



Interakcije med atributi

v strojnem učenju

Aleks Jakulin

`jakulin@acm.org`

Fakulteta za računalništvo in informatiko

Univerza v Ljubljani

Interakcije

- Večina postopkov strojnega učenja predpostavi,
 - da so vsi atributi pogojno neodvisni (naivni Bayes, logistična regresija, linearni SVM, perceptron),
 - ali da so vsi atributi pogojno odvisni (klasifikacijska drevesa, konstruktivna indukcija, pravila, metode z jedri, metode najbližjih sosed).
- Zakaj pa dosežemo najboljše rezultate z mešanico teh dveh pristopov, kjer so nekatere skupine atributov neodvisne, druge pa odvisne?

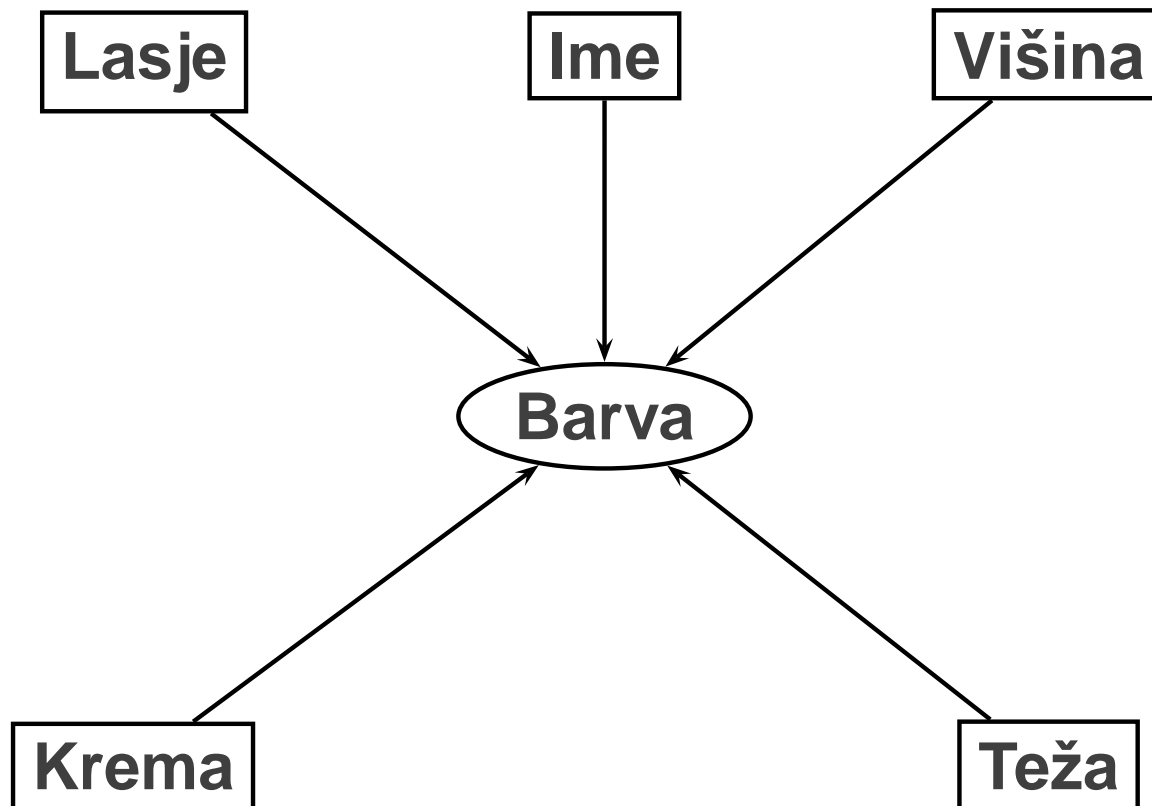
Analogije

Strojno učenje: interakcija je takrat, ko s hkratnim upoštevanjem skupine atributov dosežemo boljše rezultate kot s posamičnim.

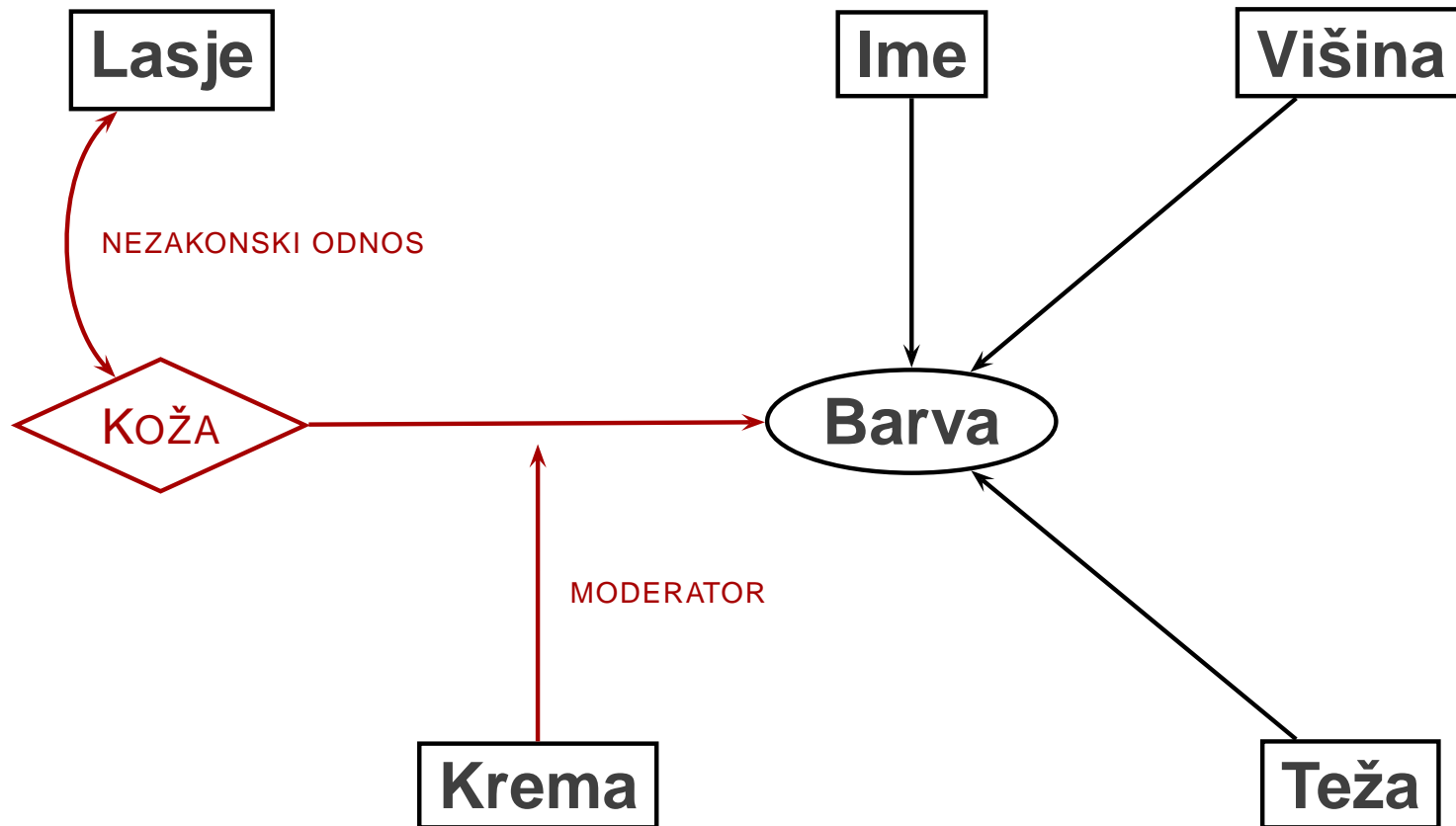
Ekonomija: dve podjetji lahko dosežeta skupaj slabše ali boljše rezultate, če se povežeta. Dva izdelka se lahko prodajata skupaj bolje (suknjič in hlače) ali slabše (Coca Cola in Pepsi Cola), kot ločeno.

Pravo: kako izračunamo število glasov neke države, da bo glasovanje pravično ne glede na možne koalicije med državami?

Glasovanje



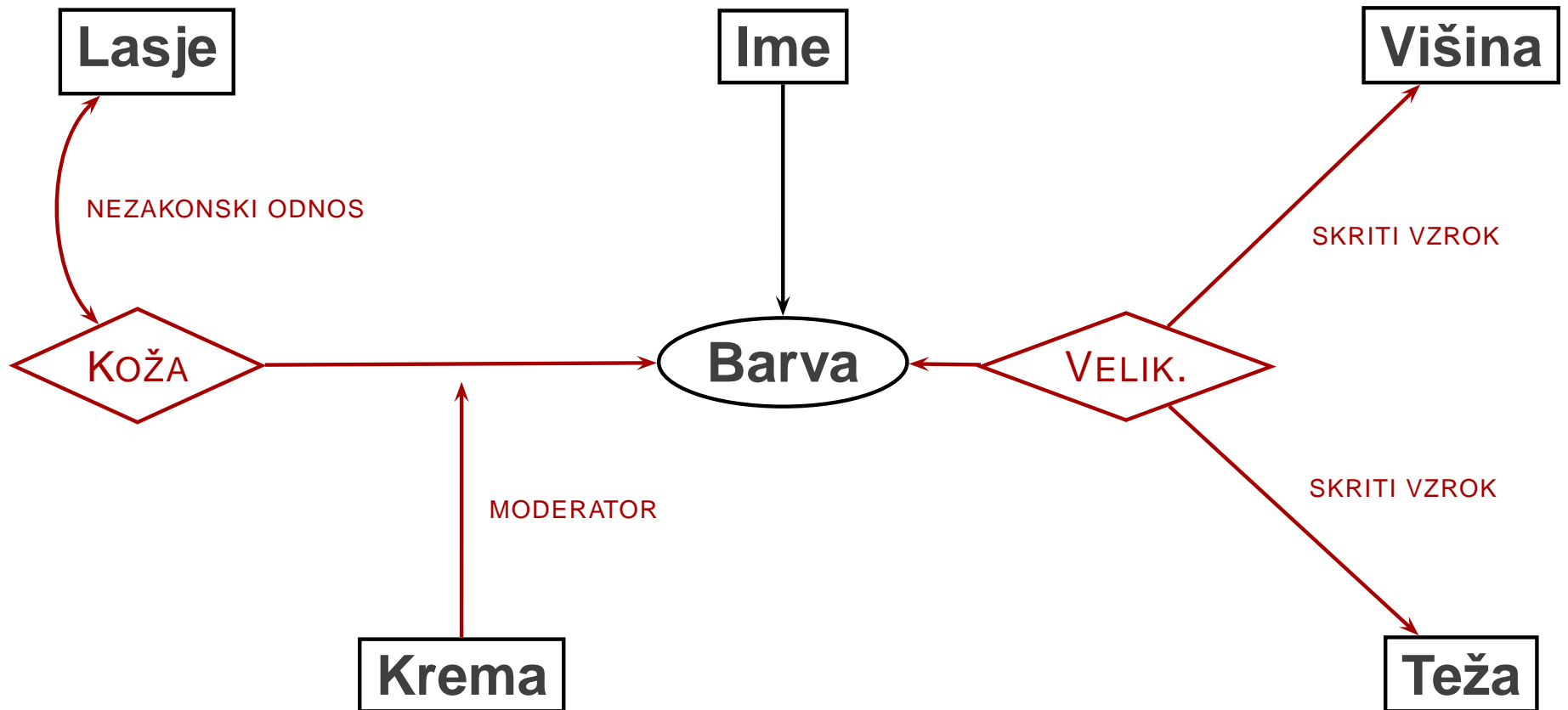
Glasovanje



TEMU
PRAVIMO:

PRAVA
INTERAKCIJA

Glasovanje

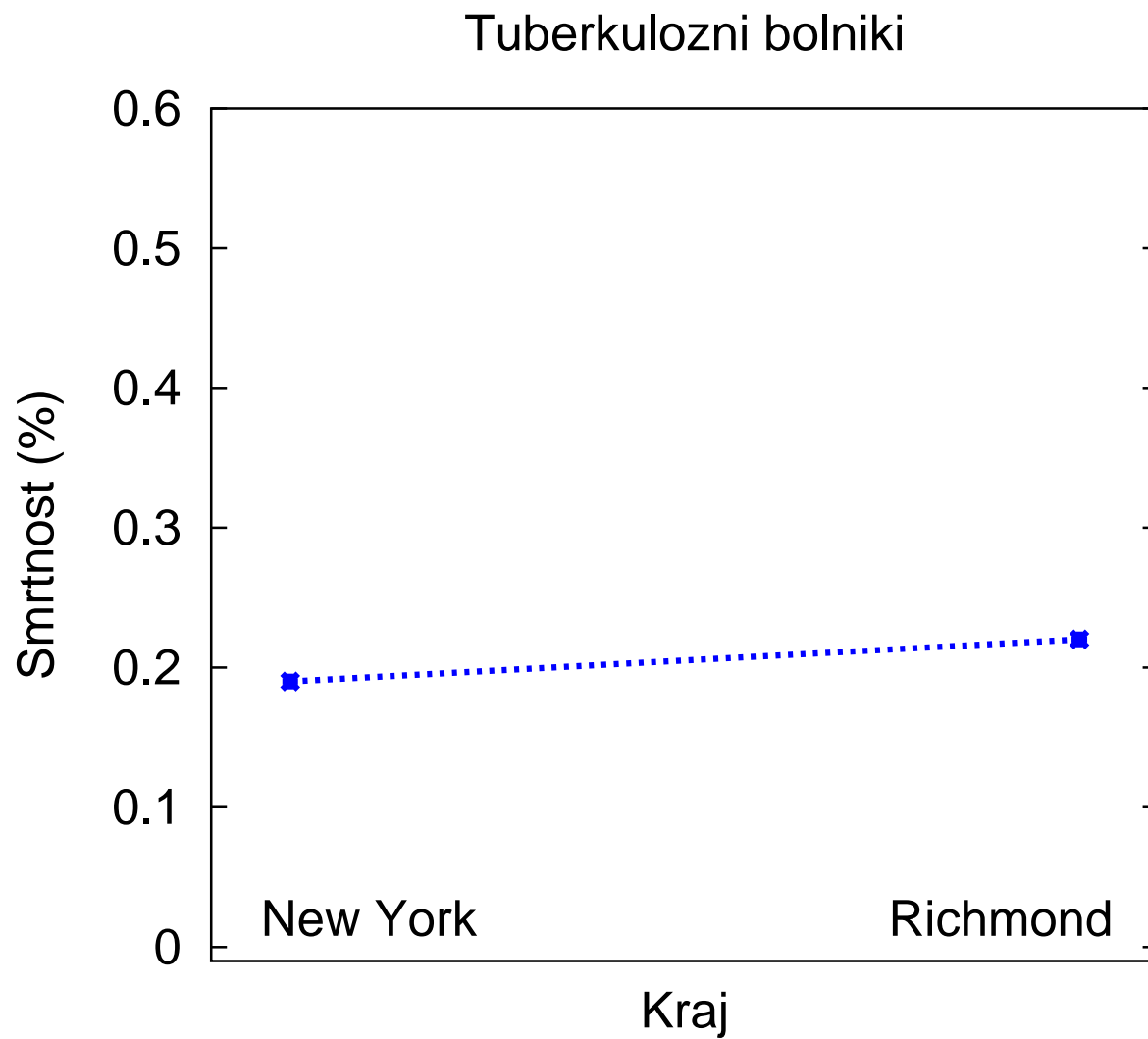


TEMU
PRAVIMO:

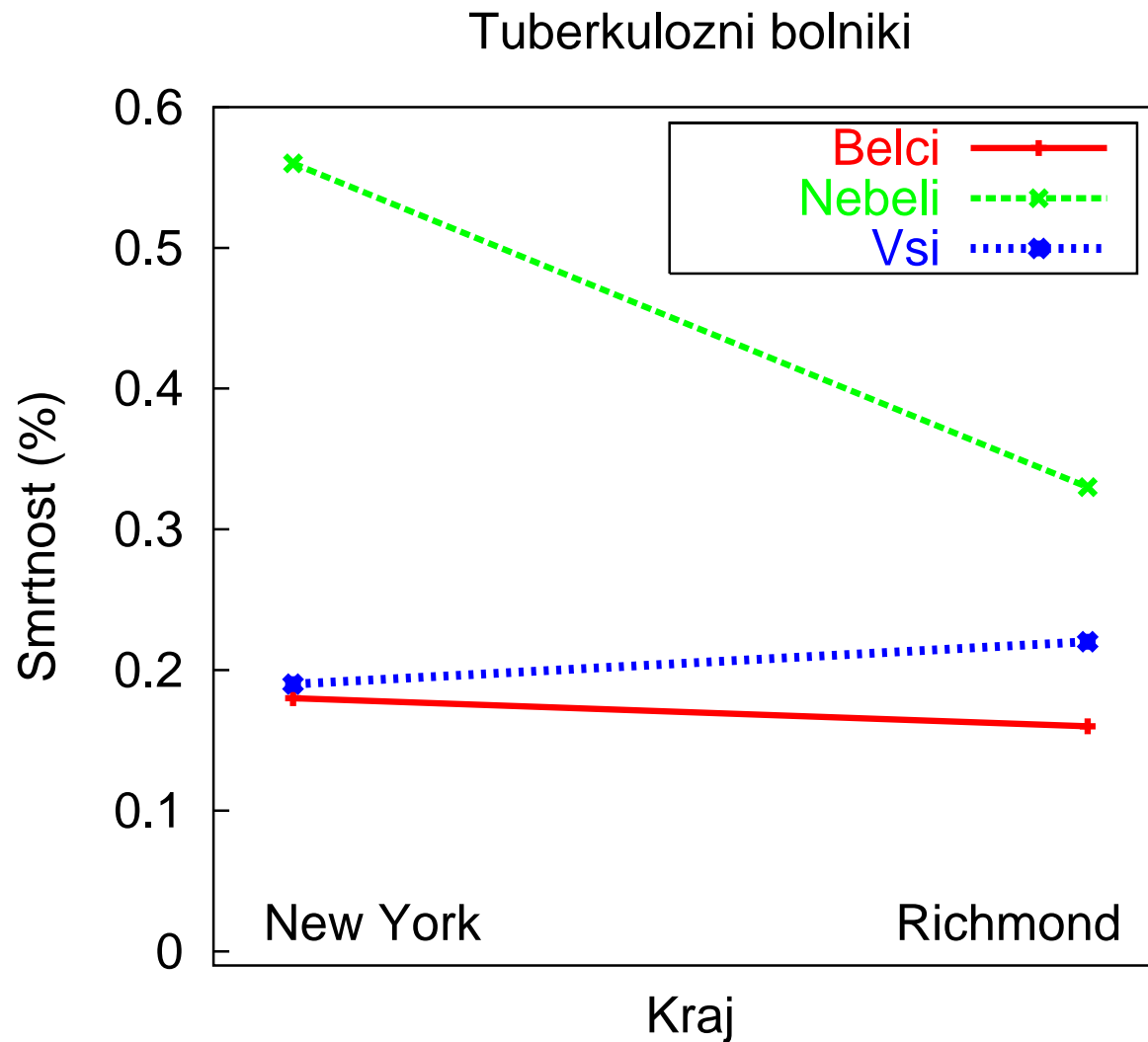
PRAVA
INTERAKCIJA

NEPRAVA
INTERAKCIJA

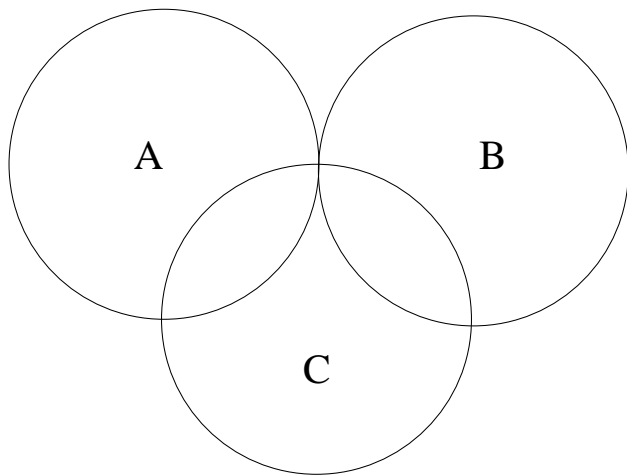
Simpsonov paradoks



Simpsonov paradoks

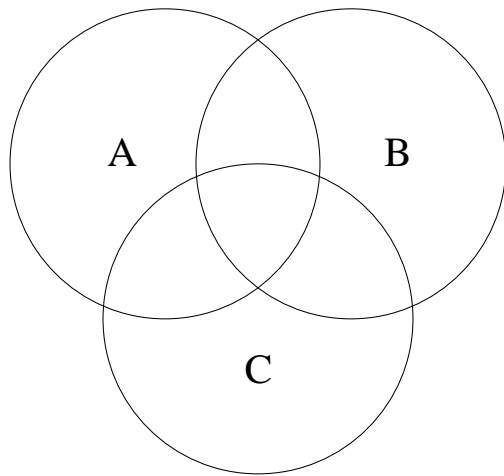


Informacijski prispevek



- Atribut in razredni atribut sta vira informacij. Želimo oceniti količino informacij, ki so skupne obem virom.
- Količino informacij o razrednem atributu C , ki jo dobimo s poznavanjem atributa A , meri *informacijski prispevek*:
$$\text{Gain}_C(A) := H(A) + H(C) - H(AC).$$
- Razlaga: naše nepoznavanje neznanega C se zmanjša za $\text{Gain}_C(A)$, ko poznamo A .
- Informacijski prispevek pove vse, če so vsi atributi pogojno neodvisni glede na razred in če so interakcije le reda 2.

Interakcijski prispevek

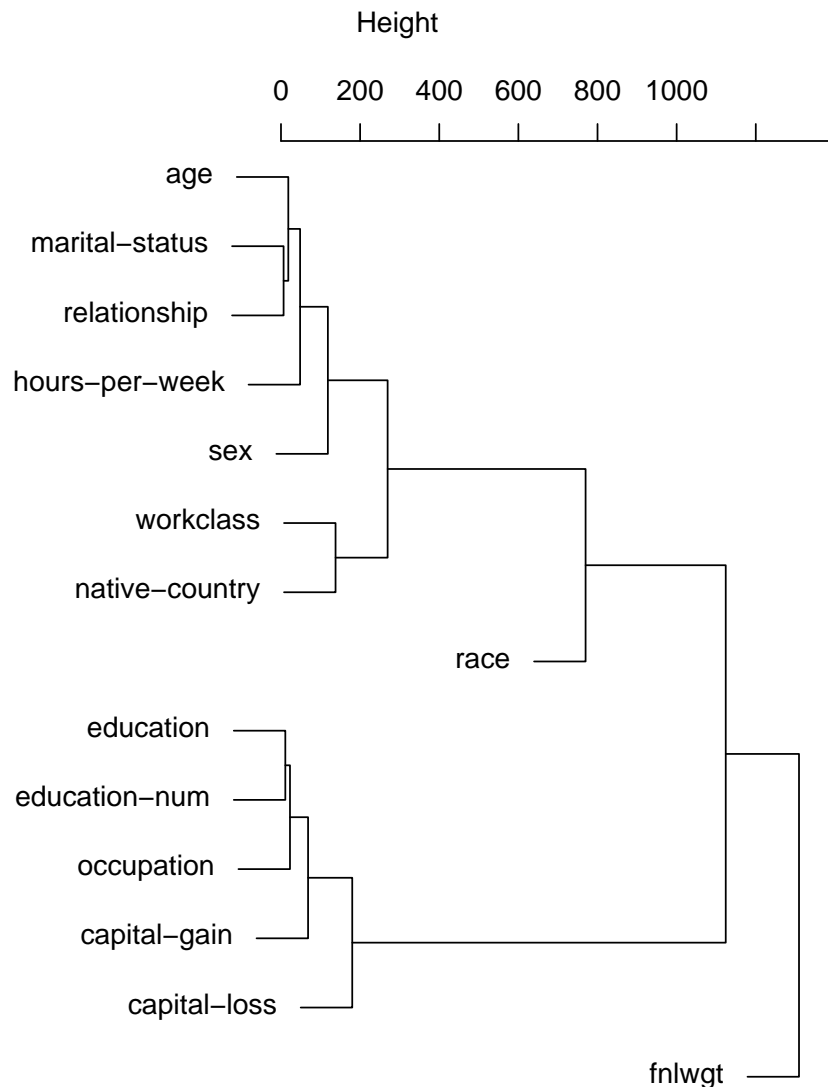


- Kako ocenimo informacijsko vsebino, ki je skupna trem virom informacij?
- Posplošitev informacijskega prispevka za interakcije reda 3 je *interakcijski prispevek*:

$$\begin{aligned}IG_3(ABC) &:= H(AB) + H(AC) + H(BC) - H(A) \\ &\quad - H(B) - H(C) - H(ABC) \\ &= \text{Gain}_C(AB) - \text{Gain}_C(A) - \text{Gain}_C(B).\end{aligned}$$

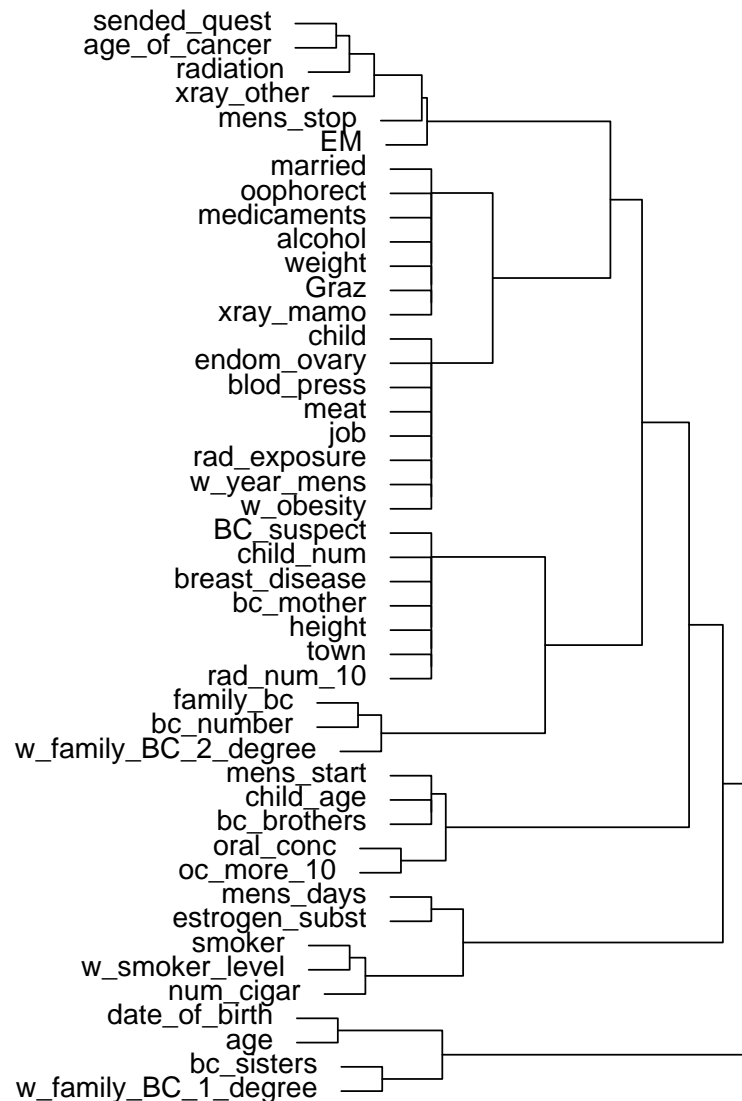
- Ko je IG negativen: soodvisnost
- Ko je IG pozitiven: sodejavnost
- Ko je IG nič: interakcija reda 3 med temi atributi ne obstaja.

Analiza soodvisnosti



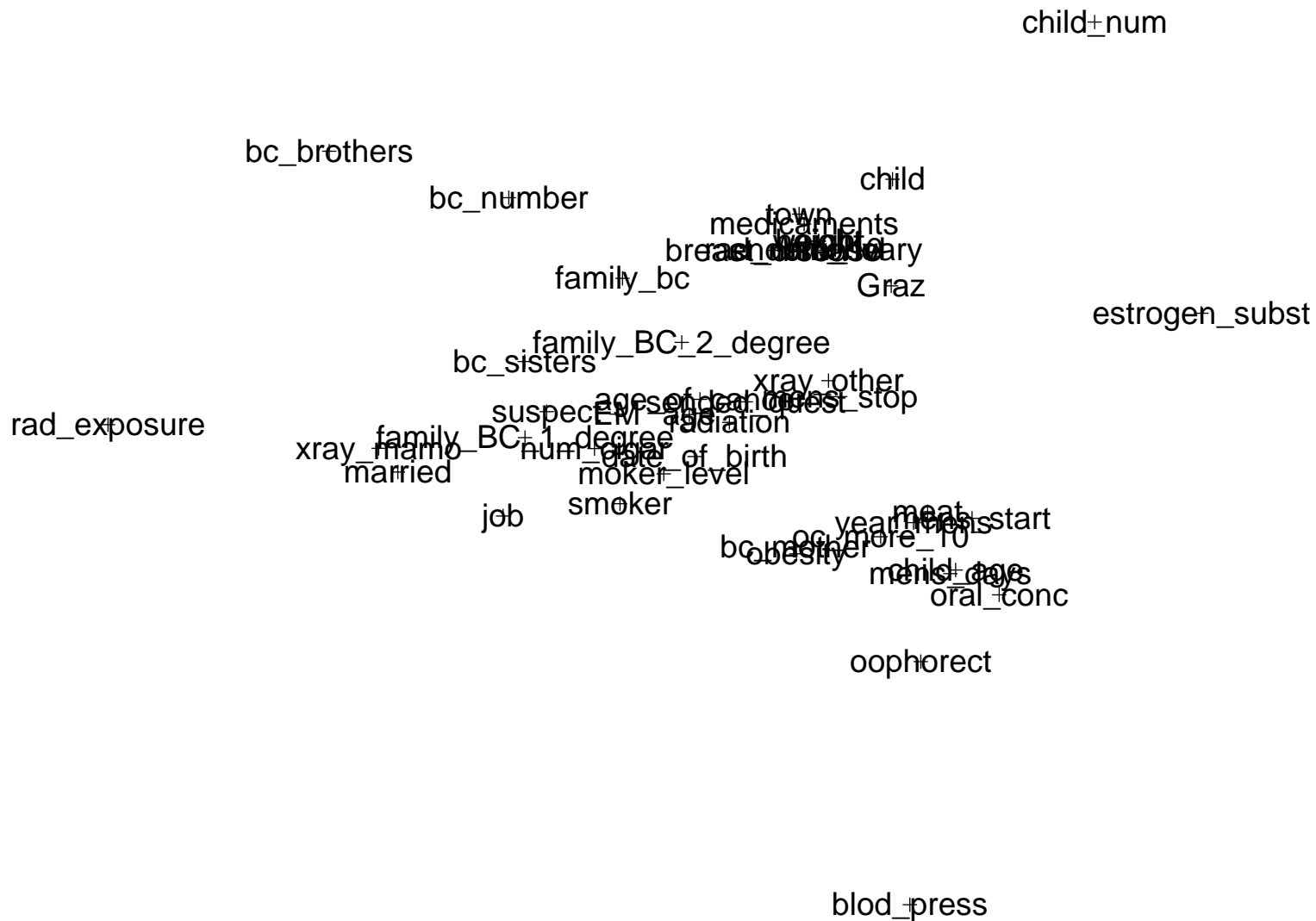
- Domena Census/Adult iz UCI, 2 razreda posameznikov, bogati in revni.
- Podobnost med atributoma je proporcionalna negiranemu interakcijskemu prispevku reda 3 med njima in razredom.
- Upoštevamo le soodvisnosti.
- Za prikaz matrike podobnosti smo uporabili hierarhično razvrščanje z združevanjem.

Analiza soodvisnosti

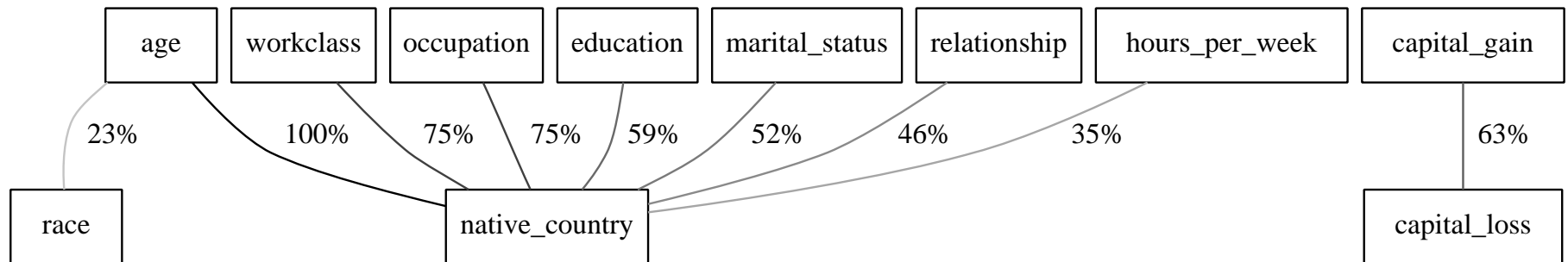


- Domena o raku dojke z Graza.
- Izrazita skupina atributov, ki opisujejo že zdravljene paciente.
- Več skupin atributov, ki označujejo nekatere rizične faktorje: družina, hormoni, kajenje, starost.

Analiza soodvisnosti

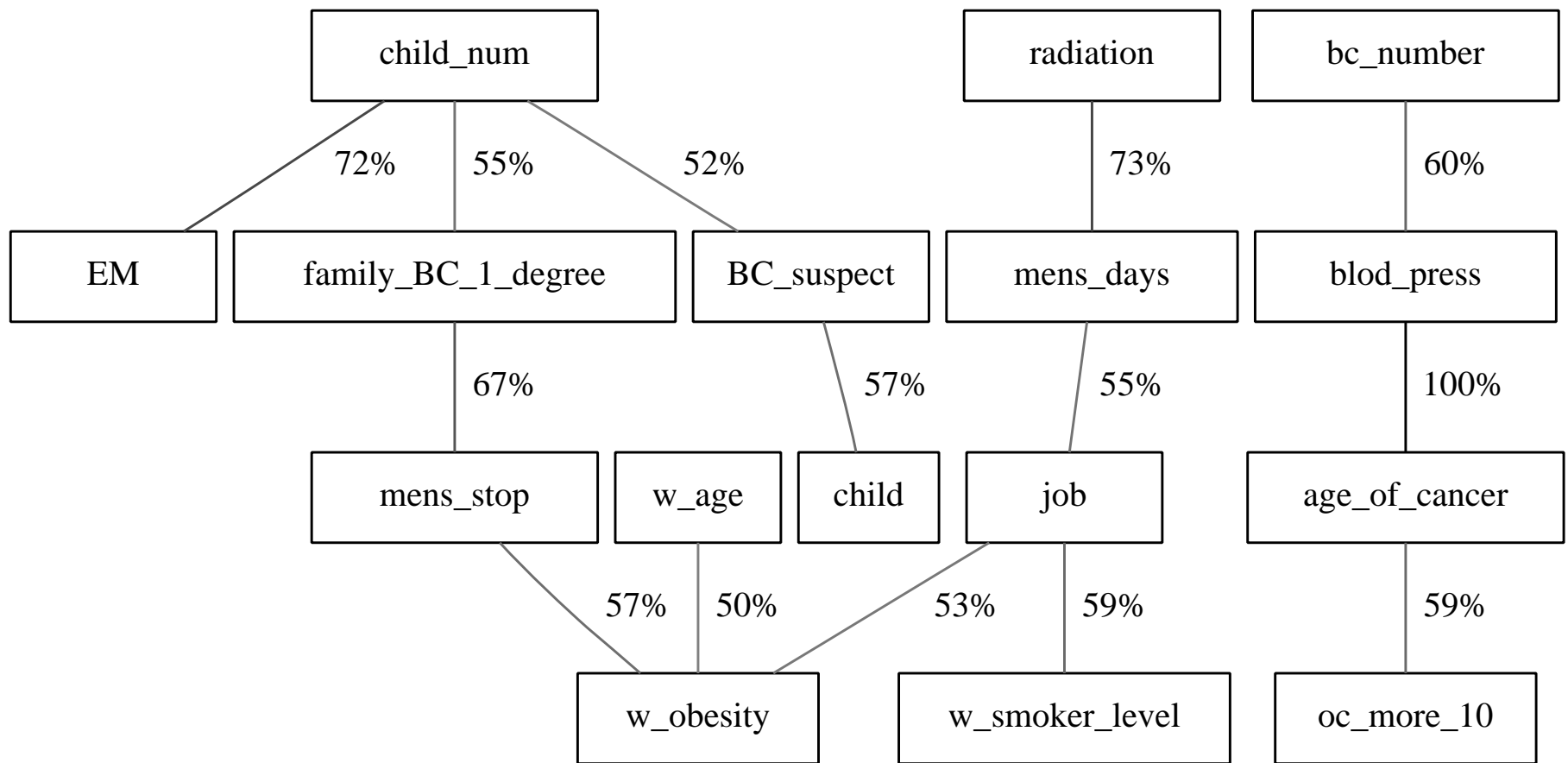


Analiza sodejavnosti



- Procent na povezavi označuje relativno moč sodejavnosti napram najmočnejši.
- Dežela izvora je močan moderator za veliko število interakcij reda 2.
- Sodejavnosti so redko tranzitivne.
- Sodejavnosti so gozd in ne drevo.

Analiza sodejavnosti



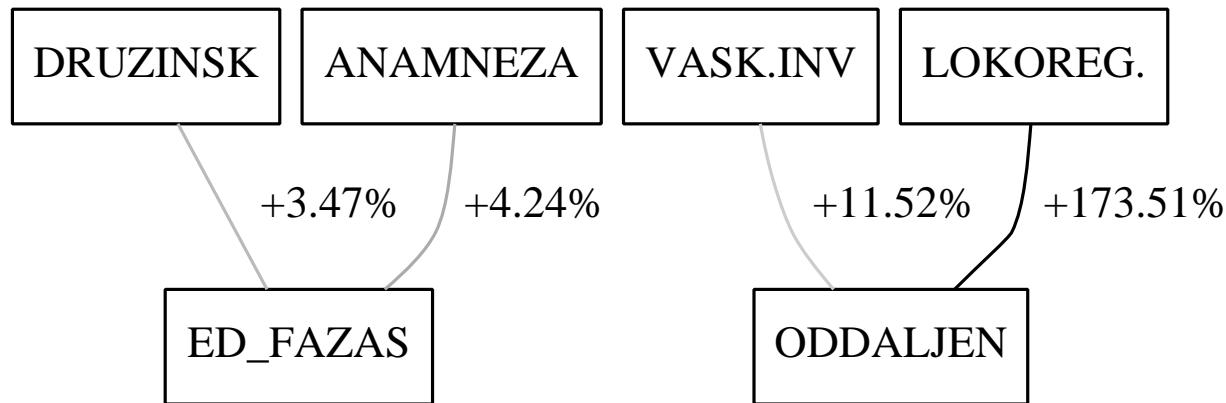
- Domena o raku dojke z Graza.

Pomembnost interakcij (1)

Kdaj je interakcija pomembna?

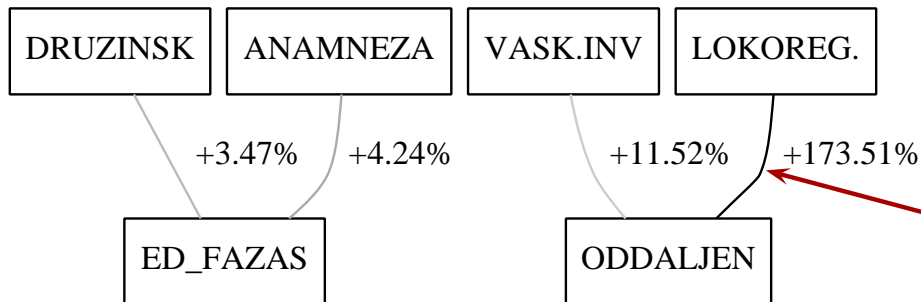
- Posebni testi za pogojno odvisnost in neodvisnost, npr. Cochran-Mantel-Haenszel.
- Ocena kvalitete naslednjih dveh klasifikatorjev na testni množici:
 - Glasovalni klasifikator, ki predpostavlja neodvisnost med atributoma.
 - Segmentacijski klasifikator, ki izkorišča odvisnost med atributoma z odpravljanjem interakcij.

Pomembnost interakcij (2)



- Ponavadi je pomembnih interakcij malo:
 - Manjša zanesljivost statistik.
 - Razdrobljenost primerov.
 - Neučinkovitost kartezičnega produkta.

Pomembnost interakcij (2)

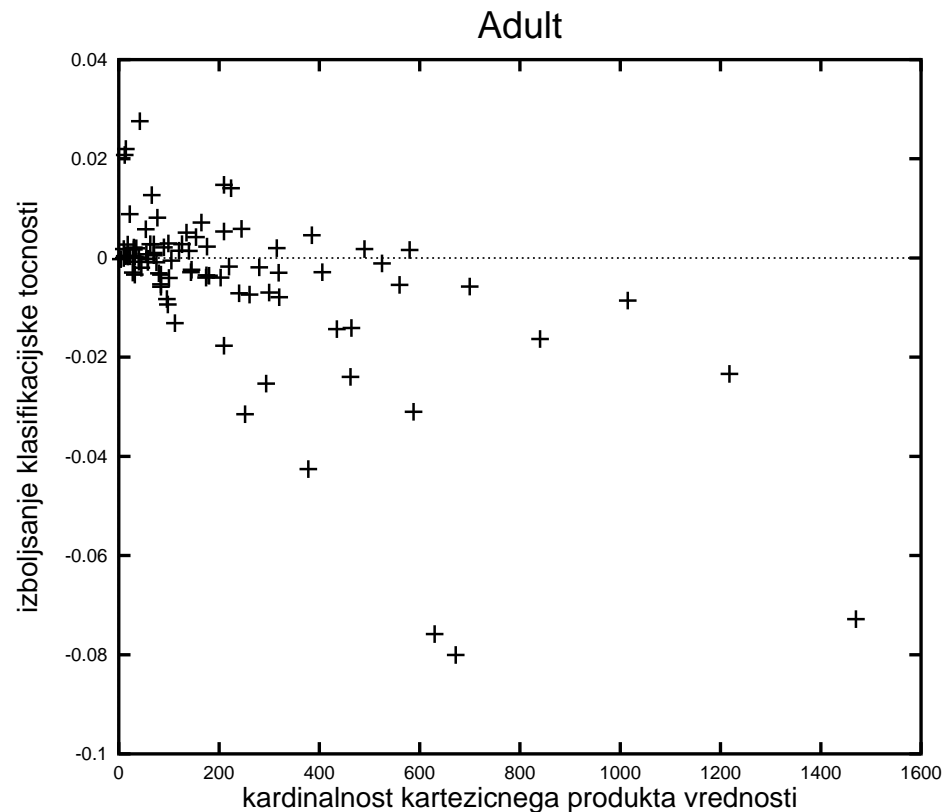


```
ODDALJEN > 0: y
ODDALJEN <= 0:
: . . . LOKOREG. <= 0: n
      LOKOREG. > 0: y
```

POPOLNO KLASIFIKACIJSKO DREVO,
DOBLJENO S C4.5.

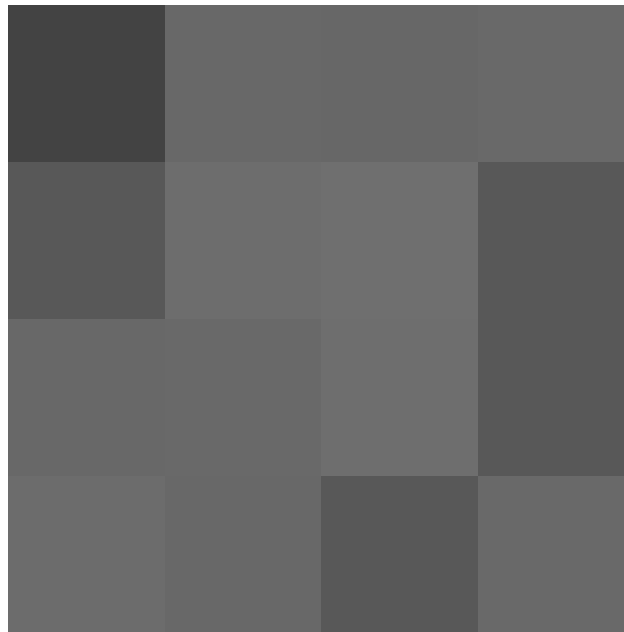
- Vendar so te važne: izbira in vrednotenje atributov, kombiniranje atributov v pravilih, drevesih in konstruktivni indukciji.

Kardinalnost atributov



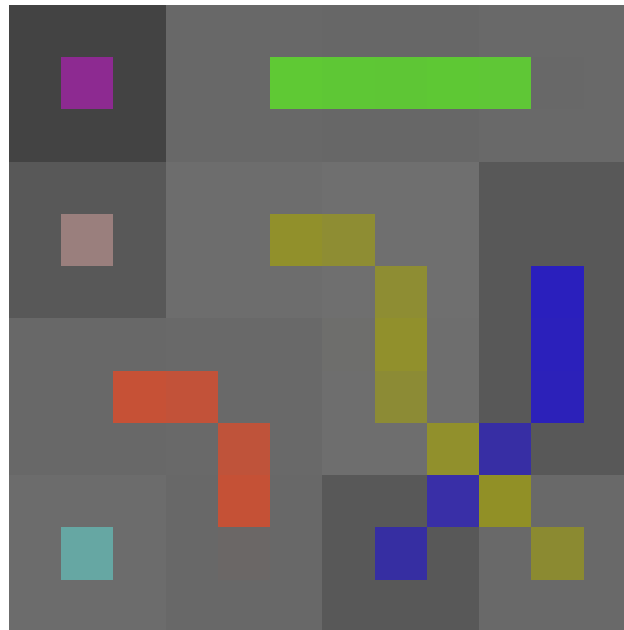
Večje, kot je število vrednosti atributov, manjša je možnost, da bo interakcija med njima pomembna.

Poenostavljanje atributov (1)



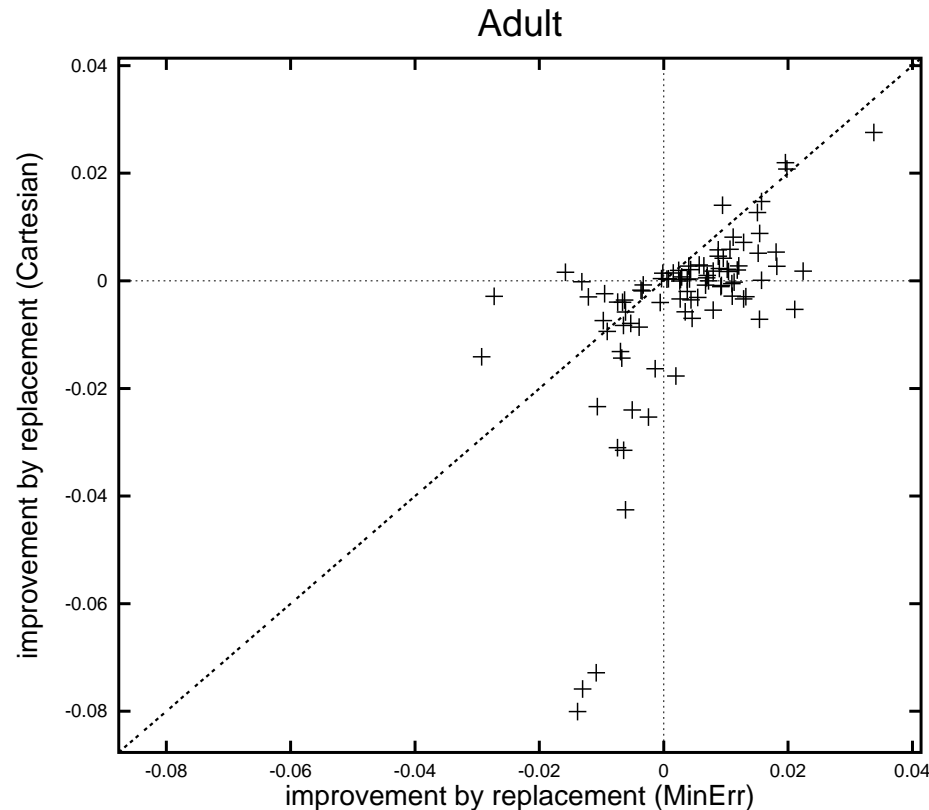
Dva atributa združimo v njunem kartezičnem produktu, a tega lahko poenostavimo z združevanjem podobnih vrednosti.

Poenostavljanje atributov (1)



Dva atributa združimo v njunem kartezičnem produktu, a tega lahko poenostavimo z združevanjem podobnih vrednosti.

Poenostavljanje atributov (2)



Odpravljanje interakcij z minimizacijo napake ponavadi da boljše rezultate kot nepoenostavljeno združevanje atributov v kartezičnem produktu.

Izboljšanje klasifikatorjev

'adult'	Orig.	Soodv.	Sodej.
NBC	0.416	0.352	0.392
LR	1.562	0.418	1.564
SVM	—	—	—

'breast'	Orig.	Soodv.	Sodej.
NBC	0.262	0.187	0.171
LR	0.016	0.016	0.016
SVM	0.032	0.032	0.016

Merska enota: Kullback-Leiblerjeva divergenca v natih.

Moderiranje klasifikatorjev z $\epsilon = 10^{-5}$.

- Iskanje po prostoru atributov, usmerjanje z interakcijskim prispevkom, ocenjevanje z ovojnico (wrapper).
- Uporabljeno odpravljanje interakcij z minimizacijo napake.
- Podatkov mora biti dovolj!

Algoritem ovojnice

```
 $f \leftarrow 0$  {Števec zmot}  
 $\mathcal{H} \leftarrow \{A_1, A_2, \dots, A_n\}$   
 $b \leftarrow q(L(\mathcal{H}))$  {Osnovni rezultat}  
 $\mathcal{I} \leftarrow \mathcal{H} \times \mathcal{H}$  {Vsi pari atributov}  
while  $f < N \wedge \mathcal{I} \neq \emptyset$  do  
   $\langle A, B \rangle \leftarrow \arg \max_{I \in \mathcal{I}} IG_3(I, C)$   
   $\mathcal{I} \leftarrow \mathcal{I} \setminus \{\langle A, B \rangle\}$  {Odpravi to interakcijo z novim atributom}  
   $\hat{\mathcal{H}} \leftarrow R(\mathcal{H}, \langle A, B \rangle)$   
   $\hat{b} \leftarrow \hat{q}(L, \hat{\mathcal{H}})$   
  if  $\hat{b} > b$  then {Je rezultat te množice boljši?}  
     $f \leftarrow 0$   
     $\mathcal{H} \leftarrow \hat{\mathcal{H}}$   
     $b \leftarrow \hat{b}$   
  else  
     $f \leftarrow f + 1$   
  end if  
end while  
return  $L(\mathcal{H})$ 
```


Uporaba interakcij

● Napovedovanje:

- Odpravljanje pomembnih interakcij prispeva k boljši klasifikacijski točnosti.
- Interakcije omejujejo in preprečujejo kratkovidnost pri diskretizaciji in izbiri atributov.
- Interakcije upravičijo konstruktivno indukcijo.

● Preučevanje:

Interakcije so zanimive, še posebej tiste nepričakovane: interakcije med simptomi, zdravili, ipd.

Povzetek

- Interakcije delimo na sodejavnosti (prave) in soodvisnosti (neprave). Interakcijski prispevek odkriva interakcije reda 3 ter njihov tip.
- Pragmatični test pomembnosti interakcij, ki temelji na ocenjevanju kvalitete dveh različnih klasifikatorjev na testnih podatkih.
- Metodologija za analizo interakcij v domeni, z grafi sodejavnosti in dendrogrami soodvisnosti.
- Izboljšani rezultati algoritmov strojnega učenja z odpravljanjem interakcij.

Nadaljnje delo

- Orodje za interakcijsko analizo.
- Podpora numeričnim in urejenim atributom.
- Podpora interakcijam višjih redov.
- Izboljšane metode za odpravljanje interakcij, še posebej soodvisnosti.
- Raziskava vpliva interakcij na izbiro delitev v klasifikacijskih drevesih ter pri diskretizaciji.
- Uporaba.



Negotovost

v strojnem učenju

Aleks Jakulin

`jakulin@acm.org`

Fakulteta za računalništvo in informatiko

Univerza v Ljubljani

Nedeterminizem

- Negotovost
- Neznanje
- Nepredvidljivost

Vse oblike nedeterminizma opišemo z verjetnostnimi porazdelitvami. Če sprejmemo nefrekvenčno interpretacijo verjetnosti, ne potrebujemo mehke logike. Popolni klasifikatorji za nedeterministične domene so negotovi.

**KLASIFIKATORJEVA NEGOTOVOST NAJ SE KOT OCENA
ČIMBOLJ PRIBLIŽA NJEGOVEMU DEJANSKEMU NEZNANJU!**

Odločanje z negotovostjo

Optimalna odločitev z najmanjšim tveganjem:

$$\arg \min_{\tilde{c} \in \mathcal{D}_C} \sum_{\hat{c} \in \mathcal{D}_C} \Pr\{d(\mathbf{i}) = \hat{c}\} M(\tilde{c}, \hat{c}).$$

i testni primer, $M(\textit{predicted}, \textit{actual})$ cenovna matrika, d stohastični oz. negotovi klasifikator, \mathcal{D}_C množica razredov.

Ne potrebujemo posebnih cenovno-občutljivih postopkov učenja: zadoščajo le negotovi klasifikatorji.

Verjetnostne porazdelitve

Model preslika nek deskriptor v verjetnost:

- Neparometrični diskretni histogrami
- Parametrične porazdelitve

Ko je možnih porazdelitev več:

- MaxEnt: najvarnejša izmed danih porazdelitev.
- MinEnt: najprofitabilnejša izmed danih porazdelitev.
- Bayesovsko združevanje porazdelitev na podlagi vnaprejšnjih (apriori) porazdelitev modelov.
- Novo: porazdelitve višjih redov, porazdelitve porazdelitev.

Funkcije cenilke (1)

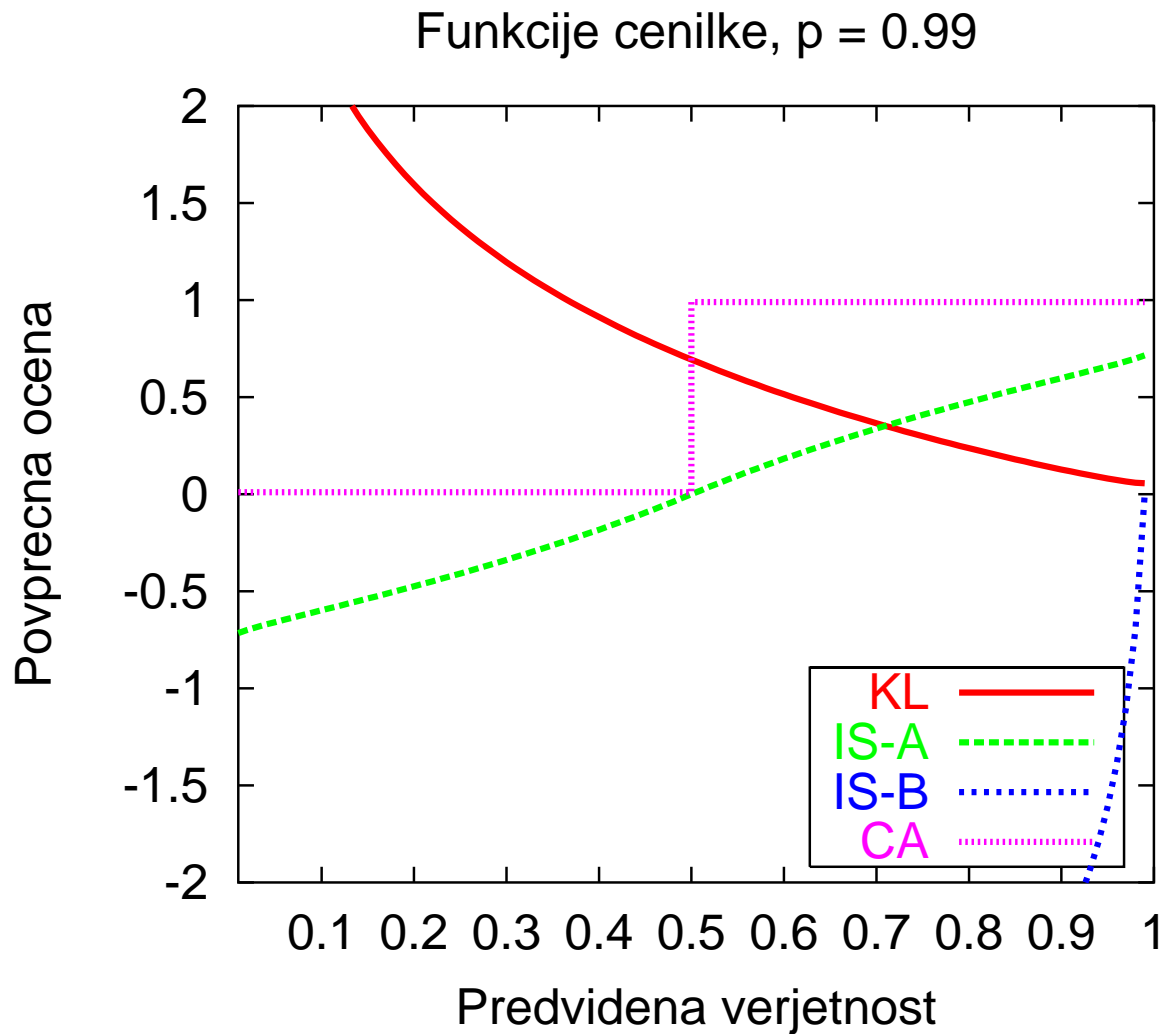
Iščemo najboljši negotovi klasifikator, ne da bi vedeli za funkcijo koristnosti. Najbolj negotovi so nekoristni. Najbolj gotovi so tvegani.

Možnost je *Kullback-Leiblerjeva divergenca*:

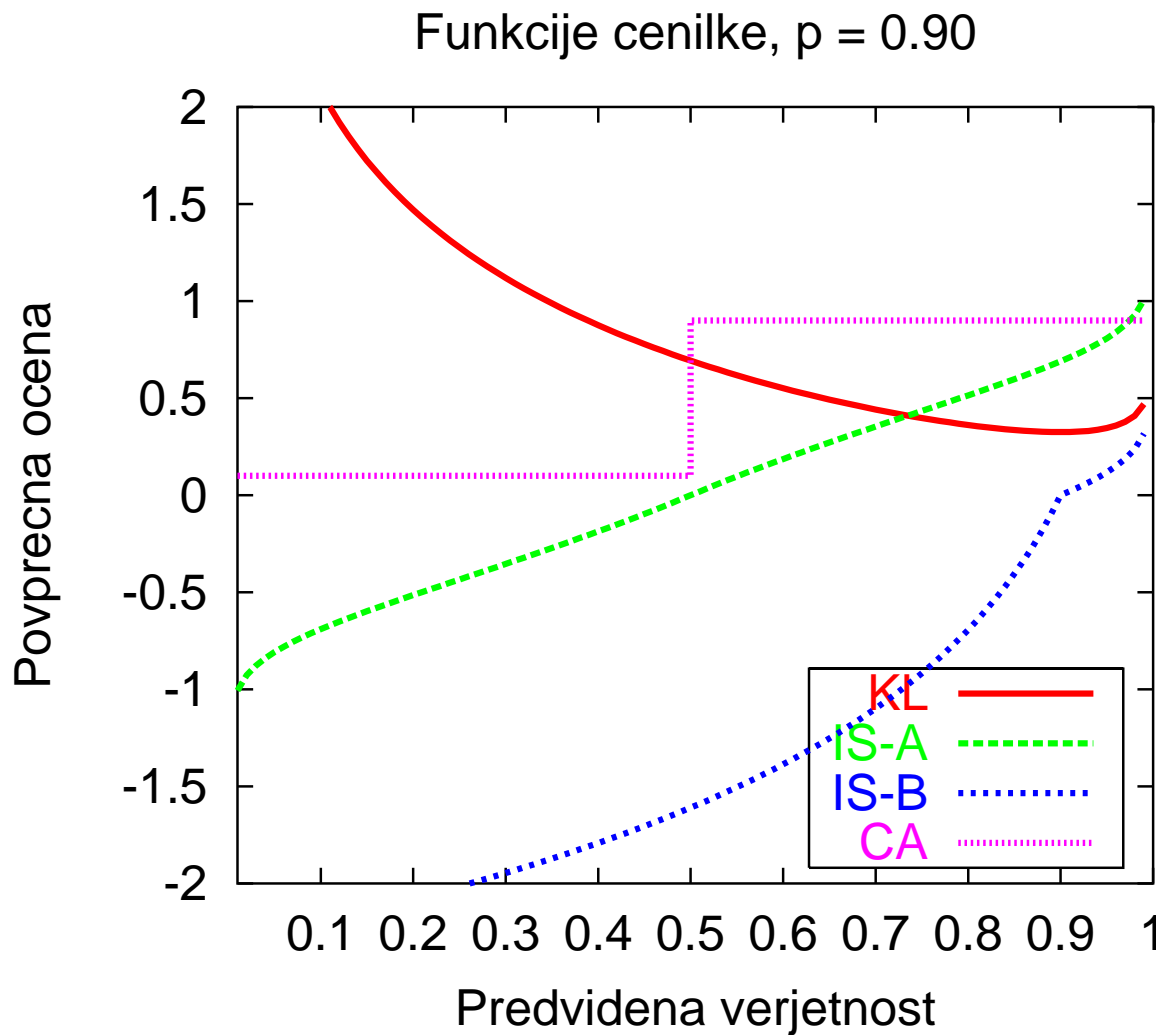
$$D(P||Q) := \sum_{c \in \mathcal{D}_C} P(C(\mathbf{i}) = c) \log \frac{P(C(\mathbf{i}) = c)}{\Pr\{d(\mathbf{i}) = c\}}.$$

Dejanska verjetnostna porazdelitev razredov za dan primer \mathbf{i} je $P = P(C(\mathbf{i})|\mathbf{i})$, klasifikator d pa predvideva, da je $Q = \Pr\{d(\mathbf{i})\}$.

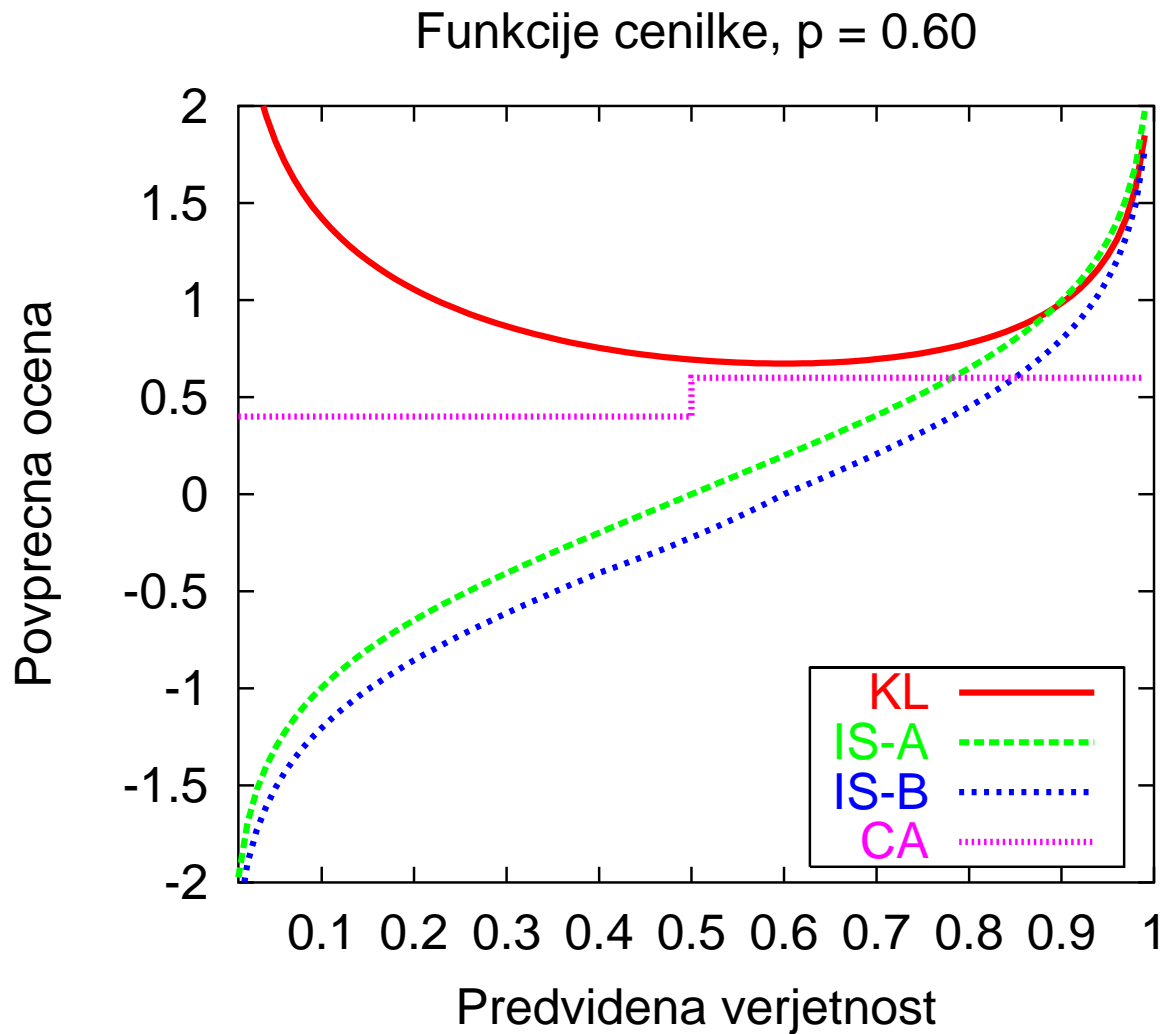
Funkcije cenilke (2)



Funkcije cenilke (2)



Funkcije cenilke (2)



Funkcije cenilke (3)

Če poznamo M , ocenimo *obžalovanje* (regret), izgubo zaradi nepopolnosti negotovosti kot ocene nepredvidljivosti:

$$R(d, \mathbf{i}) := \left(\sum_{c \in \mathcal{D}_C} P(C(\mathbf{i}) = c) M \left[\left(\arg \min_{\tilde{c} \in \mathcal{D}_C} \sum_{\hat{c} \in \mathcal{D}_C} \Pr\{d(\mathbf{i}) = \hat{c}\} \hat{M}(\tilde{c}, \hat{c}) \right), c \right] \right) - \min_{\tilde{c} \in \mathcal{D}_C} \sum_{c \in \mathcal{D}_C} P(C(\mathbf{i}) = c) M(\tilde{c}, c).$$

Enostavnejša oblika za situacije, kjer je razred znan (Expected Value of Perfect Information):

$$\text{EVPI} := M(d_o(\mathbf{i}), C(\mathbf{i})) - M(C(\mathbf{i}), C(\mathbf{i})).$$

Gradniki klasifikatorjev

- Ocenjevanje
- Delitev
- Projekcija
- Glasovanje

Delitev:

Prostor učnih primerov razdelimo na več disjunktnih področij.

- Naivni Bayes: delimo po vsakem atributu. Delitev je kolikor je atributov.
- Klasifikacijska drevesa: delimo po vozliščih. Delitev je le ena.
- Pravila: vsako pravilo pokriva del prostora učnih primerov.

Ocenjevanje:

Za neko skupino primerov ocenimo verjetnostno porazdelitev razredov.

- Klasifikacijska drevesa: v vsakem listu.
- Pravila: za vsako pravilo.
- Naivni Bayes: za vsako vrednost atributa.
- Diskretizacija: za vsako podskupino vrednosti nekega zveznega atributa.

Glasovanje:

Če imamo več verjetnostnih porazdelitev razredov, jih združimo v eno samo. Lahko izberemo eno od porazdelitev, ali jih več združimo v eno samo.

- Naivni Bayes: vsak atribut ponudi porazdelitev.
- Klasifikacijski gozdovi: vsako drevo ponudi porazdelitev.
- Skupine klasifikatorjev: vsak klasifikator ponudi porazdelitev.
- Izbira atributov: nekatere attribute izločimo iz glasovanja.
- Tehtanje atributov: atributom pripišemo število glasov, ki jih imajo.

Projekcija:

Vrednosti skupine atributov opišemo z zveznim deskriptorjem. Porazdelitev razredov v odvisnosti od deskriptorja lahko opišemo tudi s parametrično verjetnostno porazdelitvijo.

- Linearna diskriminanta: hiperravnina ločuje primere različnih razredov.
 - Podporni vektorji: uporabimo stopničasto porazdelitev.
 - Logistična regresija: uporabimo logistično porazdelitev.
- Najbližji sosedi: deskriptor je oddaljenost testnega primera do učnih primerov

Naivni Bayesov klasifikator

brez interakcij

$$V \begin{pmatrix} E[\mathcal{T}, C, S(A_1)] \\ E[\mathcal{T}, C, S(A_2)] \\ E[\mathcal{T}, C, S(A_3)] \end{pmatrix}$$

z interakcijo med A_1, A_2

$$V \begin{pmatrix} E[\mathcal{T}, C, S(A_1, A_2)] \\ E[\mathcal{T}, C, S(A_3)] \end{pmatrix}$$

- Za negotove klasifikatorje so interakcije definirane s potrebo po uporabi delitve,
- za diskriminativne klasifikatorje pa s pojavom linearne neseparabilnosti.

Povzetek

- Klasifikatorji naj ocenijo svojo negotovost.
- Funkcije cenilke ocenjujejo kvaliteto negotovih klasifikatorjev. Če imamo podan M , uporabimo obžalovanje, sicer pa KL divergenco.
- Klasifikatorje gradimo z uporabo funkcij ocenjevanja, delitve, projekcije in glasovanja.
- Interakcije so primer prekletstva dimenzionalnosti: prekletstvo hkratnega obravnavanja skupine atributov.