Uvod v odkrivanje znanj iz podatkov (Poslovna inteligenca)

2. izpitni rok

6. februar 2018

Priimek in ime (tiskano): _____

Vpisna številka: _____

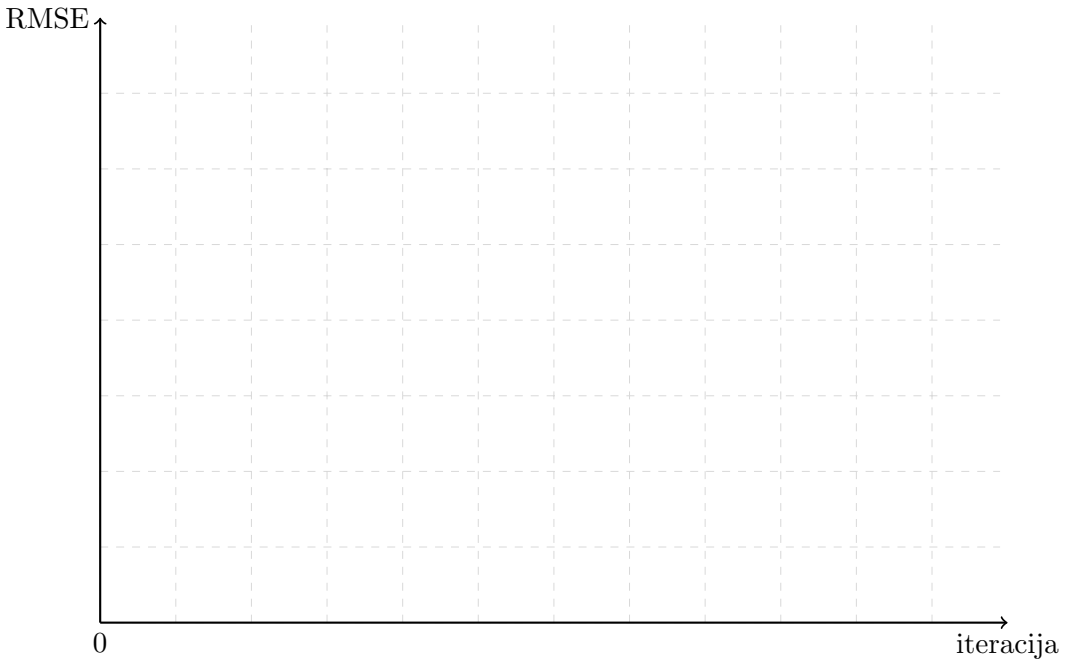
Naloga	1	2	3	4	5	Vsota
Vrednost	5	3	5	6	6	25
Točk						

[5] 1. Priporočilni sistem, ki deluje na podlagi matričnega razcepa, smo pognali na podatkih o ocenah knjig. V podatkih se pojavlja 2000 uporabnikov, ki so z ocenami med 0 in 10 ocenili (nekatere od) 15000 knjig. Skupno imamo v učni množici 100000 ocen. Število latentnih faktorjev smo nastavili na $k = 50$, stopnjo učenja na 0.01. Optimizacijo ustavimo, ko se RMSE na učni množici med iteracijama razlikuje za manj kot 0.1 odstotka.

Poskusili smo več različnih stopenj regularizacije λ. Na testni množici je z RMSE 1.8 zmagala regularizacija $\lambda = 1.2$.

V spodnji koordinatni sistem vrišite krivulji RMSE na učnih podatkih glede na iteracijo optimizacije za dve stopnji regularizacije: za $\lambda = 0$ in za $\lambda = 1.2$. Obakrat začnite z isto začetno rešitvijo (istima matrikama P in Q). Krivulji začnite z RMSE pred prvo iteracijo in jih rišite do ustavitve.

Vaši krivulji seveda ne moreta biti popolnoma natančni, a naj prikazujeta lastnosti dane situacije.

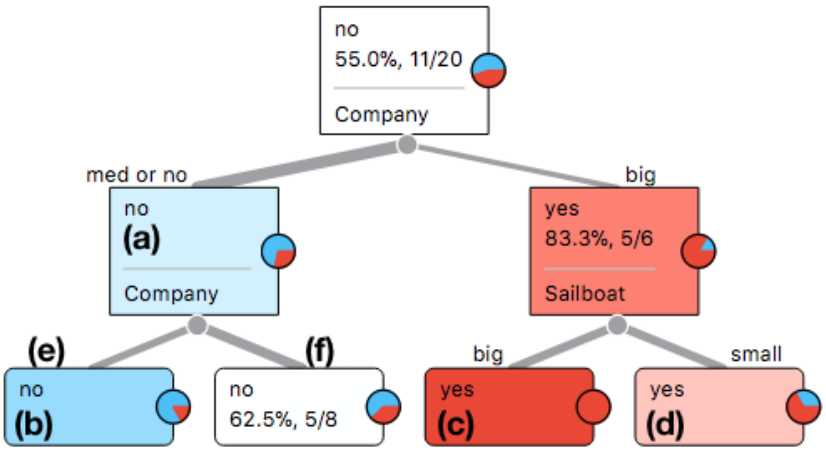


Solution: Krivulji začneta v isti točki. Obe načeloma padata (na začetku so lahko čudni artefakti). RMSE tiste z regularizacijo pada počasneje in se nekje skoraj ustavi. Ustavi se blizu ali pod 1.8, nikakor pa ne veliko nad tem.

2. Dobili smo spodnje podatke, ki poročajo, ali bo prijateljica šla jadrat glede na podatke o vremenski napovedi, velikosti družbe, ki bi šla zraven in velikosti barke, ki je na voljo. V koloni "Sail" je označeno, ali je na jadranje šla ("yes") ali ne ("no").

	Sail	Outlook	Company ▲	Sailboat
1	yes	rainy	big	big
2	yes	rainy	big	small
5	yes	sunny	big	big
6	yes	sunny	big	small
17	yes	sunny	big	big
18	no	sunny	big	small
3	no	rainy	med	big
4	no	rainy	med	small
7	yes	sunny	med	big
8	yes	sunny	med	big
9	yes	sunny	med	small
12	no	rainy	med	big
19	no	sunny	med	big
20	no	sunny	med	big
10	yes	sunny	no	small
11	no	sunny	no	big
13	no	rainy	no	big
14	no	rainy	no	big
15	no	rainy	no	small
16	no	rainy	no	small

Iz podatkov smo zgradili klasifikacijsko drevo.



Nekaj podatkov v drevesu manjka: označili smo jih z (a) do (f). Dopolni manjkajoče skladno z ostalimi oznakami v drevesu in skladno s podano učno množico. Odgovore zapiši na spodnje alineje:

[½] (a) _____

[½] (b) _____

[½] (c) _____

[½] (d) _____

[½] (e) _____

[½] (f) _____

Solution: **a**: 71.4%, 10/14, **b**: 83.3%, 5/6, **c**: 100%, 3/3, **d**: 66.7%, 2/3, **e**: no, **f**: med.

3. Dana je funkcija $y(\theta_0, \theta_1) = (1 + 2\theta_0\theta_1)^2 + (1 - 3\theta_1)^2$.

[1] (a) Izpelji gradient funkcije $y(\theta_0, \theta_1)$ (podaj enačbo gradienta).

[1] (b) Izračunaj gradient funkcije $y(\theta_0, \theta_1)$ v točki $[\theta_0, \theta_1]^{\mathrm{T}} = [2, 1]^{\mathrm{T}}$.

[1] (c) Preveri zgoraj izračunano vrednost gradienta tako, da zanj izračunaš numerični približek. Pri tem uporabi samo enačbo funkcije in ne izpeljano enačbo gradienta. Dober približek na primer dobiš z uporabo $\epsilon = \pm 0.01$.

[1] (d) Z gradientnim sestopom iščemo vrednosti parametrov funkcije $y(\theta_0, \theta_1)$, pri katerih ima ta funkcija minimum. Začetne vrednosti parametrov nastavimo na $[\theta_0, \theta_1]^{\mathrm{T}} = [2, 1]^{\mathrm{T}}$. Uporabimo stopnjo učenja 0.1. Kakšna je vrednost parametrov po prvem koraku gradientnega sestopa, torej po tem, ko z gradientnim sestopom prvič osvežimo vrednost parametrov.

[1] (e) Kakšna je vrednost parametrov po drugem koraku gradientnega sestopa?

Solution: **a**: $[4\theta_1(1 + 2\theta_0\theta_1), 4\theta_0(1 + 2\theta_0\theta_1) - 6(1 - 3\theta_1)]$, **b**: $[20, 52]^{\mathrm{T}}$, **c**: enako, **d**: $[0., -4.2]^{\mathrm{T}}$, **e**: $[1.68, 3.96]^{\mathrm{T}}$.

Stran je prazna, da lahko nanjo rešujete nalogo.

4. V matriki ocen $R \in \mathbb{R}^{m \times n}$ vsaka vrstica predstavlja enega od m uporabnikov, vsak stolpec pa enega od n predmetov (ali izdelkov). Matrika R je redka matrika, kar pomeni, da večina njenih vrednosti ni določenih. Matriko R približno predstavimo z matrikama $P \in \mathbb{R}^{m \times k}$ in $Q \in \mathbb{R}^{k \times n}$ (tako, da je $r_{ui} \approx \hat{r}_{ui} = p_u q_i^T$). Naj bodo konkretne vrednosti teh matrik:

$$P = \begin{bmatrix} 1 & 0 \\ 2 & 2 \\ 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$Q = \begin{bmatrix} 2 & 0 & 1 & 2 & 2 \\ 1 & 2 & 2 & 1 & 0 \end{bmatrix}$$

[1] (a) Kaj je predstavljeno z matriko P?

[1] (b) Kaj je predstavljeno z matriko Q?

[2] (c) Izdelke po vrsti označimo z i_j, kjer je i oznaka za izdelek in je $j = 1 \ldots n$ zaporedna številka izdelka. Na osnovi podanega predlagaj dve skupini izdelkov tako, da navedeš, kateri izdelki so v prvi in kateri so v drugi skupini.

[1] (d) V priporočilnih sistemih $\hat{r}_{ui}$ uporabimo kot oceno, ki jo je napovedal naš model. Izračunajte napovedane ocene za vse uporabnike in vse predmete – torej, izračunajte vse elemente matrike $\hat{R}$.

[1] (e) Metoda, ki priporočilni model gradi z matrično faktorizacijo (npr. algoritem ISMF) v vsaki iteraciji spremeni matriki P in Q tako, da dobimo boljši približek podatkov v matriki R. Kako merimo kakovost razcepa matrike R v njena faktorja P in Q? Opišite z besedami in podajte funkcijo, ki meri kakovost razcepa na učnih podatkih.

Solution:

a) vsaka vrstica 1 uporabnika

b) vsak stolpec 1 izdelek

c) $(1, 4, 5), (2, 3)$

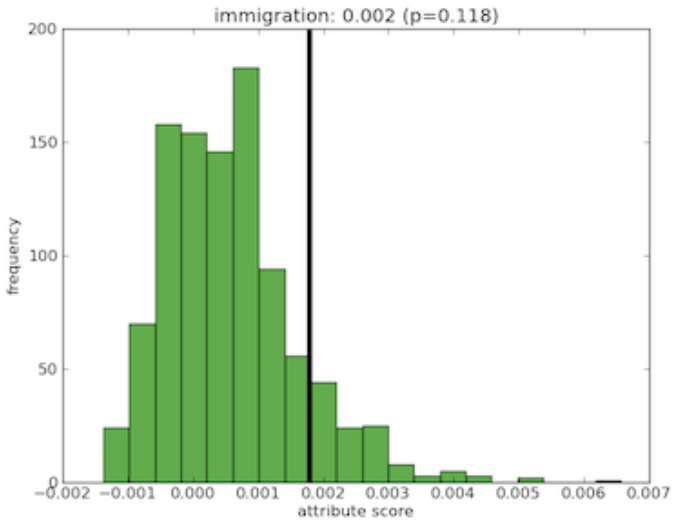
d)

$$P = \begin{bmatrix} 2 & 0 & 1 & 2 & 2 \\ 6 & 4 & 6 & 6 & 4 \\ 5 & 2 & 4 & 5 & 4 \\ 4 & 4 & 5 & 4 & 2 \end{bmatrix}$$

e) za podani primer, kvadrat razlike napovedi in znane vrednosti

Stran je prazna, da lahko nanjo rešujete nalogo.

5. Koristnost atributov za neki klasifikacijski problem smo ocenjevali tako, da smo za vsak posamezen atribut s permutacijskim testom pridobili referenčno (ničelno) porazdelitev ocene koristnosti tega atributa. Koristnost smo ocenjevali z informacijskim prispevkom. Primer take porazdelitve za atribut "immigration" kaže spodnji histogram. Uporabljena ocena atributa — informacijski prispevek (na grafu označen z "atribute score") — je taka, da je atribut pri njenih višjih vrednostih bolj informativen oziroma bolje koreliran z razredno spremenljivko. V spodnjo porazdelitev smo vrisali tudi oceno informativnosti tega atributa na originalnih (nepermutiranih) podatkih. Njegova ocena je znašala 0.002.



[2] (a) Kaj smo pravzaprav naredili s permutacijskim testom oziroma kako smo pridobili podatke za izris ničelne porazdelitve ocene atributa?

[1] (b) Oceni, kolikokrat smo morali ponoviti permutacijski poskus, da smo dobili podatke za prikazani histogram. Z drugimi besedami, kolikokrat smo morali v ta namen izmeriti informativnost atributa "immigration".

[2] (c) Kako bi vrednost p=0.118 razbrali iz slike?

[1] (d) Bi bilo ta atribut smiselno obdržati v učni množici? Zakaj?