Završni ispit iz Strojnog učenja (ak. god. 2020./2021.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **24 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 4A: Procjena parametara i Bayesov klasifikator *(9 pitanja)*

1 (T) MAP-procjenitelj definiramo kao $\hat{\boldsymbol{\theta}}_{\text{MAP}} = \text{argmax}_{\boldsymbol{\theta}} p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})$. Pri odabiru apriorne distribucije $p(\boldsymbol{\theta})$, nastojimo da je to neka standardna teorijska distribucija i da je konjugatna distribucija za izglednost $\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})$. **Što to znači i zašto to želimo?**

- A To znači da je aposteriorna distribucija parametara ista kao izglednost parametara, pa primjenom Bayesovog pravila možemo izračunati apriornu vjerojatnost parametara te, nakon zanemarivanja nazivnika koji je za fiksiran skup podataka konstantan, pronaći parametre koji maksimiziraju aposterionu vjerojatnost
- B To znači da je apriorna distribucija upravljana hiperparametrima kojima možemo ugoditi vjerojatnost parametara slučajne varijable koju procjenjujemo, tj. parametri apriorne distribucije i parametri izglednosti su identični, što nam omogućava da te dvije distribucije pomnožimo i zatim nađemo maksimizator
- C To znači da je apriorna distribucija ista vrsta distribucije kao i vjerojatnost podataka uz dane parametre, tj. izglednost parametara, pa će njihov umnožak biti distribucija koja je proporcionalna aposteriornoj distribuciji i čiji ćemo maksimum moći izračunati Bayesovim pravilom
- D To znači da će umnožak izglednosti i apriorne distribucije dati distribuciju koja je iste vrste kao i apriorna distribucija, a ako je riječ o standardnoj teorijskoj distribuciji iz eksponencijalne familije, njezin mod (maksimizator) postoji u zatvorenoj formi, što nam omogućava da procjenitelj izračunamo analitički

2 (N) Raspolažemo sljedećim skupom označenih primjera:

$$\mathcal{D} = \{x^{(i)}, y^{(i)}\} = \{(4,1), (-3,1), (-2,0), (1,0), (0,1), (-8,1)\}$$

Na ovom skupu treniramo univarijatni Bayesov klasifikator, za što trebamo procijeniti izglednosti klasa $p(x|y)$. Te su izglednosti definirane Gaussovom gustoćom vjerojatnosti:

$$p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{-\frac{(x-\mu)^2}{2\sigma^2}\Big\}$$

Parametre μ i σ^2 gustoće vjerojatnosti $p(x|y)$ procjenjujemo MLE-om. Neka su μ_1 i σ_1^2 parametri gustoće vjerojatnosti $p(x|y=1)$ dobiveni MLE-om na podskupu primjera $\mathcal{D}_{y=1}$. **Koliko iznosi log-izglednost $\mathcal{L}(\mu_1, \sigma_1^2|\mathcal{D}_{y=1})$?**

A -9.32 B -11.58 C -7.42 D -10.47

3 (N) U beta-Bernoullijevom modelu, apriornu vjerojatnost parametra μ modeliramo beta-distribucijom. Gustoća vjerojatnosti i mod (maksimizator) beta-distribucije su:

$$p(\mu|\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\mu^{\alpha-1}(1-\mu)^{\beta-1} \qquad \mu^* = \frac{\alpha-1}{\alpha+\beta-2}$$

Neka $\alpha = \beta = 2$. Računamo MAP-procjenu za parametar μ Bernoulijeve varijable. To radimo na dva uzorka, $\mathcal{D}_1 = (N_1, m_1)$ i $\mathcal{D}_2 = (N_2, m_2)$, koji nam pristižu jedan za drugim. Pritom koristimo svojstvo konjugatnosti, na način da aposteriornu gustoću vjerojatnosti izračunatu na temelju prvog uzorka koristimo kao apriornu gustoću vjerojatnosti pri procjeni na temelju drugog uzorka. U prvom uzorku, veličine $N_1 = 80$, Bernoullijeva varijabla realizirana je s vrijednošću $y = 1$ ukupno $m_1 = 42$ puta. U drugom uzorku, veličine $N_2 = 20$, Bernoullijeva varijabla realizirana je s vrijednošću $y = 1$ ukupno $m_2 = 2$ puta. Izračunajte MAP-procjene za parametar μ na temelju ova dva uzorka. **Koliko iznosi promjena u procjeni za μ između prve i druge procjene?**

A $+0.072$ B -0.083 C -0.191 D $+0.169$

4 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 3$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene kovarijacijske matrice. Model je definiran kao $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y)$. Prisjetimo se da je izglednost klase s oznakom $y = j$ kod Gaussovog Bayesovog klasifikatora definirana multivarijantnom Gaussovom gustoćom vjerojatnosti:

$$p(\mathbf{x}|y=j) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}_j|^{1/2}} \exp\Big\{ -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^{\mathrm{T}}\mathbf{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)\Big\}$$

gdje je Σ_j matrica kovarijacije za klasu j. Treniranjem modela dobili smo sljedeće procjene za parametre:

$$\hat{\mu}_1 = 0.1, \quad \hat{\boldsymbol{\mu}}_1 = (1, 0, -2), \quad \hat{\Sigma}_1 = \begin{pmatrix} 5 & 2 & 4 \\ 2 & 5 & 3 \\ 4 & 3 & 6 \end{pmatrix}, \quad \hat{\mu}_2 = 0.9, \quad \hat{\boldsymbol{\mu}}_2 = (2, -1, 5), \quad \hat{\Sigma}_2 = \begin{pmatrix} 5.0 & -0.222 & -0.444 \\ -0.222 & 2.778 & -0.333 \\ -0.444 & -0.333 & 0.444 \end{pmatrix}$$

Iz ovoga smo zatim procijenili dijeljenu kovarijacijsku matricu $\hat{\Sigma}$ definiranu kao težinski prosjek kovarijacijskih matrica $\hat{\Sigma}_j$, $j = 1, 2$. Zanima nas klasifikacija modela za primjer $\mathbf{x} = (1, 0, 1)$. **Koliko iznosi predikcija modela za klasu $y = 1$ za taj primjer, $h_1(\mathbf{x})$?**

[A] -10.913 [B] -12.788 [C] -8.513 [D] -5.902

5 (T) Bayesov klasifikator definirali smo na sljedeći način: $h_j(\mathbf{x}; \boldsymbol{\theta}) = P(y = j|\mathbf{x}) = \frac{p(\mathbf{x}|y)P(y)}{\sum_{y'} p(\mathbf{x}|y')P(y')}$. Neka je broj klasa jednak dva, a značajke neka su binarne i viševrijednosne diskretne vrijednosti. **Koje teorijske distribucije ćemo koristiti za $P(y)$ i $P(\mathbf{x}|y)$?**

[A] Gaussovu distribuciju za $P(y)$ i multinulijevu distribuciju za $P(\mathbf{x}|y)$

[B] Kategoričku distribuciju za $P(y)$ i Gaussovu distribuciju za $P(\mathbf{x}|y)$

[C] Bernoullijevu distribuciju za $P(y)$ i multinulijevu distribuciju za $P(\mathbf{x}|y)$

[D] Bernoullijevu distribuciju za $P(y)$ i Gaussovu distribuciju za $P(\mathbf{x}|y)$

6 (P) Gaussovim Bayesovim klasifikatorom rješavamo problem klasifikacije u $K = 10$ klasa sa $n = 100$ značajki. Prisjetite se da kod Gaussovog Bayesovog klasifikatora uvođenjem odgovarajućih pretpostavki na kovarijacijsku matricu Σ možemo utjecati na broj parametara modela a time onda i na složenost modela. Razmatramo tri modela s kovarijacijskim matricama u koje smo ugradili sljedeće pretpostavke:

$\mathcal{H}_1$: Značajke imaju različite varijance, ali iste za sve klase, te nisu korelirane

$\mathcal{H}_2$: Značajke nisu korelirane, imaju jednaku varijancu unutar svake klase, no različitu za svaku klasu

$\mathcal{H}_3$: Između značajki postoje korelacije, ali se one ne razlikuju između klasa

Neka '$\supset$' označava relaciju "složeniji od", a neka '$>$' označava relaciju "ima više parametara od". **Što možemo zaključiti o složenosti i broju parametara za gornja četiri modela?**

[A] $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$, $\mathcal{H}_3 \supset \mathcal{H}_1$ [C] $\mathcal{H}_1 > \mathcal{H}_3 > \mathcal{H}_2$, $\mathcal{H}_1 \supset \mathcal{H}_2 \supset \mathcal{H}_3$

[B] $\mathcal{H}_1 > \mathcal{H}_2 > \mathcal{H}_3$, $\mathcal{H}_1 \supset \mathcal{H}_2$ [D] $\mathcal{H}_3 > \mathcal{H}_2 > \mathcal{H}_1$, $\mathcal{H}_3 \supset \mathcal{H}_1 \supset \mathcal{H}_2$

7 (N) Na sljedećem skupu treniramo naivan Bayesov model za binarnu klasifikaciju *"Skupo ljetovanje na Jadranu"*:

i	Mjesto	Otok	Smještaj	Prijevoz	$y^{(i)}$
1	Kvarner	da	privatni	auto	1
2	Kvarner	ne	kamp	bus	0
3	Dalmacija	da	hotel	avion	1
4	Dalmacija	ne	privatni	avion	0
5	Istra	da	kamp	auto	0
6	Istra	ne	kamp	bus	1
7	Dalmacija	da	hotel	auto	1

Procjene radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija primjera $\mathbf{x} =$ (Dalmacija, ne, kamp, bus). **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

[A] 0.706 [B] 0.237 [C] 0.318 [D] 0.685

8 (N) Treniramo polunaivan Bayesov klasifikator sa $n = 3$ binarne varijable, x_1, x_2 i x_3. Zajednička vjerojatnost tih triju varijabli definirana je sljedećom tablicom:

	$x_3 = 0$		$x_3 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_1 = 0$	0.3	0.1	0.0	0.1
$x_1 = 1$	0.2	0.0	0.2	0.1

Prije treniranja klasifikatora, koristimo uzajamnu informaciju kako bismo procijenili koje su varijable najviše statistički zavisne, jer se te varijable isplati združiti u zajednički faktor. Odlučili smo združiti onaj par varijabli koje imaju uzajamnu informaciju veću od 0.01. Ako to vrijedi za dva para varijabli, onda ćemo sve tri varijable združiti u jedan faktor. Izračunajte uzajamne informacije između svih parova varijabli te odredite koje varijable ćemo združiti u zajedničke faktore prema gornjem pravilu. **Kako glasi faktorizacija zajedničke vjerojatnosti tog polunaivnog Bayesovog klasifikatora?**

A $P(y)P(x_1, x_2|y)P(x_3|y)$
C $P(y)P(x_1|y)P(x_2|y)P(x_3|y)$
B $P(y)P(x_1, x_3|y)P(x_2|y)$
D $P(y)P(x_1, x_2, x_3|y)$

9 (P) Naivan Bayesov klasifikator pretpostavlja uvjetnu nezavisnost značajki unutar neke klase, to jest $x_j \perp x_k | y$. Međutim, u stvarnosti ta pretpostavka rijetko kada vrijedi. Kao primjer, razmotrite model za klasifikaciju novinskih članaka, čija je zadaća odrediti je li tema članka Brexit ili ne. Model koristi binarne značajke koje indiciraju pojavljivanje određene riječi u novinskom članku. Na primjer, izglednost $P(\text{uk}|y = 1)$ jest vjerojatnost da se u članku koji je na temu Brexita pojavi riječ "uk" (Ujedinjeno Kraljevstvo). Razmotrite sljedeće četiri riječi koje se općenito mogu pojaviti u novinskim člancima: "trgovina", "uk", "brexit" i "konzum". **Za koju od sljedećih jednakosti općenito očekujemo da *ne vrijedi* i da se time onda narušava pretpostavka naivnog Bayesovog klasifikatora?**

A $P(\text{uk}|y = 1) = P(\text{uk}|\text{brexit}, y = 1)$
C $P(\text{uk}|y = 1) = P(\text{trgovina}|y = 1)$
B $P(\text{brexit}|y = 0) = P(\text{brexit}|\text{konzum}, y = 0)$
D $P(\text{uk}|y = 1) = P(\text{uk}|\text{konzum}, y = 1)$

Cjelina 4B: Probabilistički grafički modeli *(6 pitanja)*

10 (T) Bayesove mreže na sažet način prikazuju zajedničku distribuciju te kodiraju uvjetne stohastičke nezavisnosti između varijabli. No, kao i svaki model strojnog učenja, tako se i Bayesove mreže mogu prenaučiti. **Koja je veza između uvjetnih nezavisnosti varijabli u Bayesovoj mreži i opasnosti od prenaučenosti?**

A Uvođenje pretpostavki o uvjetnoj nezavisnosti pojednostavljuje strukturu Bayesove mreže i smanjuje broj parametara, čime se smanjuje i mogućnost prenaučenosti

B Uvjetne nezavisnosti određuju stukturu mreže na način da definiraju koji su čvorovi mreže međusobno povezani, međutim to nema utjecaja na složenost modela niti na sklonost prenaučenosti

C Uvođenjem pretpostavki o uvjetnoj nezavisnosti povećava se broj čvorova mreže, a time i broj parametara, što model čini složenijim i time sklonijim prenaučenosti

D Pretpostavke o uvjetnoj nezavisnosti čine induktivnu pristranost modela, pa što je više uvjetnih nezavisnosti, to je veća pristranost i model je lako prenaučiti

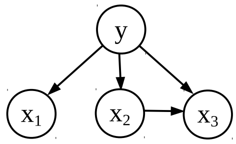
11 (P) Razmotrite Bayesovu mrežu koja zajedničku vjerojatnost faktorizira na sljedeći način:

$$P(w, x, y, z) = P(w)P(y)P(x|w, y)P(z|w)$$

Odredite topološki uređaj varijabli. Ako postoji više mogućih topoloških uređaja, izaberite onaj koji po leksičkom poretku dolazi prvi (npr. x, y, z dolazi prije x, z, y). Zatim primijenite uređajno Markovljevo svojstvo te izvedite sve uvjetne nezavisnosti koje su kodirane u ovoj Bayesovoj mreži. **Koje sve uvjetne nezavisnosti vrijede u ovoj Bayesovoj mreži?**

A $y \perp w,\ y \perp x | \{w, z\}$
B $x \perp y | z,\ z \perp w | y$
C $w \perp y,\ z \perp \{x, y\} | w$
D $w \perp y,\ z \perp x,\ x \perp w | \{z, y\}$

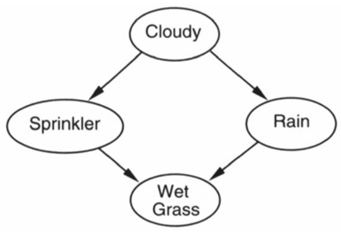
12 (P) Na slici ispod prikazana je Bayesova mreža koja odgovara polunaivnom Bayesovom klasifikatoru. Pretpostavite da su značajke x_1, x_2 i x_3 ternarne varijable te da je oznaka klase y također ternarna varijabla. Označimo ovaj model sa $\mathcal{H}_2$. Model $\mathcal{H}_2$ može se pojednostaviti ako se ukloni brid između varijabli x_2 i x_3. Označimo takav model sa $\mathcal{H}_1$. S druge strane, od modela $\mathcal{H}_2$ može se napraviti još složeniji model tako da se doda brid između varijabli x_1 i x_2. Označimo takav model sa $\mathcal{H}_3$.



Razmotrite koliko parametara imaju modeli $\mathcal{H}_1$, $\mathcal{H}_2$ i $\mathcal{H}_3$. **Koliko model $\mathcal{H}_2$ ima više parametara od modela $\mathcal{H}_1$, a koliko manje parametara od modela $\mathcal{H}_3$?**

A 12 više, 12 manje   B 12 više, 8 manje   C 2 više, 4 manje   D 4 više, 4 manje

13 (N) Na slici ispod prikazana je Bayesova mreža za problem prskalice za travu, koji smo bili koristili na predavanjima. Varijable su: C (oblačno/*cloudy*), S (prskalica/*sprinkler*), R (kiša/*rain*) i W (mokra trava/*wet grass*). Dane su i tablice uvjetnih vjerojatnosti za svaki čvor.



C	$P(C)$
0	0.5
1	0.5

S	C	$P(S\|C)$
0	0	0.5
0	1	0.9
1	0	0.5
1	1	0.1

R	C	$P(R\|C)$
0	0	0.8
0	1	0.2
1	0	0.2
1	1	0.8

W	R	S	$P(W\|R,S)$
0	0	0	1.0
0	0	1	0.9
0	1	0	0.1
0	1	1	0.01
1	0	0	0.0
1	0	1	0.1
1	1	0	0.9
1	1	1	0.99

Izračunajte aposteriornu vjerojatnost da radi prskalica ako je trava mokra i oblačno je.

A 0.459   B 0.111   C 0.400   D 0.805

14 (N) Bayesovom mrežom s četiri varijable modeliramo konstrukte pozitivne psihologije. Koristimo binarne varijable *Ljubav* (L), *Sreća* (S), *Tjeskoba* (T), s vrijednostima 0 (nema) i 1 (ima), te ternarnu varijablu *Novac* (N), s vrijednostima 0 (nema), 1 (ima malo) i 2 (ima puno). Strukturu Bayesove mreže definirali smo tako da ona zrcali sljedeće pretpostavljene kauzalne odnose: L uzrokuje S, a N uzrokuje S i T. Tako definiranu Bayesovu mrežu zatim treniramo na sljedećem skupu od $N = 7$ primjera:

L	N	S	T	L	N	S	T
1	0	1	0	1	0	1	0
0	2	0	1	1	2	1	1
1	1	1	0	0	0	0	0
0	2	1	0				

Parametre modela procjenjujemo MAP-procjeniteljem sa $\alpha = \beta = 2$ (za binarne varijable) odnosno $\alpha_k = 2$ (za ternarnu varijablu). Na kraju nas zanima koja je vjerojatnost života uz ljubav, sreću i puno novaca. Napravite potrebne MAP-procjene parametara. **Koliko iznosi zajednička vjerojatnost $P(L = 1, S = 1, N = 2)$?**

A 0.747   B 0.033   C 0.813   D 0.148

15 (P) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x, y, z) = P(x)P(y)P(z|x, y)$. Sve varijable su binarne. Vrijedi $P(x = 1) = 0.2$ i $P(y = 1) = 0.3$. Tablica uvjetne vjerojatnosti za čvor z je sljedeća:

z	x	y	$p(z\|x,y)$	z	x	y	$p(z\|x,y)$
0	0	0	0.1	1	0	0	0.9
0	0	1	0.2	1	0	1	0.8
0	1	0	0.5	1	1	0	0.5
0	1	1	0.9	1	1	1	0.1

Postupkom uzorkovanja s odbijanjem uzorkujemo iz aposteriorne distribucije $P(y|x = 0, z = 1)$. Uzorkovanje smo ponovili ukupno $N = 500$ puta. **Koliki je očekivani broj vektora koje ćemo pri uzorkovanju morati odbaciti?**

A 180   B 404   C 152   D 89

Cjelina 5: Grupiranje *(6 pitanja)*

16 (T) Konvergencija je poželjno svojstvo algoritma grupiranja. **Je li točno da algoritam k-sredina uvijek konvergira?**

- A Algoritam uvijek konvergira zato što je broj primjera N uvijek veći ili jednak broju grupa K, a kao mjera udaljenosti koristi se euklidska udaljenost, koja je nužno nenegativna
- B Da, algoritam uvijek konvergira zato što je broj particija N primjera u K skupova ograničen, a optimizacijski postupak definiran je tako da se J u svakoj iteraciji smanjuje
- C Algoritam konvergira samo ako su početna središta dobro odabrana, inače se može dogoditi da algoritam oscilira između dva rješenja
- D Kako se radi o algoritmu koji grupira primjere u vektorskom prostoru, broj rješenja je neograničen, stoga algoritam ne mora konvergirati

17 (N) Raspolažemo sljedećim neoznačenim skupom primjera: $\mathcal{D} = \{(\mathbf{x}^{(i)})\}_i = \{(1,1),(1,2),(2,2),(2,3),(4,3)\}$. Primjere grupiramo algoritmom k-sredina sa $K = 2$ grupe. Za početna središta odabrali smo primjere $\mathbf{x}^{(2)} = (1,2)$ i $\mathbf{x}^{(5)} = (4,3)$. Provedite prvu iteraciju algoritma k-sredina. **Koliko iznosi vrijednost kriterijske funkcije J nakon ažuriranja centroida?**

A 3.00 B 4.25 C 6.66 D 1.85

18 (N) Particijskim algoritmom grupiranja grupiramo $N = 1000$ primjera. Na temelju znanja o problemu zaključili smo da bi primjeri trebali formirati $K = 3$ grupe, pa smo s tim brojem grupa proveli grupiranje. Kako bismo evaluirali točnost grupiranja, slučajnim odabirom smo iz skupa primjera uzorkovali 10 primjera, ručno smo označili primjere iz tog uzorka, i zatim na tom uzorku računamo Randov indeks. Označavanje smo proveli tako da smo svakom primjeru iz uzorka dodijelili oznaku točne grupe. Oznake grupe dobivene algoritmom grupiranja y_{pred} i oznake točnih grupa y_{true} za svih deset primjera u uzorku su sljedeće:

i	1	2	3	4	5	6	7	8	9	10
$y^{(i)}_{pred}$	0	1	2	1	1	0	0	2	1	2
$y^{(i)}_{true}$	1	0	0	2	1	0	2	1	0	1

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.62 B 0.53 C 0.48 D 0.57

19 (P) Za grupiranje skupa primjera $\mathcal{D}$ koristimo algoritam GMM. Koristimo nekoliko varijanti tog modela:

$\mathcal{H}_1$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina

$\mathcal{H}_2$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

$\mathcal{H}_3$: Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom

$\mathcal{H}_4$: Model sa $K = 10$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL^0_α prosječna log-izglednost za model $\mathcal{H}_\alpha$ na početku izvođenja EM-algoritma, a neka je LL^*_α prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma. **Što možemo unaprijed zaključiti o ovim log-izglednostima?**

A $LL^0_2 \geq LL^0_4 \geq LL^0_3$, $LL^*_1 \geq LL^*_2$
C $LL^0_2 \leq LL^0_4$, $LL^*_2 \leq LL^*_1 \geq LL^*_3$
B $LL^0_2 \geq LL^0_4$, $LL^*_1 \geq LL^*_2 \geq LL^*_3$
D $LL^0_3 \geq LL^0_4$, $LL^*_1 \geq LL^*_3 \geq LL^*_4$

20 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom (matrica je simetrična, pa je donji trokut izostavljen):

$$
\begin{array}{c}
\\
\mathbf{x}^{(1)}\\ \mathbf{x}^{(2)}\\ \mathbf{x}^{(3)}\\ \mathbf{x}^{(4)}\\ \mathbf{x}^{(5)}
\end{array}
\begin{array}{c}
\begin{array}{ccccc} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \mathbf{x}^{(3)} & \mathbf{x}^{(4)} & \mathbf{x}^{(5)} \end{array}\\
\begin{pmatrix}
1 & 0.3 & 0.5 & 0.7 & 0.5\\
 & 1 & 0.9 & 0.4 & 0.6\\
 & & 1 & 0.8 & 0.1\\
 & & & 1 & 0.6\\
 & & & & 1
\end{pmatrix}
\end{array}
$$

Provedite grupiranje algoritmom HAC s jednostrukim povezivanjem te nacrtajte pripadni dendrogram. Primijetite da dendrogram odgovara binarnom stablu, s pojedinim primjerima u listovima. **Kojem binarnom stablu odgovara dobiveni dendrogram?**

A $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(1)}), (\mathbf{x}^{(4)}, \mathbf{x}^{(5)}))$ C $(((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(1)}, \mathbf{x}^{(5)}))$

B $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(5)}, \mathbf{x}^{(1)}))$ D $((((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), \mathbf{x}^{(1)}), \mathbf{x}^{(5)})$

21 (T) Broj grupa K hiperparametar je mnogih algoritama grupiranja, pa tako i algoritma GMM. Optimalan broj grupa može se odrediti na razine načine, a jedan od njih je Akaikeov kriterij. **Na kojem se principu temelji odabir broja grupa Akaikeovim kriterijem?**

A Optimalan broj grupa je onaj kod kojeg, nakon daljnjeg povećanja broja grupa, vrijednost log-izglednosti stagnira ili blago raste

B Model s optimalnim brojem grupa je onaj koji minimizira log-izglednost nepotpunih podataka, a maksimizira log-izglednost potpunih podataka

C Model s optimalnim brojem grupa je onaj koji podatke čini najvjerojatnijima, ali to čini sa što manje parametara

D Optimalan broj grupa je onaj koji maksimizira očekivanje log-izglednost modela, uz pretpostavku izotropne kovarijacijske matrice

Cjelina 6: Vrednovanje modela *(3 pitanja)*

22 (T) Procjena pogreške modela metodom unakrsne provjere omogućava nam da procijenimo prediktivnu moć modela, mjerenu kao točnost modela na neviđenom skupu primjera. Daljnja razrada te ideje je ugniježđena višestruka unakrsna provjera, koja se u praksi vrlo često koristi. **Koja je motivacija za korištenje ugniježđene višestruke unakrsne provjere, umjesto obične unakrsne provjere?**

A Provodi optimizaciju hiperparametra modela na uniji skupa za provjeru i skupa za testiranje, čime postiže bolju točnost modela jer više primjera ostaje za treniranje

B Razdvaja skup za učenje od skupa za ispitivanje te time osigurava da doista mjerimo prediktivnu moć modela, odnosno ispitnu pogrešku, a ne pogrešku učenja

C Omogućava nam da odredimo točnost modela s klasifikacijskim pragom, na način da u obzir uzimamo preciznost i odziv za različite vrijednosti klasifikacijskog praga

D Omogućava nam da procijenimo prediktivnu moć modela optimalne složenosti te maksimalno iskoristimo raspoložive podatke za učenje i ispitivanje

23 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$ i $\gamma \in \{2^{-5}, 2^{-4}, \dots, 2^{5}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 5 ponavljanja u vanjskoj petlji i 3 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 2820 B 5585 C 2980 D 4096

24 (N) Na ispitnome skupu evaluiramo klasifikator sa $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$
\begin{array}{c c}
 & \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\
\begin{array}{c} y=1 \\ y=2 \\ y=3 \end{array} & \begin{pmatrix} 5 & 13 & 1 \\ 6 & 5 & 4 \\ 4 & 2 & 16 \end{pmatrix}
\end{array}
$$

Izračunajte mikro-F1 (F_1^{μ}) i makro-F1 (F_1^{M}) mjere na ovoj matrici zabune. **Koliko iznosi razlika između vrijednosti mikro-F1 i makro-F1 mjere, $F_1^{\mu} - F_1^{M}$?**

A 0.023 B 0.114 C 0.091 D 0.059