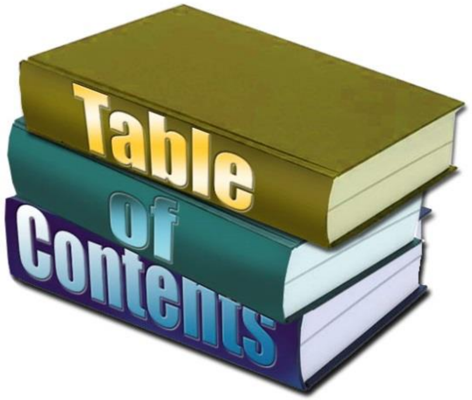
OSNOVE UMETNE INTELIGENCE 2018/19

ocenjevanje učenja

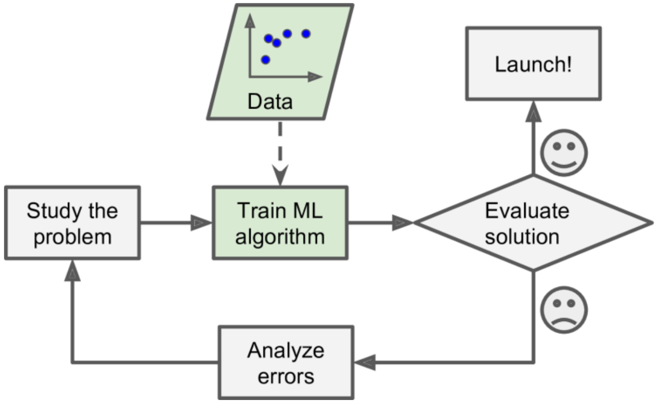
naivni Bayes

Pregled



- strojno učenje
  - ocenjevanje učenja
  - naivni Bayesov klasifikator
  - nomogrami za naivni Bayes

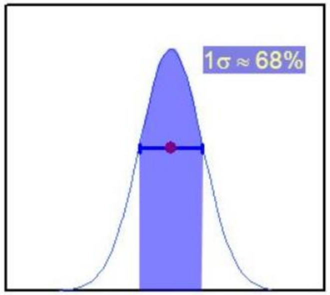
Ocenjevanje učenja

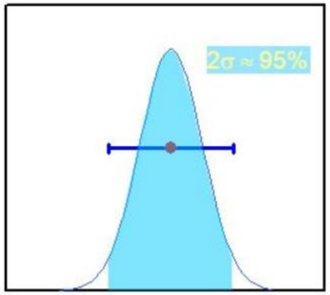


- kriteriji za ocenjevanje hipotez:
  - točnost (angl. *accuracy*)
  - kompleksnost (angl. *complexity*)
  - razumljivost (angl. *comprehensibility*) – subjektivni kriterij

- ocenjevanje točnosti:
  - na **učnih** podatkih (angl. *training set, learning set)*
  - na **testnih** podatkih (angl. *testing set, test set*)
    - izločimo del učnih podatkov, s katerimi simuliramo ne-videne podatke
    - želimo si, da je testna množica reprezentativna za nove podatke
    - uporabimo lahko **intervale zaupanja** v oceno uspešnosti na testni množici, ki upoštevajo število testnih primerov
  - na **novih** (ne-videnih) podatkih (angl. *new data, unseen data*)
    - na njih bo naučeni sistem dejansko deloval

Ocenjevanje učenja





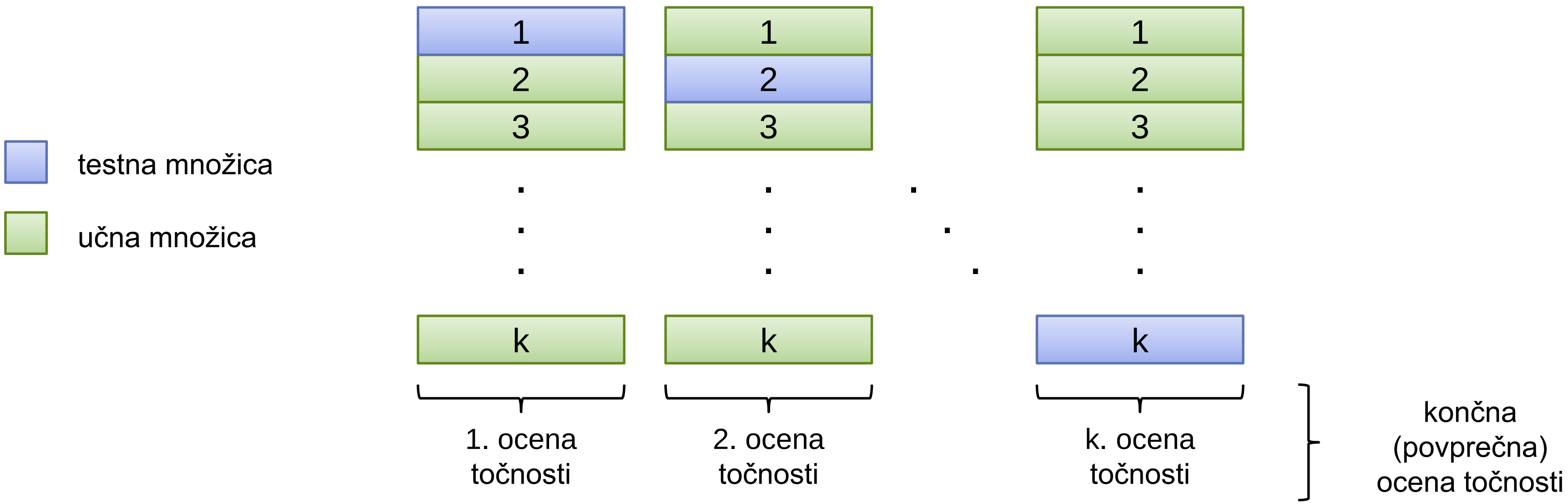
- nasprotujoča si cilja:
  - potrebujemo čim več podatkov za **uspešno učenje**
  - potrebujemo čim več podatkov za **zanesljivo ocenjevanje točnosti** (večje število testnih primerov nam daje ožji interval zaupanja v oceno točnosti)

- rešitev:
  - kadar je učnih podatkov dovolj, lahko izločimo **testno množico** (angl. *holdout test set*)
  - alternativa: **večkratne delitve** na učno in testno množico

- različni načini **vzorčenja testnih primerov**:
  - naključno, nenaključno (npr. prečno preverjanje)
  - poljubno ali stratificirano (zagotovimo enako porazdelitev razredov kot v učni množici)

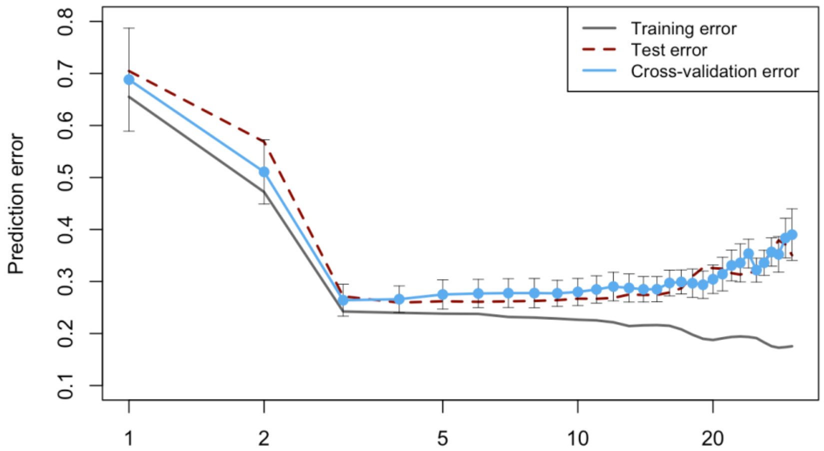
Prečno preveranje

- poseben primer večkratnega učenja in testiranja
- k-kratno prečno preverjanje (angl. *k-fold cross-validation*):
  - celo učno množico razbij na *k* disjunktnih podmnožic
  - za vsako od *k* podmnožic:
    - uporabi množico kot testno množico
    - uporabi preostalih *k-1* množic kot učno množico
  - povpreči dobljenih *k* ocen točnosti v končno oceno

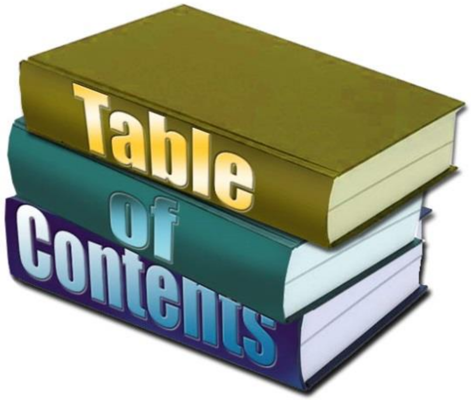


Prečno preveranje

- v praksi najpogosteje: k=10 (10-kratno prečno preverjanje)
- vplive izbranega razbitja podatkov na podmnožice lahko zmanjšamo tako, da tudi prečno preverjanje večkrat (npr. 10x) ponovimo (torej 10×10=100 izvajanj učnega algoritma) in rezultate povprečimo
- poseben primer prečnega preverjanja je metoda **izloči enega** (angl. leave-one-out, LOO)
  - *k* je enak številu primerov (vsaka testna množica ima samo en primer)
  - najbolj stabilna ocena glede učinkov razbitja na podmnožice
  - časovno zelo zamudno, primerno za manjše množice
- iz meritev na vseh podmnožicah je možno izračunati tudi varianco/ intervale zaupanja



Pregled



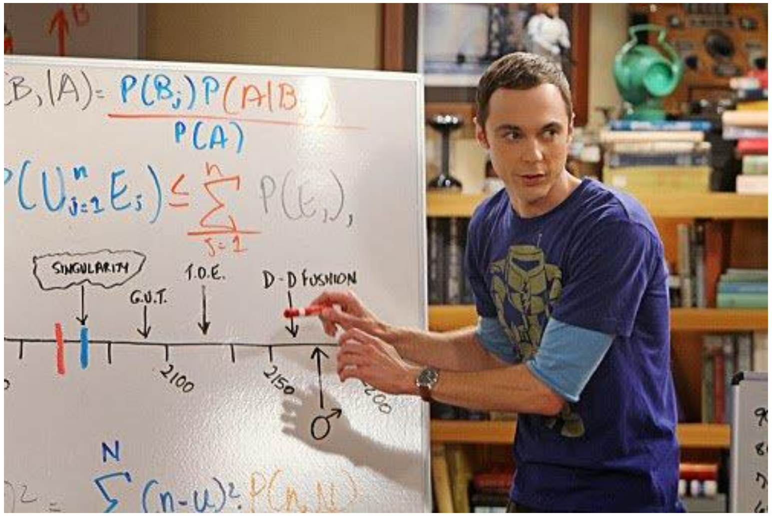
- strojno učenje
  - ocenjevanje učenja
  - naivni Bayesov klasifikator
  - nomogrami za naivni Bayes

Naivni Bayesov klasifikator

- Thomas Bayes, 1702 – 1761
- opomnik iz teorije o verjetnosti:

$$P(AB) = P(A|B) \cdot P(B)$$
$$P(AB) = P(B|A) \cdot P(A)$$
$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$
$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

Bayesovo pravilo

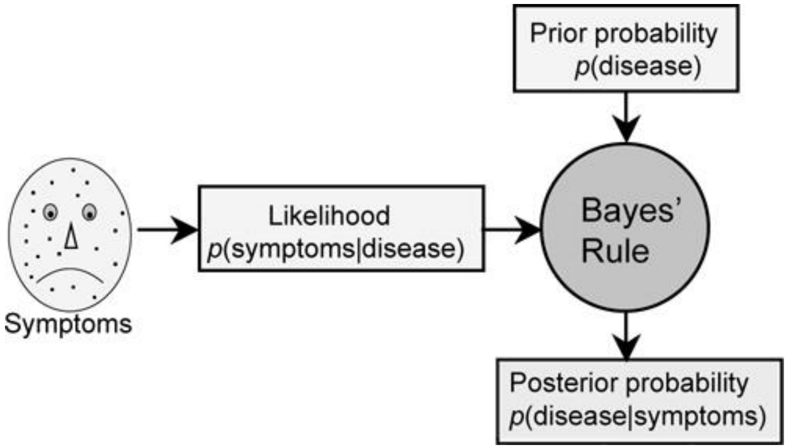


Naivni Bayesov klasifikator

- aplikacija v medicini:

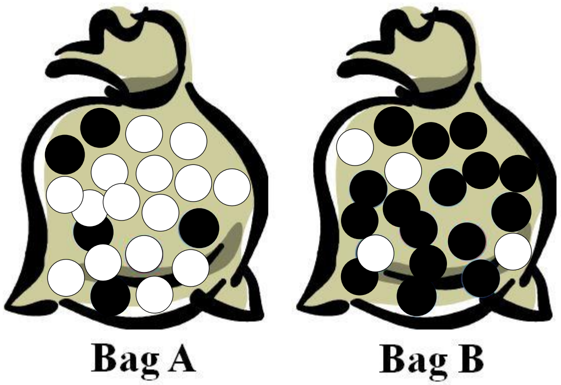
$$P(hipoteza|opažanje) = \frac{P(opažanje|hipoteza) \cdot P(hipoteza)}{P(opažanje)}$$

- zdravniki razpolagajo z vzročno in statistično informacijo:
  - verjetnost izraženih simptomov pri neki bolezni - $P(opažanje|hipoteza)$
  - verjetnost določene bolezni - $P(hipoteza)$
  - verjetnost določenega simptoma - $P(opažanje)$
- Bayesovo pravilo nam izraža **diagnostično pogojno verjetnost** $P(hipoteza|opažanje)$ na podlagi **vzročne pogojne verjetnosti** $P(opažanje|hipoteza)$



Vaja

- dve vrsti vrečk s frnikulami:
  - 4 vrečke tipa A (vsaka 5 črnih, 15 belih frnikul)
  - 1 vrečka tipa B (16 črnih, 4 bele frnikule)



- možna vprašanja:
  - Kakšna je verjetnost, da je to vrečka tipa B?
    $P(B) = ?$
  - Kakšna je verjetnost, da naključno izberemo črno frnikulo, če izbiramo iz vrečke tipa B?
    $P(\check{C}|B) = ?$
  - Naključno izberemo eno izmed vrečk in iz nje naključno izberemo frnikulo. Kakšna je verjetnost, da smo izbrali črno frnikulo iz vrečke tipa B?
    $P(B\check{C}) = P(B) \cdot P(\check{C}|B) = ?$
  - Naključno izberemo eno izmed vrečk in iz nje naključno izberemo frnikulo. Kakšna je verjetnost, da smo izbrali črno frnikulo?
    $P(\check{C}) = P(B) \cdot P(\check{C}|B) + P(A) \cdot P(\check{C}|A)$

Vaja

- Ena vrečka ima poškodovan ovoj tako, da se skozi njega vidi črna frnikula. Kakšna je verjetnost, da je to vrečka tipa B?
  $P\left(B \middle| \check{C}\right) = ?$
- B = hipoteza, Č = evidenca, opažanje
- verjetnost $P(B|\check{C})$ lahko določimo iz drugih bolj očitnih verjetnosti z Bayesovo formulo:

$$P\left(B \middle| \check{C}\right) = \frac{P(B) \cdot P\left(\check{C} \middle| B\right)}{P\left(\check{C}\right)}$$

- $P(B) = \frac{1}{5} = 0{,}2$
  $P\left(\check{C} \middle| B\right) = \frac{16}{20} = 0{,}8$
  $P\left(\check{C}\right) = \frac{4 \cdot 5 + 1 \cdot 16}{5 \cdot 20} = 0{,}444$
- $P\left(B \middle| \check{C}\right) = \frac{0{,}2 \cdot 0{,}8}{0{,}444} = 0{,}360$

dve vrsti vrečk s frnikulami:

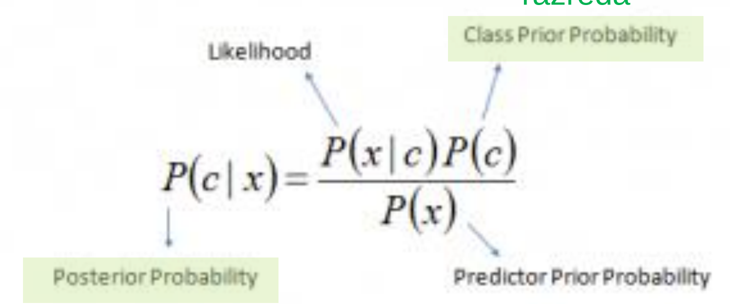
- 4 vrečke tipa A (vsaka 5 črnih, 15 belih frnikul)
- 1 vrečka tipa B (16 črnih, 4 bele frnikule)

Naivni Bayes v strojnem učenju

- evidenca → atributi
  hipoteza → razred
- zanima nas, kakšna je verjetnost razreda C pri podanih vrednostih atributov $A_1 = X_1, A_2 = X_2, \dots, A_n = X_n$:

$$P(C|X_1X_2 \dots X_n) = \frac{P(C) \cdot P(X_1X_2 \dots X_n|C)}{P(X_1X_2 \dots X_n)}$$

apriorna verjetnost razreda



aposteriorna verjetnost napovedi

Naivni Bayes v strojnem učenju

$$P(C|X_1X_2\ldots X_n) = \frac{P(C)\cdot P(X_1X_2\ldots X_n|C)}{P(X_1X_2\ldots X_n)}$$

- $P(X_1X_2\ldots X_n|C) = P(X_1|C)\cdot P(X_2\ldots X_n|X_1C) =$
  $= P(X_1|C)\cdot P(X_2|X_1C)\cdot P(X_3\ldots X_n|X_1X_2C) =$
  $= P(X_1|C)\cdot P(X_2|X_1C)\cdot P(X_3|X_1X_2C)\cdot\ldots\cdot P(X_n|X_1X_2\ldots X_{n-1}C)$
- $P(X_1X_2\ldots X_n) = P(X_1|X_2\ldots X_n)\cdot P(X_2|X_3\ldots X_n)\cdot\ldots\cdot P(X_{n-1}|X_n)\cdot P(X_n)$

- potrebujemo veliko število pogojnih verjetnosti, katerih poznavanje je v praksi težavno
- število kombinacij pogojnih verjetnosti je glede na zaloge vrednosti atributov $X_1X_2\ldots X_n$ eksponentno
- praktična rešitev: naivni Bayesov klasifikator

Naivni Bayes v strojnem učenju

- predpostavimo, da so atributi med seboj **verjetnostno neodvisni** in poenostavimo:
  $P(X_1 X_2 \ldots X_n|C) = P(X_1|C) \cdot P(X_2|X_1 C) \cdot \ldots \cdot P(X_n|X_1 X_2 \ldots X_{n-1} C)$
  $P(X_1 X_2 \ldots X_n) = P(X_1|X_2 \ldots X_n) \cdot P(X_2|X_3 \ldots X_n) \cdot \ldots \cdot P(X_{n-1}|X_n) \cdot P(X_n)$

$$\Downarrow$$

$P(X_1 X_2 \ldots X_n|C) \approx P(X_1|C) \cdot P(X_2|C) \cdot \ldots \cdot P(X_n|C)$
$P(X_1 X_2 \ldots X_n) \approx P(X_1) \cdot P(X_2) \cdot \ldots \cdot P(X_{n-1}) \cdot P(X_n)$

- približki so dobri, če so atributi med seboj dovolj neodvisni
- velja torej:

$$P(C|X_1 X_2 \ldots X_n) = \frac{P(C) \cdot P(X_1 X_2 \ldots X_n|C)}{P(X_1 X_2 \ldots X_n)} = \frac{P(C) \cdot \prod_i P(X_i|C)}{\prod_i P(X_i)}$$

konstanten člen, ki je neodvisen od ciljne spremenljivke (če opazujemo samo relativne velikosti napovedi različnih razredov, ga lahko izpustimo)

Naivni Bayes v strojnem učenju

- Bayesov klasifikator: primer **klasificiramo v razred, ki je najbolj verjeten**:

$$h(X_1 X_2 \dots X_n) = \mathrm{argmax}_k\, P(C_k) \cdot \prod_{i=1}^{n} P(X_i | C_k)$$

- **učenje**: ocenimo verjetnosti $P(C_k)$ in $P(X_i|C_k)$ za vse razrede C_k in vrednosti atributov X_i
- **napovedovanje**: uporabimo zgornjo enačbo za napovedovanje razreda novim primerom

- *opomba: s poenostavitvijo formule in izpustitvijo imenovalca izgubimo verjetnostno interpretacijo (verjetnosti razredov se ne seštejejo več v 1). Problem rešujemo npr. z normalizacijo rezultatov.*

Primer

- Zajeli smo podatke za 1000 sadežev, ki so lahko bodisi: *banana*, *pomaranča* ali *drugi sadež* (= vrednosti **razreda**). Za vsakega izmed sadežov smo izmerili, ali je *podolgovat*, *sladek* in *rumen* (= **atributi**). Meritve smo zapisali v tabelo:

sadež	**podolgovat**		**sladek**		**rumen**		**skupaj**
	da	**ne**	**da**	**ne**	**da**	**ne**	
banana	400	100	350	150	450	50	500
pomaranča	0	300	150	150	300	0	300
drugo	100	100	150	50	50	150	200
	500	500	650	350	800	200	1000

- iz tabele lahko razberemo različne verjetnosti, npr.:
  - verjetnosti razredov: $P(banana) = \frac{500}{1000} = 0{,}5$, $P(pomaranča) = 0{,}3$, $P(drugo) = 0{,}2$
  - pogojne verjetnosti: $P(dolg|banana) = \frac{4}{5} = 0{,}8$

    $P(sladek|banana) = 0{,}7$

    $P(rumen|banana) = 0{,}9$

A simple explanation of Naive Bayes Classification, https://stackoverflow.com/questions/10059594/a-simple-explanation-of-naive-bayes-classification

Primer

sadež	podolgovat		sladek		rumen		skupaj
	da	ne	da	ne	da	ne	
banana	400	100	350	150	450	50	500
pomaranča	0	300	150	150	300	0	300
drugo	100	100	150	50	50	150	200
	500	500	650	350	800	200	1000

- Imamo sadež, ki ni podolgovat, ni sladek, je pa rumen. Kateri sadež je to?

$$P(banana|neP, neS, daR)$$
$$= P(banana) \cdot P(neP|banana) \cdot P(neS|banana) \cdot P(daR|banana) =$$
$$= \frac{500}{1000} \cdot \frac{100}{500} \cdot \frac{150}{500} \cdot \frac{450}{500} = 0{,}5 \cdot 0{,}2 \cdot 0{,}3 \cdot 0{,}9 = 0{,}027$$

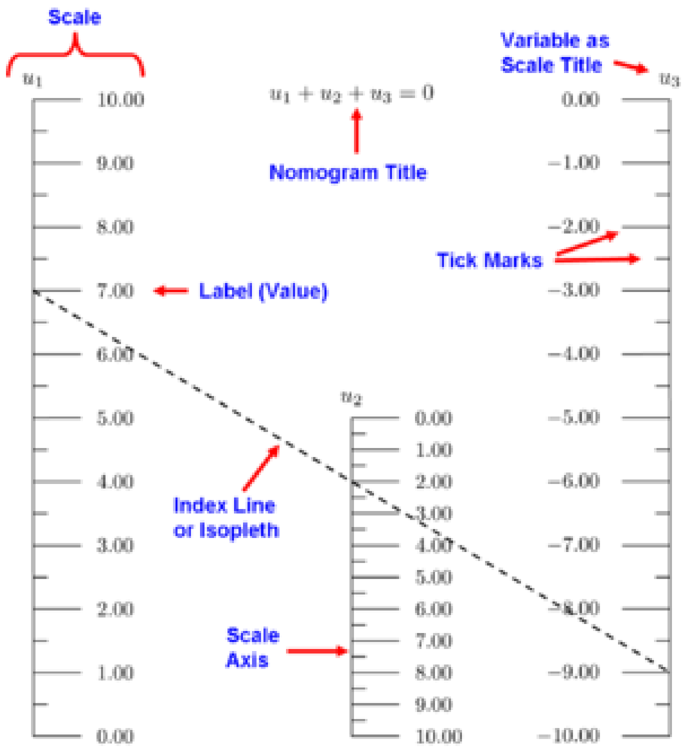
$$P(pomaranča|neP, neS, daR) = 0{,}3 \cdot 1 \cdot 0{,}5 \cdot 1 = 0{,}15 \longleftarrow$$

ta sadež je najverjetneje pomaranča

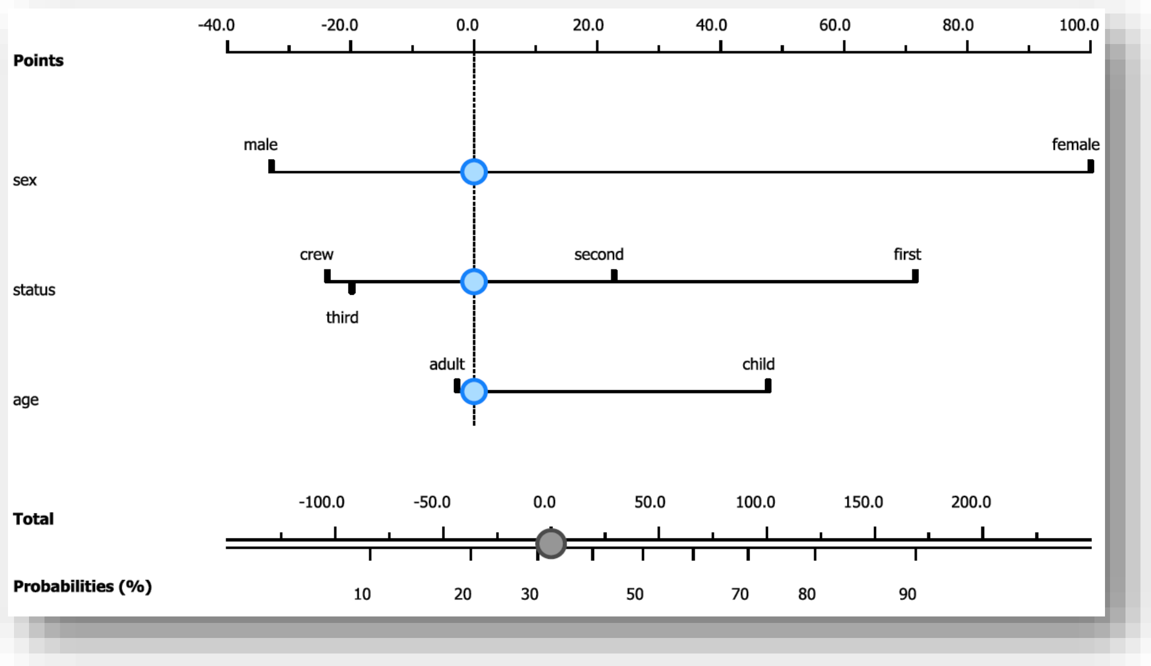
$$P(drugo|neP, neS, daR) = 0{,}2 \cdot 0{,}5 \cdot 0{,}25 \cdot 0{,}25 = 0{,}00625$$

Nomogrami

- pristop za vizualizacijo naivnega Bayesovega modela
- prikazuje:
  - pomembnost posameznih vrednosti vsakega atributa na ciljni razred
  - pomembnost posameznih atributov na ciljni razred
  - vizualno razlago napovedanih verjetnosti (brez kalkulatorja)
- "nomogram":
  - je *grafična upodobitev numeričnih odnosov med spremenljivkami*
  - omogoča uporabniku grafično pridobiti rezultat brez računanja
- uporaba:
  - matematika (iskanje vrednosti funkcij)
  - zdravniki v medicini (napovedovanje bolezni – npr. infarkta ali raka na podlagi vhodnih atributov)



Primer - ideja



- vsaka vrednost atributa doprinaša določeno število točk k ciljnemu razredu
- točke vseh vrednosti atributov seštejemo v skupno vsoto točk, ki je povezana z verjetnostjo ciljnega razreda
- razpon posameznih točk vsakega atributa govori o pomembnosti atributa za napovedovanje ciljnega razreda (zgoraj urejeni od najbolj do najmanj pomembnega)

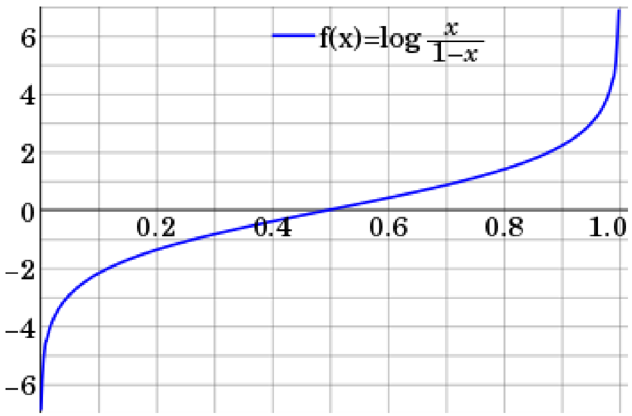
Izračun nomograma

- verjetje razreda pri naivnem Bayesu:

$$h(C|X_1X_2 \dots X_n) = P(C) \cdot \prod_{i=1}^{n} P(X_i|C)$$

- na zgornjem pravilu uporabimo logistično funkcijo (verjetnosti z intervala [0,1] preslikamo na interval $[-\infty, \infty]$, uporabimo logaritme)

$$\text{logit}\, P = \log \frac{P}{1-P}$$



- $\text{logit}\, h(C|X_1X_2 \dots X_n) =$

$$\log \frac{P(C) \cdot \prod_{i=1}^{n} P(X_i|C)}{1 - P(C) \cdot \prod_{i=1}^{n} P(X_i|C)} = \log \frac{P(C) \cdot \prod_{i=1}^{n} P(X_i|C)}{P(\bar{C}) \cdot \prod_{i=1}^{n} P(X_i|\bar{C})} = log \frac{P(C)}{P(\bar{C})} + \log \frac{\prod_{i=1}^{n} P(X_i|C)}{\prod_{i=1}^{n} P(X_i|\bar{C})}$$

$$= logit\; P(C) + \sum_{i} \log \frac{P(X_i|C)}{P(X_i|\bar{C})}$$

razmerje verjetja
(odds ratio)

Izračun nomograma

razmerje verjetja (odds ratio)

- $\operatorname{logit} P(C) + \sum_i \log \frac{P(X_i|C)}{P(X_i|\bar{C})} = \operatorname{logit} P(C) + \sum_i \log OR(X_i)$

- edino razmerje verjetja je odvisno od vrednosti atributov X_i, torej ga lahko uporabimo za "točkovanje" doprinosa atributa:

$$točke(C|X_i) = \log OR(X_i) = \log \frac{P(X_i|C)}{P(X_i|\bar{C})}$$

- skupno število točk za verjetnost celotnega primera:

$$točke(C|X_1 X_2 \dots X_n) = \sum_i \log OR(X_i) = \sum_i \log \frac{P(X_i|C)}{P(X_i|\bar{C})}$$

- Kako izračunati $OR(X_i)$? Po Bayesovem pravilu (znova) velja:

$$\frac{P(X_i|C)}{P(X_i|\bar{C})} = \frac{\frac{P(C|X_i) \cdot P(X_i)}{P(C)}}{\frac{P(\bar{C}|X_i) \cdot P(X_i)}{P(\bar{C})}} = \frac{\frac{P(C|X_i)}{P(C)}}{\frac{P(\bar{C}|X_i)}{P(\bar{C})}} = \frac{\frac{P(C|X_i)}{P(\bar{C}|X_i)}}{\frac{P(C)}{P(\bar{C})}}$$

Primer

- učna množica titanic, 2201 učnih primerov (711 preživelih – razred YES, 1490 umrlih – razred NO)

atribut	vrednost	razred = YES	razred = NO
status	first	203	122
	second	118	167
	third	178	528
	crew	212	673
age	adult	654	1438
	child	57	52
sex	male	367	1364
	female	344	126
razred		711	1490

- Kako konstruiramo nomogram?
- Kako lahko vizualiziramo odločitev za **odraslega moškega**, ki je potoval v **drugem** razredu?

primer iz Možina, Martin, et al. "Nomograms for visualization of naive Bayesian classifier." European Conference on Principles of Data Mining and Knowledge Discovery. Springer, Berlin, Heidelberg, 2004.

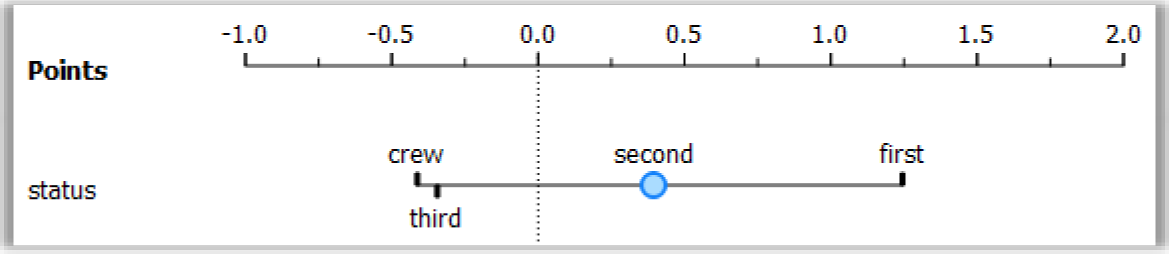
Konstrukcija nomograma

$$točke(yes|status = first) = \log\frac{\frac{P(first|yes)}{P(first|no)}}{\frac{P(yes)}{P(no)}} = \log\frac{\frac{203}{122}}{\frac{711}{1490}} = \log\frac{1{,}66}{0{,}48} = 1{,}25$$

$$točke(yes|status = second) = \log\frac{\frac{P(second|yes)}{P(second|no)}}{\frac{P(yes)}{P(no)}} = \log\frac{\frac{118}{167}}{\frac{711}{1490}} = \log\frac{0{,}71}{0{,}48} = 0{,}39$$

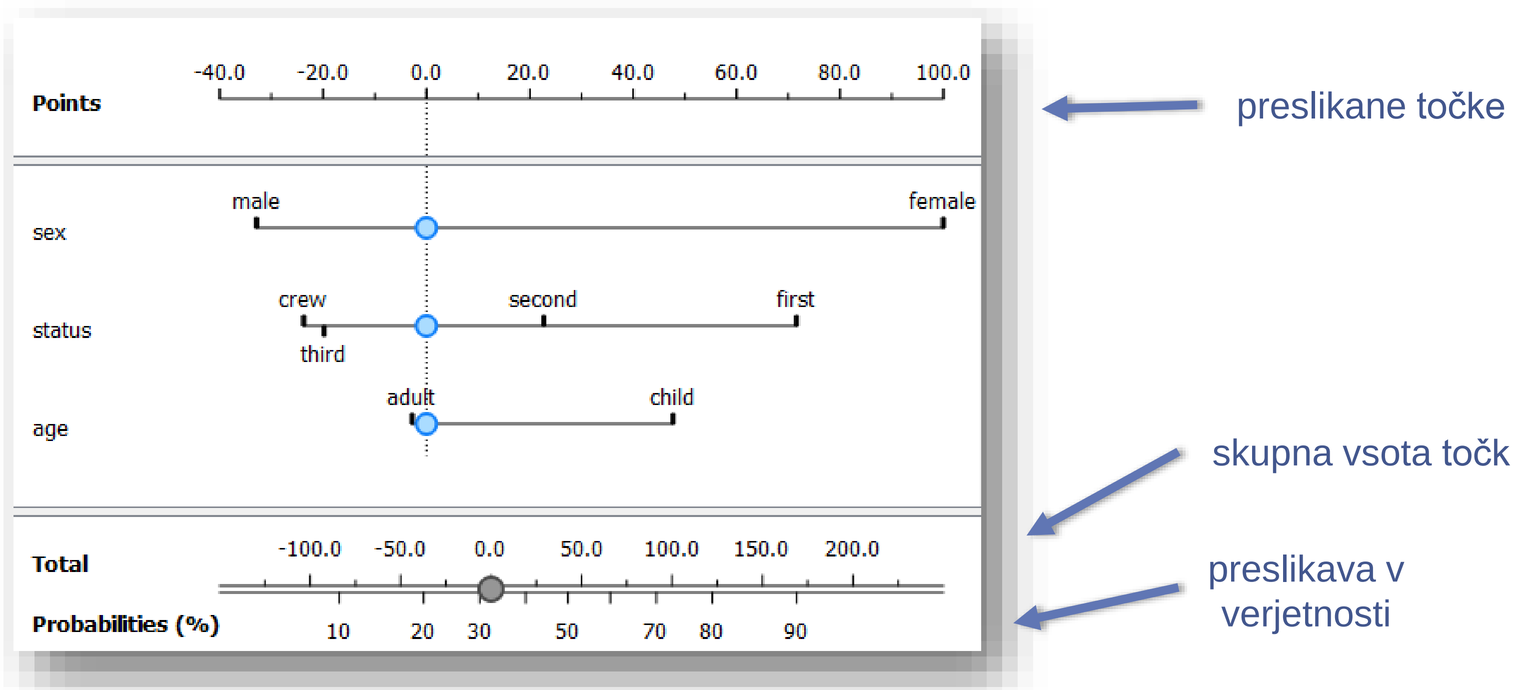
$$točke(yes|status = third) = \log\frac{\frac{P(third|yes)}{P(third|no)}}{\frac{P(yes)}{P(no)}} = \log\frac{\frac{178}{528}}{\frac{711}{1490}} = \log\frac{0{,}34}{0{,}48} = -0{,}35$$

$$točke(yes|status = crew) = \log\frac{\frac{P(crew|yes)}{P(crew|no)}}{\frac{P(yes)}{P(no)}} = \log\frac{\frac{212}{673}}{\frac{711}{1490}} = \log\frac{0{,}32}{0{,}48} = -0{,}42$$



Konstrukcija nomograma

- osi ostalih atributov poravnamo glede na ničelno vrednost prispevka atributa
- prikažemo lahko tudi skupno skalo za celotno napoved (vsoto točk)
- točke posameznih vrednosti atributov (log OR) lahko skaliramo v skalo točk, kjer s 100 točkami predstavimo prispevek največje vrednosti atributa
- skupne točke lahko preslikamo nazaj v verjetnosti*



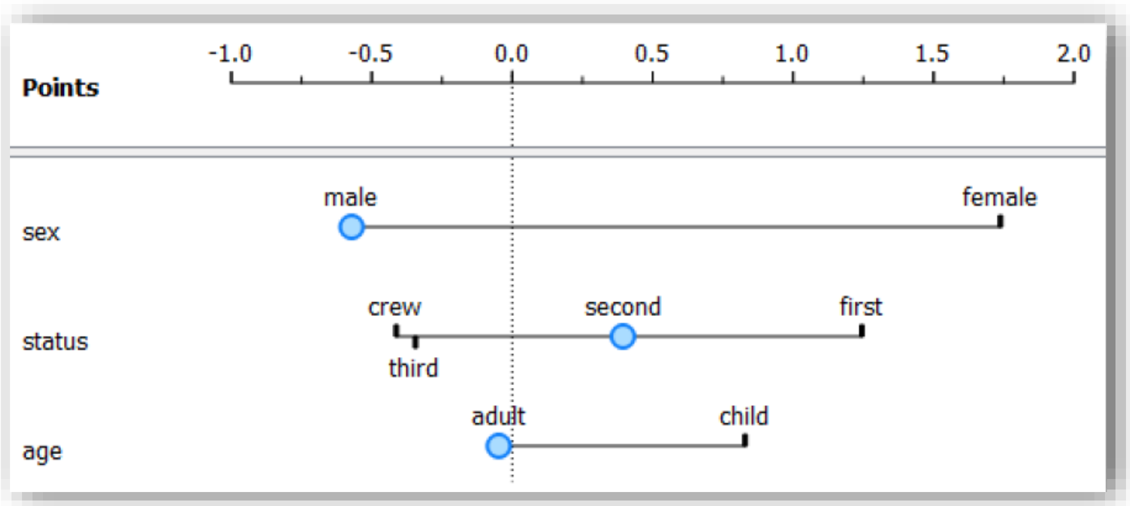
Primer

- Kako lahko pojasnimo odločitev, da je **odrasli moški**, ki je potoval v **drugem** razredu, **preživel**?

- $$točke(yes|age = adult) = \log\frac{\frac{P(adult|yes)}{P(adult|no)}}{\frac{P(yes)}{P(no)}} = \log\frac{\frac{654}{1438}}{\frac{711}{1490}} = \log\frac{0,45}{0,48} = -0,05$$

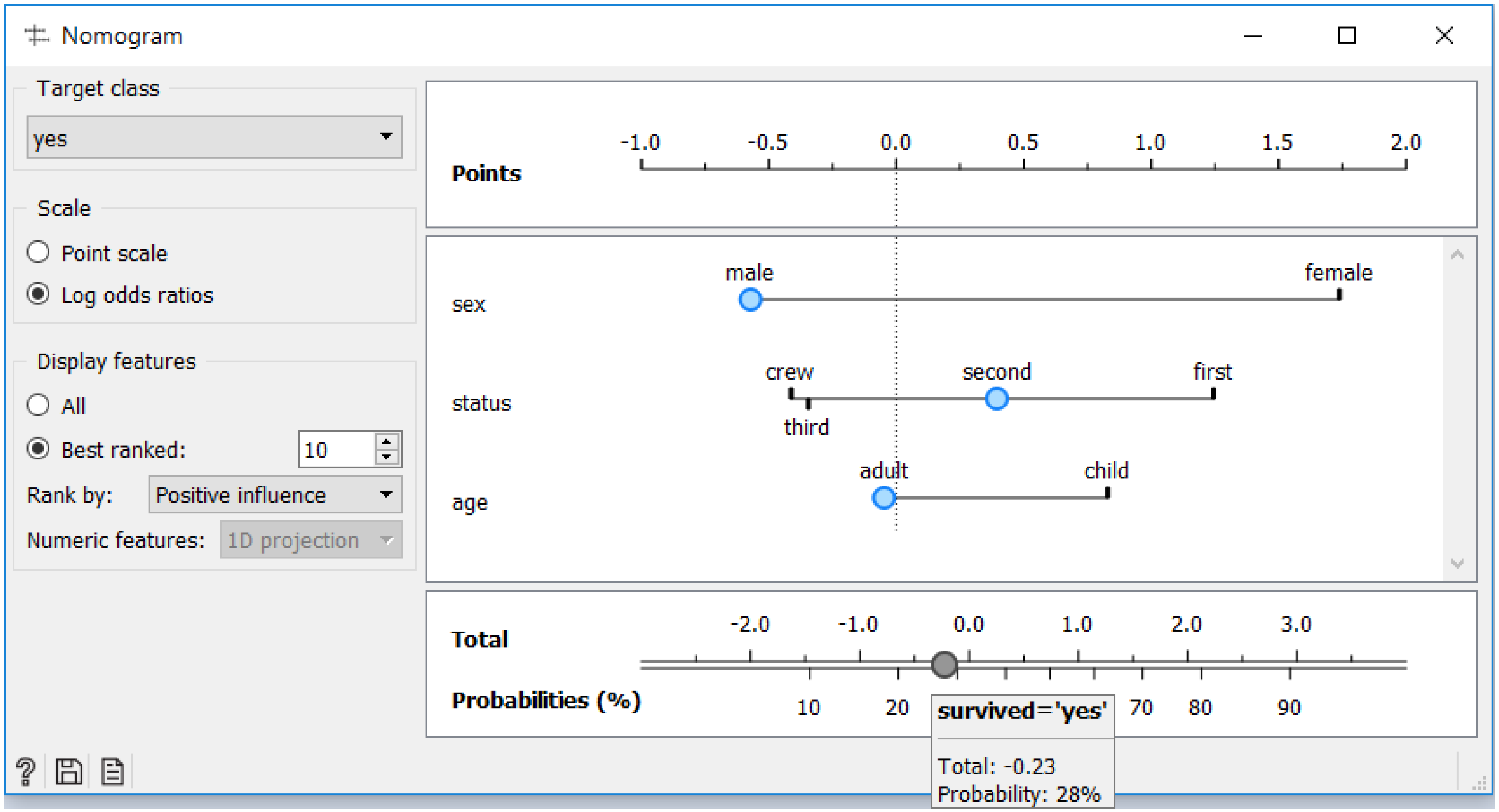
$$točke(yes|sex = male) = \log\frac{\frac{P(male|yes)}{P(male|no)}}{\frac{P(yes)}{P(no)}} = \log\frac{\frac{367}{1364}}{\frac{711}{1490}} = \log\frac{0,27}{0,48} = -0,57$$

$$točke(yes|status = second) = \log\frac{\frac{P(second|yes)}{P(second|no)}}{\frac{P(yes)}{P(no)}} = \log\frac{\frac{118}{167}}{\frac{711}{1490}} = \log\frac{0,71}{0,48} = 0,39$$



Primer

- Kako lahko pojasnimo odločitev, da je **odrasli moški**, ki je potoval v **drugem** razredu, **preživel**?
- $točke(yes|adult, male, second) = -0{,}05 - 0{,}57 + 0{,}39 = -0{,}23$




Nenadzorovano učenje