

Poslovna inteligenca

1. izpitni rok

27. januar 2014

Priimek in ime (tiskano): _____

Vpisna številka: _____

Naloga	1	2	3	4	5	Vsota
Vrednost	9	6	6	6	6	33
Točk						

1. Gimnazijski sošolci se odločajo za izbor gostilne za obletnico mature. Samo štirje so oddali svoje glasove:

Gostilne				
	Pr'Metki	Pr'Janezu	Pr'Lojzki	Pr'Petru
Žan	3	5	4	8
Zarja	2	4	6	3
Taras	5	5	6	4
Eva	4	4	1	4

Vsako od predlaganih gostiln so ocenjevali na lestvici od 0 do 10, kjer je 10 najboljša ocena.

- [3] (a) Poiščite pareto-optimalne in sub-optimalne gostilne.
- [2] (b) Katero gostilno bi izbrali, če bi se odločalo po metodi Harsany-ja?
- [2] (c) Katera gostilna bi bila najbolj primerna po metodi Nash-a?
- [2] (d) Če bi Zarja, kot glavna organizatorica, imela dvakratno utež kot ostali (ostali pa imajo enake uteži), katera gostilna bi bila zmagovalna?

- [2] 2. (a) Zakaj pri napovedovanju uporabljamo postopek "bagging"?
- [2] (b) Kaj imata skupnega "bagging" in naključni gozdovi?
- [1] (c) Katero slabost ima "bagging" (poleg počasnosti)?
- [1] (d) Katero slabost imajo naključni gozdovi (poleg počasnosti)?

[6] 3. Dani so transakcijski podatki v obliki nakupovalnih košaric:

ID	kupljeni izdelki
1	$\{a, b, d, e\}$
2	$\{b, c, d\}$
3	$\{a, b, d, e\}$
4	$\{a, c, d, e\}$
5	$\{b, c, d, e\}$
6	$\{b, d, e\}$
7	$\{c, d\}$
8	$\{a, b, c\}$
9	$\{a, d, e\}$
10	$\{a, b, c, d, e\}$

Poiščite vsa pravila, ki vključujejo natanko vse izdelke nabora $\{a, b, d, e\}$ z zaupanjem vsaj 0.7. Pri tem ne računajte zaupanja za pravila, za katera veste, da bo zaupanje premajhno - to jasno označite ter na kratko argumentirajte.

$$\sigma(X) = |\{t_i | X \subseteq t_i, t_i \in T\}| \quad s(X \rightarrow Y) = \sigma(X \cup Y)/N \quad c(X \rightarrow Y) = \sigma(X \cup Y)/\sigma(X)$$

Stran je prazna, da lahko nanjo rešujete nalogo.

4. Tabela spodaj prikazuje pravi razred in verjetnost za razred a, ki nam jo je vrnil nek klasifikator.

Pravi razred	Verjetnost razreda a
a	0.91
a	0.89
a	0.73
a	0.51
a	0.40
b	0.80
b	0.73
b	0.73
b	0.72
b	0.42

- [2] (a) Mejo za napoved razreda (nad njo napovemo razred a, pod njo pa razred b) lahko poljubno premikamo. Poročaj število primerov, katerih napovedi so TP, FP, TN, FN za vse možne meje (intervale, kjer se nič ne spremeni združite). Rezultate predstavite v tabeli s petimi stolpci (meja, TP, FP, TN, FN).
- [2] (b) Narišite krivuljo ROC.
- [2] (c) Kako lahko interpretiramo oziroma kaj pomeni površina pod krivuljo ROC (AUC)?

Stran je prazna, da lahko nanjo rešujete nalogo.

- [6] 5. Spodnji razsevni diagrami prikazujejo podatke o vožnjah avtobusa številka 9. Atributa sta dva, dan (ponedeljek, torek, sredo, četrtek, petek, sobota, nedelja) in ura odhoda z začetne postaje, ciljna spremenljivka pa je čas vožnje do končne postaje. Predlagajte, kako naj predelamo izvorna atributa, da bo linearna regresija na teh podatkih čim bolj napovedovala. Vaš predlog tudi utemeljite.

