

Odgovarjajte kratko in jedrnat, točno na zastavljena vprašanja. Vse odgovore pišite na črto pod vprašanji in **izključno na ta list**, ki ga edinega oddate na koncu! Čas pisanja: 45 minut.

(1)

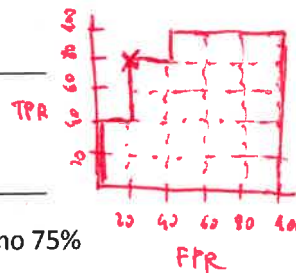
Za deset oseb je nek model strojnega učenja podal verjetnosti, da so zeleni vesoljci (razred »V«). Urejeni po verjetnosti za razred V so takole: #1 (99%, V), #2 (93%, V), #3 (83%, Z), #4 (77%, V), #5 (62%, V), #6 (50%, Z), #7 (49%, V), #8 (33%, Z), #9 (25%, Z) in #10 (11%, Z). V oklepaju je poleg verjetnosti podan še pravi razred, ki so ga določili z odvzemom krvi (razred »Z« pomeni zemljan). Nasvet: za vprašanji (a) in (b) narišite ROC krivuljo in iz nje razberite odgovora.

(a) V kateri točki (x,y) koordinatnega sistema smo, ko narišemo prvih pet primerov na ROC krivuljo? Koordinati x in y izrazite v odstotkih!

(x=20%, y=80%)

(b) Kakšna je vrednost AUROC, torej ploščine pod ROC krivuljo?

AUC = 84% (21/25)



(c) Recimo, da si za mejo, nad katero bomo napovedali, da je nekdo zelen vesoljec, vzamemo 75% verjetnost. Kakšni so priklic, natančnost in mera F1 glede na podanih 10 primerov?

priklic: 60%<sup>3/5</sup>; natančnost: 75%<sup>3/4</sup>; F1 =  $\frac{2PR}{P+R} = 67%$

(d) Kakšna je pri isti meji klasifikacijska točnost tega modela, če je le-ta dvorazreden?

$\frac{7}{10} = 70%$

(2)

V tabeli je podan model, ki smo ga za naš problem iskanja zelenih vesoljcev zgradili s pomočjo naivnega Bayesa. Uporablja le dva atributa in sicer ali ima sumljiva oseba zelenkasto polt in ali ima morda šest prstov na roki ali nogi. Napovedali bomo preprosto v bolj verjeten razred.

		zelenkasta polt				šest prstov			
p(V)	p(Z)	p(da V)	p(ne V)	p(da Z)	p(ne Z)	p(da V)	p(ne V)	p(da Z)	p(ne Z)
10%	90%	90%	10%	20%	80%	5%	95%	1%	99%

(a) Prvi opaženi osumljenec ima zelenkasto polt in videti je, da ima tudi šest prstov na rokah. V kateri razred ga bo dal naš model?

V:  $\frac{1 \cdot 9 \cdot 5}{10 \cdot 10 \cdot 100} = \frac{45}{10.000}$ ; Z:  $\frac{9 \cdot 2 \cdot 1}{10 \cdot 10 \cdot 100} = \frac{18}{10.000} \rightarrow$  razred: V

(b) Drugi osumljenec je bil prehter in zanj smo uspeli samo opaziti, da je verjetno zelene polti, prstov pa nismo mogli prešteti. Kaj bo zanj rekel naš model, če sploh lahko da napoved? Lahko da, preprosto ne upoštevaj prstov! Je lažje!

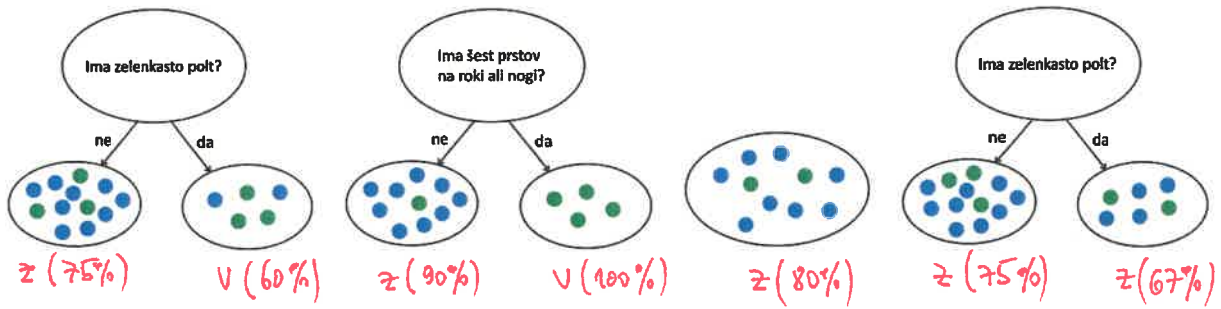
V:  $\frac{1 \cdot 9}{10 \cdot 10} = \frac{9}{100}$ ; Z:  $\frac{9 \cdot 2}{10 \cdot 10} = \frac{18}{100} \rightarrow$  razred: Z

(c) Recimo, da nočemo zbujati lažne pozornosti, da smo res našli zelenega vesoljca in zato dvignemo naš prag za napoved. Zato potrebujemo več kot samo primerjati razmerje obeh razredov, ampak potrebujemo dejansko verjetnost, ki jo vrne model. Kakšna je le-ta za našega prvega osumljenca, torej tistega z zeleno poltjo in šestimi prsti na rokah?

razmerje V:Z = 45:18 = 5:2  $\Rightarrow$  verjetnost(V) =  $\frac{5}{5+2} = \frac{5}{7} = 71%$

(3)

Spodaj je podan naključni gozd, sestavljen iz štirih odločitvenih dreves. Tretje drevo je preprosto en sam list. Problem so, seveda, še vedno zeleni vesoljci. Zemljani so označeni z modro barvo, zeleni vesoljci pa, kot se spodobi, z zeleno barvo.



(a) Kolega za vami ima sumljivo zeleno polt, res pa je, da nima šest prstov na rokah ali nogah. V kateri razred ga bo dal ta naključni gozd, če uporablja preprosto glasovanje?

glasovanje: (V, Z, Z, Z) → razred: Z

(b) Kaj pa naš prvi osumljenec iz prejšnje naloge, ki ima šest prstov na rokah in je zelene polti? Tokrat naključni gozd deluje po principu verjetnosti (soft max)? Kakšna je povprečna verjetnost, da je ta osumljenec zeleni vesoljec? ODG:  $AVG: 213\%/4 = 53,25\%$

verjetnosti za Z: (40%, 100%, 80%, 67%); verjetnosti za V: (60%, 100%, 20%, 33%)  
AVG:  $183\%/4 = 46,75\%$

(c) Naključni gozdovi so zelo uporabni tudi za ocenjevanje pomembnosti atributov. Kateri atribut bi kot najpomembnejšega označil naš naključni gozd?

V povprečju najbolje "čisti" nečistočo atribut "ima šest prstov?".

(4)

Za preiskovalne algoritme smo rekli, da kot rešitev vrnejo pot od začetnega do nekega končnega vozlišča.

(a) Kaj kot rešitev vračajo algoritmi spodbujevanega učenja (angl. *reinforcement learning*)?

STRATEGIJO

(b) Kaj se lahko zgodi, če so algoritmi spodbujevanega učenja požrešni in ne sledijo principu GLIE (angl. *greedy in the limit of infinite exploration*)?

Učajo se <sup>večinoma</sup> neoptimalne strategije.

(5)

Na množici imen {Nik, Luka, Martin, Anamarija, Barbara} naredi hierarhično gručenje; razdalja med imenom je definirana kot  $d(x,y) = |\text{length}(x) - \text{length}(y)|$ . Uporabi popolno povezanost. Napiši vse gruče velikosti >1 na posameznih nivojih dendrograma. Npr. če je  $d(\text{Vid}, \text{Miha})=1$ ,  $d(\text{Martina}, \text{Martin})=1$ , potem sta na nivoju 1 dve gruči velikosti dve: 1: {Vid, Miha}, {Martina, Martin}.

1: {N, L}, {M, B}

3: {{N, B}, A}

6: {{N, L}, {{M, B}, A}}

