

[25] 1. Podjetje Megahard, ki je ravno dobro izdalo nov operacijski sistem Korčulata, zaskrbljeno ugotavlja, da mu stranke uhajajo h konkurenčnemu podjetju Pear. Da bi pravilno usmerili reklamno kampanjo, jih za

	starost	Američan	programer	multimedija	ubežnik
začetek zanima, ali znajo	< 20	ne	da	da	da
napovedati, kakšne vrste	< 20	ne	da	da	da
strank so tipični prebežniki. V	< 20	ne	da	da	da
ta namen so zbrali nekaj	< 20	da	ne	da	da
podatkov. Zanimala jih je	20-40	da	ne	ne	da
starost anketiranca, zabeležili	> 40	da	ne	ne	da
so, ali gre za Američana ali ne,	> 40	da	ne	ne	da
poleg tega pa še, ali se	> 40	da	ne	ne	da
profesionalno ukvarja s	< 20	da	da	da	ne
programiranjem in ali ga	< 20	da	da	ne	ne
zanima obdelava slik, filmov in	< 20	da	da	ne	ne
zvoka. Na koncu so seveda	20-40	ne	ne	ne	ne
pribežili, ali namerava oseba	> 40	ne	ne	ne	ne
prebegniti h konkurenci (ali pa	> 40	ne	ne	ne	ne
je to celo že storila).					

- [..8] a) Za dane podatke izračunajte prva dva nivoja klasifikacijskega drevesa.
- [..8] b) Kakšna bi bila namerjena klasifikacijska točnost drevesa, če bi ga testirali na učnih podatkih?
- [..4] c) Drevesa (in druge klasifikatorje) navadno testiramo na ločeni testni množici. Zakaj?
- [..5] d) Razlaga, ki ste jo napisali v prejšnji točki, za ta konkretni primer morda ne velja (če ste seveda napisali pravilno razlago). Zakaj?

Pomoč: tabela z entropijami binarnih porazdelitev. Npr. entropija pa 2:3 je enaka 0.971.

	1	2	3	4	5	6	7	8	9	10	11	12
1	1.000	.918	.811	.722	.650	.592	.544	.503	.469	.439	.414	.391
2	.918	1.000	.971	.918	.863	.811	.764	.722	.684	.650	.619	.592
3	.811	.971	1.000	.985	.954	.918	.881	.845	.811	.779	.750	.722
4	.722	.918	.985	1.000	.991	.971	.946	.918	.890	.863	.837	.811
5	.650	.863	.954	.991	1.000	.994	.980	.961	.940	.918	.896	.874
6	.592	.811	.918	.971	.994	1.000	.996	.985	.971	.954	.937	.918
7	.544	.764	.881	.946	.980	.996	1.000	.997	.989	.977	.964	.949
8	.503	.722	.845	.918	.961	.985	.997	1.000	.998	.991	.982	.971
9	.469	.684	.811	.890	.940	.971	.989	.998	1.000	.998	.993	.985
10	.439	.650	.779	.863	.918	.954	.977	.991	.998	1.000	.998	.994

- [25] 2. Nad taiste podatke je Megahard spustil še naivni Bayesov klasifikator.
- [..10] a) Skicirajte gornji del naivnega Bayesovega nomograma.
- [..5] b) Je vrstni red pomembnosti atributov enak pri obeh klasifikatorjih, drevesu in Bayesu?
- [..5] c) Razložite pomen posameznih vrednosti atributov.
- [..5] d) Opišite, kdo je najverjetnejši kandidat za prebeg in kdo je najzvestejša stranka?

1.

(a) Predstavitev problemov z AND/OR grafi temelji na dekompoziciji problemov. Kaj je pogoj za primernost uporabe AND/OR grafov? Kako se glasi šibka verzija tega pogoja. Skicirajte na primeru.

(b) Katere neinformirane metode preiskovanja vrnejo optimalne rešitve v smislu dolžine (števila korakov) le-te? Katera izmed njih pri tem razvije najmanj vozlišč?

(c) Kateri algoritem v splošnem razvije manj vozlišč, A* ali IDA*? Kaj so dobre lastnosti enega in drugega?

(d) Kako IDA* izbere naslednjo mejo, če s trenutno mejo f ni našel rešitve?

2.

(a) Vzemimo, da je pri regresiranju ciljev G skozi akcijo A planer dobil regresirane cilje RG , pri čemer je bilo:

$$G = \{a,b,c,d\}, \quad RG = \{c,d,e,f\}$$

Recimo, da definicije akcije A ne poznamo, lahko pa na osnovi tega rezultata nekaj sklepamo o množicah $add(A)$, $del(A)$ ter $can(A)$. Za vsako od teh množic določi glede na gornji rezultat regresiranja, katere elemente *mora nujno* vsebovati, katerih sploh *ne sme* vsebovati ter katere *lahko* vsebuje, vendar ne nujno.

(b) Kje lahko vključimo hevristično znanje pri planiranju po principu sredstev in ciljev? Razložite kako lahko to znanje koristi!