Završni ispit iz Strojnog učenja 1 (ak. god. 2021./2022.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 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 1: Procjena parametara i Bayesov klasifikator (8 pitanja)

1 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 2$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene i dijagonalne kovarijacijske matrice. Izglednost klase s oznakom $y = j$ definirana je multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

Model treniramo na skup podataka od $N = 7$ primjera:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((-1,-2),0),((0,0),0),((1,2),0),((3,-1),1),((4,-1),1),((4,1),1),((5,1),1)\}$$

Procijenite parametre modela na ovom skupu primjera. Budući da je skup primjera malen, za procjenu kovarijacijske matrice koristite nepristran procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. **Koliko iznosi $h_0(\mathbf{x})$ za primjer $\mathbf{x} = (0, 0)$?**

A -3.84 B -4.13 C -2.75 D -3.03

2 (P) MAP-procjeniteljem na skupu $\mathcal{D}$ procjenjujemo parametre $\boldsymbol{\mu}$ kategoričke (multinulijeve) varijable $\mathbf{x}$, tj. procjenjujemo parametre multinulijeve distribucije $P(\mathbf{x}|\boldsymbol{\mu})$. Varijabla $\mathbf{x}$ može proprimiti tri moguće vrijednosti ($K = 3$). Apriorna Dirichletova distribucija $P(\boldsymbol{\mu}|\boldsymbol{\alpha})$ definirana je parametrima $\boldsymbol{\alpha}$ definiranima kao $\boldsymbol{\alpha} = \alpha \cdot (1, 4, 2)$, gdje faktor α određuje vršnost Dirichletove distribucije, $\alpha \geqslant 1$. U skupu podataka $\mathcal{D}$ realizirane su sve tri vrijednosti kategoričke varijable $\mathbf{x}$, i to sa sljedećim brojem realizacija: $N_1 = 50$, $N_2 = 29$, $N_3 = 40$. **Koliko mora iznositi faktor α, a da bi vrijednost x_2 bila najvjerojatnija vrijednost kategoričke varijable $\mathbf{x}$ prema procijenjenoj distribuciji $P(\mathbf{x}|\boldsymbol{\mu})$?**

A $\alpha > 2$ B $\alpha > 5$ C $\alpha > 8$ D $\alpha > 7$

3 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{-3, -2, 1, 4, 4\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Koristimo MLE kako bismo na uzorku $\mathcal{D}$ procijenili parametre $\hat{\mu}_{\mathrm{MLE}}$ i $\hat{\sigma}^2_{\mathrm{MLE}}$. Usporedbe radi, parametar σ^2 dodatno procjenjujemo nepristranim procjeniteljem $\hat{\sigma}^2_{\mathrm{UB}}$. Izračunajte log-izglednost parametara $(\hat{\mu}_{\mathrm{MLE}}, \hat{\sigma}^2_{\mathrm{MLE}})$ te log-izglednost parametara $(\hat{\mu}_{\mathrm{MLE}}, \hat{\sigma}^2_{\mathrm{UB}})$. **Koliko je prva log-izglednost veća od druge log-izglednosti?**

A 0.095 B 0.058 C 0.075 D 0.018

4 (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:

	$x_3 = 0$		$x_3 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_1 = 0$	0.1	0.0	0.0	0.1
$x_1 = 1$	0.1	0.4	0.1	0.2

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 $P(y)P(x_1, x_2, x_3|y)$ 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 $P(y)P(x_1, x_2|y)P(x_3|y)$

5 (T) Parametre modela možemo procijeniti pomoću procjenitelja najveće izglednosti (MLE) ili procjenitelja najveće aposteriorne vjerojatnosti (MAP). Općenito, MLE i MAP daju različite procjene, no u nekim slučajevima mogu dati jednake procjene. **Kada će MLE i MAP dati jednake procjene?**

A Kada je aposteriorna distribucija $p(\boldsymbol{\theta}|\mathcal{D})$ unimodalna

B Kada broj primjera N teži prema beskonačno

C Kada je $p(\boldsymbol{\theta})$ konjugatna izglednosti $p(\mathcal{D}|\boldsymbol{\theta})$

D Kada broj primjera N teži prema nuli

6 (P) Naivan Bayesov klasifikator pretpostavlja uvjetnu nezavisnost značajki unutar neke klase, premda u stvarnosti ta pretpostavka rijetko kada vrijedi. Razmotrite klasifikator poruka e-pošte čija je zadaća odrediti je li poruka prijevara (*scam*). Neka $y = 1$ označava da je poruka prijevara, a $y = 0$ da ona to nije. Model koristi binarne značajke koje indiciraju pojavljivanje određene riječi u tekstu poruke. Na primjer, izglednost $P(\text{lottery}|y = 1)$ jest vjerojatnost da se u prijevarnoj poruci pojavi riječ "lottery". Razmotrite sljedeće četiri riječi koje se općenito mogu pojaviti porukama e-pošte: "lottery", "winner", "prince" i "faithfully". Primijetite da mnoge prijevarne poruke od žrtve traže uplatu predujma. Takve poruke tipično govore o ostavštini nigerijskog princa ili o dobitku na lutriji (nikada oboje!) te završavaju kićenim pozdravom "yours faithfully". **Za koju od sljedećih jednakosti očekujemo da tipično *ne vrijedi* i da je time narušena pretpostavka naivnog Bayesovog klasifikatora?**

A $P(\text{prince}|y = 1) = P(\text{prince}|\text{lottery}, y = 1)$

B $P(\text{prince}|y = 1) = P(\text{prince}|\text{faithfully}, y = 1)$

C $P(\text{faithfully}|y = 0) = P(\text{faithfully}|\text{winner}, y = 0)$

D $P(\text{faithfully}|y = 1) = P(\text{faithfully}|\text{lottery}, y = 1)$

7 (T) Jedan od nedostataka generativnih modela u odnosu na diskriminativne modele jest nepotrebna složenost modeliranja. **Što to zapravo znači?**

A Generativni modeli modeliraju distribuciju $p(y|\mathbf{x})$, što iziskuje više parametara nego li modeliranje granice između klasa

B Za klasifikaciju nam je potrebna samo distribucija $p(y|\mathbf{x})$, i ona se može modelirati sa manje parametara od zajedničke distribucije $p(\mathbf{x}, y)$

C Zajednička distribucija $p(\mathbf{x}, y)$ može se faktorizirati kao $p(y|\mathbf{x})p(\mathbf{x})$, no takva faktorizacija ima više parametara

D Za razliku od diskriminativnih modela, generativni modeli distribuciju $p(y|\mathbf{x})$ definiraju zasebno za svaku klasu, pa stoga imaju više parametara

8 (P) Gaussov Bayesov klasifikator koristimo za klasfikaciju jednodimenzijskih podataka u tri klase. Procijenjene izglednosti klasa su $p(x|y = 1) = \mathcal{N}(-10, 2)$, $p(x|y = 2) = \mathcal{N}(2, 2)$ i $p(x|y = 3) = \mathcal{N}(8, 2)$, a procijenjene apriorne vjerojatnosti klasa su $P(y = 1) = P(y = 2) = 1/5$ i $P(y = 3) = 3/5$. Međutim, nakon što smo naučili ovaj model, zaključili smo da na ispitnom skupu postoji pomak u distribuciji podataka u odnosu na skup za učenje te da zbog toga model ne generalizira dobro. Zaključili smo da se ovo može ispraviti tako da se naučeni model malo izmijeni, i to tako da se varijanca izglednosti klase $y = 1$ postavi na 1 i da se apriorne vjerojatnosti klasa ujednače, $P(y = 1) = P(y = 2) = P(y = 3) = 1/3$. Skicirajte gustoće zajedničke vjerojatnosti naučenog i izmijenjenog modela. Neka su h_1 i h_2 MAP-hipoteze prvog i drugog modela, te neka su a i b pozitivne konstante. Razmotrite segment ulaznog prostora za koji vrijedi $-10 \leqslant x \leqslant 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

A $[-4, -4 + a] \cup [5, 5 + b]$ B $[-4 - a, -4] \cup [5 - b, 5]$ C $[-4 - a, 5 + b]$ D $[-4 - a, -4 + b]$

Cjelina 2: Probabilistički grafički modeli (5 pitanja)

9 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Sve varijable su binarne, osim varijabli v i y, koje su ternarne. Uz navedeni topološki uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

$$\{v, w\} \perp y | x \qquad \{v, x\} \perp z | \{w, y\}$$

Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

A 24 B 21 C 23 D 18

10 (P) Bayesovom mrežom sa sedam binarnih varijabli modeliramo prometne (ne)prilike u gradu Zagrebu. U našoj mreži, strelice modeliraju pretpostavljeni smjer kauzalnosti. Jutarnje doba dana (J) ili loše vrijeme (V) mogu uzrokovati nastanak prometne gužve (G). Loše vrijeme također može uzrokovati nastanak prometne nesreće (N). Nadalje, nastanak prometne nesreće može uzrokovati nastanak prometne gužve. Loše vrijeme također može uzrokovati zastoj tramvaja (T), no i nestanak struje (S) može uzrokovati zastoj tramvaja. Konačno, zastoj tramvaja može uzrokovati masovno pješačenje putnika (P), što opet može uzrokovati prometnu nesreću. Primjenom algoritma d-odvajanja ispitujemo stohastičku zavisnost između jutarnjeg doba dana (J) i nestanka struje (S). **Koje minimalan skup varijabli treba biti opažen, a da vjerojatnost od S ne ovisi o vrijednosti J?**

A V, N, T B T C G, T D V, P

11 (T) Čest način probabilističkog zaključivanja kod Bayesovih mreža jest izračunavanje "aposteriornog upita". Kod te vrste upita zanima nas distribucija nekih varijabli (varijable upita) na temelju zadanih varijabli (opažene varijable). Međutim, Bayesova mreža kodira zajedničku vjerojatnost svih varijabli mreže, uključivo i varijabli koje nisu niti varijable upita niti opažene varijable (varijable smetnje). **Na koji način izračunavamo aposteriorni upit?**

A Kao umnožak vjerojatnosti varijable upita uvjetovane na opažene varijable i zajedničke vjerojatnosti marginalizirane po varijablama smetnje

B Kao umnožak zajedničke vjerojatnosti marginalizirane po opaženim varijablama i zajedničke vjerojatnosti marginalizirane po varijablama smetnje

C Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama smetnje i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i varijablama upita

D Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama upita i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i opaženim varijablama

12 (N) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x, y, z, w) = P(z)P(w)P(y|z, w)P(x|y)$. Sve varijable su binarne. Vrijedi $P(z = 1) = 0.2$, $P(w = 1) = 0.2$, $P(x = 1|y = 0) = 0.2$ i $P(x = 1|y = 1) = 0.6$. Tablica uvjetne vjerojatnosti za čvor y je sljedeća:

z	w	$p(y = 1\|z, w)$
0	0	0
0	1	0.4
1	0	0.2
1	1	0.7

Postupkom uzorkovanja s odbijanjem uzorkujemo iz aposteriorne distribucije $P(w|x = 0, z = 1)$. Uzorkovanje smo ponovili ukupno $N = 10000$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N. **U uzorku neodbačenih vektora, koliko je očekivano više vektora sa $w = 0$ od vektora sa $w = 1$?**

A 596 B 472 C 348 D 944

13 (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 modelira 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 0.033 B 0.148 C 0.747 D 0.813

Cjelina 3: Grupiranje (6 pitanja)

14 (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
$y_{pred}^{(i)}$	0	1	2	1	1	0	0	2	1	2
$y_{true}^{(i)}$	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 0.58 D 0.49

15 (P) Skup neoznačenih primjera u dvodimenzijskome ulaznom prostoru neka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{5} = \{(0,0), (0,4), (2,0), (2,4), (4,2)\}$$

Primjere grupiramo algoritmom K-sredina sa $K = 3$ grupe. Za početna središta grupa odaberemo nasumično primjere iz $\mathcal{D}$, pri čemu, naravno, pazimo da odaberemo različita središta. Ishod grupiranja i konačan iznos kriterijske funkcije J ovisit će o odabiru početnih središta. Neka je J^* vrijednost kriterijske funkcije u točki globalnog minimuma, dakle vrijednost koja odgovara najboljem grupiranju. Neka je J^+ vrijednost kriterijske funkcije u točki lokalnog minimuma, i to onoj točki lokalnog minimuma s najvećom vrijednošću funkcije J. **Koliko iznosi razlika $J^+ - J^*$?**

A 8 B 6 C 12 D 4

16 (P) Algoritmom GMM grupiramo primjere u dvodimenzijskome ulaznom prostoru. Skup podataka u stvarnosti je uzorkovan iz zajedničke distribucije koja se može opisati sljedećim mješavinskim modelom:

$$p(\mathbf{x}) = \sum_{j=1}^{3} \frac{1}{3}\mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \mu_1 = (5,5), \mu_2 = (5,10), \mu_3 = (-10,-10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

Skup $\mathcal{D}$ grupiramo u $K = 2$ grupe. Pritom isprobavamo tri modela, koji se međusobno razlikuju po pretpostavkama na kovarijacijsku matricu. Konkretno: dijeljena i puna kovarijacijska matrica ($\mathcal{H}_1$), nedijeljena i dijagonalna kovarijacijska matrica ($\mathcal{H}_2$) i nedijeljena i izotropna kovarijacijska matrica ($\mathcal{H}_3$). Neka je $\mathcal{L}_i$ izglednost parametara dobivena modelom $\mathcal{H}_i$ nakon konvergencije algoritma. Za inicijalizaciju središta koristi se algoritam K-means++. **Što su očekivani odnosi između izglednosti za ova tri modela?**

A $\mathcal{L}_1 > \mathcal{L}_2 > \mathcal{L}_3$ B $\mathcal{L}_1 = \mathcal{L}_2 > \mathcal{L}_3$ C $\mathcal{L}_1 > \mathcal{L}_3$, $\mathcal{L}_2 > \mathcal{L}_3$ D $\mathcal{L}_2 > \mathcal{L}_1$, $\mathcal{L}_2 > \mathcal{L}_3$

17 (T) Algoritam K-means++ proširenje je algoritma K-sredina heurističkim odabirom početnih središta grupa. **Koja je glavna ideja odabira početnih središta grupa kod algoritma K-means++?**

- A Vjerojatnost da je neki primjer središte grupe proporcionalna je s brojem primjera u toj grupi i udaljenosti tih primjera od centroida skupa podataka
- B Središta grupe su srednje vrijednosti grupa projiciranih na pravac u smjeru prve komponente rastava skupa podataka na glavne komponente (PCA)
- C Najvjerojatnije središte grupe jest primjer koji je najviše udaljen od njemu najbližeg središta
- D Središta grupa su mjesta na kojima graf kriterijske funkcije u ovisnosti o broju grupa naglo opada pa stagnira

18 (T) Model miješane gustoće sa K komponenti vjerojatnosnu distribuciju neoznačenih podaka definira kao $p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \, p(\mathbf{x}|\boldsymbol{\theta}_k)$. Što modelira uvjetna vjerojatnost $p(\mathbf{x}|\boldsymbol{\theta}_k)$?

- A Vjerojatnost da primjer $\mathbf{x}$ pripada grupi k
- B Gustoću vjerojatnosti primjera $\mathbf{x}$ unutar grupe k
- C Stupanj pripadnosti grupe k primjeru $\mathbf{x}$
- D Stupanj pripadnosti primjera $\mathbf{x}$ grupi k

19 (N) Algoritmom HAC grupiramo riječi engleskog jezika. Neoznačeni skup podataka sastoji se od sljedećih riječi:

$$\mathcal{D} = \{\text{“water”}, \text{“watering”}, \text{“earth”}, \text{“air”}\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“watering”}) = 5/8 = 0.625$. Provedite prve dvije iteracije grupiranja algoritmom HAC uz prosječno povezivanje. **Na kojoj se razini sličnosti spajaju grupe u drugoj iteraciji algoritma HAC?**

A 0.292 B 0.478 C 0.583 D 0.535

Cjelina 4: Vrednovanje modela (3 pitanja)

20 (T) Za vrednovanje višeklasnog klasifikatora često se koriste mjere F_1^{μ} (mikro F1-mjera) i F_1^{M} (makro F1-mjera). **Koji je očekivani odnos između vrijednosti tih mjera, i zašto?**

- A $F_1^M > F_1^{\mu}$, jer kod mikro F1-mjere zbrajamo matrice zabune kroz sve klase, a primjera iz manjih klasa ima manje, pa manje doprinose pogrešci
- B $F_1^M < F_1^{\mu}$, jer kod makro F1-mjere računamo prosjek F1-mjere kroz sve klase, a klasifikator na manjim klasama više griješi
- C $F_1^M > F_1^{\mu}$, jer kod makro F1-mjere zbrajamo matrice zabune kroz sve klase, a primjera iz većih klasa ima više, pa više doprinose pogrešci
- D $F_1^M < F_1^{\mu}$, jer kod mikro F1-mjere zbrajamo matrice zabune kroz sve klase, pa je ukupan broj pogrešaka veći nego kad računamo prosječnu pogrešku kroz klase

21 (N) Na ispitnome skupu evaluiramo model multinomijalne logističke regresije (MLR) za klasifikaciju u $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{c c} & \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{array}{c} y=1 \\ y=2 \\ y=3 \end{array} & \begin{pmatrix} 35 & 8 & 3 \\ 10 & 25 & 5 \\ 0 & 12 & 2 \end{pmatrix} \end{array}$$

Klasifikator MLR uspoređujemo s klasifikatorom RAND koji primjere klasificira nasumično, i to tako da oznaku $y = j$ dodjeljuje s vjerojatnošću proporcionalnoj udjelu klase j u ispitnome skupu. Izračunajte mikro-F_1 za klasifikator MLR i očekivani mikro-F_1 za klasifikator RAND. **Koliko iznosi očekivana razlika u vrijednostima mikro-F_1 klasifikatora MLR i RAND?**

A 0.155 B 0.185 C 0.085 D 0.205

22 (P) Na istom skupu označenih primjera vrednujemo četiri binarna klasifikatora s vjerojatnosnim izlazima. Za vrednovanje koristimo krivulju ROC. Svaki smo klasifikator ispitali s tri vrijednosti klasifikacijskog praga te smo za te vrijednosti izračunali FPR (stopa lažnog alarma) i TPR (odziv). Dobiveni parovi vrijednosti (FPR, TPR) za sva četiri klasifikatora su sljedeći:

$$h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) \qquad h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8)$$
$$h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) \qquad h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0)$$

Skicirajte odgovarajuće krivulje ROC, linearno interpolirajući između izmjerenih točaka, a dodajte i točke koje odgovaraju krajnjim vrijednostima klasifikacijskog praga (0 i 1). Pritom, ako je neki klasifikator lošiji od nasumičnog klasifikatora, umjesto tog klasifikatora razmatrajte njegovu popravljenu varijantu koju ćete dobiti invertiranjem izlaza ($1 - h(\mathbf{x})$ umjesto $h(\mathbf{x})$). **Koji su od ispitanih klasifikatora (eventualno nakon popravka) najbolji prema krivulji ROC?**

A h_3 i h_4 B h_1 i h_2 C h_1 i h_4 D h_1 i h_3

Završni ispit iz Strojnog učenja 1 (ak. god. 2021./2022.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 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 1: Procjena parametara i Bayesov klasifikator (8 pitanja)

1 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{-3, -2, 1, 4, 4\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Koristimo MLE kako bismo na uzorku $\mathcal{D}$ procijenili parametre $\hat{\mu}_{\text{MLE}}$ i $\hat{\sigma}^2_{\text{MLE}}$. Usporedbe radi, parametar σ^2 dodatno procjenjujemo nepristranim procjeniteljem $\hat{\sigma}^2_{\text{UB}}$. Izračunajte log-izglednost parametara $(\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{MLE}})$ te log-izglednost parametara $(\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{UB}})$. **Koliko je prva log-izglednost veća od druge log-izglednosti?**

[A] 0.018 [B] 0.058 [C] 0.095 [D] 0.075

2 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 2$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene i dijagonalne kovarijacijske matrice. Izglednost klase s oznakom $y = j$ definirana je multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

Model treniramo na skup podataka od $N = 7$ primjera:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((-1,-2),0), ((0,0),0), ((1,2),0), ((3,-1),1), ((4,-1),1), ((4,1),1), ((5,1),1)\}$$

Procijenite parametre modela na ovom skupu primjera. Budući da je skup primjera malen, za procjenu kovarijacijske matrice koristite nepristran procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. **Koliko iznosi $h_1(\mathbf{x})$ za primjer $\mathbf{x} = (4, 0)$?**

[A] -3.84 [B] -4.13 [C] -3.03 [D] -2.75

3 (T) Parametre modela možemo procijeniti pomoću procjenitelja najveće izglednosti (MLE) ili procjenitelja najveće aposteriorne vjerojatnosti (MAP). Općenito, MLE i MAP daju različite procjene, no u nekim slučajevima mogu dati jednake procjene. **Kada će MLE i MAP dati jednake procjene?**

[A] Kada je $p(\mathcal{D}|\boldsymbol{\theta})$ uniformna distribucija

[B] Kada koristimo nepristran procjenitelj za $\boldsymbol{\theta}$

[C] Kada se u $\mathcal{D}$ realizirala svaka vrijednost slučajne varijable

[D] Kada broj primjera N teži prema beskonačno

4 (T) Jedan od nedostataka generativnih modela u odnosu na diskriminativne modele jest nepotrebna složenost modeliranja. **Što to zapravo znači?**

[A] Za klasifikaciju nam je potrebna samo distribucija $p(y|\mathbf{x})$, i ona se može modelirati sa manje parametara od zajedničke distribucije $p(\mathbf{x}, y)$

[B] Generativni modeli modeliraju distribuciju $p(y|\mathbf{x})$, što iziskuje više parametara nego li modeliranje granice između klasa

[C] Zajednička distribucija $p(\mathbf{x}, y)$ može se faktorizirati kao $p(y|\mathbf{x})p(\mathbf{x})$, no takva faktorizacija ima više parametara

[D] Za razliku od diskriminativnih modela, generativni modeli distribuciju $p(y|\mathbf{x})$ definiraju zasebno za svaku klasu, pa stoga imaju više parametara

5 (P) MAP-procjeniteljem na skupu $\mathcal{D}$ procjenjujemo parametre $\boldsymbol{\mu}$ kategoričke (multinulijeve) varijable $\mathbf{x}$, tj. procjenjujemo parametre multinulijeve distribucije $P(\mathbf{x}|\boldsymbol{\mu})$. Varijabla $\mathbf{x}$ može proprimiti tri moguće vrijednosti ($K = 3$). Apriorna Dirichletova distribucija $P(\boldsymbol{\mu}|\boldsymbol{\alpha})$ definirana je parametrima $\boldsymbol{\alpha}$ definiranima kao $\boldsymbol{\alpha} = \alpha \cdot (1, 4, 2)$, gdje faktor α određuje vršnost Dirichletove distribucije, $\alpha \geqslant 1$. U skupu podataka $\mathcal{D}$ realizirane su sve tri vrijednosti kategoričke varijable $\mathbf{x}$, i to sa sljedećim brojem realizacija: $N_1 = 50$, $N_2 = 29$, $N_3 = 40$. **Koliko mora iznositi faktor α, a da bi vrijednost x_2 bila najvjerojatnija vrijednost kategoričke varijable x prema procijenjenoj distribuciji $P(\mathbf{x}|\boldsymbol{\mu})$?**

A $\alpha > 8$ B $\alpha > 5$ C $\alpha > 2$ D $\alpha > 7$

6 (P) Gaussov Bayesov klasifikator koristimo za klasfikaciju jednodimenzijskih podataka u tri klase. Procijenjene izglednosti klasa su $p(x|y = 1) = \mathcal{N}(-10, 2)$, $p(x|y = 2) = \mathcal{N}(2, 2)$ i $p(x|y = 3) = \mathcal{N}(8, 2)$, a procijenjene apriorne vjerojatnosti klasa su $P(y = 1) = P(y = 2) = 2/5$ i $P(y = 3) = 1/5$. Međutim, nakon što smo naučili ovaj model, zaključili smo da na ispitnom skupu postoji pomak u distribuciji podataka u odnosu na skup za učenje te da zbog toga model ne generalizira dobro. Zaključili smo da se ovo može ispraviti tako da se naučeni model malo izmijeni, i to tako da se varijanca izglednosti klase $y = 1$ postavi na 5 i da se apriorne vjerojatnosti klasa ujednače, $P(y = 1) = P(y = 2) = P(y = 3) = 1/3$. Skicirajte gustoće zajedničke vjerojatnosti naučenog i izmijenjenog modela. Neka su h_1 i h_2 MAP-hipoteze prvog i drugog modela, te neka su a i b pozitivne konstante. Razmotrite segment ulaznog prostora za koji vrijedi $-10 \leqslant x \leqslant 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

A $[-4 - a, -4] \cup [5 - b, 5]$ B $[-4 - a, -4 + b]$ C $[-4 - a, 5 + b]$ D $[-4, -4 + a] \cup [5, 5 + b]$

7 (P) Naivan Bayesov klasifikator pretpostavlja uvjetnu nezavisnost značajki unutar neke klase, premda u stvarnosti ta pretpostavka rijetko kada vrijedi. Razmotrite klasifikator poruka e-pošte čija je zadaća odrediti je li poruka prijevara (*scam*). Neka $y = 1$ označava da je poruka prijevara, a $y = 0$ da ona to nije. Model koristi binarne značajke koje indiciraju pojavljivanje određene riječi u tekstu poruke. Na primjer, izglednost $P(\text{lottery}|y = 1)$ jest vjerojatnost da se u prijevarnoj poruci pojavi riječ "lottery". Razmotrite sljedeće četiri riječi koje se općenito mogu pojaviti porukama e-pošte: "lottery", "winner", "prince" i "faithfully". Primijetite da mnoge prijevarne poruke od žrtve traže uplatu predujma. Takve poruke tipično govore o ostavštini nigerijskog princa ili o dobitku na lutriji (nikada oboje!) te završavaju kićenim pozdravom "yours faithfully". **Za koju od sljedećih jednakosti očekujemo da tipično *ne vrijedi* i da je time narušena pretpostavka naivnog Bayesovog klasifikatora?**

A $P(\text{faithfully}|y = 0) = P(\text{faithfully}|\text{lottery}, y = 0)$

B $P(\text{faithfully}|y = 1) = P(\text{faithfully}|\text{prince}, y = 1)$

C $P(\text{faithfully}|y = 1) = P(\text{faithfully}|\text{winner}, y = 1)$

D $P(\text{winner}|y = 1) = P(\text{winner}|\text{prince}, y = 1)$

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 $P(y)P(x_1, x_2, x_3|y)$ C $P(y)P(x_1, x_2|y)P(x_3|y)$

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

Cjelina 2: Probabilistički grafički modeli (5 pitanja)

9 (N) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x, y, z, w) = P(z)P(w)P(y|z, w)P(x|y)$. Sve varijable su binarne. Vrijedi $P(z = 1) = 0.1$, $P(w = 1) = 0.1$, $P(x = 1|y = 0) = 0.2$ i $P(x = 1|y = 1) = 0.6$. Tablica uvjetne vjerojatnosti za čvor y je sljedeća:

z	w	$p(y = 1|z, w)$
0	0	0
0	1	0.4
1	0	0.2
1	1	0.7

Postupkom uzorkovanja s odbijanjem uzorkujemo iz aposteriorne distribucije $P(w|x = 0, z = 1)$. Uzorkovanje smo ponovili ukupno $N = 10000$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N. **U uzorku neodbačenih vektora, koliko je očekivano više vektora sa $w = 0$ od vektora sa $w = 1$?**

A 348   B 944   C 472   D 596

10 (T) Čest način probabilističkog zaključivanja kod Bayesovih mreža jest izračunavanje "aposteriornog upita". Kod te vrste upita zanima nas distribucija nekih varijabli (varijable upita) na temelju zadanih varijabli (opažene varijable). Međutim, Bayesova mreža kodira zajedničku vjerojatnost svih varijabli mreže, uključivo i varijabli koje nisu niti varijable upita niti opažene varijable (varijable smetnje). **Na koji način izračunavamo aposteriorni upit?**

A Kao umnožak zajedničke vjerojatnosti marginalizirane po opaženim varijablama i zajedničke vjerojatnosti marginalizirane po varijablama smetnje

B Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama upita i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i opaženim varijablama

C Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama smetnje i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i varijablama upita

D Kao umnožak vjerojatnosti varijable upita uvjetovane na opažene varijable i zajedničke vjerojatnosti marginalizirane po varijablama smetnje

11 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Sve varijable su binarne, osim varijabli v i x, koje su ternarne. Uz navedeni topološki uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

$$\{v, w\}\perp y|x \qquad \{v, x\}\perp z|\{w, y\}$$

Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

A 24   B 23   C 18   D 21

12 (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 modelira 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 0.813   B 0.033   C 0.747   D 0.148

13 (P) Bayesovom mrežom sa sedam binarnih varijabli modeliramo prometne (ne)prilike u gradu Zagrebu. U našoj mreži, strelice modeliraju pretpostavljeni smjer kauzalnosti. Jutarnje doba dana (J) ili loše vrijeme (V) mogu uzrokovati nastanak prometne gužve (G). Loše vrijeme također može uzrokovati nastanak prometne nesreće (N). Nadalje, nastanak prometne nesreće može uzrokovati nastanak prometne gužve. Loše vrijeme također može uzrokovati zastoj tramvaja (T), no i nestanak struje (S) može uzrokovati zastoj tramvaja. Konačno, zastoj tramvaja može uzrokovati masovno pješačenje putnika (P), što opet može uzrokovati prometnu nesreću. Primjenom algoritma d-odvajanja ispitujemo stohastičku zavisnost između jutarnjeg doba dana (J) i nestanka struje (S). **Koje minimalan skup varijabli treba biti opažen, a da vjerojatnost od S ne ovisi o vrijednosti J?**

A N, T B T C V, T D V, P

Cjelina 3: Grupiranje (6 pitanja)

14 (N) Algoritmom HAC grupiramo riječi engleskog jezika. Neoznačeni skup podataka sastoji se od sljedećih riječi:

$$\mathcal{D} = \{\text{“water”}, \text{“watering”}, \text{“earth”}, \text{“air”}\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“watering”}) = 5/8 = 0.625$. Provedite prve dvije iteracije grupiranja algoritmom HAC uz prosječno povezivanje. **Na kojoj se razini sličnosti spajaju grupe u drugoj iteraciji algoritma HAC?**

A 0.478 B 0.535 C 0.292 D 0.583

15 (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	1
$y^{(i)}_{true}$	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 0.58 D 0.49

16 (P) Skup neoznačenih primjera u dvodimenzijskome ulaznom prostoru neka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{5} = \{(0,0), (0,4), (2,0), (2,4), (4,2)\}$$

Primjere grupiramo algoritmom K-sredina sa $K = 3$ grupe. Za početna središta grupa odaberemo nasumično primjere iz $\mathcal{D}$, pri čemu, naravno, pazimo da odaberemo različita središta. Ishod grupiranja i konačan iznos kriterijske funkcije J ovisit će o odabiru početnih središta. Neka je J^* vrijednost kriterijske funkcije u točki globalnog minimuma, dakle vrijednost koja odgovara najboljem grupiranju. Neka je J^+ vrijednost kriterijske funkcije u točki lokalnog minimuma, i to onoj točki lokalnog minimuma s najvećom vrijednošću funkcije J. **Koliko iznosi razlika** $J^+ - J^*$?

A 8 B 6 C 12 D 4

17 (T) Model miješane gustoće sa K komponenti vjerojatnosnu distribuciju neoznačenih podaka definira kao $p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \, p(\mathbf{x}|\boldsymbol{\theta}_k)$. Što modelira uvjetna vjerojatnost $p(\mathbf{x}|\boldsymbol{\theta}_k)$?

A Vjerojatnost da primjer $\mathbf{x}$ pripada grupi k
B Gustoću vjerojatnosti grupe k za primjer $\mathbf{x}$
C Stupanj pripadnosti grupe k primjeru $\mathbf{x}$
D Gustoću vjerojatnosti primjera $\mathbf{x}$ unutar grupe k

18 (P) Algoritmom GMM grupiramo primjere u dvodimenzijskome ulaznom prostoru. Skup podataka u stvarnosti je uzorkovan iz zajedničke distribucije koja se može opisati sljedećim mješavinskim modelom:

$$p(\mathbf{x}) = \sum_{j=1}^{3} \frac{1}{3}\mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \mu_1 = (5,5), \mu_2 = (5,10), \mu_3 = (-10,-10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

Skup $\mathcal{D}$ grupiramo u $K = 2$ grupe. Pritom isprobavamo tri modela, koji se međusobno razlikuju po pretpostavkama na kovarijacijsku matricu. Konkretno: dijeljena i puna kovarijacijska matrica ($\mathcal{H}_1$), nedijeljena i dijagonalna kovarijacijska matrica ($\mathcal{H}_2$) i nedijeljena i izotropna kovarijacijska matrica ($\mathcal{H}_3$). Neka je $\mathcal{L}_i$ izglednost parametara dobivena modelom $\mathcal{H}_i$ nakon konvergencije algoritma. Za inicijalizaciju središta koristi se algoritam K-means++. **Što su očekivani odnosi između izglednosti za ova tri modela?**

A $\mathcal{L}_2 > \mathcal{L}_1$, $\mathcal{L}_2 > \mathcal{L}_3$ B $\mathcal{L}_1 = \mathcal{L}_2 > \mathcal{L}_3$ C $\mathcal{L}_1 > \mathcal{L}_2 > \mathcal{L}_3$ D $\mathcal{L}_1 > \mathcal{L}_3$, $\mathcal{L}_2 > \mathcal{L}_3$

19 (T) Algoritam K-means++ proširenje je algoritma K-sredina heurističkim odabirom početnih središta grupa. **Koja je glavna ideja odabira početnih središta grupa kod algoritma K-means++**?

A Središta grupe su srednje vrijednosti grupa projiciranih na pravac u smjeru prve komponente rastava skupa podataka na glavne komponente (PCA)

B Najvjerojatnije središte grupe jest primjer koji je najviše udaljen od njemu najbližeg središta

C Vjerojatnost da je neki primjer središte grupe proporcionalna je s brojem primjera u toj grupi i udaljenosti tih primjera od centroida skupa podataka

D Središta grupa su mjesta na kojima graf kriterijske funkcije u ovisnosti o broju grupa naglo opada pa stagnira

Cjelina 4: Vrednovanje modela (3 pitanja)

20 (N) Na ispitnome skupu evaluiramo model multinomijalne logističke regresije (MLR) za klasifikaciju u $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 35 & 8 & 3 \\ 10 & 25 & 5 \\ 0 & 12 & 2 \end{pmatrix} \end{array}$$

Klasifikator MLR uspoređujemo s klasifikatorom RAND koji primjere klasificira nasumično, i to tako da oznaku $y = j$ dodjeljuje s vjerojatnošću proporcionalnoj udjelu klase j u ispitnome skupu. Izračunajte mikro-F_1 za klasifikator MLR i očekivani mikro-F_1 za klasifikator RAND. **Koliko iznosi očekivana razlika u vrijednostima mikro-F_1 klasifikatora MLR i RAND?**

A 0.205 B 0.155 C 0.185 D 0.085

21 (P) Na istom skupu označenih primjera vrednujemo četiri binarna klasifikatora s vjerojatnosnim izlazima. Za vrednovanje koristimo krivulju ROC. Svaki smo klasifikator ispitali s tri vrijednosti klasifikacijskog praga te smo za te vrijednosti izračunali FPR (stopa lažnog alarma) i TPR (odziv). Dobiveni parovi vrijednosti (FPR, TPR) za sva četiri klasifikatora su sljedeći:

$h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4)$ $h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8)$
$h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6)$ $h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0)$

Skicirajte odgovarajuće krivulje ROC, linearno interpolirajući između izmjerenih točaka, a dodajte i točke koje odgovaraju krajnjim vrijednostima klasifikacijskog praga (0 i 1). Pritom, ako je neki klasifikator lošiji od nasumičnog klasifikatora, umjesto tog klasifikatora razmatrajte njegovu popravljenu varijantu koju ćete dobiti invertiranjem izlaza $(1 - h(\mathbf{x})$ umjesto $h(\mathbf{x}))$. **Koji su od ispitanih klasifikatora (eventualno nakon popravka) najbolji prema krivulji ROC?**

A h_1 i h_2 B h_2 i h_3 C h_3 i h_4 D h_1 i h_4

22 (T) Za vrednovanje višeklasnog klasifikatora često se koriste mjere F_1^μ (mikro F1-mjera) i F_1^M (makro F1-mjera). **Koji je očekivani odnos između vrijednosti tih mjera, i zašto?**

- A $F_1^M < F_1^\mu$, jer kod mikro F1-mjere zbrajamo matrice zabune kroz sve klase, pa je ukupan broj pogrešaka veći nego kad računamo prosječnu pogrešku kroz klase
- B $F_1^M > F_1^\mu$, jer kod mikro F1-mjere zbrajamo matrice zabune kroz sve klase, a primjera iz manjih klasa ima manje, pa manje doprinose pogrešci
- C $F_1^M < F_1^\mu$, jer kod mikro F1-mjere računamo prosjek F1-mjere kroz sve klase, a klasifikator na većima klasama manje griješi
- D $F_1^M < F_1^\mu$, jer kod makro F1-mjere računamo prosjek F1-mjere kroz sve klase, a klasifikator na manjim klasama više griješi

Završni ispit iz Strojnog učenja 1 (ak. god. 2021./2022.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 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 1: Procjena parametara i Bayesov klasifikator (8 pitanja)

1 (P) MAP-procjeniteljem na skupu $\mathcal{D}$ procjenjujemo parametre $\boldsymbol{\mu}$ kategoričke (multinulijeve) varijable $\mathbf{x}$, tj. procjenjujemo parametre multinulijeve distribucije $P(\mathbf{x}|\boldsymbol{\mu})$. Varijabla $\mathbf{x}$ može proprimiti tri moguće vrijednosti ($K = 3$). Apriorna Dirichletova distribucija $P(\boldsymbol{\mu}|\boldsymbol{\alpha})$ definirana je parametrima $\boldsymbol{\alpha}$ definiranima kao $\boldsymbol{\alpha} = \alpha \cdot (2, 1, 4)$, gdje faktor α određuje vršnost Dirichletove distribucije, $\alpha \geqslant 1$. U skupu podataka $\mathcal{D}$ realizirane su sve tri vrijednosti kategoričke varijable $\mathbf{x}$, i to sa sljedećim brojem realizacija: $N_1 = 50$, $N_2 = 64$, $N_3 = 40$. **Koliko mora iznositi faktor α, a da bi vrijednost x_2 bila najvjerojatnija vrijednost kategoričke varijable x prema procijenjenoj distribuciji $P(\mathbf{x}|\boldsymbol{\mu})$?**

A $\alpha > 2$ B $\alpha > 8$ C $\alpha > 5$ D $\alpha > 7$

2 (T) Jedan od nedostataka generativnih modela u odnosu na diskriminativne modele jest nepotrebna složenost modeliranja. **Što to zapravo znači?**

A Zajednička distribucija $p(\mathbf{x}, y)$ može se faktorizirati kao $p(y|\mathbf{x})p(\mathbf{x})$, no takva faktorizacija ima više parametara

B Za razliku od diskriminativnih modela, generativni modeli distribuciju $p(y|\mathbf{x})$ definiraju zasebno za svaku klasu, pa stoga imaju više parametara

C Generativni modeli modeliraju distribuciju $p(y|\mathbf{x})$, što iziskuje više parametara nego li modeliranje granice između klasa

D Za klasifikaciju nam je potrebna samo distribucija $p(y|\mathbf{x})$, i ona se može modelirati sa manje parametara od zajedničke distribucije $p(\mathbf{x}, y)$

3 (P) Naivan Bayesov klasifikator pretpostavlja uvjetnu nezavisnost značajki unutar neke klase, premda u stvarnosti ta pretpostavka rijetko kada vrijedi. Razmotrite klasifikator poruka e-pošte čija je zadaća odrediti je li poruka prijevara (*scam*). Neka $y = 1$ označava da je poruka prijevara, a $y = 0$ da ona to nije. Model koristi binarne značajke koje indiciraju pojavljivanje određene riječi u tekstu poruke. Na primjer, izglednost $P(\text{lottery}|y = 1)$ jest vjerojatnost da se u prijevarnoj poruci pojavi riječ "lottery". Razmotrite sljedeće četiri riječi koje se općenito mogu pojaviti porukama e-pošte: "lottery", "winner", "prince" i "faithfully". Primijetite da mnoge prijevarne poruke od žrtve traže uplatu predujma. Takve poruke tipično govore o ostavštini nigerijskog princa ili o dobitku na lutriji (nikada oboje!) te završavaju kićenim pozdravom "yours faithfully". **Za koju od sljedećih jednakosti očekujemo da tipično *ne vrijedi* i da je time narušena pretpostavka naivnog Bayesovog klasifikatora?**

A $P(\text{faithfully}|y = 1) = P(\text{faithfully}|\text{lottery}, y = 1)$

B $P(\text{faithfully}|y = 1) = P(\text{faithfully}|\text{prince}, y = 1)$

C $P(\text{winner}|y = 1) = P(\text{winner}|\text{prince}, y = 1)$

D $P(\text{prince}|y = 1) = P(\text{prince}|\text{faithfully}, y = 1)$

4 (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:

	$x_3 = 0$		$x_3 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_1 = 0$	0.1	0.0	0.0	0.1
$x_1 = 1$	0.1	0.4	0.1	0.2

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

5 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 2$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene i dijagonalne kovarijacijske matrice. Izglednost klase s oznakom $y = j$ definirana je multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

Model treniramo na skup podataka od $N = 7$ primjera:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)}))\}_i = \{((-1, -2), 0), ((0, 0), 0), ((1, 2), 0), ((3, -1), 1), ((4, -1), 1), ((4, 1), 1), ((5, 1), 1)\}$$

Procijenite parametre modela na ovom skupu primjera. Budući da je skup primjera malen, za procjenu kovarijacijske matrice koristite nepristran procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. **Koliko iznosi $h_0(\mathbf{x})$ za primjer $\mathbf{x} = (0, 0)$?**

A -3.03 B -4.13 C -3.84 D -2.75

6 (T) Parametre modela možemo procijeniti pomoću procjenitelja najveće izglednosti (MLE) ili procjenitelja najveće aposteriorne vjerojatnosti (MAP). Općenito, MLE i MAP daju različite procjene, no u nekim slučajevima mogu dati jednake procjene. **Kada će MLE i MAP dati jednake procjene?**

A Kada broj primjera N teži prema nuli C Kada broj primjera N teži prema beskonačno
B Kada koristimo nepristran procjenitelj za $\boldsymbol{\theta}$ D Kada su $p(\boldsymbol{\theta})$ i $p(\boldsymbol{\theta}|\mathcal{D})$ različite distribucije

7 (P) Gaussov Bayesov klasifikator koristimo za klasfikaciju jednodimenzijskih podataka u tri klase. Procijenjene izglednosti klasa su $p(x|y = 1) = \mathcal{N}(-10, 2)$, $p(x|y = 2) = \mathcal{N}(2, 2)$ i $p(x|y = 3) = \mathcal{N}(8, 2)$, a procijenjene apriorne vjerojatnosti klasa su $P(y = 1) = P(y = 2) = 2/5$ i $P(y = 3) = 1/5$. Međutim, nakon što smo naučili ovaj model, zaključili smo da na ispitnom skupu postoji pomak u distribuciji podataka u odnosu na skup za učenje te da zbog toga model ne generalizira dobro. Zaključili smo da se ovo može ispraviti tako da se naučeni model malo izmijeni, i to tako da se varijanca izglednosti klase $y = 1$ postavi na 1 i da se apriorne vjerojatnosti klasa ujednače, $P(y = 1) = P(y = 2) = P(y = 3) = 1/3$. Skicirajte gustoće zajedničke vjerojatnosti naučenog i izmijenjenog modela. Neka su h_1 i h_2 MAP-hipoteze prvog i drugog modela, te neka su a i b pozitivne konstante. Razmotrite segment ulaznog prostora za koji vrijedi $-10 \leqslant x \leqslant 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

A $[-4 - a, -4] \cup [5, 5 + b]$ B $[-4 - a, 5 + b]$ C $[-4, -4 + a] \cup [5 - b, 5]$ D $[-4 - a, -4 + b]$

8 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{-3, -2, 1, 4\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Koristimo MLE kako bismo na uzorku $\mathcal{D}$ procijenili parametre $\hat{\mu}_{\mathrm{MLE}}$ i $\hat{\sigma}^2_{\mathrm{MLE}}$. Usporedbe radi, parametar σ^2 dodatno procjenjujemo nepristranim procjeniteljem $\hat{\sigma}^2_{\mathrm{UB}}$. Izračunajte log-izglednost parametara $(\hat{\mu}_{\mathrm{MLE}}, \hat{\sigma}^2_{\mathrm{MLE}})$ te log-izglednost parametara $(\hat{\mu}_{\mathrm{MLE}}, \hat{\sigma}^2_{\mathrm{UB}})$. **Koliko je prva log-izglednost veća od druge log-izglednosti?**

A 0.018 B 0.075 C 0.058 D 0.095

Cjelina 2: Probabilistički grafički modeli (5 pitanja)

9 (N) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x, y, z, w) = P(z)P(w)P(y|z, w)P(x|y)$. Sve varijable su binarne. Vrijedi $P(z = 1) = 0.2$, $P(w = 1) = 0.2$, $P(x = 1|y = 0) = 0.2$ i $P(x = 1|y = 1) = 0.6$. Tablica uvjetne vjerojatnosti za čvor y je sljedeća:

z	w	$p(y = 1\|z, w)$
0	0	0
0	1	0.4
1	0	0.2
1	1	0.7

Postupkom uzorkovanja s odbijanjem uzorkujemo iz aposteriorne distribucije $P(w|x = 0, z = 1)$. Uzorkovanje smo ponovili ukupno $N = 10000$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N. **U uzorku neodbačenih vektora, koliko je očekivano više vektora sa $w = 0$ od vektora sa $w = 1$?**

A 348 B 596 C 472 D 944

10 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Sve varijable su binarne, osim varijabli v i x, koje su ternarne. Uz navedeni topološki uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

$$\{v, w\}\perp y|x \qquad \{v, x\}\perp z|\{w, y\}$$

Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

A 23 B 18 C 24 D 21

11 (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 modelira 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 0.813 B 0.148 C 0.033 D 0.747

12 (P) Bayesovom mrežom sa sedam binarnih varijabli modeliramo prometne (ne)prilike u gradu Zagrebu. U našoj mreži, strelice modeliraju pretpostavljeni smjer kauzalnosti. Jutarnje doba dana (J) ili loše vrijeme (V) mogu uzrokovati nastanak prometne gužve (G). Loše vrijeme također može uzrokovati nastanak prometne nesreće (N). Nadalje, nastanak prometne nesreće može uzrokovati nastanak prometne gužve. Loše vrijeme također može uzrokovati zastoj tramvaja (T), no i nestanak struje (S) može uzrokovati zastoj tramvaja. Konačno, zastoj tramvaja može uzrokovati masovno pješačenje putnika (P), što opet može uzrokovati prometnu nesreću. Primjenom algoritma d-odvajanja ispitujemo stohastičku zavisnost između jutarnjeg doba dana (J) i nestanka struje (S). **Koje minimalan skup varijabli treba biti opažen, a da vjerojatnost od S ne ovisi o vrijednosti J?**

A N, T B T C V, P D G, T

13 (T) Čest način probabilističkog zaključivanja kod Bayesovih mreža jest izračunavanje "aposteriornog upita". Kod te vrste upita zanima nas distribucija nekih varijabli (varijable upita) na temelju zadanih varijabli (opažene varijable). Međutim, Bayesova mreža kodira zajedničku vjerojatnost svih varijabli mreže, uključivo i varijabli koje nisu niti varijable upita niti opažene varijable (varijable smetnje). **Na koji način izračunavamo aposteriorni upit?**

A Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama upita i zajedničke vjerojatnosti marginalizirane po opaženim varijablama

B Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama smetnje i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i varijablama upita

C Kao umnožak vjerojatnosti varijable upita uvjetovane na opažene varijable i zajedničke vjerojatnosti marginalizirane po varijablama smetnje

D Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama upita i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i opaženim varijablama

Cjelina 3: Grupiranje (6 pitanja)

14 (P) Algoritmom GMM grupiramo primjere u dvodimenzijskome ulaznom prostoru. Skup podataka u stvarnosti je uzorkovan iz zajedničke distribucije koja se može opisati sljedećim mješavinskim modelom:

$$p(\mathbf{x}) = \sum_{j=1}^{3} \frac{1}{3}\mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \mu_1 = (5,5), \mu_2 = (5,10), \mu_3 = (-10,-10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

Skup $\mathcal{D}$ grupiramo u $K = 2$ grupe. Pritom isprobavamo tri modela, koji se međusobno razlikuju po pretpostavkama na kovarijacijsku matricu. Konkretno: dijeljena i puna kovarijacijska matrica ($\mathcal{H}_1$), nedijeljena i dijagonalna kovarijacijska matrica ($\mathcal{H}_2$) i nedijeljena i izotropna kovarijacijska matrica ($\mathcal{H}_3$). Neka je $\mathcal{L}_i$ izglednost parametara dobivena modelom $\mathcal{H}_i$ nakon konvergencije algoritma. Za inicijalizaciju središta koristi se algoritam K-means++. **Što su očekivani odnosi između izglednosti za ova tri modela?**

A $\mathcal{L}_1 = \mathcal{L}_2 > \mathcal{L}_3$ B $\mathcal{L}_2 > \mathcal{L}_1$, $\mathcal{L}_2 > \mathcal{L}_3$ C $\mathcal{L}_1 > \mathcal{L}_3$, $\mathcal{L}_2 > \mathcal{L}_3$ D $\mathcal{L}_1 > \mathcal{L}_2 > \mathcal{L}_3$

15 (T) Algoritam K-means++ proširenje je algoritma K-sredina heurističkim odabirom početnih središta grupa. **Koja je glavna ideja odabira početnih središta grupa kod algoritma K-means++**?

A Najvjerojatnije središte grupe jest primjer koji je najviše udaljen od njemu najbližeg središta

B Središta grupe su srednje vrijednosti grupa projiciranih na pravac u smjeru prve komponente rastava skupa podataka na glavne komponente (PCA)

C Središta grupa su mjesta na kojima graf kriterijske funkcije u ovisnosti o broju grupa naglo opada pa stagnira

D Vjerojatnost da je neki primjer središte grupe proporcionalna je s brojem primjera u toj grupi i udaljenosti tih primjera od centroida skupa podataka

16 (N) Algoritmom HAC grupiramo riječi engleskog jezika. Neoznačeni skup podataka sastoji se od sljedećih riječi:

$$\mathcal{D} = \{\text{“water”}, \text{“waterloo”}, \text{“earth”}, \text{“air”}\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2| / |\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“waterloo”}) = 5/7 = 0.714$. Provedite prve dvije iteracije grupiranja algoritmom HAC uz prosječno povezivanje. **Na kojoj se razini sličnosti spajaju grupe u drugoj iteraciji algoritma HAC?**

A 0.535 B 0.292 C 0.583 D 0.354

17 (P) Skup neoznačenih primjera u dvodimenzijskome ulaznom prostoru neka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{5} = \{(0,0), (0,4), (2,0), (2,4), (4,2)\}$$

Primjere grupiramo algoritmom K-sredina sa $K = 3$ grupe. Za početna središta grupa odaberemo nasumično primjere iz $\mathcal{D}$, pri čemu, naravno, pazimo da odaberemo različita središta. Ishod grupiranja i konačan iznos kriterijske funkcije J ovisit će o odabiru početnih središta. Neka je J^* vrijednost kriterijske funkcije u točki globalnog minimuma, dakle vrijednost koja odgovara najboljem grupiranju. Neka je J^+ vrijednost kriterijske funkcije u točki lokalnog minimuma, i to onoj točki lokalnog minimuma s najvećom vrijednošću funkcije J. **Koliko iznosi razlika** $J^+ - J^*$?

A 12 B 8 C 6 D 4

18 (T) Model miješane gustoće sa K komponenti vjerojatnosnu distribuciju neoznačenih podaka definira kao $p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k\, p(\mathbf{x}|\boldsymbol{\theta}_k)$. Što modelira uvjetna vjerojatnost $p(\mathbf{x}|\boldsymbol{\theta}_k)$?

[A] Gustoću vjerojatnosti primjera $\mathbf{x}$ unutar grupe k

[B] Stupanj pripadnosti primjera $\mathbf{x}$ grupi k

[C] Vjerojatnost da primjer $\mathbf{x}$ pripada grupi k

[D] Stupanj pripadnosti grupe k primjeru $\mathbf{x}$

19 (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
$y_{pred}^{(i)}$	0	0	2	1	1	0	0	2	1	1
$y_{true}^{(i)}$	1	0	0	2	1	0	2	1	1	1

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

[A] 0.49 [B] 0.53 [C] 0.62 [D] 0.58

Cjelina 4: Vrednovanje modela (3 pitanja)

20 (T) Za vrednovanje višeklasnog klasifikatora često se koriste mjere F_1^μ (mikro F1-mjera) i F_1^M (makro F1-mjera). **Koji je očekivani odnos između vrijednosti tih mjera, i zašto?**

[A] $F_1^M > F_1^\mu$, jer kod mikro F1-mjere zbrajamo matrice zabune kroz sve klase, a primjera iz manjih klasa ima manje, pa manje doprinose pogrešci

[B] $F_1^M < F_1^\mu$, jer kod makro F1-mjere računamo prosjek F1-mjere kroz sve klase, a klasifikator na manjim klasama više griješi

[C] $F_1^M < F_1^\mu$, jer kod mikro F1-mjere zbrajamo matrice zabune kroz sve klase, pa je ukupan broj pogrešaka veći nego kad računamo prosječnu pogrešku kroz klase

[D] $F_1^M > F_1^\mu$, jer kod makro F1-mjere zbrajamo matrice zabune kroz sve klase, a primjera iz većih klasa ima više, pa više doprinose pogrešci

21 (P) Na istom skupu označenih primjera vrednujemo četiri binarna klasifikatora s vjerojatnosnim izlazima. Za vrednovanje koristimo krivulju ROC. Svaki smo klasifikator ispitali s tri vrijednosti klasifikacijskog praga te smo za te vrijednosti izračunali FPR (stopa lažnog alarma) i TPR (odziv). Dobiveni parovi vrijednosti (FPR, TPR) za sva četiri klasifikatora su sljedeći:

$$h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) \qquad h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8)$$
$$h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) \qquad h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0)$$

Skicirajte odgovarajuće krivulje ROC, linearno interpolirajući između izmjerenih točaka, a dodajte i točke koje odgovaraju krajnjim vrijednostima klasifikacijskog praga (0 i 1). Pritom, ako je neki klasifikator lošiji od nasumičnog klasifikatora, umjesto tog klasifikatora razmatrajte njegovu popravljenu varijantu koju ćete dobiti invertiranjem izlaza ($1 - h(\mathbf{x})$ umjesto $h(\mathbf{x})$). **Koji su od ispitanih klasifikatora (eventualno nakon popravka) najbolji prema krivulji ROC?**

[A] h_1 i h_2 [B] h_2 i h_3 [C] h_3 i h_4 [D] h_1 i h_4

22 (N) Na ispitnome skupu evaluiramo model multinomijalne logističke regresije (MLR) za klasifikaciju u $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 20 & 18 & 3 \\ 21 & 25 & 2 \\ 4 & 2 & 5 \end{pmatrix} \end{array}$$

Klasifikator MLR uspoređujemo s klasifikatorom RAND koji primjere klasificira nasumično, i to tako da oznaku $y = j$ dodjeljuje s vjerojatnošću proporcionalnoj udjelu klase j u ispitnome skupu. Izračunajte mikro-F_1 za klasifikator MLR i očekivani mikro-F_1 za klasifikator RAND. **Koliko iznosi očekivana razlika u vrijednostima mikro-F_1 klasifikatora MLR i RAND?**

A 0.185 B 0.155 C 0.205 D 0.085

Završni ispit iz Strojnog učenja 1 (ak. god. 2021./2022.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 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 1: Procjena parametara i Bayesov klasifikator (8 pitanja)

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

2 (P) Naivan Bayesov klasifikator pretpostavlja uvjetnu nezavisnost značajki unutar neke klase, premda u stvarnosti ta pretpostavka rijetko kada vrijedi. Razmotrite klasifikator poruka e-pošte čija je zadaća odrediti je li poruka prijevara (*scam*). Neka $y = 1$ označava da je poruka prijevara, a $y = 0$ da ona to nije. Model koristi binarne značajke koje indiciraju pojavljivanje određene riječi u tekstu poruke. Na primjer, izglednost $P(\text{lottery}|y = 1)$ jest vjerojatnost da se u prijevarnoj poruci pojavi riječ "lottery". Razmotrite sljedeće četiri riječi koje se općenito mogu pojaviti porukama e-pošte: "lottery", "winner", "prince" i "faithfully". Primijetite da mnoge prijevarne poruke od žrtve traže uplatu predujma. Takve poruke tipično govore o ostavštini nigerijskog princa ili o dobitku na lutriji (nikada oboje!) te završavaju kićenim pozdravom "yours faithfully". **Za koju od sljedećih jednakosti očekujemo da tipično *ne vrijedi* i da je time narušena pretpostavka naivnog Bayesovog klasifikatora?**

A $P(\text{lottery}|y = 1) = P(\text{lottery}|\text{winner}, y = 1)$
B $P(\text{faithfully}|y = 0) = P(\text{faithfully}|\text{winner}, y = 0)$
C $P(\text{prince}|y = 1) = P(\text{prince}|\text{faithfully}, y = 1)$
D $P(\text{faithfully}|y = 1) = P(\text{faithfully}|\text{winner}, y = 1)$

3 (P) MAP-procjeniteljem na skupu $\mathcal{D}$ procjenjujemo parametre $\boldsymbol{\mu}$ kategoričke (multinulijeve) varijable $\mathbf{x}$, tj. procjenjujemo parametre multinulijeve distribucije $P(\mathbf{x}|\boldsymbol{\mu})$. Varijabla $\mathbf{x}$ može poprimiti tri moguće vrijednosti ($K = 3$). Apriorna Dirichletova distribucija $P(\boldsymbol{\mu}|\boldsymbol{\alpha})$ definirana je parametrima $\boldsymbol{\alpha}$ definiranima kao $\boldsymbol{\alpha} = \alpha \cdot (1, 4, 2)$, gdje faktor α određuje vršnost Dirichletove distribucije, $\alpha \geqslant 1$. U skupu podataka $\mathcal{D}$ realizirane su sve tri vrijednosti kategoričke varijable $\mathbf{x}$, i to sa sljedećim brojem realizacija: $N_1 = 50$, $N_2 = 29$, $N_3 = 40$. **Koliko mora iznositi faktor α, a da bi vrijednost x_2 bila najvjerojatnija vrijednost kategoričke varijable x prema procijenjenoj distribuciji $P(\mathbf{x}|\boldsymbol{\mu})$?**

A $\alpha > 7$ B $\alpha > 2$ C $\alpha > 5$ D $\alpha > 8$

4 (T) Jedan od nedostataka generativnih modela u odnosu na diskriminativne modele jest nepotrebna složenost modeliranja. **Što to zapravo znači?**

A Za razliku od diskriminativnih modela, generativni modeli distribuciju $p(y|\mathbf{x})$ definiraju zasebno za svaku klasu, pa stoga imaju više parametara

B Za klasifikaciju nam je potrebna samo distribucija $p(y|\mathbf{x})$, i ona se može modelirati sa manje parametara od zajedničke distribucije $p(\mathbf{x}, y)$

C Zajednička distribucija $p(\mathbf{x}, y)$ može se faktorizirati kao $p(y|\mathbf{x})p(\mathbf{x})$, no takva faktorizacija ima više parametara

D Generativni modeli modeliraju distribuciju $p(y|\mathbf{x})$, što iziskuje više parametara nego li modeliranje granice između klasa

5 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{-3, -2, 1, 4\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Koristimo MLE kako bismo na uzorku $\mathcal{D}$ procijenili parametre $\hat{\mu}_{\mathrm{MLE}}$ i $\hat{\sigma}^2_{\mathrm{MLE}}$. Usporedbe radi, parametar σ^2 dodatno procjenjujemo nepristranim procjeniteljem $\hat{\sigma}^2_{\mathrm{UB}}$. Izračunajte log-izglednost parametara $(\hat{\mu}_{\mathrm{MLE}}, \hat{\sigma}^2_{\mathrm{MLE}})$ te log-izglednost parametara $(\hat{\mu}_{\mathrm{MLE}}, \hat{\sigma}^2_{\mathrm{UB}})$. **Koliko je prva log-izglednost veća od druge log-izglednosti?**

A 0.095 B 0.075 C 0.018 D 0.058

6 (T) Parametre modela možemo procijeniti pomoću procjenitelja najveće izglednosti (MLE) ili procjenitelja najveće aposteriorne vjerojatnosti (MAP). Općenito, MLE i MAP daju različite procjene, no u nekim slučajevima mogu dati jednake procjene. **Kada će MLE i MAP dati jednake procjene?**

A Kada su $p(\boldsymbol{\theta})$ i $p(\boldsymbol{\theta}|\mathcal{D})$ različite distribucije

B Kada $p(\boldsymbol{\theta})$ definiramo kao unimodalnu distribuciju

C Kada broj primjera N teži prema beskonačno

D Kada je $p(\mathcal{D}|\boldsymbol{\theta})$ uniformna distribucija

7 (P) Gaussov Bayesov klasifikator koristimo za klasfikaciju jednodimenzijskih podataka u tri klase. Procijenjene izglednosti klasa su $p(x|y=1) = \mathcal{N}(-10, 2)$, $p(x|y=2) = \mathcal{N}(2, 2)$ i $p(x|y=3) = \mathcal{N}(8, 2)$, a procijenjene apriorne vjerojatnosti klasa su $P(y=1) = P(y=2) = 2/5$ i $P(y=3) = 1/5$. Međutim, nakon što smo naučili ovaj model, zaključili smo da na ispitnom skupu postoji pomak u distribuciji podataka u odnosu na skup za učenje te da zbog toga model ne generalizira dobro. Zaključili smo da se ovo može ispraviti tako da se naučeni model malo izmijeni, i to tako da se varijanca izglednosti klase $y = 1$ postavi na 5 i da se apriorne vjerojatnosti klasa ujednače, $P(y=1) = P(y=2) = P(y=3) = 1/3$. Skicirajte gustoće zajedničke vjerojatnosti naučenog i izmijenjenog modela. Neka su h_1 i h_2 MAP-hipoteze prvog i drugog modela, te neka su a i b pozitivne konstante. Razmotrite segment ulaznog prostora za koji vrijedi $-10 \leqslant x \leqslant 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

A $[-4-a, -4] \cup [5-b, 5]$ B $[-4-a, 5+b]$ C $[-4-a, -4+b]$ D $[-4, -4+a] \cup [5, 5+b]$

8 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 2$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene i dijagonalne kovarijacijske matrice. Izglednost klase s oznakom $y = j$ definirana je multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

Model treniramo na skup podataka od $N = 7$ primjera:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((-1,-2),0), ((0,0),0), ((1,2),0), ((3,-1),1), ((4,-1),1), ((4,1),1), ((5,1),1)\}$$

Procijenite parametre modela na ovom skupu primjera. Budući da je skup primjera malen, za procjenu kovarijacijske matrice koristite nepristran procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y=j)$. **Koliko iznosi $h_1(\mathbf{x})$ za primjer $\mathbf{x} = (4, 0)$?**

A -4.13 B -2.75 C -3.03 D -3.84

Cjelina 2: Probabilistički grafički modeli (5 pitanja)

9 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Sve varijable su binarne, osim varijabli v i x, koje su ternarne. Uz navedeni topološki uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

$$\{v, w\} \perp y | x \qquad \{v, x\} \perp z | \{w, y\}$$

Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

A 18 B 21 C 23 D 24

10 (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 modelira 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 0.813 C 0.148 D 0.033

11 (P) Bayesovom mrežom sa sedam binarnih varijabli modeliramo prometne (ne)prilike u gradu Zagrebu. U našoj mreži, strelice modeliraju pretpostavljeni smjer kauzalnosti. Jutarnje doba dana (J) ili loše vrijeme (V) mogu uzrokovati nastanak prometne gužve (G). Loše vrijeme također može uzrokovati nastanak prometne nesreće (N). Nadalje, nastanak prometne nesreće može uzrokovati nastanak prometne gužve. Loše vrijeme također može uzrokovati zastoj tramvaja (T), no i nestanak struje (S) može uzrokovati zastoj tramvaja. Konačno, zastoj tramvaja može uzrokovati masovno pješačenje putnika (P), što opet može uzrokovati prometnu nesreću. Primjenom algoritma d-odvajanja ispitujemo stohastičku zavisnost između jutarnjeg doba dana (J) i nestanka struje (S). **Koje minimalan skup varijabli treba biti opažen, a da vjerojatnost od S ne ovisi o vrijednosti J?**

A V, P B G, T C T D N, T

12 (T) Čest način probabilističkog zaključivanja kod Bayesovih mreža jest izračunavanje "aposteriornog upita". Kod te vrste upita zanima nas distribucija nekih varijabli (varijable upita) na temelju zadanih varijabli (opažene varijable). Međutim, Bayesova mreža kodira zajedničku vjerojatnost svih varijabli mreže, uključivo i varijabli koje nisu niti varijable upita niti opažene varijable (varijable smetnje). **Na koji način izračunavamo aposteriorni upit?**

A Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama upita i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i opaženim varijablama

B Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama smetnje i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i varijablama upita

C Kao umnožak zajedničke vjerojatnosti marginalizirane po opaženim varijablama i zajedničke vjerojatnosti marginalizirane po varijablama smetnje

D Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama upita i zajedničke vjerojatnosti marginalizirane po opaženim varijablama

13 (N) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x, y, z, w) = P(z)P(w)P(y|z, w)P(x|y)$. Sve varijable su binarne. Vrijedi $P(z = 1) = 0.1$, $P(w = 1) = 0.1$, $P(x = 1|y = 0) = 0.2$ i $P(x = 1|y = 1) = 0.6$. Tablica uvjetne vjerojatnosti za čvor y je sljedeća:

z	w	$p(y = 1|z, w)$
0	0	0
0	1	0.4
1	0	0.2
1	1	0.7

Postupkom uzorkovanja s odbijanjem uzorkujemo iz aposteriorne distribucije $P(w|x=0,z=1)$. Uzorkovanje smo ponovili ukupno $N=10000$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N. **U uzorku neodbačenih vektora, koliko je očekivano više vektora sa $w=0$ od vektora sa $w=1$?**

A 348 B 596 C 472 D 944

Cjelina 3: Grupiranje (6 pitanja)

14 (P) Skup neoznačenih primjera u dvodimenzijskome ulaznom prostoru neka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{5} = \{(0,0),(0,4),(2,0),(2,4),(4,2)\}$$

Primjere grupiramo algoritmom K-sredina sa $K=3$ grupe. Za početna središta grupa odaberemo nasumično primjere iz $\mathcal{D}$, pri čemu, naravno, pazimo da odaberemo različita središta. Ishod grupiranja i konačan iznos kriterijske funkcije J ovisit će o odabiru početnih središta. Neka je J^* vrijednost kriterijske funkcije u točki globalnog minimuma, dakle vrijednost koja odgovara najboljem grupiranju. Neka je J^+ vrijednost kriterijske funkcije u točki lokalnog minimuma, i to onoj točki lokalnog minimuma s najvećom vrijednošću funkcije J. **Koliko iznosi razlika $J^+ - J^*$?**

A 8 B 4 C 6 D 12

15 (P) Algoritmom GMM grupiramo primjere u dvodimenzijskome ulaznom prostoru. Skup podataka u stvarnosti je uzorkovan iz zajedničke distribucije koja se može opisati sljedećim mješavinskim modelom:

$$p(\mathbf{x}) = \sum_{j=1}^{3} \frac{1}{3}\mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \mu_1 = (5,5), \mu_2 = (5,10), \mu_3 = (-10,-10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

Skup $\mathcal{D}$ grupiramo u $K=2$ grupe. Pritom isprobavamo tri modela, koji se međusobno razlikuju po pretpostavkama na kovarijacijsku matricu. Konkretno: dijeljena i puna kovarijacijska matrica ($\mathcal{H}_1$), nedijeljena i dijagonalna kovarijacijska matrica ($\mathcal{H}_2$) i nedijeljena i izotropna kovarijacijska matrica ($\mathcal{H}_3$). Neka je $\mathcal{L}_i$ izglednost parametara dobivena modelom $\mathcal{H}_i$ nakon konvergencije algoritma. Za inicijalizaciju središta koristi se algoritam K-means++. **Što su očekivani odnosi između izglednosti za ova tri modela?**

A $\mathcal{L}_1 > \mathcal{L}_2 > \mathcal{L}_3$ B $\mathcal{L}_1 > \mathcal{L}_3, \mathcal{L}_2 > \mathcal{L}_3$ C $\mathcal{L}_1 = \mathcal{L}_2 > \mathcal{L}_3$ D $\mathcal{L}_2 > \mathcal{L}_1, \mathcal{L}_2 > \mathcal{L}_3$

16 (T) Algoritam K-means++ proširenje je algoritma K-sredina heurističkim odabirom početnih središta grupa. **Koja je glavna ideja odabira početnih središta grupa kod algoritma K-means++?**

A Vjerojatnost da je neki primjer središte grupe proporcionalna je s brojem primjera u toj grupi i udaljenosti tih primjera od centroida skupa podataka

B Središta grupe su srednje vrijednosti grupa projiciranih na pravac u smjeru prve komponente rastava skupa podataka na glavne komponente (PCA)

C Najvjerojatnije središte grupe jest primjer koji je najviše udaljen od njemu najbližeg središta

D Središta grupa su mjesta na kojima graf kriterijske funkcije u ovisnosti o broju grupa naglo opada pa stagnira

17 (T) Model miješane gustoće sa K komponenti vjerojatnosnu distribuciju neoznačenih podaka definira kao $p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \, p(\mathbf{x}|\boldsymbol{\theta}_k)$. Što modelira uvjetna vjerojatnost $p(\mathbf{x}|\boldsymbol{\theta}_k)$?

A Stupanj pripadnosti primjera $\mathbf{x}$ grupi k

B Gustoću vjerojatnosti primjera $\mathbf{x}$ unutar grupe k

C Vjerojatnost da primjer $\mathbf{x}$ pripada grupi k

D Gustoću vjerojatnosti grupe k za primjer $\mathbf{x}$

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
$y_{pred}^{(i)}$	0	1	2	1	1	0	0	2	1	1
$y_{true}^{(i)}$	1	0	0	2	1	0	2	1	0	1

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.53   B 0.62   C 0.49   D 0.58

19 (N) Algoritmom HAC grupiramo riječi engleskog jezika. Neoznačeni skup podataka sastoji se od sljedećih riječi:

$$\mathcal{D} = \{\text{“water”}, \text{“watering”}, \text{“earth”}, \text{“air”}\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2| / |\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“watering”}) = 5/8 = 0.625$. Provedite prve dvije iteracije grupiranja algoritmom HAC uz prosječno povezivanje. **Na kojoj se razini sličnosti spajaju grupe u drugoj iteraciji algoritma HAC?**

A 0.478   B 0.583   C 0.292   D 0.535

Cjelina 4: Vrednovanje modela (3 pitanja)

20 (P) Na istom skupu označenih primjera vrednujemo četiri binarna klasifikatora s vjerojatnosnim izlazima. Za vrednovanje koristimo krivulju ROC. Svaki smo klasifikator ispitali s tri vrijednosti klasifikacijskog praga te smo za te vrijednosti izračunali FPR (stopa lažnog alarma) i TPR (odziv). Dobiveni parovi vrijednosti (FPR, TPR) za sva četiri klasifikatora su sljedeći:

$$h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) \qquad h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8)$$
$$h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) \qquad h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0)$$

Skicirajte odgovarajuće krivulje ROC, linearno interpolirajući između izmjerenih točaka, a dodajte i točke koje odgovaraju krajnjim vrijednostima klasifikacijskog praga (0 i 1). Pritom, ako je neki klasifikator lošiji od nasumičnog klasifikatora, umjesto tog klasifikatora razmatrajte njegovu popravljenu varijantu koju ćete dobiti invertiranjem izlaza ($1 - h(\mathbf{x})$ umjesto $h(\mathbf{x})$). **Koji su od ispitanih klasifikatora (eventualno nakon popravka) najbolji prema krivulji ROC?**

A h_1 i h_2 B h_3 i h_4 C h_1 i h_4 D h_1 i h_3

21 (N) Na ispitnome skupu evaluiramo model multinomijalne logističke regresije (MLR) za klasifikaciju u $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{c c} & \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{array}{c} y=1 \\ y=2 \\ y=3 \end{array} & \begin{pmatrix} 20 & 18 & 3 \\ 21 & 25 & 2 \\ 4 & 2 & 5 \end{pmatrix} \end{array}$$

Klasifikator MLR uspoređujemo s klasifikatorom RAND koji primjere klasificira nasumično, i to tako da oznaku $y = j$ dodjeljuje s vjerojatnošću proporcionalnoj udjelu klase j u ispitnome skupu. Izračunajte mikro-F_1 za klasifikator MLR i očekivani mikro-F_1 za klasifikator RAND. **Koliko iznosi očekivana razlika u vrijednostima mikro-F_1 klasifikatora MLR i RAND?**

A 0.155   B 0.205   C 0.085   D 0.185

22 (T) Za vrednovanje višeklasnog klasifikatora često se koriste mjere F_1^μ (mikro F1-mjera) i F_1^M (makro F1-mjera). **Koji je očekivani odnos između vrijednosti tih mjera, i zašto?**

A $F_1^M < F_1^\mu$, jer kod mikro F1-mjere računamo prosjek F1-mjere kroz sve klase, a klasifikator na većima klasama manje griješi

B $F_1^M < F_1^\mu$, jer kod makro F1-mjere računamo prosjek F1-mjere kroz sve klase, a klasifikator na manjim klasama više griješi

C $F_1^M > F_1^\mu$, jer kod mikro F1-mjere zbrajamo matrice zabune kroz sve klase, a primjera iz manjih klasa ima manje, pa manje doprinose pogrešci

D $F_1^M < F_1^\mu$, jer kod mikro F1-mjere zbrajamo matrice zabune kroz sve klase, pa je ukupan broj pogrešaka veći nego kad računamo prosječnu pogrešku kroz klase

Završni ispit iz Strojnog učenja 1 (ak. god. 2021./2022.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 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 1: Procjena parametara i Bayesov klasifikator (8 pitanja)

1 (P) Naivan Bayesov klasifikator pretpostavlja uvjetnu nezavisnost značajki unutar neke klase, premda u stvarnosti ta pretpostavka rijetko kada vrijedi. Razmotrite klasifikator poruka e-pošte čija je zadaća odrediti je li poruka prijevara (*scam*). Neka $y = 1$ označava da je poruka prijevara, a $y = 0$ da ona to nije. Model koristi binarne značajke koje indiciraju pojavljivanje određene riječi u tekstu poruke. Na primjer, izglednost $P(\text{lottery}|y = 1)$ jest vjerojatnost da se u prijevarnoj poruci pojavi riječ "lottery". Razmotrite sljedeće četiri riječi koje se općenito mogu pojaviti porukama e-pošte: "lottery", "winner", "prince" i "faithfully". Primijetite da mnoge prijevarne poruke od žrtve traže uplatu predujma. Takve poruke tipično govore o ostavštini nigerijskog princa ili o dobitku na lutriji (nikada oboje!) te završavaju kićenim pozdravom "yours faithfully". **Za koju od sljedećih jednakosti očekujemo da tipično *ne vrijedi* i da je time narušena pretpostavka naivnog Bayesovog klasifikatora?**

- A $P(\text{faithfully}|y = 1) = P(\text{faithfully}|\text{winner}, y = 1)$
- B $P(\text{winner}|y = 1) = P(\text{winner}|\text{faithfully}, y = 1)$
- C $P(\text{faithfully}|y = 1) = P(\text{faithfully}|\text{prince}, y = 1)$
- D $P(\text{winner}|y = 1) = P(\text{winner}|\text{prince}, y = 1)$

2 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{-3, -2, 1, 4\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Koristimo MLE kako bismo na uzorku $\mathcal{D}$ procijenili parametre $\hat{\mu}_{\text{MLE}}$ i $\hat{\sigma}^2_{\text{MLE}}$. Usporedbe radi, parametar σ^2 dodatno procjenjujemo nepristranim procjeniteljem $\hat{\sigma}^2_{\text{UB}}$. Izračunajte log-izglednost parametara $(\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{MLE}})$ te log-izglednost parametara $(\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{UB}})$. **Koliko je prva log-izglednost veća od druge log-izglednosti?**

A 0.075 B 0.095 C 0.018 D 0.058

3 (T) Jedan od nedostataka generativnih modela u odnosu na diskriminativne modele jest nepotrebna složenost modeliranja. **Što to zapravo znači?**

- A Za klasifikaciju nam je potrebna samo distribucija $p(y|\mathbf{x})$, i ona se može modelirati sa manje parametara od zajedničke distribucije $p(\mathbf{x}, y)$
- B Generativni modeli modeliraju distribuciju $p(y|\mathbf{x})$, što iziskuje više parametara nego li modeliranje granice između klasa
- C Za razliku od diskriminativnih modela, generativni modeli distribuciju $p(y|\mathbf{x})$ definiraju zasebno za svaku klasu, pa stoga imaju više parametara
- D Zajednička distribucija $p(\mathbf{x}, y)$ može se faktorizirati kao $p(y|\mathbf{x})p(\mathbf{x})$, no takva faktorizacija ima više parametara

4 (T) Parametre modela možemo procijeniti pomoću procjenitelja najveće izglednosti (MLE) ili procjenitelja najveće aposteriorne vjerojatnosti (MAP). Općenito, MLE i MAP daju različite procjene, no u nekim slučajevima mogu dati jednake procjene. **Kada će MLE i MAP dati jednake procjene?**

- A Kada se u $\mathcal{D}$ realizirala svaka vrijednost slučajne varijable
- B Kada $p(\boldsymbol{\theta})$ definiramo kao uniformnu distribuciju
- C Kada je $p(\boldsymbol{\theta})$ konjugatna izglednosti $p(\mathcal{D}|\boldsymbol{\theta})$
- D Kada koristimo nepristran procjenitelj za $\boldsymbol{\theta}$

5 (P) MAP-procjeniteljem na skupu $\mathcal{D}$ procjenjujemo parametre $\boldsymbol{\mu}$ kategoričke (multinulijeve) varijable $\mathbf{x}$, tj. procjenjujemo parametre multinulijeve distribucije $P(\mathbf{x}|\boldsymbol{\mu})$. Varijabla $\mathbf{x}$ može proprimiti tri moguće vrijednosti ($K = 3$). Apriorna Dirichletova distribucija $P(\boldsymbol{\mu}|\boldsymbol{\alpha})$ definirana je parametrima $\boldsymbol{\alpha}$ definiranima kao $\boldsymbol{\alpha} = \alpha \cdot (1, 4, 2)$, gdje faktor α određuje vršnost Dirichletove distribucije, $\alpha \geqslant 1$. U skupu podataka $\mathcal{D}$ realizirane su sve tri vrijednosti kategoričke varijable $\mathbf{x}$, i to sa sljedećim brojem realizacija: $N_1 = 50$, $N_2 = 29$, $N_3 = 40$. **Koliko mora iznositi faktor α, a da bi vrijednost x_2 bila najvjerojatnija vrijednost kategoričke varijable x prema procijenjenoj distribuciji $P(\mathbf{x}|\boldsymbol{\mu})$?**

A $\alpha > 5$ B $\alpha > 7$ C $\alpha > 8$ D $\alpha > 2$

6 (P) Gaussov Bayesov klasifikator koristimo za klasfikaciju jednodimenzijskih podataka u tri klase. Procijenjene izglednosti klasa su $p(x|y = 1) = \mathcal{N}(-10, 2)$, $p(x|y = 2) = \mathcal{N}(2, 2)$ i $p(x|y = 3) = \mathcal{N}(8, 2)$, a procijenjene apriorne vjerojatnosti klasa su $P(y = 1) = P(y = 2) = 1/5$ i $P(y = 3) = 3/5$. Međutim, nakon što smo naučili ovaj model, zaključili smo da na ispitnom skupu postoji pomak u distribuciji podataka u odnosu na skup za učenje te da zbog toga model ne generalizira dobro. Zaključili smo da se ovo može ispraviti tako da se naučeni model malo izmijeni, i to tako da se varijanca izglednosti klase $y = 1$ postavi na 5 i da se apriorne vjerojatnosti klasa ujednače, $P(y = 1) = P(y = 2) = P(y = 3) = 1/3$. Skicirajte gustoće zajedničke vjerojatnosti naučenog i izmijenjenog modela. Neka su h_1 i h_2 MAP-hipoteze prvog i drugog modela, te neka su a i b pozitivne konstante. Razmotrite segment ulaznog prostora za koji vrijedi $-10 \leqslant x \leqslant 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

A $[-4, -4 - a] \cup [5, 5 + b]$ B $[-4 - a, 5 + b]$ C $[-4, -4 + a] \cup [5 - b, 5]$ D $[-4 - a, -4 + b]$

7 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 2$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene i dijagonalne kovarijacijske matrice. Izglednost klase s oznakom $y = j$ definirana je multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

Model treniramo na skup podataka od $N = 7$ primjera:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)}))\}_i = \{((-1, -2), 0), ((0, 0), 0), ((1, 2), 0), ((3, -1), 1), ((4, -1), 1), ((4, 1), 1), ((5, 1), 1)\}$$

Procijenite parametre modela na ovom skupu primjera. Budući da je skup primjera malen, za procjenu kovarijacijske matrice koristite nepristran procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. **Koliko iznosi $h_1(\mathbf{x})$ za primjer $\mathbf{x} = (4, 0)$?**

A -3.03 B -4.13 C -2.75 D -3.84

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:

	$x_3 = 0$		$x_3 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_1 = 0$	0.1	0.0	0.0	0.1
$x_1 = 1$	0.1	0.4	0.1	0.2

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

Cjelina 2: Probabilistički grafički modeli (5 pitanja)

9 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Sve varijable su binarne, osim varijabli v i x, koje su ternarne. Uz navedeni topološki uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

$$\{v, w\} \perp y | x \qquad \{v, x\} \perp z | \{w, y\}$$

Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

A 23 B 18 C 21 D 24

10 (P) Bayesovom mrežom sa sedam binarnih varijabli modeliramo prometne (ne)prilike u gradu Zagrebu. U našoj mreži, strelice modeliraju pretpostavljeni smjer kauzalnosti. Jutarnje doba dana (J) ili loše vrijeme (V) mogu uzrokovati nastanak prometne gužve (G). Loše vrijeme također može uzrokovati nastanak prometne nesreće (N). Nadalje, nastanak prometne nesreće može uzrokovati nastanak prometne gužve. Loše vrijeme također može uzrokovati zastoj tramvaja (T), no i nestanak struje (S) može uzrokovati zastoj tramvaja. Konačno, zastoj tramvaja može uzrokovati masovno pješačenje putnika (P), što opet može uzrokovati prometnu nesreću. Primjenom algoritma d-odvajanja ispitujemo stohastičku zavisnost između jutarnjeg doba dana (J) i nestanka struje (S). **Koje minimalan skup varijabli treba biti opažen, a da vjerojatnost od S ne ovisi o vrijednosti J?**

A V, P B N, T C T D G, T

11 (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 modelira 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 0.148 C 0.813 D 0.033

12 (N) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x, y, z, w) = P(z)P(w)P(y|z, w)P(x|y)$. Sve varijable su binarne. Vrijedi $P(z = 1) = 0.2$, $P(w = 1) = 0.2$, $P(x = 1|y = 0) = 0.2$ i $P(x = 1|y = 1) = 0.6$. Tablica uvjetne vjerojatnosti za čvor y je sljedeća:

z	w	$p(y = 1\|z, w)$
0	0	0
0	1	0.4
1	0	0.2
1	1	0.7

Postupkom uzorkovanja s odbijanjem uzorkujemo iz aposteriorne distribucije $P(w|x = 0, z = 1)$. Uzorkovanje smo ponovili ukupno $N = 10000$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N. **U uzorku neodbačenih vektora, koliko je očekivano više vektora sa $w = 0$ od vektora sa $w = 1$?**

A 348 B 472 C 944 D 596

13 (T) Čest način probabilističkog zaključivanja kod Bayesovih mreža jest izračunavanje "aposteriornog upita". Kod te vrste upita zanima nas distribucija nekih varijabli (varijable upita) na temelju zadanih varijabli (opažene varijable). Međutim, Bayesova mreža kodira zajedničku vjerojatnost svih varijabli mreže, uključivo i varijabli koje nisu niti varijable upita niti opažene varijable (varijable smetnje). **Na koji način izračunavamo aposteriorni upit?**

A Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama upita i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i opaženim varijablama

B Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama upita i zajedničke vjerojatnosti marginalizirane po opaženim varijablama

C Kao umnožak vjerojatnosti varijable upita uvjetovane na opažene varijable i zajedničke vjerojatnosti marginalizirane po varijablama smetnje

D Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama smetnje i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i varijablama upita

Cjelina 3: Grupiranje (6 pitanja)

14 (N) Algoritmom HAC grupiramo riječi engleskog jezika. Neoznačeni skup podataka sastoji se od sljedećih riječi:

$$\mathcal{D} = \{\text{"water"}, \text{"watering"}, \text{"earth"}, \text{"air"}\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{"water"}, \text{"watering"}) = 5/8 = 0.625$. Provedite prve dvije iteracije grupiranja algoritmom HAC uz prosječno povezivanje. **Na kojoj se razini sličnosti spajaju grupe u drugoj iteraciji algoritma HAC?**

A 0.535 B 0.583 C 0.292 D 0.478

15 (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
$y_{pred}^{(i)}$	0	1	2	1	1	0	0	2	1	1
$y_{true}^{(i)}$	1	0	0	2	1	0	2	1	0	1

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.62 B 0.58 C 0.53 D 0.49

16 (T) Model miješane gustoće sa K komponenti vjerojatnosnu distribuciju neoznačenih podaka definira kao $p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \, p(\mathbf{x}|\boldsymbol{\theta}_k)$. Što modelira uvjetna vjerojatnost $p(\mathbf{x}|\boldsymbol{\theta}_k)$?

A Stupanj pripadnosti primjera $\mathbf{x}$ grupi k

B Stupanj pripadnosti grupe k primjeru $\mathbf{x}$

C Gustoću vjerojatnosti grupe k za primjer $\mathbf{x}$

D Gustoću vjerojatnosti primjera $\mathbf{x}$ unutar grupe k

17 (P) Algoritmom GMM grupiramo primjere u dvodimenzijskome ulaznom prostoru. Skup podataka u stvarnosti je uzorkovan iz zajedničke distribucije koja se može opisati sljedećim mješavinskim modelom:

$$p(\mathbf{x}) = \sum_{j=1}^{3} \frac{1}{3} \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \mu_1 = (5,5), \mu_2 = (5,10), \mu_3 = (-10,-10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

Skup $\mathcal{D}$ grupiramo u $K = 2$ grupe. Pritom isprobavamo tri modela, koji se međusobno razlikuju po pretpostavkama na kovarijacijsku matricu. Konkretno: dijeljena i puna kovarijacijska matrica ($\mathcal{H}_1$), nedijeljena i dijagonalna kovarijacijska matrica ($\mathcal{H}_2$) i nedijeljena i izotropna kovarijacijska matrica ($\mathcal{H}_3$). Neka je $\mathcal{L}_i$ izglednost parametara dobivena modelom $\mathcal{H}_i$ nakon konvergencije algoritma. Za inicijalizaciju središta koristi se algoritam K-means++. **Što su očekivani odnosi između izglednosti za ova tri modela?**

A $\mathcal{L}_1 > \mathcal{L}_2 > \mathcal{L}_3$ B $\mathcal{L}_1 > \mathcal{L}_3$, $\mathcal{L}_2 > \mathcal{L}_3$ C $\mathcal{L}_2 > \mathcal{L}_1$, $\mathcal{L}_2 > \mathcal{L}_3$ D $\mathcal{L}_1 = \mathcal{L}_2 > \mathcal{L}_3$

18 (T) Algoritam K-means++ proširenje je algoritma K-sredina heurističkim odabirom početnih središta grupa. **Koja je glavna ideja odabira početnih središta grupa kod algoritma K-means++**?

A Središta grupa su mjesta na kojima graf kriterijske funkcije u ovisnosti o broju grupa naglo opada pa stagnira

B Vjerojatnost da je neki primjer središte grupe proporcionalna je s brojem primjera u toj grupi i udaljenosti tih primjera od centroida skupa podataka

C Središta grupe su srednje vrijednosti grupa projiciranih na pravac u smjeru prve komponente rastava skupa podataka na glavne komponente (PCA)

D Najvjerojatnije središte grupe jest primjer koji je najviše udaljen od njemu najbližeg središta

19 (P) Skup neoznačenih primjera u dvodimenzijskome ulaznom prostoru neka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{5} = \{(0,0),(0,4),(2,0),(2,4),(4,2)\}$$

Primjere grupiramo algoritmom K-sredina sa $K = 3$ grupe. Za početna središta grupa odaberemo nasumično primjere iz $\mathcal{D}$, pri čemu, naravno, pazimo da odaberemo različita središta. Ishod grupiranja i konačan iznos kriterijske funkcije J ovisit će o odabiru početnih središta. Neka je J^* vrijednost kriterijske funkcije u točki globalnog minimuma, dakle vrijednost koja odgovara najboljem grupiranju. Neka je J^+ vrijednost kriterijske funkcije u točki lokalnog minimuma, i to onoj točki lokalnog minimuma s najvećom vrijednošću funkcije J. **Koliko iznosi razlika** $J^+ - J^*$?

A 8 B 12 C 4 D 6

Cjelina 4: Vrednovanje modela (3 pitanja)

20 (T) Za vrednovanje višeklasnog klasifikatora često se koriste mjere F_1^μ (mikro F1-mjera) i F_1^M (makro F1-mjera). **Koji je očekivani odnos između vrijednosti tih mjera, i zašto?**

A $F_1^M < F_1^\mu$, jer kod makro F1-mjere računamo prosjek F1-mjere kroz sve klase, a klasifikator na manjim klasama više griješi

B $F_1^M < F_1^\mu$, jer kod mikro F1-mjere računamo prosjek F1-mjere kroz sve klase, a klasifikator na većima klasama manje griješi

C $F_1^M > F_1^\mu$, jer kod makro F1-mjere zbrajamo matrice zabune kroz sve klase, a primjera iz većih klasa ima više, pa više doprinose pogrešci

D $F_1^M < F_1^\mu$, jer kod mikro F1-mjere zbrajamo matrice zabune kroz sve klase, pa je ukupan broj pogrešaka veći nego kad računamo prosječnu pogrešku kroz klase

21 (P) Na istom skupu označenih primjera vrednujemo četiri binarna klasifikatora s vjerojatnosnim izlazima. Za vrednovanje koristimo krivulju ROC. Svaki smo klasifikator ispitali s tri vrijednosti klasifikacijskog praga te smo za te vrijednosti izračunali FPR (stopa lažnog alarma) i TPR (odziv). Dobiveni parovi vrijednosti (FPR, TPR) za sva četiri klasifikatora su sljedeći:

$$h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) \qquad h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8)$$
$$h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) \qquad h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0)$$

Skicirajte odgovarajuće krivulje ROC, linearno interpolirajući između izmjerenih točaka, a dodajte i točke koje odgovaraju krajnjim vrijednostima klasifikacijskog praga (0 i 1). Pritom, ako je neki klasifikator lošiji od nasumičnog klasifikatora, umjesto tog klasifikatora razmatrajte njegovu popravljenu varijantu koju ćete dobiti invertiranjem izlaza ($1 - h(\mathbf{x})$ umjesto $h(\mathbf{x})$). **Koji su od ispitanih klasifikatora (eventualno nakon popravka) najbolji prema krivulji ROC?**

A h_1 i h_4 B h_1 i h_2 C h_2 i h_3 D h_3 i h_4

22 (N) Na ispitnome skupu evaluiramo model multinomijalne logističke regresije (MLR) za klasifikaciju u $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 35 & 8 & 3 \\ 10 & 25 & 5 \\ 0 & 12 & 2 \end{pmatrix} \end{array}$$

Klasifikator MLR uspoređujemo s klasifikatorom RAND koji primjere klasificira nasumično, i to tako da oznaku $y = j$ dodjeljuje s vjerojatnošću proporcionalnoj udjelu klase j u ispitnome skupu. Izračunajte mikro-F_1 za klasifikator MLR i očekivani mikro-F_1 za klasifikator RAND. **Koliko iznosi očekivana razlika u vrijednostima mikro-F_1 klasifikatora MLR i RAND?**

A 0.205 B 0.085 C 0.185 D 0.155

Završni ispit iz Strojnog učenja 1 (ak. god. 2021./2022.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 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 1: Procjena parametara i Bayesov klasifikator (8 pitanja)

1 (P) Naivan Bayesov klasifikator pretpostavlja uvjetnu nezavisnost značajki unutar neke klase, premda u stvarnosti ta pretpostavka rijetko kada vrijedi. Razmotrite klasifikator poruka e-pošte čija je zadaća odrediti je li poruka prijevara (*scam*). Neka $y = 1$ označava da je poruka prijevara, a $y = 0$ da ona to nije. Model koristi binarne značajke koje indiciraju pojavljivanje određene riječi u tekstu poruke. Na primjer, izglednost $P(\text{lottery}|y = 1)$ jest vjerojatnost da se u prijevarnoj poruci pojavi riječ "lottery". Razmotrite sljedeće četiri riječi koje se općenito mogu pojaviti porukama e-pošte: "lottery", "winner", "prince" i "faithfully". Primijetite da mnoge prijevarne poruke od žrtve traže uplatu predujma. Takve poruke tipično govore o ostavštini nigerijskog princa ili o dobitku na lutriji (nikada oboje!) te završavaju kićenim pozdravom "yours faithfully". **Za koju od sljedećih jednakosti očekujemo da tipično *ne vrijedi* i da je time narušena pretpostavka naivnog Bayesovog klasifikatora?**

A $P(\text{faithfully}|y = 0) = P(\text{faithfully}|\text{winner}, y = 0)$

B $P(\text{faithfully}|y = 1) = P(\text{faithfully}|\text{winner}, y = 1)$

C $P(\text{winner}|y = 1) = P(\text{winner}|\text{faithfully}, y = 1)$

D $P(\text{winner}|y = 1) = P(\text{winner}|\text{lottery}, y = 1)$

2 (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 $P(y)P(x_1, x_3|y)P(x_2|y)$

B $P(y)P(x_1|y)P(x_2|y)P(x_3|y)$

C $P(y)P(x_1, x_2, x_3|y)$

D $P(y)P(x_1, x_2|y)P(x_3|y)$

3 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{-3, -2, 1, 4\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Koristimo MLE kako bismo na uzorku $\mathcal{D}$ procijenili parametre $\hat{\mu}_{\text{MLE}}$ i $\hat{\sigma}^2_{\text{MLE}}$. Usporedbe radi, parametar σ^2 dodatno procjenjujemo nepristranim procjeniteljem $\hat{\sigma}^2_{\text{UB}}$. Izračunajte log-izglednost parametara $(\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{MLE}})$ te log-izglednost parametara $(\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{UB}})$. **Koliko je prva log-izglednost veća od druge log-izglednosti?**

A 0.075 B 0.018 C 0.095 D 0.058

4 (P) MAP-procjeniteljem na skupu $\mathcal{D}$ procjenjujemo parametre $\boldsymbol{\mu}$ kategoričke (multinulijeve) varijable $\mathbf{x}$, tj. procjenjujemo parametre multinulijeve distribucije $P(\mathbf{x}|\boldsymbol{\mu})$. Varijabla $\mathbf{x}$ može proprimiti tri moguće vrijednosti ($K = 3$). Apriorna Dirichletova distribucija $P(\boldsymbol{\mu}|\boldsymbol{\alpha})$ definirana je parametrima $\boldsymbol{\alpha}$ definiranima kao $\boldsymbol{\alpha} = \alpha \cdot (2, 1, 4)$, gdje faktor α određuje vršnost Dirichletove distribucije, $\alpha \geqslant 1$. U skupu podataka $\mathcal{D}$ realizirane su sve tri vrijednosti kategoričke varijable $\mathbf{x}$, i to sa sljedećim brojem realizacija: $N_1 = 50$, $N_2 = 64$, $N_3 = 40$. **Koliko mora iznositi faktor α, a da bi vrijednost x_2 bila najvjerojatnija vrijednost kategoričke varijable x prema procijenjenoj distribuciji $P(\mathbf{x}|\boldsymbol{\mu})$?**

A $\alpha > 5$ B $\alpha > 7$ C $\alpha > 2$ D $\alpha > 8$

5 (T) Parametre modela možemo procijeniti pomoću procjenitelja najveće izglednosti (MLE) ili procjenitelja najveće aposteriorne vjerojatnosti (MAP). Općenito, MLE i MAP daju različite procjene, no u nekim slučajevima mogu dati jednake procjene. **Kada će MLE i MAP dati jednake procjene?**

A Kada su $p(\boldsymbol{\theta})$ i $p(\boldsymbol{\theta}|\mathcal{D})$ različite distribucije

B Kada $p(\boldsymbol{\theta})$ definiramo kao uniformnu distribuciju

C Kada se u $\mathcal{D}$ realizirala svaka vrijednost slučajne varijable

D Kada je aposteriorna distribucija $p(\boldsymbol{\theta}|\mathcal{D})$ unimodalna

6 (T) Jedan od nedostataka generativnih modela u odnosu na diskriminativne modele jest nepotrebna složenost modeliranja. **Što to zapravo znači?**

A Zajednička distribucija $p(\mathbf{x}, y)$ može se faktorizirati kao $p(y|\mathbf{x})p(\mathbf{x})$, no takva faktorizacija ima više parametara

B Generativni modeli modeliraju distribuciju $p(y|\mathbf{x})$, što iziskuje više parametara nego li modeliranje granice između klasa

C Za klasifikaciju nam je potrebna samo distribucija $p(y|\mathbf{x})$, i ona se može modelirati sa manje parametara od zajedničke distribucije $p(\mathbf{x}, y)$

D Za razliku od diskriminativnih modela, generativni modeli distribuciju $p(y|\mathbf{x})$ definiraju zasebno za svaku klasu, pa stoga imaju više parametara

7 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 2$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene i dijagonalne kovarijacijske matrice. Izglednost klase s oznakom $y = j$ definirana je multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

Model treniramo na skup podataka od $N = 7$ primjera:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)}))\}_i = \{((-4, -2), 0), ((0, 0), 0), ((4, 2), 0), ((3, -1), 1), ((4, -1), 1), ((4, 1), 1), ((5, 1), 1)\}$$

Procijenite parametre modela na ovom skupu primjera. Budući da je skup primjera malen, za procjenu kovarijacijske matrice koristite nepristran procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. **Koliko iznosi $h_1(\mathbf{x})$ za primjer $\mathbf{x} = (4, 0)$?**

A -4.13 B -3.84 C -2.75 D -3.03

8 (P) Gaussov Bayesov klasifikator koristimo za klasfikaciju jednodimenzijskih podataka u tri klase. Procijenjene izglednosti klasa su $p(x|y = 1) = \mathcal{N}(-10, 2)$, $p(x|y = 2) = \mathcal{N}(2, 2)$ i $p(x|y = 3) = \mathcal{N}(8, 2)$, a procijenjene apriorne vjerojatnosti klasa su $P(y = 1) = P(y = 2) = 2/5$ i $P(y = 3) = 1/5$. Međutim, nakon što smo naučili ovaj model, zaključili smo da na ispitnom skupu postoji pomak u distribuciji podataka u odnosu na skup za učenje te da zbog toga model ne generalizira dobro. Zaključili smo da se ovo može ispraviti tako da se naučeni model malo izmijeni, i to tako da se varijanca izglednosti klase $y = 1$ postavi na 5 i da se apriorne vjerojatnosti klasa ujednače, $P(y = 1) = P(y = 2) = P(y = 3) = 1/3$. Skicirajte gustoće zajedničke vjerojatnosti naučenog i izmijenjenog modela. Neka su h_1 i h_2 MAP-hipoteze prvog i drugog modela, te neka su a i b pozitivne konstante. Razmotrite segment ulaznog prostora za koji vrijedi $-10 \leqslant x \leqslant 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

A $[-4 - a, -4] \cup [5 - b, 5]$ B $[-4 - a, -4 + b]$ C $[-4, -4 + a] \cup [5, 5 + b]$ D $[-4 - a, 5 + b]$

Cjelina 2: Probabilistički grafički modeli (5 pitanja)

9 (N) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x,y,z,w) = P(z)P(w)P(y|z,w)P(x|y)$. Sve varijable su binarne. Vrijedi $P(z=1) = 0.1$, $P(w=1) = 0.3$, $P(x=1|y=0) = 0.2$ i $P(x=1|y=1) = 0.6$. Tablica uvjetne vjerojatnosti za čvor y je sljedeća:

z	w	$p(y=1\|z,w)$
0	0	0
0	1	0.4
1	0	0.2
1	1	0.7

Postupkom uzorkovanja s odbijanjem uzorkujemo iz aposteriorne distribucije $P(w|x=0,z=1)$. Uzorkovanje smo ponovili ukupno $N = 10000$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N. **U uzorku neodbačenih vektora, koliko je očekivano više vektora sa $w = 0$ od vektora sa $w = 1$?**

A 472 B 596 C 944 D 348

10 (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 modelira 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 0.813 B 0.747 C 0.033 D 0.148

11 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Sve varijable su binarne, osim varijabli v i y, koje su ternarne. Uz navedeni topološki uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

$$\{v,w\} \perp y | x \qquad \{v,x\} \perp z | \{w,y\}$$

Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

A 21 B 23 C 18 D 24

12 (T) Čest način probabilističkog zaključivanja kod Bayesovih mreža jest izračunavanje "aposteriornog upita". Kod te vrste upita zanima nas distribucija nekih varijabli (varijable upita) na temelju zadanih varijabli (opažene varijable). Međutim, Bayesova mreža kodira zajedničku vjerojatnost svih varijabli mreže, uključivo i varijabli koje nisu niti varijable upita niti opažene varijable (varijable smetnje). **Na koji način izračunavamo aposteriorni upit?**

A Kao umnožak zajedničke vjerojatnosti marginalizirane po opaženim varijablama i zajedničke vjerojatnosti marginalizirane po varijablama smetnje

B Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama upita i zajedničke vjerojatnosti marginalizirane po opaženim varijablama

C Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama upita i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i opaženim varijablama

D Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama smetnje i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i varijablama upita

13 (P) Bayesovom mrežom sa sedam binarnih varijabli modeliramo prometne (ne)prilike u gradu Zagrebu. U našoj mreži, strelice modeliraju pretpostavljeni smjer kauzalnosti. Jutarnje doba dana (J) ili loše vrijeme (V) mogu uzrokovati nastanak prometne gužve (G). Loše vrijeme također može uzrokovati nastanak prometne nesreće (N). Nadalje, nastanak prometne nesreće može uzrokovati nastanak prometne gužve. Loše vrijeme također može uzrokovati zastoj tramvaja (T), no i nestanak struje (S) može uzrokovati zastoj tramvaja. Konačno, zastoj tramvaja može uzrokovati masovno pješačenje putnika (P), što opet može uzrokovati prometnu nesreću. Primjenom algoritma d-odvajanja ispitujemo stohastičku zavisnost između jutarnjeg doba dana (J) i nestanka struje (S). **Koje minimalan skup varijabli treba biti opažen, a da vjerojatnost od S ne ovisi o vrijednosti J?**

[A] V, P, T [B] V, P [C] V, T [D] T

Cjelina 3: Grupiranje (6 pitanja)

14 (T) Algoritam K-means++ proširenje je algoritma K-sredina heurističkim odabirom početnih središta grupa. **Koja je glavna ideja odabira početnih središta grupa kod algoritma K-means++?**

[A] Najvjerojatnije središte grupe jest primjer koji je najviše udaljen od njemu najbližeg središta

[B] Središta grupa su mjesta na kojima graf kriterijske funkcije u ovisnosti o broju grupa naglo opada pa stagnira

[C] Vjerojatnost da je neki primjer središte grupe proporcionalna je s brojem primjera u toj grupi i udaljenosti tih primjera od centroida skupa podataka

[D] Središta grupe su srednje vrijednosti grupa projiciranih na pravac u smjeru prve komponente rastava skupa podataka na glavne komponente (PCA)

15 (P) Skup neoznačenih primjera u dvodimenzijskome ulaznom prostoru neka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{5} = \{(0,0), (0,4), (2,0), (2,4), (4,2)\}$$

Primjere grupiramo algoritmom K-sredina sa $K = 3$ grupe. Za početna središta grupa odaberemo nasumično primjere iz $\mathcal{D}$, pri čemu, naravno, pazimo da odaberemo različita središta. Ishod grupiranja i konačan iznos kriterijske funkcije J ovisit će o odabiru početnih središta. Neka je J^* vrijednost kriterijske funkcije u točki globalnog minimuma, dakle vrijednost koja odgovara najboljem grupiranju. Neka je J^+ vrijednost kriterijske funkcije u točki lokalnog minimuma, i to onoj točki lokalnog minimuma s najvećom vrijednošću funkcije J. **Koliko iznosi razlika** $J^+ - J^*$?

[A] 6 [B] 8 [C] 4 [D] 12

16 (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
$y_{pred}^{(i)}$	0	0	2	1	1	0	0	2	1	1
$y_{true}^{(i)}$	1	0	0	2	1	0	2	1	1	1

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

[A] 0.62 [B] 0.58 [C] 0.53 [D] 0.49

17 (P) Algoritmom GMM grupiramo primjere u dvodimenzijskome ulaznom prostoru. Skup podataka u stvarnosti je uzorkovan iz zajedničke distribucije koja se može opisati sljedećim mješavinskim modelom:

$$p(\mathbf{x}) = \sum_{j=1}^{3} \frac{1}{3}\mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \mu_1 = (5,5), \mu_2 = (5,10), \mu_3 = (-10,-10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

Skup $\mathcal{D}$ grupiramo u $K = 2$ grupe. Pritom isprobavamo tri modela, koji se međusobno razlikuju po pretpostavkama na kovarijacijsku matricu. Konkretno: dijeljena i puna kovarijacijska matrica ($\mathcal{H}_1$), nedijeljena i dijagonalna kovarijacijska matrica ($\mathcal{H}_2$) i nedijeljena i izotropna kovarijacijska matrica ($\mathcal{H}_3$). Neka je $\mathcal{L}_i$ izglednost parametara dobivena modelom $\mathcal{H}_i$ nakon konvergencije algoritma. Za inicijalizaciju središta koristi se algoritam K-means++. **Što su očekivani odnosi između izglednosti za ova tri modela?**

A $\mathcal{L}_2 > \mathcal{L}_1$, $\mathcal{L}_2 > \mathcal{L}_3$ B $\mathcal{L}_1 > \mathcal{L}_2 > \mathcal{L}_3$ C $\mathcal{L}_1 > \mathcal{L}_3$, $\mathcal{L}_2 > \mathcal{L}_3$ D $\mathcal{L}_1 = \mathcal{L}_2 > \mathcal{L}_3$

18 (N) Algoritmom HAC grupiramo riječi engleskog jezika. Neoznačeni skup podataka sastoji se od sljedećih riječi:

$$\mathcal{D} = \{\text{“water”}, \text{“watering”}, \text{“moon”}, \text{“air”}\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“watering”}) = 5/8 = 0.625$. Provedite prve dvije iteracije grupiranja algoritmom HAC uz prosječno povezivanje. **Na kojoj se razini sličnosti spajaju grupe u drugoj iteraciji algoritma HAC?**

A 0.354   B 0.583   C 0.292   D 0.535

19 (T) Model miješane gustoće sa K komponenti vjerojatnosnu distribuciju neoznačenih podaka definira kao $p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \, p(\mathbf{x}|\boldsymbol{\theta}_k)$. Što modelira uvjetna vjerojatnost $p(\mathbf{x}|\boldsymbol{\theta}_k)$?

A Gustoću vjerojatnosti primjera $\mathbf{x}$ unutar grupe k

B Stupanj pripadnosti primjera $\mathbf{x}$ grupi k

C Stupanj pripadnosti grupe k primjeru $\mathbf{x}$

D Vjerojatnost da primjer $\mathbf{x}$ pripada grupi k

Cjelina 4: Vrednovanje modela (3 pitanja)

20 (P) Na istom skupu označenih primjera vrednujemo četiri binarna klasifikatora s vjerojatnosnim izlazima. Za vrednovanje koristimo krivulju ROC. Svaki smo klasifikator ispitali s tri vrijednosti klasifikacijskog praga te smo za te vrijednosti izračunali FPR (stopa lažnog alarma) i TPR (odziv). Dobiveni parovi vrijednosti (FPR, TPR) za sva četiri klasifikatora su sljedeći:

$$h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) \qquad h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8)$$
$$h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) \qquad h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0)$$

Skicirajte odgovarajuće krivulje ROC, linearno interpolirajući između izmjerenih točaka, a dodajte i točke koje odgovaraju krajnjim vrijednostima klasifikacijskog praga (0 i 1). Pritom, ako je neki klasifikator lošiji od nasumičnog klasifikatora, umjesto tog klasifikatora razmatrajte njegovu popravljenu varijantu koju ćete dobiti invertiranjem izlaza ($1 - h(\mathbf{x})$ umjesto $h(\mathbf{x})$). **Koji su od ispitanih klasifikatora (eventualno nakon popravka) najbolji prema krivulji ROC?**

A h_1 i h_2 B h_1 i h_3 C h_1 i h_4 D h_3 i h_4

21 (N) Na ispitnome skupu evaluiramo model multinomijalne logističke regresije (MLR) za klasifikaciju u $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 35 & 8 & 3 \\ 10 & 25 & 5 \\ 0 & 12 & 2 \end{pmatrix} \end{array}$$

Klasifikator MLR uspoređujemo s klasifikatorom RAND koji primjere klasificira nasumično, i to tako da oznaku $y = j$ dodjeljuje s vjerojatnošću proporcionalnoj udjelu klase j u ispitnome skupu. Izračunajte mikro-F_1 za klasifikator MLR i očekivani mikro-F_1 za klasifikator RAND. **Koliko iznosi očekivana razlika u vrijednostima mikro-F_1 klasifikatora MLR i RAND?**

A 0.085   B 0.155   C 0.205   D 0.185

22 (T) Za vrednovanje višeklasnog klasifikatora često se koriste mjere F_1^μ (mikro F1-mjera) i F_1^M (makro F1-mjera). **Koji je očekivani odnos između vrijednosti tih mjera, i zašto?**

A $F_1^M < F_1^\mu$, jer kod mikro F1-mjere zbrajamo matrice zabune kroz sve klase, pa je ukupan broj pogrešaka veći nego kad računamo prosječnu pogrešku kroz klase

B $F_1^M < F_1^\mu$, jer kod makro F1-mjere računamo prosjek F1-mjere kroz sve klase, a klasifikator na manjim klasama više griješi

C $F_1^M > F_1^\mu$, jer kod makro F1-mjere zbrajamo matrice zabune kroz sve klase, a primjera iz većih klasa ima više, pa više doprinose pogrešci

D $F_1^M < F_1^\mu$, jer kod mikro F1-mjere računamo prosjek F1-mjere kroz sve klase, a klasifikator na većima klasama manje griješi

Završni ispit iz Strojnog učenja 1 (ak. god. 2021./2022.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 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 1: Procjena parametara i Bayesov klasifikator (8 pitanja)

1 (P) Gaussov Bayesov klasifikator koristimo za klasfikaciju jednodimenzijskih podataka u tri klase. Procijenjene izglednosti klasa su $p(x|y=1) = \mathcal{N}(-10,2)$, $p(x|y=2) = \mathcal{N}(2,2)$ i $p(x|y=3) = \mathcal{N}(8,2)$, a procijenjene apriorne vjerojatnosti klasa su $P(y=1) = P(y=2) = 2/5$ i $P(y=3) = 1/5$. Međutim, nakon što smo naučili ovaj model, zaključili smo da na ispitnom skupu postoji pomak u distribuciji podataka u odnosu na skup za učenje te da zbog toga model ne generalizira dobro. Zaključili smo da se ovo može ispraviti tako da se naučeni model malo izmijeni, i to tako da se varijanca izglednosti klase $y=1$ postavi na 5 i da se apriorne vjerojatnosti klasa ujednače, $P(y=1) = P(y=2) = P(y=3) = 1/3$. Skicirajte gustoće zajedničke vjerojatnosti naučenog i izmijenjenog modela. Neka su h_1 i h_2 MAP-hipoteze prvog i drugog modela, te neka su a i b pozitivne konstante. Razmotrite segment ulaznog prostora za koji vrijedi $-10 \leqslant x \leqslant 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

A $[-4-a, 5+b]$ B $[-4-a,-4] \cup [5-b,5]$ C $[-4-a,-4+b]$ D $[-4,-4+a] \cup [5,5+b]$

2 (P) MAP-procjeniteljem na skupu $\mathcal{D}$ procjenjujemo parametre $\boldsymbol{\mu}$ kategoričke (multinulijeve) varijable $\mathbf{x}$, tj. procjenjujemo parametre multinulijeve distribucije $P(\mathbf{x}|\boldsymbol{\mu})$. Varijabla $\mathbf{x}$ može proprimiti tri moguće vrijednosti ($K=3$). Apriorna Dirichletova distribucija $P(\boldsymbol{\mu}|\boldsymbol{\alpha})$ definirana je parametrima $\boldsymbol{\alpha}$ definiranima kao $\boldsymbol{\alpha} = \alpha \cdot (2,1,4)$, gdje faktor α određuje vršnost Dirichletove distribucije, $\alpha \geqslant 1$. U skupu podataka $\mathcal{D}$ realizirane su sve tri vrijednosti kategoričke varijable $\mathbf{x}$, i to sa sljedećim brojem realizacija: $N_1 = 50$, $N_2 = 64$, $N_3 = 40$. **Koliko mora iznositi faktor α, a da bi vrijednost x_2 bila najvjerojatnija vrijednost kategoričke varijable x prema procijenjenoj distribuciji $P(\mathbf{x}|\boldsymbol{\mu})$?**

A $\alpha > 5$ B $\alpha > 7$ C $\alpha > 8$ D $\alpha > 2$

3 (T) Parametre modela možemo procijeniti pomoću procjenitelja najveće izglednosti (MLE) ili procjenitelja najveće aposteriorne vjerojatnosti (MAP). Općenito, MLE i MAP daju različite procjene, no u nekim slučajevima mogu dati jednake procjene. **Kada će MLE i MAP dati jednake procjene?**

A Kada koristimo nepristran procjenitelj za $\boldsymbol{\theta}$

B Kada broj primjera N teži prema beskonačno

C Kada je aposteriorna distribucija $p(\boldsymbol{\theta}|\mathcal{D})$ unimodalna

D Kada je $p(\boldsymbol{\theta})$ konjugatna izglednosti $p(\mathcal{D}|\boldsymbol{\theta})$

4 (P) Naivan Bayesov klasifikator pretpostavlja uvjetnu nezavisnost značajki unutar neke klase, premda u stvarnosti ta pretpostavka rijetko kada vrijedi. Razmotrite klasifikator poruka e-pošte čija je zadaća odrediti je li poruka prijevara (*scam*). Neka $y=1$ označava da je poruka prijevara, a $y=0$ da ona to nije. Model koristi binarne značajke koje indiciraju pojavljivanje određene riječi u tekstu poruke. Na primjer, izglednost $P(\text{lottery}|y=1)$ jest vjerojatnost da se u prijevarnoj poruci pojavi riječ "lottery". Razmotrite sljedeće četiri riječi koje se općenito mogu pojaviti porukama e-pošte: "lottery", "winner", "prince" i "faithfully". Primijetite da mnoge prijevarne poruke od žrtve traže uplatu predujma. Takve poruke tipično govore o ostavštini nigerijskog princa ili o dobitku na lutriji (nikada oboje!) te završavaju kićenim pozdravom "yours faithfully". **Za koju od sljedećih**

jednakosti očekujemo da tipično *ne vrijedi* i da je time narušena pretpostavka naivnog Bayesovog klasifikatora?

- A $P(\text{lottery}|y=1) = P(\text{lottery}|\text{faithfully}, y=1)$
- B $P(\text{faithfully}|y=1) = P(\text{faithfully}|\text{prince}, y=1)$
- C $P(\text{lottery}|y=1) = P(\text{lottery}|\text{winner}, y=1)$
- D $P(\text{faithfully}|y=0) = P(\text{faithfully}|\text{winner}, y=0)$

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

6 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 2$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene i dijagonalne kovarijacijske matrice. Izglednost klase s oznakom $y = j$ definirana je multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

Model treniramo na skup podataka od $N = 7$ primjera:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)}))\}_i = \{((-4,-2),0), ((0,0),0), ((4,2),0), ((3,-1),1), ((4,-1),1), ((4,1),1), ((5,1),1)\}$$

Procijenite parametre modela na ovom skupu primjera. Budući da je skup primjera malen, za procjenu kovarijacijske matrice koristite nepristran procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. **Koliko iznosi $h_0(\mathbf{x})$ za primjer $\mathbf{x} = (0, 0)$?**

A -4.13 B -3.03 C -2.75 D -3.84

7 (T) Jedan od nedostataka generativnih modela u odnosu na diskriminativne modele jest nepotrebna složenost modeliranja. **Što to zapravo znači?**

- A Zajednička distribucija $p(\mathbf{x}, y)$ može se faktorizirati kao $p(y|\mathbf{x})p(\mathbf{x})$, no takva faktorizacija ima više parametara
- B Za klasifikaciju nam je potrebna samo distribucija $p(y|\mathbf{x})$, i ona se može modelirati sa manje parametara od zajedničke distribucije $p(\mathbf{x}, y)$
- C Generativni modeli modeliraju distribuciju $p(y|\mathbf{x})$, što iziskuje više parametara nego li modeliranje granice između klasa
- D Za razliku od diskriminativnih modela, generativni modeli distribuciju $p(y|\mathbf{x})$ definiraju zasebno za svaku klasu, pa stoga imaju više parametara

8 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{-3, -2, 1, 4, 4\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Koristimo MLE kako bismo na uzorku $\mathcal{D}$ procijenili parametre $\hat{\mu}_{\mathrm{MLE}}$ i $\hat{\sigma}^2_{\mathrm{MLE}}$. Usporedbe radi, parametar σ^2 dodatno procjenjujemo nepristranim procjeniteljem $\hat{\sigma}^2_{\mathrm{UB}}$. Izračunajte log-izglednost parametara $(\hat{\mu}_{\mathrm{MLE}}, \hat{\sigma}^2_{\mathrm{MLE}})$ te log-izglednost parametara $(\hat{\mu}_{\mathrm{MLE}}, \hat{\sigma}^2_{\mathrm{UB}})$. **Koliko je prva log-izglednost veća od druge log-izglednosti?**

A 0.075 B 0.018 C 0.058 D 0.095

Cjelina 2: Probabilistički grafički modeli (5 pitanja)

9 (P) Bayesovom mrežom sa sedam binarnih varijabli modeliramo prometne (ne)prilike u gradu Zagrebu. U našoj mreži, strelice modeliraju pretpostavljeni smjer kauzalnosti. Jutarnje doba dana (J) ili loše vrijeme (V) mogu uzrokovati nastanak prometne gužve (G). Loše vrijeme također može uzrokovati nastanak prometne nesreće (N). Nadalje, nastanak prometne nesreće može uzrokovati nastanak prometne gužve. Loše vrijeme također može uzrokovati zastoj tramvaja (T), no i nestanak struje (S) može uzrokovati zastoj tramvaja. Konačno, zastoj tramvaja može uzrokovati masovno pješačenje putnika (P), što opet može uzrokovati prometnu nesreću. Primjenom algoritma d-odvajanja ispitujemo stohastičku zavisnost između jutarnjeg doba dana (J) i nestanka struje (S). **Koje minimalan skup varijabli treba biti opažen, a da vjerojatnost od S ne ovisi o vrijednosti J?**

A N, T B V, T C V, P D T

10 (T) Čest način probabilističkog zaključivanja kod Bayesovih mreža jest izračunavanje "aposteriornog upita". Kod te vrste upita zanima nas distribucija nekih varijabli (varijable upita) na temelju zadanih varijabli (opažene varijable). Međutim, Bayesova mreža kodira zajedničku vjerojatnost svih varijabli mreže, uključivo i varijabli koje nisu niti varijable upita niti opažene varijable (varijable smetnje). **Na koji način izračunavamo aposteriorni upit?**

A Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama smetnje i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i varijablama upita

B Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama upita i zajedničke vjerojatnosti marginalizirane po opaženim varijablama

C Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama upita i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i opaženim varijablama

D Kao umnožak zajedničke vjerojatnosti marginalizirane po opaženim varijablama i zajedničke vjerojatnosti marginalizirane po varijablama smetnje

11 (N) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x, y, z, w) = P(z)P(w)P(y|z, w)P(x|y)$. Sve varijable su binarne. Vrijedi $P(z = 1) = 0.1$, $P(w = 1) = 0.2$, $P(x = 1|y = 0) = 0.2$ i $P(x = 1|y = 1) = 0.6$. Tablica uvjetne vjerojatnosti za čvor y je sljedeća:

z	w	$p(y = 1\|z, w)$
0	0	0
0	1	0.4
1	0	0.2
1	1	0.7

Postupkom uzorkovanja s odbijanjem uzorkujemo iz aposteriorne distribucije $P(w|x = 0, z = 1)$. Uzorkovanje smo ponovili ukupno $N = 10000$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N. **U uzorku neodbačenih vektora, koliko je očekivano više vektora sa $w = 0$ od vektora sa $w = 1$?**

A 472 B 944 C 348 D 596

12 (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 modelira 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 0.148 B 0.747 C 0.033 D 0.813

13 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Sve varijable su binarne, osim varijabli v i x, koje su ternarne. Uz navedeni topološki uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

$$\{v, w\} \perp y | x \qquad \{v, x\} \perp z | \{w, y\}$$

Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

[A] 23 [B] 21 [C] 24 [D] 18

Cjelina 3: Grupiranje (6 pitanja)

14 (P) Skup neoznačenih primjera u dvodimenzijskome ulaznom prostoru neka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{5} = \{(0,0), (0,4), (2,0), (2,4), (4,2)\}$$

Primjere grupiramo algoritmom K-sredina sa $K = 3$ grupe. Za početna središta grupa odaberemo nasumično primjere iz $\mathcal{D}$, pri čemu, naravno, pazimo da odaberemo različita središta. Ishod grupiranja i konačan iznos kriterijske funkcije J ovisit će o odabiru početnih središta. Neka je J^* vrijednost kriterijske funkcije u točki globalnog minimuma, dakle vrijednost koja odgovara najboljem grupiranju. Neka je J^+ vrijednost kriterijske funkcije u točki lokalnog minimuma, i to onoj točki lokalnog minimuma s najvećom vrijednošću funkcije J. **Koliko iznosi razlika** $J^+ - J^*$?

[A] 4 [B] 12 [C] 6 [D] 8

15 (T) Model miješane gustoće sa K komponenti vjerojatnosnu distribuciju neoznačenih podaka definira kao $p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \, p(\mathbf{x}|\boldsymbol{\theta}_k)$. Što modelira uvjetna vjerojatnost $p(\mathbf{x}|\boldsymbol{\theta}_k)$?

[A] Vjerojatnost da primjer $\mathbf{x}$ pripada grupi k
[B] Gustoću vjerojatnosti grupe k za primjer $\mathbf{x}$
[C] Gustoću vjerojatnosti primjera $\mathbf{x}$ unutar grupe k
[D] Stupanj pripadnosti primjera $\mathbf{x}$ grupi k

16 (T) Algoritam K-means++ proširenje je algoritma K-sredina heurističkim odabirom početnih središta grupa. **Koja je glavna ideja odabira početnih središta grupa kod algoritma K-means++**?

[A] Najvjerojatnije središte grupe jest primjer koji je najviše udaljen od njemu najbližeg središta
[B] Središta grupa su mjesta na kojima graf kriterijske funkcije u ovisnosti o broju grupa naglo opada pa stagnira
[C] Središta grupe su srednje vrijednosti grupa projiciranih na pravac u smjeru prve komponente rastava skupa podataka na glavne komponente (PCA)
[D] Vjerojatnost da je neki primjer središte grupe proporcionalna je s brojem primjera u toj grupi i udaljenosti tih primjera od centroida skupa podataka

17 (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
$y_{pred}^{(i)}$	0	0	2	1	1	0	0	2	1	1
$y_{true}^{(i)}$	1	0	0	2	1	0	2	1	1	1

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

[A] 0.58 [B] 0.53 [C] 0.62 [D] 0.49

18 (N) Algoritmom HAC grupiramo riječi engleskog jezika. Neoznačeni skup podataka sastoji se od sljedećih riječi:

$$\mathcal{D} = \{\text{“water”}, \text{“waterloo”}, \text{“earth”}, \text{“air”}\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“waterloo”}) = 5/7 = 0.714$. Provedite prve dvije iteracije grupiranja algoritmom HAC uz prosječno povezivanje. **Na kojoj se razini sličnosti spajaju grupe u drugoj iteraciji algoritma HAC?**

A 0.292 B 0.583 C 0.354 D 0.535

19 (P) Algoritmom GMM grupiramo primjere u dvodimenzijskome ulaznom prostoru. Skup podataka u stvarnosti je uzorkovan iz zajedničke distribucije koja se može opisati sljedećim mješavinskim modelom:

$$p(\mathbf{x}) = \sum_{j=1}^{3} \frac{1}{3}\mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \mu_1 = (5,5), \mu_2 = (5,10), \mu_3 = (-10,-10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

Skup $\mathcal{D}$ grupiramo u $K = 2$ grupe. Pritom isprobavamo tri modela, koji se međusobno razlikuju po pretpostavkama na kovarijacijsku matricu. Konkretno: dijeljena i puna kovarijacijska matrica ($\mathcal{H}_1$), nedijeljena i dijagonalna kovarijacijska matrica ($\mathcal{H}_2$) i nedijeljena i izotropna kovarijacijska matrica ($\mathcal{H}_3$). Neka je $\mathcal{L}_i$ izglednost parametara dobivena modelom $\mathcal{H}_i$ nakon konvergencije algoritma. Za inicijalizaciju središta koristi se algoritam K-means++. **Što su očekivani odnosi između izglednosti za ova tri modela?**

A $\mathcal{L}_1 > \mathcal{L}_3$, $\mathcal{L}_2 > \mathcal{L}_3$ B $\mathcal{L}_2 > \mathcal{L}_1$, $\mathcal{L}_2 > \mathcal{L}_3$ C $\mathcal{L}_1 > \mathcal{L}_2 > \mathcal{L}_3$ D $\mathcal{L}_1 = \mathcal{L}_2 > \mathcal{L}_3$

Cjelina 4: Vrednovanje modela (3 pitanja)

20 (N) Na ispitnome skupu evaluiramo model multinomijalne logističke regresije (MLR) za klasifikaciju u $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{c c} & \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{array}{c} y=1 \\ y=2 \\ y=3 \end{array} & \begin{pmatrix} 20 & 18 & 3 \\ 21 & 25 & 2 \\ 4 & 2 & 5 \end{pmatrix} \end{array}$$

Klasifikator MLR uspoređujemo s klasifikatorom RAND koji primjere klasificira nasumično, i to tako da oznaku $y = j$ dodjeljuje s vjerojatnošću proporcionalnoj udjelu klase j u ispitnome skupu. Izračunajte mikro-F_1 za klasifikator MLR i očekivani mikro-F_1 za klasifikator RAND. **Koliko iznosi očekivana razlika u vrijednostima mikro-F_1 klasifikatora MLR i RAND?**

A 0.185 B 0.155 C 0.085 D 0.205

21 (T) Za vrednovanje višeklasnog klasifikatora često se koriste mjere F_1^μ (mikro F1-mjera) i F_1^M (makro F1-mjera). **Koji je očekivani odnos između vrijednosti tih mjera, i zašto?**

A $F_1^M < F_1^\mu$, jer kod makro F1-mjere računamo prosjek F1-mjere kroz sve klase, a klasifikator na manjim klasama više griješi

B $F_1^M > F_1^\mu$, jer kod makro F1-mjere zbrajamo matrice zabune kroz sve klase, a primjera iz većih klasa ima više, pa više doprinose pogrešci

C $F_1^M < F_1^\mu$, jer kod mikro F1-mjere računamo prosjek F1-mjere kroz sve klase, a klasifikator na većima klasama manje griješi

D $F_1^M < F_1^\mu$, jer kod mikro F1-mjere zbrajamo matrice zabune kroz sve klase, pa je ukupan broj pogrešaka veći nego kad računamo prosječnu pogrešku kroz klase

22 (P) Na istom skupu označenih primjera vrednujemo četiri binarna klasifikatora s vjerojatnosnim izlazima. Za vrednovanje koristimo krivulju ROC. Svaki smo klasifikator ispitali s tri vrijednosti klasifikacijskog praga te smo za te vrijednosti izračunali FPR (stopa lažnog alarma) i TPR (odziv). Dobiveni parovi vrijednosti (FPR, TPR) za sva četiri klasifikatora su sljedeći:

$h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4)$

$h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6)$

$h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8)$

$h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0)$

Skicirajte odgovarajuće krivulje ROC, linearno interpolirajući između izmjerenih točaka, a dodajte i točke koje odgovaraju krajnjim vrijednostima klasifikacijskog praga (0 i 1). Pritom, ako je neki klasifikator lošiji od nasumičnog klasifikatora, umjesto tog klasifikatora razmatrajte njegovu popravljenu varijantu koju ćete dobiti invertiranjem izlaza ($1 - h(\mathbf{x})$ umjesto $h(\mathbf{x})$). **Koji su od ispitanih klasifikatora (eventualno nakon popravka) najbolji prema krivulji ROC?**

A h_1 i h_2 B h_2 i h_3 C h_1 i h_4 D h_3 i h_4

Završni ispit iz Strojnog učenja 1 (ak. god. 2021./2022.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 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 1: Procjena parametara i Bayesov klasifikator (8 pitanja)

1 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 2$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene i dijagonalne kovarijacijske matrice. Izglednost klase s oznakom $y = j$ definirana je multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

Model treniramo na skup podataka od $N = 7$ primjera:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)}))\}_i = \{((-4,-2),0),((0,0),0),((4,2),0),((3,-1),1),((4,-1),1),((4,1),1),((5,1),1)\}$$

Procijenite parametre modela na ovom skupu primjera. Budući da je skup primjera malen, za procjenu kovarijacijske matrice koristite nepristran procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. **Koliko iznosi $h_0(\mathbf{x})$ za primjer $\mathbf{x} = (0, 0)$?**

A -3.84 B -3.03 C -4.13 D -2.75

2 (T) Jedan od nedostataka generativnih modela u odnosu na diskriminativne modele jest nepotrebna složenost modeliranja. **Što to zapravo znači?**

A Za klasifikaciju nam je potrebna samo distribucija $p(y|\mathbf{x})$, i ona se može modelirati sa manje parametara od zajedničke distribucije $p(\mathbf{x}, y)$

B Zajednička distribucija $p(\mathbf{x}, y)$ može se faktorizirati kao $p(y|\mathbf{x})p(\mathbf{x})$, no takva faktorizacija ima više parametara

C Generativni modeli modeliraju distribuciju $p(y|\mathbf{x})$, što iziskuje više parametara nego li modeliranje granice između klasa

D Za razliku od diskriminativnih modela, generativni modeli distribuciju $p(y|\mathbf{x})$ definiraju zasebno za svaku klasu, pa stoga imaju više parametara

3 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{-3, -2, 1, 4, 4\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Koristimo MLE kako bismo na uzorku $\mathcal{D}$ procijenili parametre $\hat{\mu}_{\text{MLE}}$ i $\hat{\sigma}^2_{\text{MLE}}$. Usporedbe radi, parametar σ^2 dodatno procjenjujemo nepristranim procjeniteljem $\hat{\sigma}^2_{\text{UB}}$. Izračunajte log-izglednost parametara $(\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{MLE}})$ te log-izglednost parametara $(\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{UB}})$. **Koliko je prva log-izglednost veća od druge log-izglednosti?**

A 0.018 B 0.075 C 0.095 D 0.058

4 (T) Parametre modela možemo procijeniti pomoću procjenitelja najveće izglednosti (MLE) ili procjenitelja najveće aposteriorne vjerojatnosti (MAP). Općenito, MLE i MAP daju različite procjene, no u nekim slučajevima mogu dati jednake procjene. **Kada će MLE i MAP dati jednake procjene?**

A Kada $p(\boldsymbol{\theta})$ definiramo kao uniformnu distribuciju

B Kada se u $\mathcal{D}$ realizirala svaka vrijednost slučajne varijable

C Kada $p(\boldsymbol{\theta})$ definiramo kao unimodalnu distribuciju

D Kada je $p(\mathcal{D}|\boldsymbol{\theta})$ uniformna distribucija

5 (P) Gaussov Bayesov klasifikator koristimo za klasfikaciju jednodimenzijskih podataka u tri klase. Procijenjene izglednosti klasa su $p(x|y=1) = \mathcal{N}(-10, 2)$, $p(x|y=2) = \mathcal{N}(2, 2)$ i $p(x|y=3) = \mathcal{N}(8, 2)$, a procijenjene apriorne vjerojatnosti klasa su $P(y=1) = P(y=2) = 2/5$ i $P(y=3) = 1/5$. Međutim, nakon što smo naučili ovaj model, zaključili smo da na ispitnom skupu postoji pomak u distribuciji podataka u odnosu na skup za učenje te da zbog toga model ne generalizira dobro. Zaključili smo da se ovo može ispraviti tako da se naučeni model malo izmijeni, i to tako da se varijanca izglednosti klase $y = 1$ postavi na 5 i da se apriorne vjerojatnosti klasa ujednače, $P(y=1) = P(y=2) = P(y=3) = 1/3$. Skicirajte gustoće zajedničke vjerojatnosti naučenog i izmijenjenog modela. Neka su h_1 i h_2 MAP-hipoteze prvog i drugog modela, te neka su a i b pozitivne konstante. Razmotrite segment ulaznog prostora za koji vrijedi $-10 \leqslant x \leqslant 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

A $[-4, -4+a] \cup [5, 5+b]$ B $[-4-a, -4] \cup [5-b, 5]$ C $[-4-a, -4+b]$ D $[-4-a, 5+b]$

6 (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:

	$x_3=0$		$x_3=1$	
	$x_2=0$	$x_2=1$	$x_2=0$	$x_2=1$
$x_1=0$	0.1	0.0	0.0	0.1
$x_1=1$	0.1	0.4	0.1	0.2

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

7 (P) Naivan Bayesov klasifikator pretpostavlja uvjetnu nezavisnost značajki unutar neke klase, premda u stvarnosti ta pretpostavka rijetko kada vrijedi. Razmotrite klasifikator poruka e-pošte čija je zadaća odrediti je li poruka prijevara (*scam*). Neka $y = 1$ označava da je poruka prijevara, a $y = 0$ da ona to nije. Model koristi binarne značajke koje indiciraju pojavljivanje određene riječi u tekstu poruke. Na primjer, izglednost $P(\text{lottery}|y=1)$ jest vjerojatnost da se u prijevarnoj poruci pojavi riječ "lottery". Razmotrite sljedeće četiri riječi koje se općenito mogu pojaviti porukama e-pošte: "lottery", "winner", "prince" i "faithfully". Primijetite da mnoge prijevarne poruke od žrtve traže uplatu predujma. Takve poruke tipično govore o ostavštini nigerijskog princa ili o dobitku na lutriji (nikada oboje!) te završavaju kićenim pozdravom "yours faithfully". **Za koju od sljedećih jednakosti očekujemo da tipično *ne vrijedi* i da je time narušena pretpostavka naivnog Bayesovog klasifikatora?**

A $P(\text{faithfully}|y=0) = P(\text{faithfully}|\text{lottery}, y=0)$

B $P(\text{faithfully}|y=1) = P(\text{faithfully}|\text{winner}, y=1)$

C $P(\text{lottery}|y=0) = P(\text{lottery}|\text{faithfully}, y=0)$

D $P(\text{lottery}|y=1) = P(\text{lottery}|\text{winner}, y=1)$

8 (P) MAP-procjeniteljem na skupu $\mathcal{D}$ procjenjujemo parametre $\boldsymbol{\mu}$ kategoričke (multinulijeve) varijable $\mathbf{x}$, tj. procjenjujemo parametre multinulijeve distribucije $P(\mathbf{x}|\boldsymbol{\mu})$. Varijabla $\mathbf{x}$ može proprimiti tri moguće vrijednosti ($K = 3$). Apriorna Dirichletova distribucija $P(\boldsymbol{\mu}|\boldsymbol{\alpha})$ definirana je parametrima $\boldsymbol{\alpha}$ definiranima kao $\boldsymbol{\alpha} = \alpha \cdot (1, 4, 2)$, gdje faktor α određuje vršnost Dirichletove distribucije, $\alpha \geqslant 1$. U skupu podataka $\mathcal{D}$ realizirane su sve tri vrijednosti kategoričke varijable $\mathbf{x}$, i to sa sljedećim brojem realizacija: $N_1 = 50$, $N_2 = 29$, $N_3 = 40$. **Koliko mora iznositi faktor α, a da bi vrijednost x_2 bila najvjerojatnija vrijednost kategoričke varijable x prema procijenjenoj distribuciji $P(\mathbf{x}|\boldsymbol{\mu})$?**

A $\alpha > 5$ B $\alpha > 8$ C $\alpha > 2$ D $\alpha > 7$

Cjelina 2: Probabilistički grafički modeli (5 pitanja)

9 (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 modelira 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 0.033   B 0.148   C 0.813   D 0.747

10 (N) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x, y, z, w) = P(z)P(w)P(y|z, w)P(x|y)$. Sve varijable su binarne. Vrijedi $P(z = 1) = 0.1$, $P(w = 1) = 0.1$, $P(x = 1|y = 0) = 0.2$ i $P(x = 1|y = 1) = 0.6$. Tablica uvjetne vjerojatnosti za čvor y je sljedeća:

z	w	$p(y = 1\|z, w)$
0	0	0
0	1	0.4
1	0	0.2
1	1	0.7

Postupkom uzorkovanja s odbijanjem uzorkujemo iz aposteriorne distribucije $P(w|x = 0, z = 1)$. Uzorkovanje smo ponovili ukupno $N = 10000$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N. **U uzorku neodbačenih vektora, koliko je očekivano više vektora sa $w = 0$ od vektora sa $w = 1$?**

A 944   B 348   C 472   D 596

11 (T) Čest način probabilističkog zaključivanja kod Bayesovih mreža jest izračunavanje "aposteriornog upita". Kod te vrste upita zanima nas distribucija nekih varijabli (varijable upita) na temelju zadanih varijabli (opažene varijable). Međutim, Bayesova mreža kodira zajedničku vjerojatnost svih varijabli mreže, uključivo i varijabli koje nisu niti varijable upita niti opažene varijable (varijable smetnje). **Na koji način izračunavamo aposteriorni upit?**

A Kao umnožak vjerojatnosti varijable upita uvjetovane na opažene varijable i zajedničke vjerojatnosti marginalizirane po varijablama smetnje

B Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama upita i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i opaženim varijablama

C Kao umnožak zajedničke vjerojatnosti marginalizirane po opaženim varijablama i zajedničke vjerojatnosti marginalizirane po varijablama smetnje

D Kao omjer zajedničke vjerojatnosti marginalizirane po varijablama smetnje i zajedničke vjerojatnosti marginalizirane po varijablama smetnje i varijablama upita

12 (P) Bayesovom mrežom sa sedam binarnih varijabli modeliramo prometne (ne)prilike u gradu Zagrebu. U našoj mreži, strelice modeliraju pretpostavljeni smjer kauzalnosti. Jutarnje doba dana (J) ili loše vrijeme (V) mogu uzrokovati nastanak prometne gužve (G). Loše vrijeme također može uzrokovati nastanak prometne nesreće (N). Nadalje, nastanak prometne nesreće može uzrokovati nastanak prometne gužve. Loše vrijeme također može uzrokovati zastoj tramvaja (T), no i nestanak struje (S) može uzrokovati zastoj tramvaja. Konačno, zastoj tramvaja može uzrokovati masovno pješačenje putnika (P), što opet može uzrokovati prometnu nesreću. Primjenom algoritma d-odvajanja ispitujemo stohastičku zavisnost između jutarnjeg doba dana (J) i nestanka struje (S). **Koje minimalan skup varijabli treba biti opažen, a da vjerojatnost od S ne ovisi o vrijednosti J?**

A G, T B V, P C V, P, T D T

13 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Sve varijable su binarne, osim varijabli v i x, koje su ternarne. Uz navedeni topološki uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

$$\{v, w\} \perp y | x \qquad \{v, x\} \perp z | \{w, y\}$$

Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

A 21 B 24 C 23 D 18

Cjelina 3: Grupiranje (6 pitanja)

14 (N) Algoritmom HAC grupiramo riječi engleskog jezika. Neoznačeni skup podataka sastoji se od sljedećih riječi:

$$\mathcal{D} = \{\text{“water”}, \text{“waterloo”}, \text{“earth”}, \text{“air”}\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2| / |\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“waterloo”}) = 5/7 = 0.714$. Provedite prve dvije iteracije grupiranja algoritmom HAC uz prosječno povezivanje. **Na kojoj se razini sličnosti spajaju grupe u drugoj iteraciji algoritma HAC?**

A 0.354 B 0.292 C 0.583 D 0.535

15 (T) Algoritam K-means++ proširenje je algoritma K-sredina heurističkim odabirom početnih središta grupa. **Koja je glavna ideja odabira početnih središta grupa kod algoritma K-means++**?

A Najvjerojatnije središte grupe jest primjer koji je najviše udaljen od njemu najbližeg središta

B Vjerojatnost da je neki primjer središte grupe proporcionalna je s brojem primjera u toj grupi i udaljenosti tih primjera od centroida skupa podataka

C Središta grupe su srednje vrijednosti grupa projiciranih na pravac u smjeru prve komponente rastava skupa podataka na glavne komponente (PCA)

D Središta grupa su mjesta na kojima graf kriterijske funkcije u ovisnosti o broju grupa naglo opada pa stagnira

16 (P) Algoritmom GMM grupiramo primjere u dvodimenzijskome ulaznom prostoru. Skup podataka u stvarnosti je uzorkovan iz zajedničke distribucije koja se može opisati sljedećim mješavinskim modelom:

$$p(\mathbf{x}) = \sum_{j=1}^{3} \frac{1}{3} \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \mu_1 = (5,5), \mu_2 = (5,10), \mu_3 = (-10,-10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

Skup $\mathcal{D}$ grupiramo u $K = 2$ grupe. Pritom isprobavamo tri modela, koji se međusobno razlikuju po pretpostavkama na kovarijacijsku matricu. Konkretno: dijeljena i puna kovarijacijska matrica ($\mathcal{H}_1$), nedijeljena i dijagonalna kovarijacijska matrica ($\mathcal{H}_2$) i nedijeljena i izotropna kovarijacijska matrica ($\mathcal{H}_3$). Neka je $\mathcal{L}_i$ izglednost parametara dobivena modelom $\mathcal{H}_i$ nakon konvergencije algoritma. Za inicijalizaciju središta koristi se algoritam K-means++. **Što su očekivani odnosi između izglednosti za ova tri modela?**

A $\mathcal{L}_1 = \mathcal{L}_2 > \mathcal{L}_3$ B $\mathcal{L}_1 > \mathcal{L}_2 > \mathcal{L}_3$ C $\mathcal{L}_1 > \mathcal{L}_3$, $\mathcal{L}_2 > \mathcal{L}_3$ D $\mathcal{L}_2 > \mathcal{L}_1$, $\mathcal{L}_2 > \mathcal{L}_3$

17 (T) Model miješane gustoće sa K komponenti vjerojatnosnu distribuciju neoznačenih podaka definira kao $p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \, p(\mathbf{x}|\boldsymbol{\theta}_k)$. Što modelira uvjetna vjerojatnost $p(\mathbf{x}|\boldsymbol{\theta}_k)$?

A Stupanj pripadnosti primjera $\mathbf{x}$ grupi k

B Vjerojatnost da primjer $\mathbf{x}$ pripada grupi k

C Gustoću vjerojatnosti primjera $\mathbf{x}$ unutar grupe k

D Gustoću vjerojatnosti grupe k za primjer $\mathbf{x}$

18 (P) Skup neoznačenih primjera u dvodimenzijskome ulaznom prostoru neka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{5} = \{(0,0), (0,4), (2,0), (2,4), (4,2)\}$$

Primjere grupiramo algoritmom K-sredina sa $K = 3$ grupe. Za početna središta grupa odaberemo nasumično primjere iz $\mathcal{D}$, pri čemu, naravno, pazimo da odaberemo različita središta. Ishod grupiranja i konačan iznos kriterijske funkcije J ovisit će o odabiru početnih središta. Neka je J^* vrijednost kriterijske funkcije u točki globalnog minimuma, dakle vrijednost koja odgovara najboljem grupiranju. Neka je J^+ vrijednost kriterijske funkcije u točki lokalnog minimuma, i to onoj točki lokalnog minimuma s najvećom vrijednošću funkcije J. **Koliko iznosi razlika $J^+ - J^*$?**

A 6 B 12 C 4 D 8

19 (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	0	2	1	1	0	0	2	1	1
$y^{(i)}_{true}$	1	0	0	2	1	0	2	1	1	1

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.49 B 0.62 C 0.58 D 0.53

Cjelina 4: Vrednovanje modela (3 pitanja)

20 (T) Za vrednovanje višeklasnog klasifikatora često se koriste mjere F_1^μ (mikro F1-mjera) i F_1^M (makro F1-mjera). **Koji je očekivani odnos između vrijednosti tih mjera, i zašto?**

A $F_1^M < F_1^\mu$, jer kod mikro F1-mjere računamo prosjek F1-mjere kroz sve klase, a klasifikator na većima klasama manje griješi

B $F_1^M < F_1^\mu$, jer kod mikro F1-mjere zbrajamo matrice zabune kroz sve klase, pa je ukupan broj pogrešaka veći nego kad računamo prosječnu pogrešku kroz klase

C $F_1^M > F_1^\mu$, jer kod mikro F1-mjere zbrajamo matrice zabune kroz sve klase, a primjera iz manjih klasa ima manje, pa manje doprinose pogrešci

D $F_1^M < F_1^\mu$, jer kod makro F1-mjere računamo prosjek F1-mjere kroz sve klase, a klasifikator na manjim klasama više griješi

21 (N) Na ispitnome skupu evaluiramo model multinomijalne logističke regresije (MLR) za klasifikaciju u $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$
\begin{array}{c c}
 & \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\
\begin{array}{c} y=1 \\ y=2 \\ y=3 \end{array} & \begin{pmatrix} 30 & 8 & 3 \\ 11 & 25 & 5 \\ 4 & 12 & 2 \end{pmatrix}
\end{array}
$$

Klasifikator MLR uspoređujemo s klasifikatorom RAND koji primjere klasificira nasumično, i to tako da oznaku $y = j$ dodjeljuje s vjerojatnošću proporcionalnoj udjelu klase j u ispitnome skupu. Izračunajte mikro-F_1 za klasifikator MLR i očekivani mikro-F_1 za klasifikator RAND. **Koliko iznosi očekivana razlika u vrijednostima mikro-F_1 klasifikatora MLR i RAND?**

A 0.085 B 0.205 C 0.155 D 0.185

22 (P) Na istom skupu označenih primjera vrednujemo četiri binarna klasifikatora s vjerojatnosnim izlazima. Za vrednovanje koristimo krivulju ROC. Svaki smo klasifikator ispitali s tri vrijednosti klasifikacijskog praga te smo za te vrijednosti izračunali FPR (stopa lažnog alarma) i TPR (odziv). Dobiveni parovi vrijednosti (FPR, TPR) za sva četiri klasifikatora su sljedeći:

$$h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) \qquad h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8)$$
$$h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) \qquad h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0)$$

Skicirajte odgovarajuće krivulje ROC, linearno interpolirajući između izmjerenih točaka, a dodajte i točke koje odgovaraju krajnjim vrijednostima klasifikacijskog praga (0 i 1). Pritom, ako je neki klasifikator lošiji od nasumičnog klasifikatora, umjesto tog klasifikatora razmatrajte njegovu popravljenu varijantu koju ćete dobiti invertiranjem izlaza ($1 - h(\mathbf{x})$ umjesto $h(\mathbf{x})$). **Koji su od ispitanih klasifikatora (eventualno nakon popravka) najbolji prema krivulji ROC?**

A h_1 i h_4 B h_1 i h_3 C h_3 i h_4 D h_1 i h_2

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22
Grupa A	D	D	B	D	B	A	B	B	B	D	C	D	B	B	A	D	C	B	D	B	D	C
Grupa B	B	D	D	A	D	D	D	A	D	C	A	D	D	B	D	A	D	A	B	A	D	D
Grupa C	B	D	C	A	A	C	A	B	D	C	B	C	B	B	A	C	B	A	D	B	D	D
Grupa D	A	A	A	B	B	C	D	B	D	C	A	B	B	A	D	C	B	C	D	C	C	B
Grupa E	D	A	A	B	B	C	C	A	D	A	B	C	D	A	D	D	C	D	A	A	A	A
Grupa F	D	C	A	D	B	C	B	C	D	D	A	D	B	A	B	B	A	A	A	C	C	B
Grupa G	D	C	B	C	A	A	B	C	C	A	A	A	C	D	C	A	A	B	B	C	A	C
Grupa H	C	A	D	A	A	B	D	D	B	D	D	B	B	C	A	D	C	D	C	D	C	A