 podajamo primerjavo deležev (v odstotkih), v katerih je vsak izmed postopkov samodejne izbire ocene zanesljivosti predlagal posamezne ocene.

Iz tabele lahko vidimo, da je pristop z metaučenjem najpogosteje predlagal ocene *BVCK*, *BAGV* in *CNK-a*, ki so najbolj uspešne kot samostojne ocene. Kljub temu, da algoritem notranjega prečnega preverjanja ni najpogosteje predlagal istih ocen, pa lahko iz pregleda tabel A.2–A.10 v dodatku A vidimo, da je v veliki večini najpogosteje predlagal oceno, ki je najboljše pozitivno korelirala s primeri tudi pri testiranju samostojnih ocen. Na osnovi tega lokalnega prilagajanja domeni namesto induciranja splošnega pravila za izbiro ocene pa je dosegel tudi večje število značilnih pozitivnih korelacij.

5. SAMODEJNA IZBIRA NAJBOLJ USTREZNE OCENE ZANESLJIVOSTI

Tabela 5.7: Odstotek samodejno izbranih ocen pri uporabi metaučenja in postopka notranjega prečnega preverjanja.

	metaučenje	notranje prečno preverjanje
Savar	11	12
SAbias-p	14	5
SAbias-a	0	3
BAGV	15	10
LCV	2	17
DENS	8	8
CNK-p	11	22
CNK-a	19	9
BVCK	20	14

Poglavje 6

Aplikacija ocen zanesljivosti na realni domeni

Predlagane ocene zanesljivosti smo z uporabo pristopa dinamične izbire z meta-učanjem testirali tudi na realni domeni. Množica podatkov je vsebovala primere 1035 pacientk, ki so bile v letih 1983–1987 operativno zdravljene za rakom dojke na Onkološkem inštitutu Kliničnega centra v Ljubljani. Učni primeri so bili opisani z demografskimi podatki pacientk, znanimi prognostičnimi dejavniki za morebitno ponovitev bolezni in s podatki o prejetih terapijah in načinih zdravljenja. Cilj raziskave je bil za vsako pacientko napovedati čas morebitne ponovitve bolezni v letih po operativnem posegu. Pretekle raziskave so pokazale, da je napovedovanje ponovitve

bolezni težek problem, saj do ponovitve lahko pride v širokem časovnem obdobju – času približno 20 let po zdravljenju. Ker natančni vzroki za natančen čas ponovitve boleznii niso znani niti strokovnjakom s področja, smo napovedim časa ponovitve regresijskega sistema dodali še ocene zanesljivosti napovedi. Z njimi smo zdravnikom želeli posredovati dodatno informacijo za validacijo pravilnosti regresijskih napovedi.

6.1 Množica podatkov

Primere v učni množici opisuje množica 24 atributov in dve odvisni spremenljivki, potrebni za reševanje opisanega regresijskega problema:

1. Binarna diskretna spremenljivka *dog_TTP* hrani informacijo o tem, ali je pri pacientki prišlo do ponovitve boleznii (vrednost 1) ali ne (vrednost 0).
2. Zvezna spremenljivka *TTP* hrani podatek o času do naslednjega relevantnega medicinskega dogodka. Če ima *dog_TTP* vrednost 1, gre za čas do ponovitve boleznii, če pa ima vrednost 0, gre za čas do zaključka opazovanja pacientke, ki je lahko njena smrt ali zaključek raziskave.

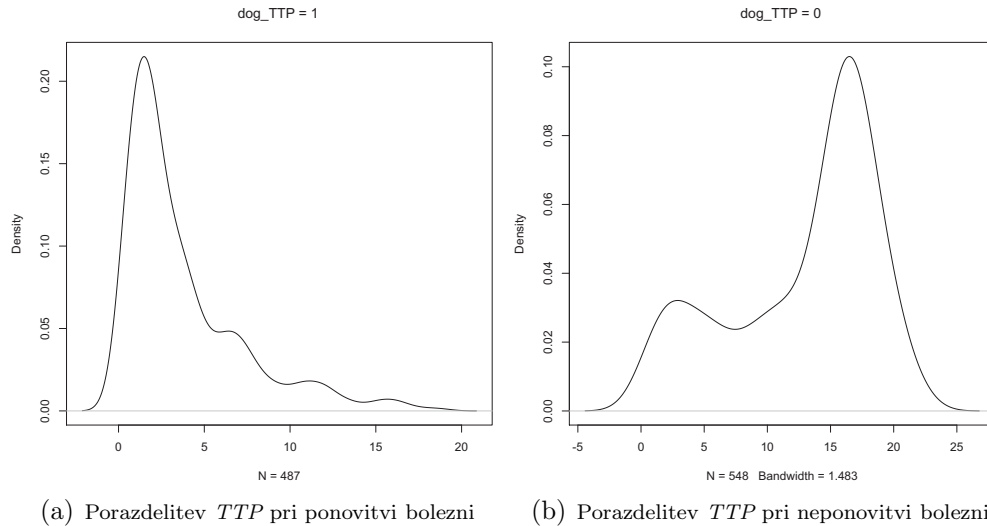
Grafa na sliki 6.1 prikazujeta gostoto porazdelitve spremenljivke *TTP* pri pacientkah s ponovitvijo boleznii (6.2(a)) in pri pacientkah brez ponovitve (6.2(b)).

Iz grafa 6.2(a) vidimo, da prihaja do ponovitve boleznii v času do 20 let po zdravljenju, verjetnost za ponovitev boleznii pa izstopa v lokalnih maksimumih krivulje, ki so v približno 3., 7. in 11. letu po zdravljenju.

6.2 Izbira napovednega modela

Preslikava učnih podatkov. Iz opisa podatkov in cilja raziskave vidimo, da je problem napovedovanja pravzaprav dvoplasten:

1. Napovedati je potrebno, ali bo prišlo do ponovitve boleznii ali ne.

Slika 6.1: Porazdelitev odvisne spremenljivke TTP v primeru ponovitve in neponovitve bolezni.

2. V primeru napovedi, da bo prišlo do ponovitve bolezni, je potrebno napovedati čas ponovitve.

Za reševanje te težave smo uporabili naslednje možne pristope:

1. gradnja regresijskega modela na množici podatkov, ki smo ji v primeru neponovitve bolezni nastavili vrednost označbe TTP na enotno, vnaprej določeno vrednost, ki je izven pričakovanega intervala ponovitve bolezni (neponovitev bolezni smo izenačili s ponovitvijo po x letih, $x \in \{21, 50\}$). Pri napovedovanju smo uporabili le odvisno spremenljivko TTP , dog_TTP smo odstranili;
2. uporaba dvonivojskega sistema klasifikator-regresor. Klasifikator je bil namenjen napovedovanju odvisne spremenljivke dog_TTP , torej, ali do ponovitve bolezni pride, in, v primeru pozitivnega odgovora, regresor za napovedovanje časa ponovitve bolezni, torej spremenljivke TTP ;
3. uporaba regresijskega modela na učni množici z dinamično določitvijo označb za primere brez ponovitve bolezni. V prvem koraku definiranja spremenjene učne množice smo primere brez ponovitve bolezni obravnavali kot neoznačene učne primere [Dempster in sod., 1977] tako, da smo jim z uporabo lokalno

utežene regresije na učnih primerih s ponovitvijo bolezni napovedali vrednost odvisne spremenljivke. Nato smo s prištevanjem 21 let vrednosti teh napovedi premaknili preko meje zaloge vrednosti spremenljivke TTP in s tem definirali te primere kot primere brez ponovitve bolezni, upoštevajoč medsebojne razvrstitve njihovih regresijskih napovedi.

Še pred uporabo ocen zanesljivosti smo za vsakega od navedenih pristopov z uporabo desetkratnega prečnega preverjanja ocenili točnost regresijskih modelov. Ocene točnosti RMSE so prikazane v tabeli 6.1. Iz tabele lahko vidimo, da smo z izjemo pri regresijskih drevesih najmanjšo napako pri napovedi dosegli pri pristopu z dinamično določitvijo označb primerov, pri katerih do ponovitve bolezni ni prišlo. Na podlagi tega smo se odločili ta pristop uporabiti tudi pri nadaljevanju raziskave.

Tabela 6.1: Točnost (RMSE - relativna srednja kvadratna napaka) regresijskih modelov pri različnih pristopih oblikovanja učne množice.

	RMSE				
	neponovitev = labela 50	neponovitev = labela 21	dvofazni sistem		dinamična določitev label
			klasifikacijska točnost dre- vesa	točnost regresorja	
RT	0.885	0.863	0.640	1.154	0.867
LR	0.877	0.857		1.035	0.816
NN	2.059	1.998		0.916	1.910
BAG	0.852	0.840		0.953	0.799
SVM	1.011	0.952		1.005	0.896
LWR	0.884	0.852		0.949	0.825
RF	0.864	0.867		0.934	0.814
GAM	0.878	0.859		0.994	0.829

Izbira regresijskega modela in ocene zanesljivosti. Čeprav smo pri uporabi opisanega pristopa najnižjo RMSE napako dosegli pri modelu bagging in modelu naključnih gozdov, ju zaradi prevelike časovne zahtevnosti pri izračunu ocen zanesljivosti nismo uporabili (izračun ocen zanesljivosti za vsakega od teh modelov bi trajal približno 48 ur na procesorju Intel Core2 Duo 6600 2.40 GHz). Med naslednjima najboljšima modeloma, linearno regresijo in lokalno uteženo regresijo, ki imata primerljivi napaki, smo se odločili za uporabo lokalno utežene regresije.

Pred nadaljnjim postopkom dela s podatki smo iz množice 1035 pacientk izločili 15 primerov in jih shranili za kasnejše ovrednotenje uporabljene metodologije, ki smo ga izvedli v sodelovanju s strokovnjaki področja. Na preostalih podatkih (učni množici) smo nato, enako kot pri testiranju ocen zanesljivosti, z metodo izloči-enega izračunali napake primerov in vrednosti 9 ocen zanesljivosti, prikazanih v tabeli 4.1. Iz analize korelacij ocen zanesljivosti z napako, prikazane v tabeli 6.2, smo ugotovili, da z napako značilno pozitivno korelira le ocena *BVCK*, kar nam je dalo vedeti, da spada problemska domena med težavne v smislu ocenjevanja zanesljivosti, kot je to prikazano v tabeli 5.2. Ker je metaprediktor ravno tako kot najbolj primerno oceno napovedal oceno *BVCK* in ker je izbor ocene z najvišjim korelacijskim koeficientom na učnih podatkih v skladu tudi s postopkom notranjega prečnega preverjanja, smo se to oceno odločili uporabiti pri testiranju.

Tabela 6.2: Odstotek značilnih pozitivnih in negativnih korelacij med napako pri napovedi in oceno zanesljivosti, ki je bila napovedana z notranjim prečnim preverjanjem.

	korelacijski koeficient	stopnja značilnosti
SAvar	-0.046	0.140
SAbias-p	0.033	0.298
SAbias-a	0.001	0.985
BAGV	0.059	0.060
LCV	0.008	0.793
DENS	-0.074	0.018
CNK-p	-0.042	0.179
CNK-a	0.034	0.282
BVCK	0.071	0.024

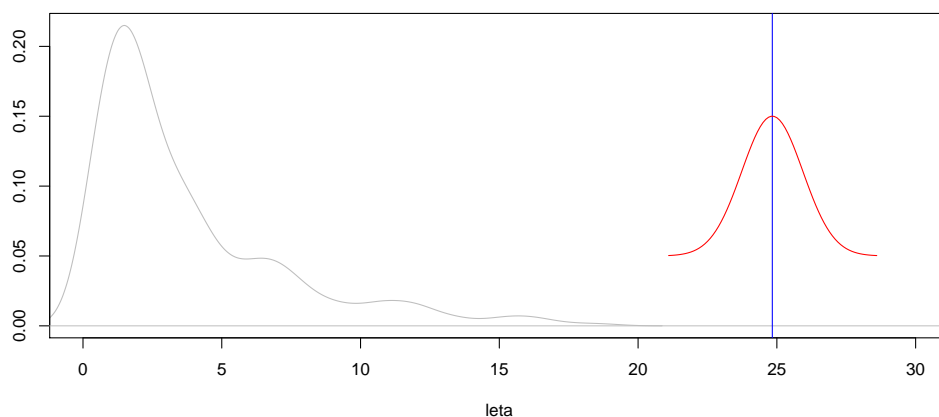
6.3 Testiranje ocen zanesljivosti v praksi

V okviru praktičnega testiranja ocenjevanja zanesljivosti smo za 15 primerov iz testne množice napovedali čase ponovitve bolezni skupaj z vrednostmi ocene *BVCK*. Te smo nato za vsakega izmed pacientov grafično predstavili, kot je to prikazano na sliki 6.2. Ocene zanesljivosti smo grafično prikazali kot zvončaste krivulje s sredino v vrednosti

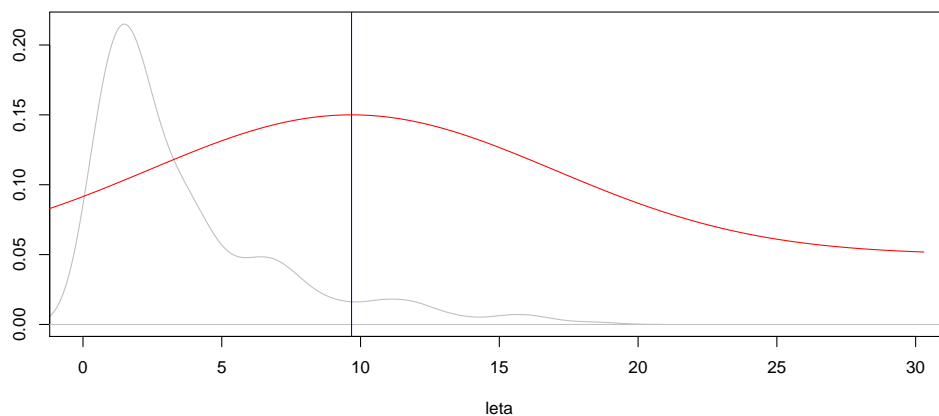
6. APLIKACIJA OCEN ZANESLJIVOSTI NA REALNI DOMENI

napovedi, katerih širina predstavlja manjšo (široka krivulja) ali visoko (ozka krivulja) zanesljivost.

Slika 6.2: Grafični prikaz napovedi časa ponovitve (navpična črta) in ocene zanesljivosti (širina zvončaste funkcije okoli napovedi).



(a) Napoved ponovitve bolezni z nizko oceno napake



(b) Napoved ponovitve bolezni z visoko oceno napake

Če vrednost napovedi označimo s K , smo obliko krivulje zanesljivosti definirali kot funkcijo normalne porazdelitve s parametroma $N(K, \sigma)$, kjer smo σ definirali s preslikavo ocene zanesljivosti v standardizirano normalno porazdelitev $N(0, 1)$, nato pa s preslikavo te v porazdelitev napake primerov v učni množici:

$$\sigma = \frac{O - \text{avg}(O)}{\text{sd}(O)} \cdot \text{sd}_{i \in L}(C_i - K_i) + \text{avg}_{i \in L}(C_i - K_i) \quad (6.1)$$

V zgornji enačbi predstavlja O vrednost ocene zanesljivosti danega testnega primera, $avg()$ funkcijo povprečne vrednosti, $sd()$ funkcijo standardne deviacije, L pa učno množico primerov.

Grafične napovedi za 15 primerov smo skupaj z njihovimi atributi (brez prave vrednosti označbe) posredovali ekspertom področja in jih prosili, naj na lestvici od -5 do 5 ocenijo njihovo strinjanje z napovedmi, pri čemer naj bi ocena -5 pomenila največje nestrinjanje, 5 pa največje strinjanje. Njihove povratne informacije, vrednosti ocen zanesljivosti in dejansko napako pri napovedi prikazuje tabela 6.3.

Analiza prikazanih rezultatov je pokazala, da je višina korelacije ocene točnosti napovedi ekspertov s pravo napako napovedi enaka -0.267 (pričakujemo negativno korelacijo, saj je ocena točnosti napovedi obratna oceni napake; korelacija ocene napake eksperta z napako pri napovedi je torej 0.267) pri stopnji značilnosti 0.336 , višina korelacije ocene *BVCK* z napako pa 0.197 (pričakujemo pozitivno korelacijo, saj z oceno zanesljivosti ocenjujemo napako) pri stopnji značilnosti 0.483 . S stališča višine in značilnosti korelacijskih koeficientov torej ne moremo sklepati, da sta ocena ekspertov in ocena *BVCK* kakovostni oceni napake napovedi. Vendar pa, če primerjamo statistično razliko obeh korelacijskih koeficientov [Blalock, 1972], ugotovimo, da sta koeficienta, izračunana na 15 parov vrednosti, statistično enaka s stopnjo značilnosti 0.857 , kar pomeni, da med njima statistično gledano ni razlike.

Čeprav ocena *BVCK* ni dosegla značilne korelacije, dobljen rezultat pomeni, da enako dobro ocenjuje zanesljivost napovedi kot eksperti. Če se omejimo na ta aspekt, lahko metodo v tem kontekstu označimo kot uspešno.

6. APLIKACIJA OCEN ZANESLJIVOSTI NA REALNI DOMENI

Tabela 6.3: Primerjava ocen točnosti napovedi s strani ekspertov z ocenami zanesljivosti in napako pri napovedi za 15 testnih pacientov.

primer	ocena eksperta	ocena BVCK	napaka pri napovedi
1	5	1.29	4.37
2	3	6.84	14.75
3	5	1.64	5.07
4	-5	3.75	11.79
5	-5	3.50	5.43
6	5	3.40	9.44
7	0	2.41	8.53
8	3	4.33	4.69
9	-5	4.47	17.23
10	2	5.53	6.47
11	2	3.80	4.37
12	3	1.41	17.20
13	-2	4.77	9.68
14	5	5.84	9.96
15	3	2.34	9.70

Poglavje 7

Diskusija in zaključek

7.1 Pregled metodologije in rezultatov

Ocenjevanje zanesljivosti posameznih napovedi je področje, ki je lahko pri odločanju kritičnega pomena. Obstoječe metode za ocenjevanje zanesljivosti pristopajo k tej nalogi z ocenjevanjem točnosti modela kot celote, medtem ko je ostajalo področje ocenjevanja posameznih napovedi pretežno neraziskano. Čeprav so izboljšave pogosto uporabljenih regresijskih modelov omogočile razširitev informacije o napovedi tudi z informacijo o njeni zanesljivosti, je bila zasnova splošne metode, ki ni vezana na določen regresijski model, raziskovalni izziv.

V okviru disertacije smo razvili različne ocene zanesljivosti, ki so zasnovane z upo-

rabo različnih pristopov. Kot novost na področju zasnove ocen zanesljivosti smo zasnovali in uporabili pristop z analizo občutljivosti, ki je postopek, neodvisen od učnega algoritma oziroma regresijskega modela. Z obravnavo regresijskega prediktorja kot mehanizma črne škatle smo s kontroliranim spreminjanjem učne množice opazovali spremembe v izhodnih napovedih. Na ta način smo zasnovali pojem *občutljivostnih napovedi*, ki smo jih v nadaljnjih korakih uporabili za zasnovo ocen zanesljivosti *SAbias* in *SAvar*, ki opisujeta lokalno pristranost in lokalno varianco. Z empirično analizo smo pokazali, da tehnika analize občutljivosti ni uspešna za uporabo z vsemi regresijskimi modeli. Pokazali smo tudi kriterije za presojo uspešnosti metode glede na dani regresijski model. Poleg ocen, zasnovanih z analizo občutljivosti, smo za primerjavo implementirali še 4 druge ocene (*BAGV*, *LCV*, *DENS* in *CNK*). Ker oceni *SAbias* in *CNK* nosita tudi informacijo o smeri ocenjene napake, smo ti dve oceni zasnovali v predznačeni (*SAbias-p* in *CNK-p*) in absolutni (*SAbias-a* in *CNK-a*) različici.

Rezultati testiranj na 28 domenah in 8 regresijskih modelih so v skladu s pričakovanji pokazali nadvlado števila značilnih pozitivnih korelacij ocen zanesljivosti z napako napovedi nad številom značilnih negativnih korelacij. V povprečju so se kot najboljše ocene izkazale (v tem vrstnem redu): *BAGV*, *CNK-p*, *LCV* in *SAvar*. V nadaljevanju smo pokazali, da z linearnim kombiniranjem ocen ne moremo doseči bistvene izboljšave števila pozitivnih korelacij, istočasno pa smo s postopkom kombiniranja razvili oceno *BVCK*, ki je po uspehu enakovredna najboljši oceni *BAGV* in jo pridružili naboru obstoječih ocen. Povzemamo lastnosti uporabljenih ocen zanesljivosti:

- *SAvar* na podlagi analize občutljivosti opisuje varianco napovedi, pridobljenih z različnimi občutljivostnimi modeli. Prevladujoče število značilnih pozitivnih korelacij smo z uporabo te ocene dosegli pri linearni regresiji in posplošenem aditivnem modelu (54%).
- *SAbias* meri pristranost občutljivostnih napovedi glede na začetno napoved. Izstopajoče pozitivne rezultate je dosegla le različica *SAbias-p*, in sicer pri regresijskih drevesih (82% značilnih pozitivnih korelacij).
- *BAGV* meri varianco napovedi posameznih prediktorjev v agregatu, od nevron-

skih mrež smo oceno posplošili za uporabo tudi z drugimi regresijskimi modeli. V povprečju je ocena glede na ostale dosegla najboljše rezultate. Primerjave ocen zanesljivosti kažejo, da je po številu značilnih pozitivnih korelacij (46%) *BAGV* najbolj primerna ocena zanesljivosti pri lokalno uteženi regresiji.

- *LCV* je ocena, ki meri lokalno napako z uporabo lokalnega prečnega preverjanja. V primerjavi z ostalimi ocenami je dosegla najvišja števila značilnih pozitivnih korelacij pri uporabi z metodo podpornih vektorjev (61%), z lokalno uteženo regresijo (46%) in z naključnimi gozdovi (61%).
- *DENS* je prilagojena in obrnjena ocena gostote prostora, ki pri izračunu gostote uporablja redukcijo razsežnosti. Ocena ni pri nobenem regresijskem modelu dala izstopajočih rezultatov glede na število prevladujočih značilnih pozitivnih korelacij.
- *CNK* je novozasnovana ocena, ki ocenjuje napako napovedi na podlagi lokalne informacije v okolici učnega primera. Njena različica *CNK-p* je v primerjavi z drugimi ocenami dosegla najboljše rezultate pri regresijskih drevesih (86% značilnih pozitivnih korelacij), različica *CNK-a* pa pri linearni regresiji in posplošenem aditivnem modelu (pri obeh 57% značilnih pozitivnih korelacij).
- *BVCK* je ocena, pridobljena s povprečenjem ocen *BAGV* in *CNK-a*. S kombiniranjem teh dveh ocen je ocena *BVCK* glede na število značilnih pozitivnih korelacij dosegla boljše rezultate pri nevronskih mrežah (54%) in modelu bagging (61%). Rezultati tudi kažejo, da je za ta dva modela *BVCK* najbolj primerna ocena zanesljivosti.

Z obravnavanjem dinamične izbire najbolj primerne ocene, z metaučenjem v prostoru domen in modelov ter z notranjim prečnim preverjanjem, smo pokazali možnost uporabe dveh pristopov, s katerima lahko dosežemo izboljšanje števila značilnih pozitivnih korelacij med ocenami zanesljivosti in napakami napovedi. S tehniko metaučenja smo tudi analizirali kriterije za napovedovanje ocen, ki jih uporablja metaklasifikator, in ugotovili, da igra pri izbiri ocene zanesljivosti največjo vlogo ocena uspeha celotnega regresijskega modela na učnih podatkih (RMSE), ki ji sledijo lokalne mere, ki opisujejo problemski prostor. Z analizo pokritosti prostora do-

mena/model z uspešnostjo ocen smo pokazali, da so si podmnožice ocen med seboj redundantne. Z odstranitvijo ocen, ki ne prispevajo k večjemu številu značilnih pozitivnih korelacij v prostoru domena/model, smo pri poskusih dosegli izboljšanje povprečnih rezultatov metaučenja. Najboljše rezultate smo dosegli pri samodejni izbiri optimalne ocene s postopkom notranjega prečnega preverjanja, kjer so ocene zanesljivosti značilno pozitivno korelirale z napako v povprečju pri 73% domen in značilno negativno v 0% domen.

Rezultati, ki za posamezne regresijske modele izmed vseh opravljenih testiranj najbolj izstopajo, so prikazani v tabeli 7.1. Iz povzetka rezultatov lahko vidimo, da najboljšega rezultata pri posameznih regresijskih modelov ni dosegla nobena samostojna ocena, ampak da so bili ti doseženi bodisi z uporabo linearnega kombiniranja ocen bodisi s postopkom dinamične izbire. Vidimo tudi, da izbor najboljših metod v povprečju pozitivno korelira v 73% primerov (natančneje 73.4%), kar je približno enakovredno rezultatu, ki smo ga dobili z notranjim prečnim preverjanjem (natančneje 72.5%). Na tej podlagi lahko pristop notranjega prečnega preverjanja označimo kot obetavno metodo dinamične izbire ocene zanesljivosti, ki potrjuje, da je takšen pristop kombiniranja ocen smiseln.

Pri aplikaciji metode na področju onkoloških podatkov smo obravnavali uporabnost metode v realni domeni. Ugotovili smo, da spada domena med bolj težavne tako na področju napovedovanja kot tudi ocenjevanja zanesljivosti napovedi. Točnost napovedovanja je v domeni pacientov namreč omejena z nepopolnim poznavanjem prognostičnih faktorjev za ponovitev bolezni, ravno tako pa je z napovedovanjem potrebno reševati dvojni problem (ali do ponovitve bolezni sploh pride; če da, po kolikšnem času). Domena je ravno tako težavna tudi na področju ocenjevanja zanesljivosti, saj je z napako značilno korelirala le ocena zanesljivosti *BVCK*. Z uporabo te ocene, ki je bila samodejno predlagana tudi s pristopoma metaučenja in notranjega prečnega preverjanja, smo ovrednotili točnost napovedi ponovitve bolezni. S primerjavo povratnih informacij ekspertov z oceno zanesljivosti smo ugotovili, da se le-ti značilno ne razlikujeta, kar kaže na potencial uporabe metodologije v praksi.

Tabela 7.1: Povzetek najbolj uspešnih ocen pri posameznih regresijskih modelih. Prikazan je odstotek značilnih pozitivnih in negativnih korelacij in ocene, s katerimi smo dosegli navedene rezultate. Okrajšave Mxx predstavljajo kombinacijo ocen xx (glej tabelo 5.4), obravnavano pri metaučenju, okrajšava NPP pa predstavlja izbiro ocene s postopkom notranjega prečnega preverjanja.

model	+/-	ocene
RT	89/0	(<i>SAbias-p</i> + <i>CNK-p</i>), M10, M6
LR	73/0	NPP
NN	73/0	NPP
BAG	68/0	M13
SVM	67/0	NPP
LWR	73/0	NPP
RF	64/0	M14
GAM	80/0	NPP
povprečno	73/0	

7.2 Časovne zahtevnosti algoritmov

Pomemben aspekt medsebojne primerjave ocen zanesljivosti je tudi glede na njihovo časovno zahtevnost. V nadaljevanju podajamo oceno časovne zahtevnosti izračuna ocen zanesljivosti za en testni primer. Pri tem uporabljamo naslednjo notacijo:

- n predstavlja število učnih primerov,
- a predstavlja število atributov,
- k je število najbližjih sosedov,
- p je število prediktorjev v agregatu modela bagging,
- m predstavlja število uporabljenih regresijskih modelov,
- $N = \{n_1, n_2, \dots, n_{|N|}\}$ je množica s števili primerov v uporabljenih testnih domenah,
- $A = \{a_1, a_2, \dots, a_{|N|}\}$ je množica s števili atributov v uporabljenih testnih domenah; i -ti domeni pripadata istoležni števili primerov $n_i \in N$ in atributov $a_i \in A$,

- $E = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{|E|}\}$ je množica uporabljenih vrednosti parametrov ε ,
- $M(n_u, a; n_t)$ predstavlja časovno zahtevnost gradnje poljubnega regresijskega modela na učni množici, ki jo sestavlja n_u učnih primerov z a atributi, in izračun vrednosti napovedi n_t testnim primerom,
- z $O(n^2a)$ podajamo red časovne zahtevnosti izračuna medsebojnih razdalj za n primerov (torej $n \times n$ razdalj),
- c_X predstavlja časovno zahtevnost manj zahtevnih konstantnih operacij, potrebnih za izračun ocene X , ki vključujejo predvsem množenje, seštevanje ali preslikavo (uporaba jedrne funkcije) posameznih komponent v enačbi.

Z uporabo gornje notacije podajamo oceno časovnih zahtevnosti izračuna ocen zanesljivosti za en testni primer. Poleg časovnih zahtevnosti navajamo tudi podatke o dejanskem času izračuna na primeru domene z 294 primeri in 13 atributi za linearno regresijo, regresijska drevesa, lokalno uteženo regresijo in za model bagging (časovno najbolj zahteven model). Čas izvajanja je bil merjen na računalniku s procesorjem *Intel Core2 Duo 6600 2.40 GHz* in *1GB pomnilnika*.

SAvar. Časovno zahtevnost izračuna ocene za en primer podamo z:

$$SAvar(E, n, a) = 2|E| \cdot M(n + 1, a; 1) + c_{SAvar} \quad (7.1)$$

Pri našem testiranju smo uporabili $|E| = 5$.

Časi izvajanja: LR 0.199 s, RT 0.467 s, LWR 0.660 s, BAG 4.922 s.

SAbias. Časovno zahtevnost izračuna ocene za en primer podamo z:

$$SAbias(E, n, a) = 2|E| \cdot M(n + 1, a; 1) + M(n - 1, a; 1) + c_{SAbias} \quad (7.2)$$

Pri našem testiranju smo uporabili $|E| = 5$.

Časi izvajanja: LR 0.198 s, RT 0.484 s, LWR 0.644 s, BAG 4.583 s.

BAGV. Časovno zahtevnost izračuna ocene za en primer podamo z:

$$BAGV(p, n, a) = p \cdot M(n, a; 1) + c_{BAGV} \quad (7.3)$$

Pri našem testiranju smo uporabili $p = 50$.

Časi izvajanja: LR 0.942 s, RT 1.965 s, LWR 2.976 s, BAG 21.889 s.

LCV. Časovno zahtevnost izračuna ocene za en primer podamo z:

$$LCV(k, n, a) = O(n^2a) + k \cdot M(k - 1, a; 1) + c_{LCV} \quad (7.4)$$

Pri našem testiranju smo uporabili $k = \max(5, \frac{n}{20})$.

Časi izvajanja: LR 0.255 s, RT 0.339 s, LWR 0.656 s, BAG 3.338 s.

DENS. Časovno zahtevnost izračuna ocene za en primer podamo z:

$$DENS(k, n, a) = O(n^2a) + c_{DENS} \quad (7.5)$$

Pri našem testiranju smo uporabili $k = n$.

Časi izvajanja: LR 0.286 s, RT 0.4 s, LWR 0.683 s, BAG 3.393 s.

CNK. Časovno zahtevnost izračuna ocene za en primer podamo z:

$$CNK(k, n, a) = O(n^2a) + M(n, a; 1) + c_{CNK} \quad (7.6)$$

7. DISKUSIJA IN ZAKLJUČEK

Pri našem testiranju smo uporabili $k = 5$.

Časi izvajanja: LR 0.138 s, RT 0.165 s, LWR 0.57 s, BAG 0.912 s.

BVCK. Časovno zahtevnost izračuna ocene za en primer podamo z:

$$BVCK(n, k, p, a) = O(n^2 a) + (p + 1)M(n, a; 1) + c_{BVCK} \quad (7.7)$$

Pri našem testiranju smo uporabili $p = 50$ in $k = 5$.

Časi izvajanja: LR 1.049 s, RT 2.141 s, LWR 3.503 s, BAG 22.052 s.

METAUČENJE. Časovna zahtevnost izračuna ocene za en primer:

Izračun učne množice za metaprediktor:

$$\begin{aligned} META_u(k, m, N = \{n_i\}, A = \{a_i\}) = \\ m|N| \cdot \left[10 \cdot M\left(\frac{9n_i}{10}, a_i; \frac{n_i}{10}\right) + n_i \cdot DENS(k, n_i, a_i) + \right. \\ \left. + O(n_i^2 a_i) + (k + 1) \cdot M(n_i - 1, a_i; 1) + \right. \\ \left. + (\text{izračun ocen zanesljivosti za } p_i \text{ primerov}) \right] \\ + c_{META_u} \approx \\ \approx m|N|(4|E| + 2p + 2k + 4)M(n_i, a_i, 1) \end{aligned} \quad (7.8)$$

Izračun izbrane ocene za učni primer:

$$\begin{aligned} META_p(k, m, N = \{n_i\}, A = \{a_i\}) = \\ M(m|N|, 7; 1) + (\text{izračun ocene za primer}) + c_{META_p} \end{aligned} \quad (7.9)$$

V okviru našega testiranja je veljalo $m = 8$ in $|N| = 28$.

Časi izvajanja: izračun učne množice 181.45 s, izračun ocene 0.035 s za *napoved* + čas izračuna *napovedane ocene*.

NOTRANJE PREČNO PREVERJANJE. Časovno zahtevnost izračuna ocene za en primer podamo z:

$$\begin{aligned}
 NPP(n, a) &= nM(n-1, a; 1) + \\
 &+ (\text{izračunaj vse ocene za učno množico}) + \\
 &+ (\text{izračun ocene za primer}) + c_{NPP} \approx \\
 &\approx (4|E| + 2p + k + n + 3)M(n, a; 1)
 \end{aligned}
 \tag{7.10}$$

Časi izvajanja: LR 3515.54 s (0.98 h), RT 7167.52 s (1.99 h), LWR 9834.83 s (2.73 h), BAG 76461.84 s (21.24 h).

Iz časovnih zahtevnosti in dejanskih časov izvajanj vidimo, da lahko posamezne ocene zanesljivosti izračunamo v realnem času. Kljub večji časovni zahtevnosti gradnje učne množice za metaprediktor in izbora ocene pri notranjem prečnem preverjanju pa lahko najbolj zahtevne časovne operacije izvedemo kot predpripravo in izven okvirov izvajalnega algoritma za izračun ocene zanesljivosti posameznega testnega primera. Na ta način lahko oceno zanesljivosti za testni primer tudi pri teh dveh pristopih izračunamo v realnem času, kar jima daje praktično uporabnost.

7.3 Nadaljnje delo

Ideje za nadaljnje delo vključujejo zlasti:

- povečanje interpretabilnosti in razumljivosti posameznih ocen zanesljivosti, saj ima vsaka od predstavljenih ocen sebi lastno zalogo vrednosti. Na ta način različnih ocen med seboj in med različnimi domenami ni možno primerjati. S potencialno preslikavo na interval $[0, 1]$ bi lahko dosegli večjo primerljivost vrednosti ocen in morda tudi verjetnostno interpretacijo zanesljivosti;

7. DISKUSIJA IN ZAKLJUČEK

- raziskati bi bilo potrebno, ali lahko z vrednostmi ocen zanesljivosti, ki korelirajo s predznačeno napako, dosežemo tudi popravke prvotnih napovedi regresijskega modela in s tem izboljšavo njegove točnosti;
- organizacija razvitih algoritmov v okolju R v modularno knjižnico za okolje R in prispevanje paketa v javno dostopno zbirko (*contributed packages*);
- razvoj pristopa za samodejno izbiro optimalne ocene, ki bo vezana na posamezni učni primer. Izziv pri tem cilju izvira zlasti iz težavnosti kombiniranja vrednosti ocen zanesljivosti, ki imajo različne zaloge vrednosti, znotraj ene problemske domene;
- testiranje delovanje razvite metodologije tudi na drugih realnih domenah.

Literatura

Alpaydin, E. *Introduction to machine Learning*. The MIT Press: Cambridge, Massachusetts, 2004.

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

Baluja, S. Probabilistic modeling for face orientation discrimination: Learning from labeled and unlabeled data. In *Neural Information Processing systems (NIPS '98)*, 1998.

Birattari, M., H. Bontempi, H. Bersini. Local learning for data analysis. In *Proceedings of the 8th Belgian-Dutch Conference on Machine Learning*, pages 55–61, 1998.

Blalock, H. M. *Social Statistics*, pages 406–407. McGraw-Hill: NY, 1972.

Blum, A., T. Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the 11th Annual Conference on Computational Learning Theory*, pages 92–100, 1998.

Bosnić, Z. Ocenjevanje zanesljivosti predikcije pri regresiji, diplomska naloga. Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Ljubljana, 2001.

Bosnić, Z. Uporaba transdukcije pri regresijskem napovedovanju, magistrska naloga. Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Ljubljana, 2003.

- Bosnić, Z., I. Kononenko. Estimation of individual prediction reliability using the local sensitivity analysis. *Applied Intelligence, v tisku*, 2007a.
- Bosnić, Z., I. Kononenko. Estimation of prediction reliability in regression based on a transductive approach. In *Proceedings of IICAI 2005, Pune, India*, pages 3502–3516, 2005.
- Bosnić, Z., I. Kononenko. Estimation of regressor reliability. *Journal of Intelligent Systems, v tisku*, 2007b.
- Bosnić, Z., I. Kononenko, M. Robnik-Šikonja, M. Kukar. Evaluation of prediction reliability in regression using the transduction principle. In Zajc, Baldimir, M. Tkalčič, editors, *Proceedings of Eurocon 2003*, pages 99–103, Ljubljana, 2003.
- Bousquet, O., A. Elisseeff. Algorithmic stability and generalization performance. In *NIPS*, pages 196–202, 2000.
- Bousquet, O., A. Elisseeff. Stability and generalization. In *Journal of Machine Learning Research*, volume 2, pages 499–526, 2002.
- Bousquet, O., M. Pontil. Leave-one-out error and stability of learning algorithms with applications. In al., J.A.K. Suykens, editor, *Advances in Learning Theory: Methods, Models and Applications*. IOS Press, 2003.
- Bratko, I. *Prolog programming for artificial intelligence*. Addison-Wesley Publishers Limited, third edition edition, 2001. ISBN 0201-40375-7.
- Breierova, L., M. Choudhari. An introduction to sensitivity analysis. MIT System Dynamics in Education Project, september 1996.
- Breiman, L. Pasting bites together for prediction in large data sets and on-line. Univ. of Calif., Berkeley, Dept. of Statistics technical report, 1997.
- Breiman, L. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.
- Breiman, L. Random forests. *Machine Learning*, 45(1):5–32, 2001.

- Breiman, L., J. H. Friedman, R. A. Olshen, C. J. Stone. *Classification and Regression Trees*. Wadsworth International Group, Belmont CA, 1984.
- Brierley, P., B. Batty. Data mining with neural networks - an applied example in understanding electricity consumption patterns. *Knowledge Discovery and Data Mining*, ed. M. Bramer, pages 140–303, 1999.
- Burges, C. J. C. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.
- Carney, J., P. Cunningham. Confidence and prediction intervals for neural network ensembles. In *Proceedings of IJCNN'99, The International Joint Conference on Neural Networks, Washington, USA*, pages 1215–1218, 1999.
- Chambers, J. M., T. J. Hastie, editors. *Statistical Models in S*, chapter Chapter 4, Linear models. Wadsworth & Brooks/Cole, 1992.
- Chang, C., C. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Christiannini, N., J. Shawe-Taylor. *Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.
- Cleveland, W., C. Loader. Smoothing by local regression: Principles and methods. Technical Report, AT&T Bell Laboratories, Murray Hill, NY., 1995.
- Cohn, D. A., Z. Ghahramani, M. I. Jordan. Active learning with statistical models. In Tesauro, G., D. Touretzky, T. Leen, editors, *Advances in Neural Information Processing Systems*, volume 7, pages 705–712. The MIT Press, 1995. URL citeseer.ist.psu.edu/cohn95active.html.
- Crowder, M. J., A. C. Kimber, R. L. Smith, T. J. Sweeting. *Statistical concepts in reliability. Statistical Analysis of Reliability Data*. Chapman & Hall, London, UK, 1991.

- Sa, V.de. Learning classification with unlabeled data. In Cowan, Jack D., Gerald Tesauero, Joshua Alspecter, editors, *Proc. NIPS'93, Neural Information Processing Systems*, pages 112–119, San Francisco, CA, 1993. Morgan Kaufmann Publishers.
- Dempster, A. P., N. M. Laird, D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39 (1):1–38, 1977.
- Dobnikar, A. *Nevrnske mreže: Teorija in aplikacije*. Didakta, Radovljica, 1990.
- Drucker, H. Improving regressors using boosting techniques. In *Proc. 14th International Conference on Machine Learning*, pages 107–115. Morgan Kaufmann, 1997.
- Elidan, G., M. Ninio, N. Friedman, D. Schuurmans. Data perturbation for escaping local maxima in learning, 2002.
- Gammerman, A., V. Vovk, V. Vapnik. Learning by transduction. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 148–155, Madison, Wisconsin, 1998.
- Giacinto, G., F. Roli. Dynamic classifier selection based on multiple classifier behaviour. *Pattern Recognition*, 34(9):1879–1881, 2001.
- Goldman, S., Y. Zhou. Enhancing supervised learning with unlabeled data. In *Proc. 17th International Conf. on Machine Learning*, pages 327–334. Morgan Kaufmann, San Francisco, CA, 2000.
- Härdle, W., M. Müller, S. Sperlich, A. Werwatz. *Nonparametric and Semiparametric Models*, chapter An Introduction, pages 1–18. Springer, 2004.
- Hashem, H. Sensitivity analysis for feedforward artificial neural networks with differentiable activation functions. In *Proceedings of 1992 International Joint Conference on Neural Networks IJCNN92*, volume I, pages 419–424, 1992.
- Hastie, T., R. Tibshirani. *Generalized Additive Models*. Chapman and Hall, London, 1990.

- Hastie, T., R. Tibshirani, J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2001.
- Heskes, T. Practical confidence and prediction intervals. In Mozer, Michael C., Michael I. Jordan, Thomas Petsche, editors, *Advances in Neural Information Processing Systems*, volume 9, pages 176–182. The MIT Press, 1997.
- Jamnik, R. *Verjetnostni račun in statistika*. Društvo matematikov, fizikov in astronomov Slovenije, Ljubljana, 1995.
- Jeon, B., D. A. Landgrebe. Parzen density estimation using clustering-based branch and bound. *Trans. Pattern Anal. Mach. Intell.*, pages 950–954, 1994.
- Karalič, A. Employing linear regression in regression tree leaves. In Neumann, B., editor, *Proceedings of European Conference on Artificial Intelligence (ECAI'92)*, pages 440–441. John Wiley & Sons, 1992.
- Kearns, M. J., D. Ron. Algorithmic stability and sanity-check bounds for leave-one-out cross-validation. In *Computational Learning Theory*, pages 152–162, 1997.
- Kjaerulff, U., L. C. van der Gaag. Making sensitivity analysis computationally efficient. *Submitted to UAI 2000*, 2000.
- Kleijnen, J. Experimental designs for sensitivity analysis of simulation models. *Tutorial at the Eurosim 2001 Conference*, 2001.
- Kononenko, I. *Strojno učenje. 2. popravljena in dopolnjena izdaja*. Fakulteta za računalništvo in informatiko, Ljubljana, 2005. ISBN 961-6209-52-3.
- Kononenko, I. Bayesove nevronske mreže, disertacija. Univerza v Ljubljani, Fakulteta za elektrotehniko in računalništvo, Ljubljana, 1990.
- Kononenko, I., M. Kukar. *Machine Learning and Data Mining: Introduction to Principles and Algorithms*. Horwood Publishing Limited, UK, 2007. ISBN 1-901275-21-4.

- Kukar, M. Evaluation of classification reliability and cost sensitive combining of machine learning methods, disertacija. Univerza v Ljubljani, Fakulteta za elektrotehniko in računalništvo, Ljubljana, 2001.
- Kukar, M., I. Kononenko. Reliable classifications with machine learning. In Elomaa, T., H. Manilla, H. Toivonen, editors, *Proc. Machine Learning: ECML-2002*, pages 219–231. Springer Verlag, Helsinki Finland, 2002.
- Li, F., H. Wechsler. Open set face recognition using transduction. *IEEE transactions on pattern analysis and machine intelligence*, 27(11):1686–1697, 2005.
- Li, M., P. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications*. Springer-Verlag, New York, 1993.
- Linden, A., F. Weber. Implementing inner drive by competence reflection. In *Proceedings of the 2nd International Conference on Simulation of Adaptive Behavior, Hawaii*, pages 321–326, 1992.
- McCulloch, W. S., W. Pitts. A logical calculus of the ideas imminent in nervous activity. *Bull. of Math. and Biophysics*, 5:115–133, 1943.
- Michie, D., D. J. Spiegelhalter, C. C. Taylor. *Machine Learning, Neural and Statistical Classification*. Ellis Horwood, 1994. ISBN 0-13-106360-X.
- Mitchell, T. *Machine Learning*. McGraw-Hill, New York, 1997. ISBN 0-07-042807-7.
- Mitchell, T. The role of unlabelled data in supervised learning. In *Proceedings of the 6th International Colloquium of Cognitive Science*, San Sebastian, Spain, 1999.
- Nouretdinov, I., T. Melluish, V. Vovk. Ridge regression confidence machine. In *Proc. 18th International Conf. on Machine Learning*, pages 385–392. Morgan Kaufmann, San Francisco, CA, 2001.
- Orel, B. *Osnove numerične matematike*. Fakulteta za računalništvo in informatiko, Ljubljana, 1997.

- Pavešić, N. *Razpoznavanje vzorcev: Uvod v analizo in razumevanje vidnih in slušnih signalov, 2. razširjena izdaja*. Fakulteta za elektrotehniko, Ljubljana, 2000. ISBN 961-6210-81-5.
- Press, W. H. *Numerical Recipes in C: the Art of Scientific Computing*. Cambridge University Press, 2002.
- R, 2006. *R Development Core Team. A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.
- Ridgeway, G. Prediction in the era of massive data sets. In Giudici, P., editor, *Statistical Modeling for Data Mining*, pages 109–119. Kluwer, 2000.
- Ridgeway, G., D. Madigan, T. Richardson. Boosting methodology for regression problems. In Heckerman, D., J. Whittaker, editors, *Proc. Artificial Intelligence and Statistics*, pages 152–161, 1999.
- Robnik-Šikonja, M. Lastnosti in uporaba hevristične funkcije relief v strojnem učenju, disertacija. Univerza v Ljubljani, Fakulteta za elektrotehniko in računalništvo, Ljubljana, 2001.
- Robnik-Šikonja, M., I. Kononenko. An adaptation of relief for attribute estimation in regression. In *Proc. Int. Conf. on Machine Learning ICML-97*, pages 296–304, 1997.
- Rosipal, R., M. Girolami, L. Trejo. On kernel principal component regression with covariance inflation criterion for model selection. University of Paisley, Technical report, 2000.
- Ruskey, F., M. Weston. A survey of venn diagrams. Department of Computer Science, University of Victoria, Canada, <http://www.combinatorics.org/Surveys/ds5/VennEJC.html>, 2007.
- Saltelli, A., S. Tarantola, F. Campolongo, M. Ratto. Sensitivity analysis in practice: A guide to assessing scientific models. John Wiley & Sons Ltd, England, 2003.

- Saltelli, A., M. Ratto, S. Tarantola, F. Campolongo. Sensitivity analysis for chemical models. In *Chemical Reviews*, volume 105(7), pages 2811–2828, 2005.
- Saunders, C., A. Gammerman, V. Vovk. Transduction with confidence and credibility. In *Proceedings of IJCAI'99*, volume 2, pages 722–726, 1999.
- Schaal, S., C. G. Atkeson. Constructive incremental learning from only local information. *Neural Computation*, 10(8):2047–2084, 1998.
- Schaal, S., C. G. Atkeson. Assessing the quality of learned local models. In Cowan, Jack D., Gerald Tesauro, Joshua Alspector, editors, *Advances in Neural Information Processing Systems*, volume 6, pages 160–167. Morgan Kaufmann Publishers, Inc., 1994.
- Schapire, R. E. A brief introduction to boosting. In *IJCAI*, pages 1401–1406, 1999.
- Schölkopf, D., C. J. C. Burges, A. J. Smola. *Advances in Kernel Methods: Support Vector Learning*. The MIT Press, Cambridge, Massachusetts, 1998.
- Seeger, M. Learning with labeled and unlabeled data. Technical report, <http://www.dai.ed.ac.uk/seeger/papers.html>, 2000.
- Silverman, B. W. *Density Estimation for Statistics and Data Analysis. Monographs on Statistics and Applied Probability*. Chapman and Hall, London, 1986.
- Smola, A. J., B. Schölkopf. A tutorial on support vector regression. NeuroCOLT2 Technical Report NC2-TR-1998-030, 1998.
- StatLib, 2005. Department of statistics at carnegie mellon university. statlib – data, software and news from the statistics community. <http://lib.stat.cmu.edu/>, 2005.
- Thrun, S. B., K. Möller. Active exploration in dynamic environments. In Moody, J. E., S. J. Hanson, R. P. Lippmann, editors, *Advances in Neural Information Processing Systems*, volume 4, pages 531–538. Morgan Kaufmann Publishers, Inc., 1992.

- Tibshirani, R., K. Knight. The covariance inflation criterion for adaptive model selection. *Journal of the Royal Statistical Society, Series B*, 61:529–546, 1999.
- Torgo, L. Functional models for regression tree leaves. In *Proc. 14th International Conference on Machine Learning*, pages 385–393. Morgan Kaufmann, 1997.
- Torgo, L. *Data Mining with R: learning by case studies*. University of Porto, LIACC-FEP, 2003.
- Tveter, D. R. The backprop algorithm. <http://www.dontveter.com/bpr/public2.html>, 2003.
- Vapnik, V. *The Nature of Statistical Learning Theory*. Springer, 1995.
- Venables, W. N., D. M. Smith, R Development Core Team. *An Introduction to R*. R Development Core Team, 2006. ISBN 3-9000051-12-7.
- Wand, M. P., M. C. Jones. *Kernel Smoothing*. Chapman and Hall, London, 1995.
- Weigend, A., D. Nix. Predictions with confidence intervals (local error bars). In *Proceedings of the International Conference on Neural Information Processing (ICONIP'94), Seoul, Korea*, pages 847–852, 1994.
- Wettschereck, D., D. W. Aha, T. Mohri. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review*, 11(1-5):273–314, 1997.
- Whitehead, S. D. A complexity analysis of cooperative mechanisms in reinforcement learning. In *AAAI*, pages 607–613, 1991.
- Wikipedia, the free encyclopedia. Confidence interval. http://en.wikipedia.org/wiki/Confidence_interval, august 2007a.
- Wikipedia, the free encyclopedia. Reliability. <http://en.wikipedia.org/wiki/Reliability>, august 2007b.
- Wood, S. N. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC, 2006.

LITERATURA

Woods, K., W. P. Kegelmeyer, K. Bowyer. Combination of multiple classifiers using local accuracy estimates. *IEEE Transactions on PAMI*, 19(4):405–410, 1997.

Dodatek *A*

Podrobni rezultati testiranja

A. PODROBNI REZULTATI TESTIRANJ

Tabela A.1: Lastnostni uporabljenih testnih domen

Domena	# primerov	# diskretnih atr.	# zveznih atr.
auto_price	159	1	14
auto93	93	6	16
autohorse	203	8	17
basketball	96	0	4
bodyfat	252	0	14
brainsize	20	0	8
breasttumor	286	1	8
cloud	108	2	4
cpu	209	0	6
diabetes	43	0	2
echomonths	130	3	6
elusage	55	1	1
fishcatch	158	2	5
fruitfly	125	2	2
grv	123	0	3
hungarian	294	7	6
lowbwt	189	7	2
mbagrade	61	1	1
pharynx	195	4	7
pollution	60	0	15
pwlinear	200	0	10
pyrim	74	0	27
servo	167	2	2
sleep	58	0	7
transplant	131	0	2
triazines	186	0	60
tumor	86	0	4
wdbc	198	0	32

Tabela A.2: Korelacijski koeficienti med oceno zanesljivosti *Savar* in **absolutno napako** napovedi. Senčenje celic predstavlja stopnjo značilnosti korelacijskega koeficienta. Statistično značilne korelacije ($\alpha \leq 0.05$) so označene s svetlo sivo barvo (značilne pozitivne korelacije) in temno sivo barvo (značilne negativne korelacije).

	<i>Savar</i> - absolutna napaka							
	RT	LR	NN	BAG	SVM	LWR	RF	GAM
autoprice	0.518	0.408	0.374	0.512	0.518	0.523	0.348	0.408
auto93	0.205	0.251	0.236	0.034	0.379	0.29	-0.136	0.251
autohorse	0.275	0.564	0.615	0.266	0.549	0.478	0.358	0.564
basketball	0.009	0.132	0.113	-0.127	0.013	-0.005	-0.128	0.132
bodyfat	0.06	0.14	0.201	-0.088	0.766	0.082	-0.085	0.14
brainsize	0.262	0.028	-0.259	0.217	-0.478	-0.05	-0.13	0.028
breasttumor	-0.049	-0.053	-0.066	-0.134	-0.093	-0.1	-0.09	-0.053
cloud	0.413	0.366	0.435	0.298	0.664	0.672	0.198	0.366
cpu	0.455	0.594	0.604	0.373	0.716	0.659	0.191	0.594
diabetes	0.031	0.383	0.363	-0.039	0.296	0.371	-0.105	0.383
echomonths	0.002	-0.056	-0.1	0.188	0.072	-0.078	0.001	-0.056
elusage	0.128	0.092	-0.07	0.213	0.637	0.324	0.139	0.092
fishcatch	0.407	0.13	0.115	0.363	0.63	0.149	0.047	0.13
fruitfly	-0.09	-0.051	-0.061	0.046	-0.06	0.004	0.036	-0.051
grv	0.197	0.267	0.23	0.193	0.136	0.072	0.14	0.267
hungarian	0.356	0.107	-0.124	0.597	0.319	0.219	0.4	0.107
lowbwt	0.106	-0.021	0	0.191	0.205	0.034	0.14	-0.021
mbagrade	0.163	-0.116	-0.116	-0.029	0.095	0.022	0.13	-0.116
pharynx	0.297	0.171	0.07	0.217	0.125	0.19	0.07	0.171
pollution	0.187	0.439	0.046	-0.157	0.139	0.092	-0.266	0.439
pwlinear	0.15	0.229	0.338	0.102	0.287	0.194	0	0.229
pyrim	0.045	0.671	0.524	-0.144	0.369	0.439	-0.043	0.671
servo	0.55	0.006	0.125	0.487	0.103	-0.489	-0.471	0.006
sleep	-0.215	0.538	0.231	0.034	0.004	-0.264	0.078	0.538
transplant	0.12	0.408	0.437	0.195	0.554	0.081	0.215	0.408
triazines	0.368	0.273	-0.055	0.331	0.051	-0.049	0.305	0.273
tumor	-0.141	-0.119	-0.117	0.068	0.019	-0.129	-0.064	-0.119
wpbc	0.165	-0.013	0.011	0.072	0.067	0.011	0.032	-0.013
+ :	46%	54%	39%	46%	46%	39%	25%	54%
- :	0%	0%	4%	4%	4%	7%	7%	0%

A. PODROBNI REZULTATI TESTIRANJ

Tabela A.3: Korelacijski koeficienti med oceno zanesljivosti *SAbias* in **predznačeno napako** napovedi. Senčenje celic predstavlja stopnjo značilnosti korelacijskega koeficienta. Statistično značilne korelacije ($\alpha \leq 0.05$) so označene s svetlo sivo barvo (značilne pozitivne korelacije) in temno sivo barvo (značilne negativne korelacije).

	<i>SAbias</i> - predznačena napaka							
	RT	LR	NN	BAG	SVM	LWR	RF	GAM
autoprice	0.439	0.036	0.183	0.121	0.106	0.06	-0.113	0.036
auto93	0.345	0.104	0.149	-0.132	0.156	0.085	-0.092	0.081
autohorse	0.2	-0.048	-0.002	-0.05	-0.179	0.023	-0.128	-0.048
basketball	0.433	0.134	-0.045	-0.005	-0.105	-0.028	0.141	0.134
bodyfat	0.441	-0.002	-0.005	0.054	0.205	0.059	-0.071	-0.002
brainsize	0.297	-0.171	0.217	0.186	-0.328	0.229	0.11	-0.171
breasttumor	0.151	0.061	-0.014	0.114	0.277	0.083	0.101	0.061
cloud	0.493	-0.181	-0.135	0.038	0.366	0.009	-0.129	-0.173
cpu	0.358	0.132	0.195	0.01	0.427	0.02	-0.077	0.132
diabetes	0.39	-0.04	0.094	0.127	-0.275	0.016	0.171	-0.04
echomonths	0.355	0.001	0.004	0.047	0.056	-0.095	0.193	-0.008
elusage	0.429	-0.197	-0.111	0.124	0.663	-0.368	-0.039	-0.187
fishcatch	0.401	-0.092	0.098	0.165	0.458	0.187	-0.059	-0.064
fruitfly	0.219	0.045	-0.039	0.078	0.101	-0.056	0.246	0.045
grv	0.412	0.077	-0.062	0.278	0.087	0.099	-0.009	0.077
hungarian	0.15	-0.066	0.272	0.159	0.041	-0.095	0.025	-0.066
lowbwt	0.02	0.049	0.061	0.092	0.002	0.134	-0.027	0.049
mbagrade	0.263	-0.196	-0.001	0.102	-0.264	-0.049	0.283	-0.196
pharynx	0.333	0.141	0.069	0.201	-0.039	-0.073	0.095	0.141
pollution	0.293	-0.026	-0.059	-0.008	0.345	-0.09	-0.094	-0.026
pwlinear	0.518	-0.103	0.007	0.009	-0.054	-0.024	-0.098	-0.117
pyrim	0.127	0.017	0.111	0.075	0.198	0.195	0.035	0.017
servo	0.579	0.111	-0.025	0.011	0.132	-0.402	-0.118	0.095
sleep	0.509	0.031	0.358	0.334	0.281	0.202	-0.009	0.058
transplant	0.135	0.222	-0.203	0.165	0.54	0.113	0.257	0.222
triazines	0.472	-0.016	0.025	0.066	0.049	-0.024	0.01	-0.022
tumor	0.097	0.04	0.333	0.108	0.343	0.033	0.15	0.04
wpbc	0.494	0.023	0.094	0.144	-0.074	0.032	0.074	0.023
+ :	82%	7%	18%	21%	36%	4%	14%	7%
- :	0%	0%	4%	0%	7%	7%	0%	0%

Tabela A.4: Korelacijski koeficienti med oceno zanesljivosti *SAbias* in **absolutno napako** napovedi. Senčenje celic predstavlja stopnjo značilnosti korelacijskega koeficienta. Statistično značilne korelacije ($\alpha \leq 0.05$) so označene s svetlo sivo barvo (značilne pozitivne korelacije) in temno sivo barvo (značilne negativne korelacije).

	<i>SAbias</i> - absolutna napaka							
	RT	LR	NN	BAG	SVM	LWR	RF	GAM
autoprice	0.429	-0.04	0.196	0.122	0.09	-0.032	0.069	-0.04
auto93	0.234	-0.173	-0.001	-0.078	0.156	0.012	-0.188	-0.167
autohorse	0.336	0.013	0.252	0.132	0.372	-0.047	0.08	0.013
basketball	0.046	0.098	-0.08	-0.094	-0.056	-0.092	0.07	0.098
bodyfat	0.07	0.024	0.107	0.056	0.307	-0.064	0.041	0.025
brainsize	-0.331	-0.246	-0.03	-0.046	-0.191	-0.005	0.06	-0.246
breasttumor	-0.035	-0.039	-0.024	-0.022	0.03	-0.07	-0.077	-0.039
cloud	0.333	0.093	0.153	0.005	0.364	-0.203	0.063	0.086
cpu	0.356	0.336	0.346	0.101	0.623	0.036	0.093	0.336
diabetes	0.029	-0.207	-0.027	0.169	0.165	-0.075	0.081	-0.207
echomonths	0.232	-0.053	-0.052	0.064	0.127	0.168	0.073	-0.064
elusage	0.304	-0.12	0.158	0.126	0.268	0.351	0.008	-0.109
fishcatch	0.132	-0.056	0.299	0.135	0.303	0.121	0.157	-0.093
fruitfly	-0.028	-0.06	0.126	-0.017	-0.008	-0.025	0.006	-0.06
grv	0.279	0.062	0.02	0.032	0.043	0.078	0.036	0.062
hungarian	0.129	-0.018	0.216	0.276	0.018	0.23	0.218	-0.018
lowbwt	0.028	-0.127	0.046	0.073	0.003	0.047	-0.052	-0.127
mbagrade	-0.049	0.015	-0.077	-0.056	-0.021	-0.062	0.247	0.015
pharynx	0.286	0.112	-0.007	-0.089	0.138	-0.098	0.01	0.112
pollution	0.275	-0.39	0.088	0.038	-0.043	-0.138	-0.033	-0.39
pwlinear	0.257	-0.011	-0.008	-0.009	-0.017	-0.178	0.089	-0.004
pyrim	-0.089	-0.149	0.498	0.074	-0.061	-0.195	0.011	-0.149
servo	0.525	0.039	-0.203	0.299	0.009	0.421	0.203	0.01
sleep	-0.027	-0.055	0.262	0.101	0.057	0.075	-0.12	-0.025
transplant	0.136	0.084	0.286	0.047	0.678	-0.046	0.028	0.084
triazines	0.265	0.014	-0.107	0.163	-0.066	0.027	-0.003	0.015
tumor	0.013	0.118	0.159	0.011	0.103	0.155	-0.058	0.118
wpbc	0.111	0.219	0.082	-0.06	0.04	-0.044	-0.097	0.219
+ :	50%	7%	29%	11%	25%	11%	11%	7%
- :	0%	4%	4%	0%	0%	7%	0%	4%

A. PODROBNI REZULTATI TESTIRANJ

Tabela A.5: Korelacijski koeficienti med oceno zanesljivosti *BAGV* in **absolutno napako** napovedi. Senčenje celic predstavlja stopnjo značilnosti korelacijskega koeficienta. Statistično značilne korelacije ($\alpha \leq 0.05$) so označene s svetlo sivo barvo (značilne pozitivne korelacije) in temno sivo barvo (značilne negativne korelacije).

	<i>BAGV</i> - absolutna napaka							
	RT	LR	NN	BAG	SVM	LWR	RF	GAM
autoprize	0.534	0.464	0.413	0.642	0.267	0.383	0.459	0.513
auto93	0.35	0.02	0.508	0.274	0.312	0.304	0.386	-0.016
autohorse	0.283	0.42	0.561	0.27	0.178	0.089	0.28	0.434
basketball	0.079	0.102	0.116	-0.05	0.078	0.042	-0.075	-0.001
bodyfat	0.275	0.082	0.148	0.327	0.524	0.044	0.415	0.083
brainsize	-0.127	-0.045	-0.336	-0.275	-0.253	-0.078	0.362	0.17
breasttumor	-0.079	-0.021	-0.057	-0.016	0.034	0.007	-0.055	-0.027
cloud	0.455	0.319	0.269	0.451	0.525	0.403	0.551	0.343
cpu	0.775	0.62	0.628	0.782	0.703	0.375	0.53	0.609
diabetes	0.11	0.336	0.24	0.117	0.153	0.159	0.024	0.257
echomonths	0.278	0.009	0.05	0.168	0.132	0.212	0.184	0.017
elusage	0.39	0.05	0.419	0.303	0.57	0.327	0.39	-0.104
fishcatch	0.772	0.403	0.434	0.819	0.715	0.698	0.55	0.375
fruitfly	0.033	-0.071	-0.048	-0.084	-0.069	0.082	-0.089	-0.136
grv	0.357	0.193	0.149	0.275	0.064	0.174	0.189	0.163
hungarian	0.525	0.25	-0.032	0.629	0.451	0.415	0.588	0.21
lowbwt	0.09	0.039	0.023	0.276	0.095	0.076	0.12	0.01
mbagrade	0.004	-0.121	-0.06	-0.108	0.033	0.256	0.261	-0.113
pharynx	0.456	0.204	0.102	0.309	0.058	0.203	0.225	0.331
pollution	0.258	0.422	-0.062	-0.111	-0.069	-0.13	-0.042	0.397
pwlinear	0.253	0.379	0.287	0.097	0.165	0.306	0.063	0.313
pyrim	0.255	-0.07	0.64	0.269	0.155	0.705	0.112	0.535
servo	0.639	0.272	0.418	0.608	0.508	0.76	0.765	0.372
sleep	0.065	0.588	0.361	0.012	0.078	0.224	0.224	0.685
transplant	0.382	0.441	0.576	0.477	0.721	0.059	0.425	0.392
triazines	0.547	0.248	0.178	0.502	0.077	0.053	0.469	0.199
tumor	0.195	-0.044	-0.042	0.198	0.342	-0.021	0.194	0.016
wpbc	0.133	-0.05	0.048	0.027	0.019	0.08	0.096	-0.015
+	64%	54%	50%	57%	46%	46%	57%	50%
-	0%	0%	0%	0%	0%	0%	0%	0%

Tabela A.6: Korelacijski koeficienti med oceno zanesljivosti *LCV* in **absolutno napako** napovedi. Senčenje celic predstavlja stopnjo značilnosti korelacijskega koeficienta. Statistično značilne korelacije ($\alpha \leq 0.05$) so označene s svetlo sivo barvo (značilne pozitivne korelacije) in temno sivo barvo (značilne negativne korelacije).

	<i>LCV - absolutna napaka</i>							
	RT	LR	NN	BAG	SVM	LWR	RF	GAM
autoprice	0.339	0.12	0.289	0.41	0.324	0.405	0.441	0.12
auto93	0.293	0.082	0.344	0.303	0.063	0.454	0.269	0.082
autohorse	0.095	0.13	0.119	0.044	0.138	0.134	0.098	0.13
basketball	-0.059	0.235	0.141	-0.082	0.044	-0.039	0.065	0.235
bodyfat	0.033	-0.024	0.174	0.094	0.193	0.113	0.182	-0.024
brainsize	0.013	-0.059	-0.337	0.066	-0.166	-0.144	-0.259	-0.059
breasttumor	-0.08	0.082	-0.016	-0.064	-0.013	-0.051	-0.075	0.082
cloud	0.373	0.096	0.099	0.458	0.216	0.276	0.25	0.096
cpu	0.531	0.41	0.693	0.677	0.594	0.336	0.483	0.41
diabetes	0.039	-0.15	0.187	0.064	0.152	0.058	-0.059	-0.15
echomonths	0.027	0.067	0.155	0.043	0.211	0.31	0.204	0.067
elusage	0.254	0.421	0.459	0.386	0.29	0.078	0.289	0.421
fishcatch	0.562	0.243	0.231	0.656	0.653	0.618	0.523	0.243
fruitfly	-0.041	-0.092	-0.115	-0.067	-0.079	-0.033	-0.052	-0.092
grv	0.216	0.177	0.111	0.283	0.214	0.128	0.156	0.177
hungarian	0.312	0.039	-0.099	0.275	0.27	0.382	0.324	0.039
lowbwt	0.016	-0.06	-0.005	0.02	-0.017	0.238	-0.045	-0.06
mbagrade	0.111	0.035	0.035	0.15	0.012	0.11	-0.061	0.035
pharynx	0.272	0.144	0.003	0.289	0.036	0.246	0.228	0.144
pollution	-0.118	0.312	-0.054	-0.151	-0.188	0.072	-0.129	0.312
pwlinear	0.116	-0.044	0	0.087	0.079	-0.015	0.055	-0.044
pyrim	-0.165	0.169	0.495	-0.129	0.305	0.574	0.321	0.169
servo	0.263	0.5	0.468	0.165	0.759	0.692	0.266	0.5
sleep	0.238	-0.007	0.062	0.382	0.286	0.222	0.342	-0.007
transplant	0.683	0.256	0.546	0.739	0.725	0.21	0.44	0.256
triazines	0.1	0.006	0.147	0.223	0.275	0.226	0.19	0.006
tumor	0.094	0.049	0.132	0.22	0.219	0.046	0.276	0.049
wpbc	0.058	0.072	0.013	0.13	0.164	-0.045	0.176	0.072
+ :	36%	32%	36%	50%	61%	46%	61%	32%
- :	0%	0%	0%	0%	0%	0%	0%	0%

A. PODROBNI REZULTATI TESTIRANJ

Tabela A.7: Korelacijski koeficienti med oceno zanesljivosti *DENS* in **absolutno napako** napovedi. Senčenje celic predstavlja stopnjo značilnosti korelacijskega koeficienta. Statistično značilne korelacije ($\alpha \leq 0.05$) so označene s svetlo sivo barvo (značilne pozitivne korelacije) in temno sivo barvo (značilne negativne korelacije).

	<i>DENS - absolutna napaka</i>							
	RT	LR	NN	BAG	SVM	LWR	RF	GAM
autoprice	0.244	0.258	0.174	0.335	0.263	0.384	0.36	0.258
auto93	-0.043	0.097	0.048	0.076	0.074	0.215	0.115	0.097
autohorse	0.323	0.25	0.28	0.26	0.225	0.185	0.259	0.25
basketball	0.018	0.091	0.084	0.069	0.08	0.083	0.088	0.091
bodyfat	0.128	0.099	0.096	0.067	0.31	0.283	0.24	0.099
brainsize	-0.314	-0.155	-0.111	-0.312	-0.158	-0.063	-0.282	-0.155
breasttumor	-0.079	-0.054	-0.063	-0.118	-0.111	-0.062	-0.129	-0.054
cloud	0.529	0.286	0.342	0.564	0.386	0.587	0.556	0.286
cpu	0.427	0.351	0.365	0.354	0.317	0.39	0.347	0.351
diabetes	0.375	0.419	0.425	0.414	0.451	0.439	0.225	0.419
echomonths	-0.347	-0.036	-0.043	-0.279	0.013	-0.31	-0.26	-0.036
elusage	0.248	0.167	0.205	0.306	0.261	0.13	0.155	0.167
fishcatch	0.417	0.294	0.277	0.509	0.366	0.24	0.328	0.294
fruitfly	-0.118	-0.011	-0.01	-0.11	-0.001	-0.105	-0.114	-0.011
grv	0.018	0.085	0.028	-0.011	-0.021	-0.02	-0.05	0.085
hungarian	0.099	0.192	-0.102	0.079	0.184	0.036	0.132	0.192
lowbwt	0.077	0.064	0.011	-0.008	0.117	0.063	0.036	0.064
mbagrade	0.038	-0.078	-0.077	-0.044	-0.035	-0.151	-0.045	-0.078
pharynx	0.163	0.091	0.119	0.166	0.131	0.165	0.144	0.091
pollution	0.113	0.201	0.071	0.091	0.13	0.008	0.024	0.201
pwlinear	0.164	0.096	0.122	0.095	0.153	0.242	0.124	0.096
pyrim	0.043	0.185	0.164	0.024	0.017	0.116	0.04	0.185
servo	-0.1	-0.293	-0.307	-0.111	-0.295	-0.38	-0.184	-0.293
sleep	-0.163	0.125	0.03	-0.075	0	-0.152	-0.185	0.125
transplant	0.509	0.482	0.473	0.474	0.472	0.419	0.505	0.482
triazines	0.094	0.188	0.119	0.165	0.163	0.185	0.126	0.188
tumor	-0.022	-0.057	-0.056	-0.122	-0.195	-0.204	-0.115	-0.057
wpbc	-0.089	-0.028	-0.124	-0.065	0.005	-0.02	-0.065	-0.028
+	36%	32%	25%	36%	39%	43%	32%	32%
-	4%	4%	4%	7%	4%	7%	11%	4%

Tabela A.8: Korelacijski koeficienti med oceno zanesljivosti *CNK* in **predznačeno napako** napovedi. Senčenje celic predstavlja stopnjo značilnosti korelacijskega koeficienta. Statistično značilne korelacije ($\alpha \leq 0.05$) so označene s svetlo sivo barvo (značilne pozitivne korelacije) in temno sivo barvo (značilne negativne korelacije).

	<i>CNK - predznačena napaka</i>							
	RT	LR	NN	BAG	SVM	LWR	RF	GAM
autoprice	0.498	0.223	0.218	0.161	0.122	-0.018	-0.201	0.223
auto93	0.539	0.381	0.246	0.292	-0.162	-0.018	0.282	0.381
autohorse	0.306	0.186	0.089	0.107	-0.033	-0.093	-0.029	0.186
basketball	0.432	0.085	0.074	0.105	0.098	0.058	0.207	0.085
bodyfat	0.29	0.184	0.159	0.139	0.207	0.483	-0.265	0.184
brainsize	-0.175	0.323	0.016	-0.18	-0.096	0.472	-0.262	0.323
breasttumor	0.145	0.011	-0.027	0.022	0.136	-0.018	0.035	0.011
cloud	0.239	0.051	0.01	-0.05	0.361	0.115	-0.114	0.051
cpu	0.477	0.019	0.036	0.038	0.356	0.089	-0.492	0.019
diabetes	0.332	0.223	0.226	0.04	0.132	0.448	0.09	0.223
echomonths	0.33	0.107	0.093	0.087	0.292	0.037	0.108	0.107
elusage	0.34	0.071	-0.088	-0.073	0.23	0.433	0.014	0.071
fishcatch	0.753	0.422	0.493	0.659	0.002	0.472	0.146	0.422
fruitfly	0.016	-0.17	-0.183	-0.126	-0.111	-0.245	-0.205	-0.17
grv	0.379	0.285	0.278	0.137	0.037	-0.056	0.014	0.285
hungarian	0.304	0.234	0.469	0.197	0.208	0.007	0.124	0.234
lowbwt	0.038	0.061	0.185	0.098	-0.028	0.077	0.074	0.061
mbagrade	0.205	-0.171	-0.165	0.105	-0.189	-0.02	0.083	-0.171
pharynx	0.172	-0.038	-0.055	-0.023	-0.15	0.003	-0.144	-0.038
pollution	0.483	0.488	0.189	0.199	0.071	0.053	0.097	0.488
pwlinear	0.246	0.313	0.064	0.043	-0.235	0.27	-0.197	0.313
pyrim	0.331	0.794	0.522	0.23	-0.173	0.226	0.017	0.794
servo	0.241	-0.001	-0.081	0.092	-0.507	-0.262	-0.264	-0.001
sleep	0.437	0.412	0.09	0.186	-0.001	-0.262	0.037	0.412
transplant	0.443	0.233	0.232	0.39	0.465	0.073	0.171	0.233
triazines	0.362	0.513	0.277	-0.043	0.064	0.259	-0.124	0.513
tumor	0.262	0.147	0.158	0.178	0.238	-0.026	0.062	0.147
wpbc	0.441	0.211	0.098	0.074	0.094	-0.082	0.062	0.211
+ :	86%	50%	36%	25%	29%	25%	11%	50%
- :	0%	0%	4%	0%	11%	11%	25%	0%

A. PODROBNI REZULTATI TESTIRANJ

Tabela A.9: Korelacijski koeficienti med oceno zanesljivosti *CNK* in **absolutno napako** napovedi. Senčenje celic predstavlja stopnjo značilnosti korelacijskega koeficienta. Statistično značilne korelacije ($\alpha \leq 0.05$) so označene s svetlo sivo barvo (značilne pozitivne korelacije) in temno sivo barvo (značilne negativne korelacije).

	<i>CNK - absolutna napaka</i>							
	RT	LR	NN	BAG	SVM	LWR	RF	GAM
autoprize	0.653	0.415	0.318	0.611	0.219	0.228	0.567	0.415
auto93	0.419	0.505	0.384	0.204	0.157	0.233	0.389	0.505
autohorse	0.457	0.358	0.456	0.323	0.187	0.077	0.182	0.358
basketball	0.242	-0.099	-0.065	0.015	0.133	-0.038	0.034	-0.099
bodyfat	0.2	0.518	0.515	0.124	0.318	0.109	0.021	0.518
brainsize	-0.405	-0.016	0.008	-0.371	-0.08	0.105	-0.534	-0.016
breasttumor	0.088	0.009	-0.035	0.072	-0.078	-0.04	0	0.009
cloud	0.483	0.212	0.24	0.353	0.552	0.129	0.36	0.212
cpu	0.679	0.753	0.754	0.533	0.762	0.472	0.656	0.753
diabetes	0.076	-0.086	-0.082	-0.055	-0.131	0.281	-0.175	-0.086
echomonths	0.339	0.134	0.119	0.155	0.188	0.056	0.226	0.134
elusage	0.221	-0.039	-0.273	0.32	-0.13	0.452	0.318	-0.039
fishcatch	0.668	0.498	0.502	0.597	0.262	0.646	0.324	0.498
fruitfly	0.027	0.028	-0.004	-0.034	-0.043	0.107	0.064	0.028
grv	0.303	0.108	0.158	0.229	0.014	0.005	0.068	0.108
hungarian	0.401	0.325	0.553	0.356	0.364	0.37	0.403	0.325
lowbwt	0.176	0.052	0.051	0.101	0.008	0.115	0.107	0.052
mbagrade	0.058	0.024	0.024	-0.075	0.134	-0.052	0.11	0.024
pharynx	0.278	0.261	-0.086	0.185	0.109	0.093	0.225	0.261
pollution	0.182	0.411	0.146	0.041	-0.039	0.02	0.085	0.411
pwlinear	0.217	0.194	0.109	0.129	0.129	0.236	0.121	0.194
pyrim	0.307	0.849	0.636	0.153	0.401	0.797	0.037	0.849
servo	0.278	0.315	0.281	0.148	-0.025	0.179	0.315	0.315
sleep	0.049	0.565	0.259	-0.045	0.169	-0.116	-0.084	0.565
transplant	0.359	0.387	0.357	0.431	0.495	0.12	0.507	0.387
triazines	0.313	0.468	0.135	0.301	0.089	0.071	0.308	0.468
tumor	0.215	-0.033	-0.061	0.03	-0.029	-0.128	-0.012	-0.033
wpbc	0.039	0.053	0.019	0.008	0.094	0.111	0.003	0.053
+ :	68%	57%	39%	46%	36%	32%	46%	57%
- :	0%	0%	4%	0%	0%	0%	4%	0%

Tabela A.10: Korelacijski koeficienti med oceno zanesljivosti *BVCK* in **absolutno napako** napovedi. Senčenje celic predstavlja stopnjo značilnosti korelacijskega koeficienta. Statistično značilne korelacije ($\alpha \leq 0.05$) so označene s svetlo sivo barvo (značilne pozitivne korelacije) in temno sivo barvo (značilne negativne korelacije).

	<i>BVCK - absolutna napaka</i>							
	RT	LR	NN	BAG	SVM	LWR	RF	GAM
autoprice	0.534	0.465	0.413	0.642	0.267	0.383	0.46	0.513
auto93	0.378	0.02	0.527	0.276	0.314	0.311	0.454	-0.016
autohorse	0.296	0.436	0.591	0.299	0.19	0.089	0.287	0.451
basketball	0.242	-0.097	-0.063	0.014	0.135	-0.034	0.033	-0.098
bodyfat	0.289	0.513	0.522	0.192	0.481	0.09	0.139	0.515
brainsize	-0.449	-0.045	-0.337	-0.418	-0.246	-0.06	0.222	0.169
breasttumor	-0.052	-0.018	-0.061	0.02	0.007	-0.001	-0.049	-0.024
cloud	0.502	0.228	0.252	0.382	0.578	0.302	0.387	0.231
cpu	0.776	0.625	0.634	0.782	0.707	0.375	0.532	0.615
diabetes	0.091	-0.007	-0.026	-0.029	-0.059	0.251	-0.165	-0.029
echomonths	0.314	0.047	0.084	0.197	0.173	0.214	0.237	0.056
elusage	0.381	0.039	0.362	0.328	0.553	0.345	0.399	-0.104
fishcatch	0.773	0.44	0.473	0.821	0.715	0.701	0.551	0.417
fruitfly	0.036	-0.048	-0.038	-0.084	-0.074	0.091	-0.068	-0.092
grv	0.327	0.112	0.16	0.242	0.02	0.023	0.077	0.111
hungarian	0.47	0.331	0.553	0.406	0.401	0.401	0.437	0.33
lowbwt	0.091	0.039	0.024	0.277	0.095	0.076	0.122	0.011
mbagrade	0.051	0.018	0.021	-0.083	0.133	0.032	0.139	0.018
pharynx	0.457	0.212	0.099	0.311	0.059	0.203	0.228	0.336
pollution	0.26	0.423	-0.059	-0.11	-0.071	-0.127	-0.034	0.398
pwlinear	0.286	0.234	0.215	0.14	0.163	0.306	0.126	0.22
pyrim	0.315	-0.07	0.645	0.159	0.403	0.792	0.043	0.535
servo	0.537	0.32	0.32	0.286	0.122	0.646	0.362	0.326
sleep	0.075	0.591	0.36	-0.028	0.167	0.216	0.086	0.685
transplant	0.406	0.453	0.486	0.527	0.755	0.075	0.471	0.43
triazines	0.36	0.248	0.148	0.32	0.091	0.072	0.323	0.199
tumor	0.235	-0.043	-0.069	0.081	0.186	-0.082	0.033	-0.029
wpbc	0.132	-0.047	0.051	0.027	0.037	0.092	0.087	-0.013
+ :	71%	50%	54%	61%	46%	43%	50%	54%
- :	4%	0%	0%	0%	0%	0%	0%	0%

A. PODROBNI REZULTATI TESTIRANJ

Tabela A.11: Dinamično izbrane ocene in korelacijski koeficienti med **ocenami zanesljivosti, izbranimi z metaučenjem in napako napovedi**. Senčenje celic predstavlja stopnjo značilnosti korelacijskega koeficienta. Statistično značilne korelacije ($\alpha \leq 0.05$) so označene s svetlo sivo barvo (značilne pozitivne korelacije) in temno sivo barvo (značilne negativne korelacije).

	<i>dinamična ocena - METAUČENJE</i>							
	RT	LR	NN	BAG	SVM	LWR	RF	GAM
autoprince	BVCK	CNK-a	BVCK	BVCK	BVCK	BVCK	BVCK	CNK-a
	0.534	0.415	0.413	0.642	0.267	0.383	0.46	0.415
auto93	BVCK	SAvar	SAvar	BVCK	SAvar	SAvar	BVCK	CNK-a
	0.378	0.251	0.236	0.276	0.379	0.29	0.454	0.505
autohorse	SAvar	BVCK	BVCK	SAvar	SAvar	BVCK	CNK-a	BVCK
	0.275	0.436	0.591	0.266	0.549	0.089	0.182	0.451
basketball	CNK-p	CNK-p	CNK-p	CNK-p	CNK-p	CNK-p	CNK-p	CNK-p
	0.432	0.085	0.074	0.105	0.098	0.058	0.207	0.085
bodyfat	BAGV	BVCK	BVCK	BAGV	BAGV	BVCK	BAGV	BVCK
	0.275	0.513	0.522	0.327	0.524	0.09	0.415	0.515
brainsize	CNK-p	SAbias-p	BAGV	SAbias-p	SAbias-p	BAGV	SAbias-p	BAGV
	-0.175	-0.171	-0.336	0.186	-0.328	-0.078	0.11	0.17
breasttumor	SAbias-p	LCV	SAbias-p	SAbias-p	LCV	CNK-a	SAbias-p	BAGV
	0.151	0.082	-0.014	0.114	-0.013	-0.04	0.101	-0.027
cloud	BAGV	BAGV	BAGV	BAGV	SAvar	BVCK	BAGV	BAGV
	0.455	0.319	0.269	0.451	0.664	0.302	0.551	0.343
cpu	CNK-a	CNK-a	CNK-a	BVCK	BVCK	BVCK	CNK-a	CNK-a
	0.679	0.753	0.754	0.782	0.707	0.375	0.656	0.753
diabetes	DENS	DENS	DENS	DENS	DENS	DENS	DENS	DENS
	0.375	0.419	0.425	0.414	0.451	0.439	0.225	0.419
echomonths	SAbias-p	CNK-a	BVCK	BVCK	SAbias-p	BVCK	BAGV	CNK-a
	0.355	0.134	0.084	0.197	0.056	0.214	0.184	0.134
elusage	BAGV	BAGV	SAvar	BVCK	DENS	SAbias-p	CNK-a	BAGV
	0.39	0.05	-0.07	0.328	0.261	-0.368	0.318	-0.104
fishcatch	CNK-a	CNK-a	CNK-a	BVCK	BVCK	BVCK	CNK-a	CNK-a
	0.668	0.498	0.502	0.821	0.715	0.701	0.324	0.498
fruitfly	SAbias-p	SAbias-p	SAbias-p	SAbias-p	LCV	SAbias-p	SAbias-p	BAGV
	0.219	0.045	-0.039	0.078	-0.079	-0.056	0.246	-0.136
grv	CNK-a	CNK-a	CNK-a	BVCK	BVCK	BVCK	BAGV	CNK-a
	0.303	0.108	0.158	0.242	0.02	0.023	0.189	0.108
hungarian	BVCK	BVCK	CNK-p	BAGV	SAvar	BVCK	BAGV	BVCK
	0.47	0.331	0.469	0.629	0.319	0.401	0.588	0.33
lowbwt	BAGV	SAvar	SAvar	BAGV	SAvar	BVCK	CNK-a	SAvar
	0.09	-0.021	0	0.276	0.205	0.076	0.107	-0.021
mbagrade	SAbias-p	DENS	DENS	SAbias-p	DENS	SAbias-p	SAbias-p	DENS
	0.263	-0.078	-0.077	0.102	-0.035	-0.049	0.283	-0.078
pharynx	BVCK	SAvar	CNK-a	BVCK	SAvar	BAGV	BVCK	SAvar
	0.457	0.171	-0.086	0.311	0.125	0.203	0.228	0.171
pollution	CNK-p	CNK-p	CNK-p	BAGV	CNK-p	BAGV	CNK-p	CNK-p
	0.483	0.488	0.189	-0.111	0.071	-0.13	0.097	0.488
pwnlinear	BVCK	SAvar	BAGV	BVCK	SAvar	BVCK	BVCK	SAvar
	0.286	0.229	0.287	0.14	0.287	0.306	0.126	0.229
pyrim	CNK-a	CNK-a	CNK-a	DENS	SAvar	CNK-a	CNK-a	CNK-a
	0.307	0.849	0.636	0.024	0.369	0.797	0.037	0.849
servo	DENS	SAvar	SAvar	BAGV	SAvar	BAGV	BAGV	SAbias-p
	-0.1	0.006	0.125	0.608	0.103	0.76	0.765	0.095
sleep	LCV	SAbias-p	SAbias-p	CNK-p	SAbias-p	SAbias-p	CNK-p	SAbias-p
	0.238	0.031	0.358	0.186	0.281	0.202	0.037	0.058
transplant	DENS	DENS	DENS	DENS	BVCK	BVCK	CNK-a	CNK-a
	0.509	0.482	0.473	0.474	0.755	0.075	0.507	0.387
triazines	CNK-p	CNK-p	CNK-a	CNK-p	CNK-p	CNK-a	BAGV	CNK-p
	0.362	0.513	0.135	-0.043	0.064	0.071	0.469	0.513
tumor	SAbias-p	SAbias-p	BAGV	SAbias-p	CNK-p	SAbias-p	SAbias-p	BAGV
	0.097	0.04	-0.042	0.108	0.238	0.033	0.15	0.016
wpsc	CNK-a	CNK-a	CNK-a	CNK-a	CNK-a	CNK-a	CNK-a	CNK-a
	0.039	0.053	0.019	0.008	0.094	0.111	0.003	0.053
+	79%	54%	46%	61%	54%	43%	64%	54%
-	0%	0%	0%	0%	0%	4%	0%	0%

Tabela A.12: Dinamično izbrane ocene in korelacijski koeficienti med **ocenami zanesljivosti, izbranimi z notranjim prečnim preverjanjem in napako napovedi**. Senčenje celic predstavlja stopnjo značilnosti korelacijskega koeficienta. Statistično značilne korelacije ($\alpha \leq 0.05$) so označene s svetlo sivo barvo (značilne pozitivne korelacije) in temno sivo barvo (značilne negativne korelacije).

	<i>dinamična ocena - INTERNO PREČNO PREVERJANJE</i>							
	RT	LR	NN	BAG	SVM	LWR	RF	GAM
autoprice	CNK-a 0.624	BVCK 0.52	BVCK 0.482	BVCK 0.519	SAvar 0.517	SAvar 0.537	CNK-a 0.559	BVCK 0.438
auto93	CNK-p 0.525	CNK-a 0.465	BVCK 0.522	CNK-p 0.335	SAvar 0.287	LCV 0.436	BVCK 0.442	CNK-a 0.537
basketball	CNK-p 0.381	SAvar 0.12	SAvar 0.087	CNK-p 0.059	CNK-p 0.137	CNK-p -0.011	CNK-p 0.176	LCV 0.236
brainsize	LCV 0.017	CNK-p 0.347	SAbias-p 0.378	SAbias-a -0.193	DENS -0.152	CNK-p 0.472	SAbias-a 0.254	CNK-p 0.349
cloud	BAGV 0.461	SAvar 0.4	SAvar 0.507	DENS 0.538	SAvar 0.639	SAvar 0.654	DENS 0.518	SAvar 0.364
diabetes	DENS 0.415	DENS 0.409	DENS 0.316	DENS 0.42	DENS 0.354	CNK-p 0.375	DENS 0.174	DENS 0.382
elusage	BVCK 0.426	LCV 0.279	BAGV 0.397	BVCK 0.403	SAvar 0.537	CNK-p 0.43	BVCK 0.429	LCV 0.417
fishcatch	BVCK 0.658	CNK-a 0.515	CNK-a 0.482	BVCK 0.76	BVCK 0.749	BVCK 0.635	BVCK 0.56	CNK-a 0.518
grv	SAbias-p 0.346	CNK-p 0.235	CNK-p 0.285	LCV 0.27	LCV 0.202	BAGV 0.241	SAvar 0.196	CNK-p 0.333
mbagrade	SAbias-p 0.165	LCV 0.018	SAbias-p -0.137	LCV 0.105	SAvar 0.047	BAGV 0.365	BAGV 0.256	SAbias-a 0.139
pollution	CNK-p 0.457	CNK-p 0.65	CNK-p 0.27	CNK-p 0.137	SAbias-p 0.212	SAvar 0.043	CNK-a 0.117	CNK-p 0.505
pyrim	CNK-p 0.37	CNK-a 1	BVCK 0.566	CNK-p 0.273	BVCK 0.536	CNK-a 0.786	LCV -0.102	CNK-a 0.834
servo	BAGV 0.6	LCV 0.514	LCV 0.427	BAGV 0.391	LCV 0.748	BAGV 0.7	BAGV 0.833	LCV 0.418
sleep	SAbias-p 0.512	BAGV 0.749	BVCK 0.535	LCV 0.319	LCV 0.345	LCV 0.213	LCV 0.283	BAGV 0.767
tumor	CNK-p 0.33	CNK-p 0.022	CNK-p 0.129	LCV 0.195	BAGV 0.202	SAbias-a 0.157	LCV 0.203	CNK-p 0.2
+	87%	73%	73%	67%	67%	73%	60%	80%
-	0%	0%	0%	0%	0%	0%	0%	0%

A. PODROBNI REZULTATI TESTIRANJ

Slika A.1: Izpis klasifikacijskega drevesa za napovedovanje uspešnosti korelacije vsaj ene izmed ocen zanesljivosti z napako. Podatki v vrsticah pomenijo:

<št.vozlišča> <kriterij> <št. primerov> <izguba> <razred> <verj.razredov> Z
zvezdico na koncu vrstice so označeni listi drevesa.

n= 224

node), split, n, loss, yval, (yprob)
* denotes terminal node

- 1) root 224 42 DA (0.1875000 0.8125000)
- 2) cv.rmse>=0.8680882 79 27 DA (0.3417722 0.6582278)
- 4) cv.rmse< 1.503797 54 24 DA (0.4444444 0.5555556)
- 8) no.attr< 12 38 17 NE (0.5526316 0.4473684)
- 16) model=nn,bag,gam,lr,lwr,svm 29 10 NE (0.6551724 0.3448276) *
- 17) model=rf,rt 9 2 DA (0.2222222 0.7777778) *
- 9) no.attr>=12 16 3 DA (0.1875000 0.8125000) *
- 5) cv.rmse>=1.503797 25 3 DA (0.1200000 0.8800000) *
- 3) cv.rmse< 0.8680882 145 15 DA (0.1034483 0.8965517)
- 6) cv.rmse>=0.3339738 96 15 DA (0.1562500 0.8437500)
- 12) model=nn,gam,lr,lwr 46 12 DA (0.2608696 0.7391304)
- 24) no.examples< 192 35 12 DA (0.3428571 0.6571429)
- 48) avg.dens< 0.03624999 21 10 DA (0.4761905 0.5238095)
- 96) avg.dens>=0.00418043 11 3 NE (0.7272727 0.2727273) *
- 97) avg.dens< 0.00418043 10 2 DA (0.2000000 0.8000000) *
- 49) avg.dens>=0.03624999 14 2 DA (0.1428571 0.8571429) *
- 25) no.examples>=192 11 0 DA (0.0000000 1.0000000) *
- 13) model=bag,rf,rt,svm 50 3 DA (0.0600000 0.9400000) *
- 7) cv.rmse< 0.3339738 49 0 DA (0.0000000 1.0000000) *

Dodatek \mathcal{B}

Knjižnica algoritmov v R

Predstavljena metodologija za ocenjevanje zanesljivosti napovedi je bila implementirana v statističnem paketu R. Celotno programsko kodo sestavlja 10 programskih modulov:

MAIN.r: glavni programski modul z inicializacijo potrebnih knjižnic in primeri uporabe funkcij za ocenjevanje zanesljivosti,

branje domen.r: branje in predobdelava testnih domen s trdega diska,

error.bagv.r: izračun ocene zanesljivosti *BAGV*,

error.dens.r: izračun ocene zanesljivosti *DENS*,

error.local.r: izračun dodatnih lokalnih informacij problemskega prostora in ocene

LCV,

internal.cv.r: dinamična izbira ocene z notranjim prečnim preverjanjem,

lwr.r: algoritem lokalno utežene regresije,

meta.domain.r: metaučenje za dinamično izbiro ocen zanesljivosti,

modelling.r: gradnja regresijskih modelov in napovedovanje,

sensitivity.r: izračun ocen zanesljivosti z uporabo analize občutljivosti, izračun preostalih ocen zanesljivosti, testiranje korelacij ocen.

util.r: pomožne funkcije (statistična analiza itd.).

Za potrebe paketnega izračuna ocen zanesljivosti in izvedbo testiranj sta v globalnem okolju definirani spremenljivki *seznam.domen* in *seznam.modelov*, ki hranita seznama imen domen in modelov (znakovnih nizov), ki smo jih uporabljali v okviru testiranj.

```
> seznam.domen
[1] "auto_price" "auto93"      "autohorse"   "basketball" "bodyfat"
[6] "brainsize" "breasttumor" "cloud"       "cpu"         "diabetes"
[11] "echomonths" "elusage"     "fishcatch"   "fruitfly"    "grv"
[16] "hungarian"  "lowbwt"     "mbagrade"    "pharynx"     "pollution"
[21] "pwnlinear"  "pyrim"      "servo"       "sleep"       "transplant"
[26] "triazines"  "tumor"      "wpbc"

> seznam.modelov
[1] "rt"  "lr"  "nn" "bag" "svm" "lwr" "rf"  "gam"
```

B.1 Branje testnih podatkov

Branje testnih podatkov izvedemo z uporabo ukaza *proc.beri(domena)*, ki vrne objekt tipa *data.frame*, napolnjen s podatki iz datoteke *domena.dat*. Po-dat-kov-na datoteka mora biti v naslednji obliki:

```
<ŠT. PRIMEROV>
<OZNAČBA1> <ATRIBUT1_1> <ATRIBUT1_2> ... <ATRIBUT1_m>
<OZNAČBA2> <ATRIBUT2_1> <ATRIBUT2_2> ... <ATRIBUT2_m>
...
<OZNAČBAn> <ATRIBUTn_1> <ATRIBUTn_2> ... <ATRIBUTn_m>
```


Predobdelavi testnih podatkov za uporabo z regresijskimi modeli sta namenjeni funkciji:

`proc.fix.missing(domena)`, ki vrne objekt tipa *data.frame* z nadomestnimi vrednostmi namesto manjkajočih in

`proc.data.preprocess(domena, model)`, ki vrne objekt tipa *data.frame*, predobdelan za uporabo z regresijskim modelom *model*.

B.2 Izračun ocen zanesljivosti

Uporaba ocen zanesljivosti poteka v dveh fazah. V prvi je potrebno izračunati občutljivostne napovedi in lokalne informacije učnih primerov, ki so potrebne za izračun ocen zanesljivosti v drugi fazi.

B.2.1 Prva faza

Izračune pomožnih količin izvedemo z:

`create.sens.tables(seznam.domen, seznam.modelov)`, ki za vsako domeno iz *seznam.domen* in modela iz *seznam.modelov* izračuna tabelo občutljivostnih napovedi in jo shrani kot objekt tipa *data.frame* z nazivom *preds.model.domena* v globalno okolje *.GlobalEnv*.

`create.localinfo.tables(seznam.domen, seznam.modelov)`, ki za vsako domeno iz *seznam.domen* in modela iz *seznam.modelov* izračuna tabelo z lokalnimi informacijami primerov (DA, DK, označbe sosedov, napovedi sosedov) napovedi in jo shrani kot objekt tipa *data.frame* z nazivom *localinfo.model.domena* v globalno okolje *.GlobalEnv*.

`create.supplest.tables(seznam.domen, seznam.modelov)`, ki za vsako domeno iz *seznam.domen* in modela iz *seznam.modelov* izračuna tabelo z dodatnimi ocenami zanesljivosti (DENS, LCV, BAGV) in jo shrani kot objekt tipa *data.frame* z nazivom *supplest.model.domena* v globalno okolje *.GlobalEnv*.

B.2.2 Druga faza izračuna

Za uspešno izvedbo testiranja in druge faze, se mora za vsako obravnavano domeno in model v globalnem okolju *.GlobalEnv* nahajati tri tabele, katerih izdelava je opisana v prvi fazi postopka. Ocene lahko izračunamo in testiramo z uporabo naslednjih ukazov:

`calculate.estimate.matrix(domain.name, model.name)`, ki za domeno *domain.name* in model *model.name* izračuna tabelo ocen zanesljivosti

```
> calculate.estimate.matrix('brainsize', 'nn')
      SAvar  SAbias-p  SAbias-a      CNK-p      CNK-a      LCV ...
[1,] 18.15208  0.76626907 0.76626907 13.0173150 13.0173150 -36.6267011 ...
[2,] 20.17404  1.10433464 1.10433464  5.5185331  5.5185331 -12.1212607 ...
[3,] 15.90364 -0.05062176 0.05062176  7.8866632  7.8866632  4.4404269 ...
[4,] 17.54080  0.58810922 0.58810922  6.8227432  6.8227432 -19.5985790 ...
[5,] 29.62446 -0.54057500 0.54057500  2.6170519  2.6170519  42.4612412 ...
[6,] 30.98192 -1.12934505 1.12934505 -6.1744162  6.1744162  -2.9993876 ...
[7,] 44.72618  0.04805629 0.04805629  0.1846789  0.1846789 -54.0250113 ...
[8,] 17.71124 -0.26788507 0.26788507 -3.2196113  3.2196113  -0.7020053 ...
[9,] 22.55315  1.46973336 1.46973336 -2.3352251  2.3352251  -6.9833516 ...
[10,] 17.06348 -2.68773474 2.68773474 -8.3911325  8.3911325 -62.0834090 ...
[11,] 25.55939 -4.09225412 4.09225412  3.5567128  3.5567128  -9.8595590 ...
[12,] 33.54441 -0.84194404 0.84194404 -13.4363696 13.4363696  -6.5520971 ...
...
```

`estimate.test(seznam.domen, model, ime.ocene)`, ki za vsako domeno iz *seznam.domen* in model *model* vrne tabelo z vrednostmi korelacijskih koeficientov in njihovih stopenj značilnosti med oceno *ime.ocene* in napako na napovedih.

```
> estimate.test(seznam.domen, 'nn', 'BAGV')
      domena model estimate      corr      sig *
1  auto_price  nn SAbias-a  0.196415665 1.308727e-02 +
2    auto93    nn SAbias-a -0.000543556 9.958742e-01
3  autohorse  nn SAbias-a  0.251538658 2.944615e-04 +
4   basketball nn SAbias-a -0.080048936 4.381699e-01
5    bodyfat   nn SAbias-a  0.106606745 9.127088e-02
6  brainsize  nn SAbias-a -0.030022334 9.000118e-01
7 breasttumor nn SAbias-a -0.023901029 6.873232e-01
8      cloud   nn SAbias-a  0.152886162 1.141863e-01
9        cpu   nn SAbias-a  0.346321992 2.798466e-07 +
...
```

`estimate.tabulate.result(seznam.domen, seznam.modelov, ime.ocene)` za vsako domeno iz *seznam.domen* in model iz *seznam.modelov* vrne tabelo s številom značilnih pozitivnih in negativnih korelacij med oceno *ime.ocene* in napako pri napovedih.

```
> estimate.tabulate.result(seznam.domen, seznam.modelov, 'BAGV')
BAGV
model result n.domains
1   rt   64/0      28
2   lr   54/0      28
3   nn   50/0      28
4   bag  57/0      28
5   svm  46/0      28
6   lwr  46/0      28
7   rf   57/0      28
8   gam  50/0      28
```

B.3 Metaučenje

B.3.1 Tvorjenje učne množice

Za uspešno izvedbo metaučjenja je potrebno v globalnem okolju *.GlobalEnv* ustvariti učno množico za metaprediktor. Za uspešno izdelavo učne množice morajo za domene, ki bodo vsebovane v učnih podatkih, v globalnem okolju *.GlobalEnv* biti prisotne tabele s podatki za izračun ocen zanesljivosti (glej razdelek B.2.1). Učno množico za metaprediktor tvorimo z ukazom:

```
DOMAIN.METADATA <- generate.domain.metadata(seznam.domen,
seznam.modelov), ki za vsako domeno iz seznam.domen in model iz
seznam.modelov izračuna attribute, potrebne za metanapovedovanje ocen
zanesljivosti in jih vrne kot objekt tipa data.frame
```

```
> (DOMAIN.METADATA <- generate.domain.metadata(seznam.domen, seznam.modelov))
estimate domain model no.examples no.attr cv.rmse avg.dens ...
1   CNK-p   tumor   rt           86      4  1.1367206 0.034676482 ...
2   CNK-p   tumor   lr           86      4  1.0123449 0.034676482 ...
3  SABias-p tumor   nn           86      4  1.1464827 0.034676482 ...
4     BAGV   tumor   bag          86      4  4.4551925 0.034676482 ...
5  SABias-p tumor   svm          86      4  1.3500560 0.034676482 ...
```

```
6 SABias-a      tumor  lwr      86      4  1.7195195 0.034676482 ...
7      BAGV      tumor  rf       86      4  1.4767400 0.034676482 ...
8      CNK-p      tumor  gam      86      4  0.9959681 0.034676482 ...
9 SABias-p brainsize  rt      20      8 82.4013013 0.004633449 ...
10     CNK-p brainsize  lr      20      8  6.2543859 0.004633449 ...
...
```

B.3.2 Napovedovanje ocene

Optimalno oceno napovemo z uporabo metaprediktorja z uporabo naslednjega ukaza:

`predict.metadata.estimate(domain, model, DOMAIN.METADATA)`, ki za domeno *domena* in model *model* vrne ime napoved metaprediktorja - ime ocene kot znakovni niz. kot tretji parameter je potrebno funkciji podati ime učne množice metaprediktorja (v našem primeru *DOMAIN.METADATA*).

```
> predict.metadata.estimate('fruitfly', 'rt', DOMAIN.METADATA)
$estimate
[1] "SAbias-p"
```

B.4 Dinamična izbira ocene z notranjim prečnim preverjanjem

Postopek notranjega prečnega preverjanja za dinamično izbiro ocene poženemo s klicem:

`internal.cv.run(seznam.domen, seznam.modelov)`, ki za domeno iz *seznam.domen* in model iz *seznam.modelov* vrne ime optimalne ocene zanesljivosti kot znakovni niz. Funkcija poda tudi vrednost korelacijskega koeficienta in njegovo stopnjo značilnosti.

```
> internal.cv.run('diabetes', 'rt')
diabetes rt CNK-p 0.410 0.006 +
```

B.5 Dostopnost algoritmov

Zadnja verzija opisanih algoritmov za okolje R je dosegljiva na zahtevo preko avtorjevega elektronskega naslova *zoran.bosnic@fri.uni-lj.si*.

V okviru nadaljnjega dela je v načrtu organizacija opisanih algoritmov v standardni paket za okolje R in prispevanje paketa v javno dostopno zbirko (*contributed packages*).