

# Poslovna inteligenca

2. izpitni rok

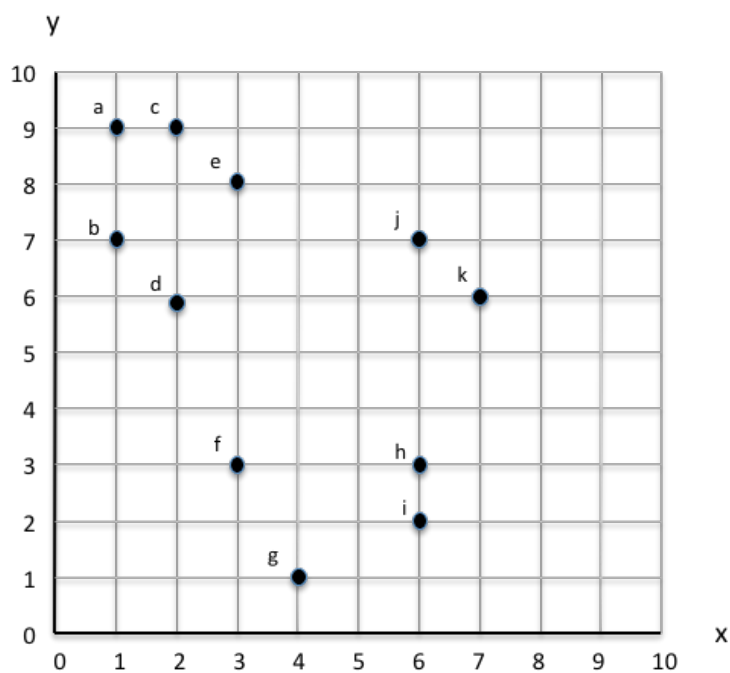
17. februar 2015

Priimek in ime (tiskano): \_\_\_\_\_

Vpisna številka: \_\_\_\_\_

Naloga	1	2	3	4	5	Vsota
Vrednost	5	6	5	6	5	27
Točk						

1. Dana je spodnja množica učnih primerov, ki smo jih opisali z dvema zveznima atributoma  $x$  in  $y$  in jih lahko predstavimo kot točke v Evklidski ravnini:



- [4] (a) Izriši dendrogram, ki ga dobiš z hierarhičnim razvrščanjem točk v skupine. Kot mero za podobnost uporabi Manhattansko razdaljo, kjer je razdalja med primeroma  $i$  in  $j$  določena kot  $d_{ij} = |x_i - x_j| + |y_i - y_j|$ . Podobnost med dvema skupinama meri s tehniko maksimalne razdalje med paroma točk iz različnih skupin (t. im. *complete linkage*).
- [1] (b) Uporabi izrisani dendrogram in na podlagi njega predlagaj razdelitev primerov v tri skupine (na dendrogramu izriši vertikalo, ki točke razdeli v tri skupine). Izpiši, kateri primeri pripadajo posamezni skupini.

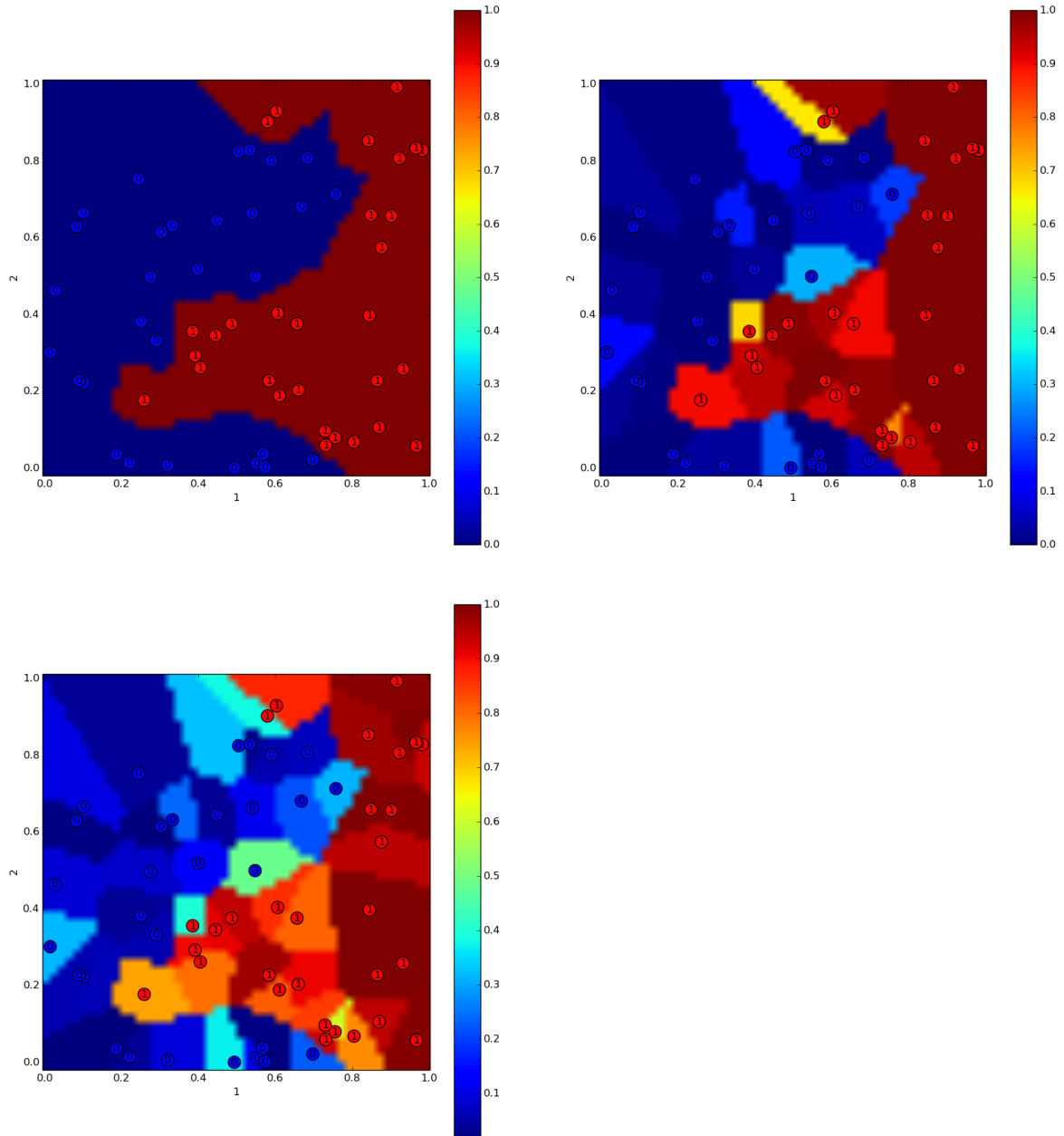
Stran je prazna, da lahko nanjo rešujete nalogo.

2. V Pythonu v nekem programu zapisali spodnjo funkcijo:

```
def j(theta, x, y, reg=0.1):  
    return -(y.dot(np.log(h(theta, x))) + (1-y).dot(np.log(1-h(theta, x))))
```

- [1] (a) Pri kateri tehniki analize podatkov smo to funkcijo uporabljali?
- [2] (b) Kaj ta funkcija izpiše (kaj so vhodni podatki in kaj je izhod)?
- [3] (c) Funkcija že vključuje argument za stopnjo regularizacije, a se tega v funkciji nismo uporabili. Dopolni funkcijo tako, da dodaš člen z regularizacijo.

- [5] 3. Za neimenovano slovensko banko želimo zgraditi odločitveni model, ki bo banki pomagal pri ocenjevanju kreditne sposobnosti kreditorejmalcev. Dobili smo podatke o podeljenih kreditih za zadnjih nekaj let. Uporabiti želimo logistično regresijo. Ker vemo, da je stopnja regularizacije kritična za kakovost modela, smo za podatke za tri stopnje regularizacije izrisali spodnje grafe. Grafi na X in Y osi prikazujejo vrednosti prvih dveh atributov, z barvo ozadja prikazujejo verjetnost klasifikacije v razred 1, točke pa so učni primeri. Ker graf zgoraj desno lepo loči oba razreda učnih podatkov in hkrati zmehta mejo med obema razredoma, se odločimo, da je stopnja regularizacije na tem grafu najustreznejša. Komentirajte postopek. Če smo kje naredili napako, predlagajte boljše rešitev.



4. Je čas kosila in odločamo se, kam bi šli jest. V bližini so 4 gostilne: “Pr’ Metki”, “Pr’ Janezu”, “Pr’ Lojzki” in “Pr’ Petru”. Odločili se bomo na podlagi okusa, cene obroka, hitrosti postrežbe in prijaznosti natakarjev.

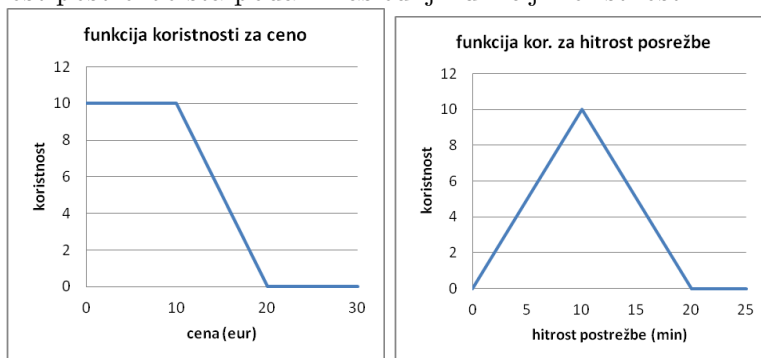
V Gurmanskem vodiču sta okus in prijaznost natakarjev za izbrane gostilne ocenjena takole (z ocenami 1-10):

	okus	prijaznost
Pr’ Metki	10	9
Pr’ Janezu	6	7
Pr’ Lojzki	10	8
Pr’ Petru	7	10

Iz preteklih izkušenj vemo, kakšne so cene obroka v posamezni gostilni in koliko časa je treba v povprečju čakati na postrežbo. Podatki so zbrani v spodnji tabeli.

	cena	hitrost
Pr’ Metki	9 EUR	12 min
Pr’ Janezu	12 EUR	13 min
Pr’ Lojzki	15 EUR	9 min
Pr’ Petru	8 EUR	10 min

Za ceno in hitrost postrežbe sta podani naslednji funkciji koristnosti:



Pri odločanju imajo kriteriji naslednje uteži: cena 30, hitrost postrežbe 20, okus 40 in prijaznost natakarjev 10.

- [4] (a) Katera gostilna najbolj ustreza našim zahtevam? Kakšen je vrstni red gostiln glede na podani model in ocene?
- [2] (b) Poiščite pare pareto-optimalnih in sub-optimalnih gostiln glede na (cena, hitrost).

Stran je prazna, da lahko nanjo rešujete nalogo.

5. Pet prijateljev je v spodnji tabeli dejavnosti označilo z ocenami od 1 do 5.

	tango	salsa	plezanje	bordanje	plavanje	kolesarjenje
Agnieszka			5		4	3
Marko	5		4	4	3	3
Nejc		1	2		3	2
Sara	2	1		5		5
Urška		2	4	2		1

Za priporočila želimo uporabiti matrično faktorizacijo z eno ( $k = 1$ ) latentno komponento.

- [2] (a) Kaj predstavljajo vrednosti obeh matrik z latentnimi faktorji, ko imamo latentno komponento ( $k = 1$ )?
- [2] (b) Faktorizirajte matriko dejavnosti s  $k = 1$  latentno komponento. Seveda ne zahtevamo natančne rešitev, le smiselna mora biti.
- [1] (c) Glede na vaš faktoriziran model Agniezski priporočite aktivnosti.