

# Poslovna inteligenca

2. izpitni rok

16. februar 2012

Ime in priimek: \_\_\_\_\_

Vpisna številka: \_\_\_\_\_

|   |
|---|
| Run $\LaTeX$ again to produce the table |
|---|

- [5] 1. Peter, vzdrževalec portala predlagam.vladi.si, želi uporabnikom portala ponuditi predloge, ki bi jih najverjetneje zanimali. Ker je pred leti napeto spremljal "NetFlix Prize", se je spomnil na način analize z matrično faktorizacijo. Za analizo ima na voljo bazo s 1500 uporabniki in 1000 predlogi. Na spletu najde nek algoritem za matrično faktorizacijo ter ga požene na bazi dosedanjih glasov (1500 uporabnikov, 1000 predlogov). Poskusi z 10 latentnimi komponentami ( $k = 10$ ), uspešnost postopka pa meri z mero RMSE in navzkrižnim preverjanjem, a ne dobi dobrih rezultatov. Zato na enak način izmeri uspešnost postopka za vse  $k$  od 1 do 150. Ker za  $k = 42$  dobi najboljši rezultat,  $RMSE=0.51$ , bo ta rezultat kot dokaz uspešnosti postopka vključil v poročilo ostalim članom skupine, ki upravlja s portalom.

Komentiraj primernost Petrovega izbora postopkov ter upravičenost njegovega zaključka. Če se s kakšnim delom opisanega postopka ne strinjaš, predlagaj alternativno rešitev.

- [10] 2. Tabela vsebuje podatke o medsebojnih dolgovih nekaterih evropskih držav. Kako bi jih čim učinkoviteje grafično predstavili? Skicirajte in pojasnite!

|          | Upniki           |           |          |           |         |        |         |             |        | skupaj    | dolž kot delež BDP | kreditni rating |
|----------|------------------|-----------|----------|-----------|---------|--------|---------|-------------|--------|-----------|--------------------|-----------------|
|          | Nemčija          | Španija   | Francija | VB        | Grčija  | Irska  | Italija | Portugalska |        |           |                    |                 |
| Dolžniki | Nemčija          |           | 60,036   | 245,629   | 230,336 | 1,884  | 60,938  | 244,956     | 3,307  | 847,086   | 83                 | AAA             |
|          | Španija          | 160,865   |          | 144,521   | 93,083  | 327    | 12,293  | 29,523      | 24,515 | 465,127   | 56                 | A               |
|          | Francija         | 217,000   | 29,419   |           | 276,013 | 1,876  | 13,136  | 48,015      | 7,030  | 592,489   | 87                 | AA+             |
|          | Velika Britanija | 509,804   | 408,875  | 237,281   |         | 12,209 | 151,947 | 48,910      | 6,047  | 1,375,073 | 81                 | AAA             |
|          | Grčija           | 18,636    | 1,048    | 47,899    | 11,546  |        | 749     | 3,164       | 8,711  | 91,753    | 166                | CC              |
|          | Irska            | 101,565   | 8,600    | 29,029    | 137,089 | 523    |         | 15,852      | 5,170  | 297,828   | 109                | BBB+            |
|          | Italija          | 144,686   | 35,512   | 372,400   | 61,268  | 457    | 11,015  |             | 2,544  | 627,882   | 121                | BBB+            |
|          | Portugalska      | 29,993    | 78,809   | 25,755    | 23,404  | 28     | 1,957   | 3,471       |        | 163,417   | 106                | BB              |
|          | skupaj           | 1,182,549 | 622,299  | 1,102,514 | 832,739 | 17,304 | 252,035 | 393,891     | 57,324 |           |                    |                 |

Pojasnilo za kreditni rating: A je boljše kot B in B je boljše kot C. Več črk pomeni boljši rating; AA+ (oz. BBB+) pomeni malo več od AA (oz. BBB).

Namig: podatke smete odšteti, sešteti in na druge načine preračunavati, če to izboljša preglednost vaše slike.

3. Tričlanska družina izbira nov družinski avto izmed petih variant. Vsak od njih je ocenil posamezen avto na absolutni lestvici od 0 do 10. Rezultati njihovega ocenjevanja so prikazani v tabeli:

|      | Clio | Yaris | Punto | Ceed | Mini |
|------|------|-------|-------|------|------|
| oče  | 4    | 3     | 5     | 6    | 7    |
| mama | 6    | 8     | 5     | 7    | 7    |
| hči  | 5    | 9     | 5     | 6    | 4    |

- [2] (a) Poiščite pare pareto-optimalnih in sub-optimalnih variant v zgornji tabeli.  
 [1] (b) Kateri avto bi bil najboljši po metodi Harsany-ja (seštevanje)?  
 [1] (c) Kateri avto bi družina kupila, če bi se odločala po metodi Nash-a (množenje)?  
 [2] (d) Kaj je prednost Nash-eve tehnike napram Harsany-jevi?  
 [2] (e) Če bi družinskim članom določili uteži (oče 50%, mama 30%, hči 20%), kateri avto bi bil zmagovalen?

**Solution:** a) Clio in Punto sta pareto-suboptimalni glede na varianto Ceed. b) Yaris c) Ceed d) Ne žrtvuje ekstremov (npr. nekoga, ki glasuje proti določeni varianti), e) Mini

Harsany 15 20 15 19 18 sum

Nash 120 216 125 252 196 product

Uteži 480 570 500 630 640 weighted sum

4. Med najbolj znane tehnike za gradnjo modelov za uvrščanje primerov (t.im. klasifikacijo) sodita gradnja klasifikacijskih dreves in tehnika gozdov klasifikacijskih dreves.

[2] (a) Zakaj pri gradnji klasifikacijskih dreves na vsakem koraku izberemo atribut z največjim informacijskim prispevkom?

[2] (b) Katere pomankljivosti klasifikacijskih dreves odpravljajo gozdovi klasifikacijskih dreves?

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

| ID | kupljeni izdelki |
|----|------------------|
| 1  | {c, 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, e}        |

Za spodnja pravila poišči njihovo podporo in zaupanje:

- $\{e\} \rightarrow \{d, b\}$
- $\{e, b\} \rightarrow \{d\}$
- $\{c\} \rightarrow \{d\}$

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

**Solution:** support, confidence

|       |       |          |
|-------|-------|----------|
| 0.500 | 0.667 | e -> d b |
| 0.500 | 0.800 | e b -> d |
| 0.625 | 1.000 | c -> d   |