

Vse naloge so enakovredne po številu točk.

1.) Dana je učna množica:

Atribut 1	Atribut 2	Razred
0	1	1
1	0	0
0	0	1

Nauči Naivni Bayesov klasifikator na tej množici in izračunaj verjetnosti obeh razredov za primer $A_1=1$ in $A_2=1$, tako da uporabiš m -oceno z vrednostmi:

- a) $m=0$
- b) $m=0.1$
- c) $m=2$

2. Na učni množici iz prejšnje naloge zgradi odločitveno drevo brez rezanja. Za ocenjevanje atributov uporabi informacijski prispevek. Kako bi to drevo klasificiralo primer, ki mu manjka vrednost A_1 in ima $A_2=1$? Odgovore utemelji!

OBRNI LIST

[22] 3. Na desni so podatki najnovejše študije o postrveh. (Za neudeleževalce vaj: A, B, C in D so štiri geni, geni A, B in C so lahko prisotni ali odsotni (0, 1), D pa ima tri možna stanja (0, 1 in 2). Naloga je s pomočjo genov ločevati med potočno in soško postrvjo.) Spodnja slika pa kaže klasifikacijsko drevo dobljeno iz teh podatkov, kot ga je sestavil računalnik z algoritmom, ki ste ga spoznali pri predmetu. Drevo je neporezano in pravilno razvrsti vse učne primere.

A	B	C	D	postrv
0	0	1	0	potočna
0	0	1	0	potočna
0	0	1	0	potočna
1	0	0	0	potočna
1	1	0	1	potočna
1	1	0	2	potočna
1	1	0	2	potočna
1	1	0	2	potočna
0	0	0	0	soška
0	1	0	0	soška
0	1	0	0	soška
1	1	1	1	soška
1	1	1	2	soška
1	1	1	2	soška

- Z metodo ostrega pogleda, brez računanja, sestavite najmanjše drevo, ki prav tako pravilno razvrsti vse učne primere.
- Ko računalnik sestavlja drevo, poskuša doseči, da je sestavljeno drevo čim manjše. Zakaj se mu je na teh podatkih to tako sfižilo?
- K vsakemu vozlišču v obeh drevesih (gornjem in tem, ki ste ga sestavili sami) pripišite, koliko učnih primerov sodi vanj. **V luči tega** in misleč na statistiko razmislite, katero od teh dreves vam je bolj všeč in zakaj?
- Kaj je Occamovo načelo (Occamova britev) in kako je povezano s to nalogo?

