Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

6. veljače 2024.

Završni ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% 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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 4: 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 $P(y=1) = P(y=2) = 1/5$ 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 $-10 \leq x \leq 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

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

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

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

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

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

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

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

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

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

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

3 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{ne}, \text{zimski}, \text{preddipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.588 B 0.488 C 0.322 D 0.741

4 (T) Polunaivan Bayesov klasifikator može ostvariti veću klasifikacijsku točnost od naivnog Bayesovog klasifikatora u situacijama kada između značajki postoji zavisnost. Jedan način određivanja zavisnosti između značajki temelji se na izračunu Kullback-Leiblerove divergencije (KL-divergencije) između distribucija dviju varijabli. **Koji je princip uporabe KL-divergencije u izgradnji polunaivnog Bayesovog klasifikatora?**

- A Što je manja KL-divergencija između $P(x_1)$ i $P(x_2)$, to su varijable x_1 i x_2 više zavisne, pa je to veća korist od njihovog združivanja u jedan faktor
- B Ako je KL-divergencija između $P(x_1)P(x_2)$ i $P(x_1, x_2)$ jednaka nuli, onda su varijable linearno zavisne i treba ih odvojiti u zasebne faktore
- C Ako je KL-divergencija između $P(x_1|x_2)$ i $P(x_2|x_1)$ vrlo mala, onda su varijable x_1 i x_2 vrlo zavisne i treba ih odvojiti u zasebne faktore
- D Što je veća KL-divergencija između $P(x_1)P(x_2)$ i $P(x_1, x_2)$, to su varijable x_1 i x_2 manje nezavisne, pa je to veća korist od njihovog združivanja u jedan faktor

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:

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

Model treniramo na skupu 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 -2.75 B -3.84 C -4.13 D -3.03

6 (T) Bayesov klasifikator temelji se na hipotezi *maximum aposteriori* (MAP). Za treniranje Bayesovog klasifikatora možemo koristiti procjenitelj MLE. **Koja je razlika između hipoteze MAP kod Bayesovog klasifikatora i procjenitelja MLE?**

- A Procjenitelj MLE određuje apriornu vjerojatnost klase, a hipoteza MAP maksimizira aposterionu vjerojatnost parametara
- B Hipoteza MAP maksimizira aposterionu vjerojatnost parametara, a procjenitelj MLE određuje aposteriornu vjerojatnost klase za zadani primjer
- C Procjenitelj MLE primjer klasificira u klasu koja je za njega navjerojatnija, a hipoteza MAP tu odluku kombinira s apriornim znanjem
- D Hipoteza MAP određuje koju klasu odabrati za primjer, a procjenitelj MLE određuje koju vrijednost odabrati za parametar

7 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{10, 5, 6, 20\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Za procjenu parametra μ koristimo procjenitelj $\hat{\mu}_{\mathrm{MLE}}$, a za procjenu parametra σ^2 procjenitelj $\hat{\sigma}^2_{\mathrm{UB}}$, koji je nepristran procjenitelj izveden odgovarajućom korekcijom iz procjenitelja $\hat{\sigma}^2_{\mathrm{MLE}}$. Izračunajte log-izglednost procijenjenih parametara na skupu $\mathcal{D}$, tj. $\ln \mathcal{L}(\hat{\mu}_{\mathrm{MLE}}, \hat{\sigma}^2_{\mathrm{UB}}|\mathcal{D})$. **Koliko iznosi ta log-izglednost?**

A -13.163 B -12.873 C -10.832 D -15.200

8 (P) Uz određene restrikcije na parametre Gaussovog Bayesovog klasifikatora, Gaussov Bayesov klasifikator i logistička regresija su generativno-diskriminativni par modela. U pravilu, logistička regresija imat će manje parametara od njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. Razmotrite slučaj binarne klasifikacije u ulaznome prostoru dimenzije $n = 300$ pomoću modela logističke regresije i njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. **Koliko će model Gaussovog Bayesovog klasifikatora imati više parametara od modela logističke regresije?**

A 45450 B 40400 C 90600 D 20300

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

9 (T) Probablistički grafički modeli na sažet način prikazuju zajedničku distribuciju n slučajnih varijabli. Ta je distribucija prikazana usmjerenim grafom. Neka graf sadrži čvor označen sa x. **Čemu odgovara taj čvor?**

A Distribuciji varijable x uvjetovane vrijednostima varijabli roditeljskih čvorova od x

B Distribuciji varijable x uz opažene sve varijable koje varijabli x prethode u topološkom uređaju

C Marginalnoj vjerojatnosti $P(x)$ ili marginalnoj gustoći vjerojatnosti $p(x)$

D Zajedničkoj vjerojatnosti varijable x i varijabli roditeljskih čvorova od x

10 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

Primjenom algoritma d-odvajanja ispitujemo zavisnosti između parova varijabli. **Koje od sljedećih tvrdnji o nezavisnosti vrijede u ovoj Bayesovoj mreži?**

A $v \perp z | x$ B $v \perp y | x$ C $v \perp y | z$ D $x \perp z | w$

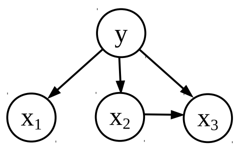
11 (N) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x, y, z, w) = P(z)P(w)P(y|z, w)P(x|y)$. Sve varijable su binarne. Vrijedi $P(z = 1) = 0.1$, $P(w = 1) = 0.1$, $P(x = 1|y = 0) = 0.2$ i $P(x = 1|y = 1) = 0.6$. Tablica uvjetne vjerojatnosti za čvor y je sljedeća:

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 konačni 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

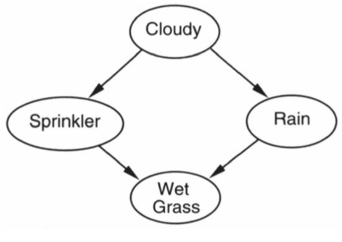
12 (P) Na slici ispod prikazana je Bayesova mreža koja odgovara polunaivnom Bayesovom klasifikatoru. Pretpostavite da su značajke x_1, x_2 i x_3 ternarne varijable te da je oznaka klase y također ternarna varijabla. Označimo ovaj model sa $\mathcal{H}_2$. Model $\mathcal{H}_2$ može se pojednostaviti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



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

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

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



C	$P(S = 1\|C)$
0	0.5
1	0.4

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

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

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

A 0.069 B 0.223 C 0.309 D 0.144

Cjelina 6: Grupiranje i vrednovanje modela (9 pitanja)

14 (T) Za procjenu parametara modela GMM (središta grupa $\boldsymbol{\mu}_k$, kovarijacijske matrice Σ_k i koeficijenti mješavine π_k) tipično se koristi algoritam maksimizacije očekivanja (EM-algoritam). Prije pokretanja EM-algoritma, parametre je potrebno inicijalizirati. U praksi se za inicijalizaciju parametara $\boldsymbol{\mu}_k$ koristi algoritam K-sredina. **Koji je očekivani efekt uporabe algoritma K-sredina za inicijalizaciju EM-algoritma?**

- A Veća potpuna log-izglednost konačnih parametara i veće vrijednosti gustoća komponenti
- B Veća log-izglednost početnih i konačnih parametara te manje iteracija do konvergencije
- C Manja log-izglednost početnih i konačnih parametara te optimalan broj grupa K
- D Manja log-izglednost početnih parametara, ali veća log-izglednost konačnih parametara

15 (T) Kod vrednovanja algoritma strojnog učenja važno je da vrednujemo model optimalne složenosti. Zbog toga označeni skup podataka dijelimo na tri disjunktna podskupa: skup za učenje, skup za provjeru i ispitni skup. Za odabir modela (tj. optimizaciju hiperparametara) koristimo skup za učenje i skup za provjeru, dok točnost optimalnog modela ispitujemo na ispitnome skupu. Zamislite da smo zaboravili upotrijebiti skup za provjeru: umjesto toga hiperparametre modela optimirali smo na skupu za učenje, a model smo potom ispitali na ispitnom skupu. **U odnosu na to da smo pravilno proveli vrednovanje algoritma, što možemo očekivati u ovom slučaju?**

- A Model će biti prenaučen i imat će manju točnost na ispitnome skupu
- B Model će biti prenaučen, ali imat će veću točnost na ispitnome skupu
- C Model će biti podnaučen, ali imat će veću točnost na ispitnome skupu
- D Model će biti optimalne složenosti i imat će veću točnost na ispitnome skupu

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

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_i = \big\{(1,1),(1,3),(1,4),(3,4)\big\}$$

Primjere grupiramo u $K = 2$ grupe algoritmom K-means++. Konačna vrijednost funkcije pogreške J bit će različita za različita pokretanja algoritma. Razmotrite sve moguće konačne vrijednosti funkcije J, tj. vrijednosti funkcije J nakon što se algoritam K-means++ zaustavi. **Koliko iznosi maksimalna moguća konačna vrijednost funkcije J?**

A 4 B 11/2 C 10/3 D 14/3

17 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

$$\begin{pmatrix} 1.0 & 0.2 & 0.8 & 0.1 & 0.4 \\ 0.2 & 1.0 & 0.3 & 0.3 & 0.7 \\ 0.8 & 0.3 & 1.0 & 0.6 & 0.5 \\ 0.1 & 0.3 & 0.6 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

Provedite grupiranje algoritmom HAC s potpunim povezivanjem. Pritom u svakoj iteraciji bilježite na kojoj razini sličnosti se odvija stapanje dviju grupa. **Koliko iznosi zbroj po svim razinama sličnosti na kojima se odvija stapanje grupa?**

A 2.4 B 1.9 C 1.8 D 2.0

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	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

19 (N) Vrednujemo višeklasni klasifikator koji koristimo kao referentni model za vrednovanje drugih klasifikatora. Klasifikator nasumično dodjeljuje oznake za tri klase ($K = 3$), i to prema distribuciji oznaka na skupu za učenje. Skup za učenje sadrži 100 primjera sa 35, 60 i 5 primjera u prvoj, drugoj odnosno trećoj klasi. S druge strane, ispitni skup sadrži 50 primjera, od kojih je po 20 u prvoj i drugoj klasi, a 10 u trećoj klasi. **Koliko iznosi očekivana makro-F1 vrijednost nasumičnog klasifikatora na ispitnome skupu?**

A 0.3111 B 0.2085 C 0.2989 D 0.2156

20 (T) Algoritam K-sredina iterativno smanjuje vrijednost funkcije pogreške J. Neovisno o broju grupa K, algoritam uvijek konvergira (završava s izvođenjem). Međutim, konačna vrijednost funkcije pogreške J može se razlikovati između dva pokretanja algoritma s istim brojem grupa K. **Zašto?**

A Zbog nelinearnosti granice između K grupa

B Zbog razlike u početnim središtima $\boldsymbol{\mu}_k$

C Zbog izotropnosti kovarijacijskih matrica Σ_k

D Zbog loše odabrane vrijednosti za broj grupa K

21 (P) Podatke u dvodimenzijskome ulaznom prostoru ($n = 2$) grupiramo u dvije grupe ($K = 2$) algoritmom Gaussove mješavine (GMM). Skup podataka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{8} = \big\{(-2,-2),(-2,-1),(-1,-2),(-1,-1),(1,2),(3,1),(3,3),(5,2)\big\}$$

Razmatramo dva modela: $\mathcal{H}_1$ i $\mathcal{H}_2$. Oba modela koriste punu kovarijacijsku matricu, međutim kod modela $\mathcal{H}_1$ kovarijacijska matrica nije dijeljena, dok je kod modela $\mathcal{H}_2$ kovarijacijska matrica dijeljena. Za svaki model razmotrite ono grupiranje koje maksimizira log-izglednost na skupu $\mathcal{D}$. Neka je $p_j(\mathbf{x}^{(i)}|y = k)$ izglednost grupe $y = k$ za primjer $\mathbf{x}^{(i)}$ prema modelu $\mathcal{H}_j$. Oznake grupa su $y = 1$ i $y = 2$, te pretpostavite da je grupiranje takvo da vrijedi $\|\boldsymbol{\mu}_1\| > \|\boldsymbol{\mu}_2\|$. **Što od sljedećega vrijedi?**

A $p_2(\mathbf{x}^{(8)}|y = 1) < p_1(\mathbf{x}^{(5)}|y = 1)$ C $p_2(\mathbf{x}^{(4)}|y = 2) > p_1(\mathbf{x}^{(3)}|y = 2)$

B $p_2(\mathbf{x}^{(6)}|y = 1) < p_1(\mathbf{x}^{(7)}|y = 1)$ D $p_2(\mathbf{x}^{(1)}|y = 2) > p_1(\mathbf{x}^{(2)}|y = 2)$

22 (P) Raspolažemo sa 1250 označenih primjera. Na tom skupu treniramo i vrednujemo algoritam SVM, optimizirajući hiperparametre C i γ. Za vrednovanje koristimo ugniježđenu unakrsnu provjeru sa 5 preklopa u vanjskoj petlji i 5 preklopa u unutarnjoj petlji. To znači da za svaku kombinaciju vrijednosti hiperparametara C i γ treniramo pet modela. Ti modeli trenirani su na skupovima za učenje koji nisu disjunktni: svaki par treniranih modela dijele određeni broj primjera za učenje. Izračunajte koliko primjera za učenje dijele svaki par modela koje treniramo u unutarnjoj petlji ugniježđene unakrsne provjere. **Za koliko bi taj broj narastao kada bismo broj preklopa u unutarnjoj petlji povećali na 10?**

A 200 B 100 C 125 D 240

Završni ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% 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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 4: Procjena parametara i Bayesov klasifikator (8 pitanja)

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

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

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

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

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

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

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

2 (T) Polunaivan Bayesov klasifikator može ostvariti veću klasifikacijsku točnost od naivnog Bayesovog klasifikatora u situacijama kada između značajki postoji zavisnost. Jedan način određivanja zavisnosti između značajki temelji se na izračunu Kullback-Leiblerove divergencije (KL-divergencije) između distribucija dviju varijabli. **Koji je princip uporabe KL-divergencije u izgradnji polunaivnog Bayesovog klasifikatora?**

A Ako je KL-divergencija između $P(x_1|x_2)$ i $P(x_2|x_1)$ vrlo mala, onda su varijable x_1 i x_2 vrlo zavisne i treba ih odvojiti u zasebne faktore

B Što je manja KL-divergencija između $P(x_1)$ i $P(x_2)$, to su varijable x_1 i x_2 više zavisne, pa je to veća korist od njihovog združivanja u jedan faktor

C Ako je KL-divergencija između $P(x_1)P(x_2)$ i $P(x_1, x_2)$ jednaka nuli, onda su varijable linearno zavisne i treba ih odvojiti u zasebne faktore

D Što je veća KL-divergencija između $P(x_1)P(x_2)$ i $P(x_1, x_2)$, to su varijable x_1 i x_2 manje nezavisne, pa je to veća korist od njihovog združivanja u jedan faktor

3 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{10, 5, 6, 20\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Za procjenu parametra μ koristimo procjenitelj $\hat{\mu}_{\text{MLE}}$, a za procjenu parametra σ^2 procjenitelj $\hat{\sigma}^2_{\text{UB}}$, koji je nepristran procjenitelj izveden odgovarajućom korekcijom iz procjenitelja $\hat{\sigma}^2_{\text{MLE}}$. Izračunajte log-izglednost procijenjenih parametara na skupu $\mathcal{D}$, tj. $\ln \mathcal{L}(\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{UB}}|\mathcal{D})$. **Koliko iznosi ta log-izglednost?**

A -10.832 B -15.200 C -13.163 D -12.873

4 (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 skupu 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 -4.13 B -3.03 C -2.75 D -3.84

5 (T) Bayesov klasifikator temelji se na hipotezi *maximum aposteriori* (MAP). Za treniranje Bayesovog klasifikatora možemo koristiti procjenitelj MLE. **Koja je razlika između hipoteze MAP kod Bayesovog klasifikatora i procjenitelja MLE?**

A Procjenitelj MLE primjer klasificira u klasu koja je za njega navjerojatnija, a hipoteza MAP tu odluku kombinira s apriornim znanjem

B Hipoteza MAP maksimizira aposterionu vjerojatnost parametara, a procjenitelj MLE određuje aposteriornu vjerojatnost klase za zadani primjer

C Hipoteza MAP određuje koju klasu odabrati za primjer, a procjenitelj MLE određuje koju vrijednost odabrati za parametar

D Procjenitelj MLE određuje apriornu vjerojatnost klase, a hipoteza MAP maksimizira aposterionu vjerojatnost parametara

6 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{ne}, \text{zimski}, \text{preddipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.741 B 0.322 C 0.488 D 0.588

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 $-10 \leq x \leq 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

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

8 (P) Uz određene restrikcije na parametre Gaussovog Bayesovog klasifikatora, Gaussov Bayesov klasifikator i logistička regresija su generativno-diskriminativni par modela. U pravilu, logistička regresija imat će manje parametara od njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. Razmotrite slučaj binarne klasifikacije u ulaznome prostoru dimenzije $n = 200$ pomoću modela logističke regresije i njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. **Koliko će model Gaussovog Bayesovog klasifikatora imati više parametara od modela logističke regresije?**

A 90600 B 45450 C 20300 D 40400

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

9 (T) Probablistički grafički modeli na sažet način prikazuju zajedničku distribuciju n slučajnih varijabli. Ta je distribucija prikazana usmjerenim grafom. Neka graf sadrži čvor označen sa x. **Čemu odgovara taj čvor?**

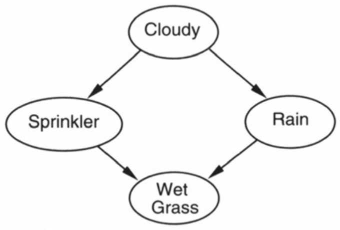
A Distribuciji varijable x uvjetovane vrijednostima varijabli roditeljskih čvorova od x

B Marginalnoj vjerojatnosti $P(x)$ ili marginalnoj gustoći vjerojatnosti $p(x)$

C Distribuciji varijable x uz opažene sve varijable koje varijabli x prethode u topološkom uređaju

D Zajedničkoj vjerojatnosti varijable x i varijabli roditeljskih čvorova od x

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



C	$P(S=1\|C)$
0	0.5
1	0.3

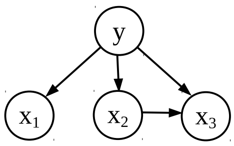
C	$P(R=1\|C)$
0	0.2
1	0.8

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

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

A 0.069 B 0.309 C 0.223 D 0.144

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



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

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

12 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

Primjenom algoritma d-odvajanja ispitujemo zavisnosti između parova varijabli. **Koje od sljedećih tvrdnji o nezavisnosti vrijede u ovoj Bayesovoj mreži?**

A $v\perp z|w$ B $v\perp z|x$ C $w\perp y|x$ D $v\perp y|z$

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 $P(z=1) = 0.1$, $P(w=1) = 0.3$, $P(x=1|y=0) = 0.2$ 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 konačni uzorak manji od N. **U uzorku neodbačenih vektora, koliko je očekivano više vektora sa $w=0$ od vektora sa $w=1$?**

A 472 B 348 C 944 D 596

Cjelina 6: Grupiranje i vrednovanje modela (9 pitanja)

14 (T) Za procjenu parametara modela GMM (središta grupa $\boldsymbol{\mu}_k$, kovarijacijske matrice Σ_k i koeficijenti mješavine π_k) tipično se koristi algoritam maksimizacije očekivanja (EM-algoritam). Prije pokretanja EM-algoritma, parametre je potrebno inicijalizirati. U praksi se za inicijalizaciju parametara $\boldsymbol{\mu}_k$ koristi algoritam K-sredina. **Koji je očekivani efekt uporabe algoritma K-sredina za inicijalizaciju EM-algoritma?**

- A Manja log-izglednost početnih i konačnih parametara te optimalan broj grupa K
- B Veća potpuna log-izglednost konačnih parametara i veće vrijednosti gustoća komponenti
- C Veća log-izglednost početnih i konačnih parametara te manje iteracija do konvergencije
- D Manja log-izglednost početnih parametara, ali veća log-izglednost konačnih parametara

15 (T) Kod vrednovanja algoritma strojnog učenja važno je da vrednujemo model optimalne složenosti. Zbog toga označeni skup podataka dijelimo na tri disjunktna podskupa: skup za učenje, skup za provjeru i ispitni skup. Za odabir modela (tj. optimizaciju hiperparametara) koristimo skup za učenje i skup za provjeru, dok točnost optimalnog modela ispitujemo na ispitnome skupu. Zamislite da smo zaboravili upotrijebiti skup za provjeru: umjesto toga hiperparametre modela optimirali smo na skupu za učenje, a model smo potom ispitali na ispitnom skupu. **U odnosu na to da smo pravilno proveli vrednovanje algoritma, što možemo očekivati u ovom slučaju?**

- A Model će biti podnaučen, ali imat će veću točnost na ispitnome skupu
- B Model će biti optimalne složenosti i imat će veću točnost na ispitnome skupu
- C Model će biti prenaučen i imat će manju točnost na ispitnome skupu
- D Model će biti prenaučen, ali imat će veću točnost na ispitnome skupu

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	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 0.49 B 0.62 C 0.53 D 0.58

17 (P) Raspolažemo sa 1500 označenih primjera. Na tom skupu treniramo i vrednujemo algoritam SVM, optimizirajući hiperparametre C i γ. Za vrednovanje koristimo ugniježđenu unakrsnu provjeru sa 5 preklopa u vanjskoj petlji i 5 preklopa u unutarnjoj petlji. To znači da za svaku kombinaciju vrijednosti hiperparametara C i γ treniramo pet modela. Ti modeli trenirani su na skupovima za učenje koji nisu disjunktni: svaki par treniranih modela dijele određeni broj primjera za učenje. Izračunajte koliko primjera za učenje dijele svaki par modela koje treniramo u unutarnjoj petlji ugniježđene unakrsne provjere. **Za koliko bi taj broj narastao kada bismo broj preklopa u unutarnjoj petlji povećali na 10?**

A 150 B 240 C 120 D 288

18 (P) Podatke u dvodimenzijskome ulaznom prostoru ($n = 2$) grupiramo u dvije grupe ($K = 2$) algoritmom Gaussove mješavine (GMM). Skup podataka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{8} = \big\{(-2,-2),(-2,-1),(-1,-2),(-1,-1),(1,2),(3,1),(3,3),(5,2)\big\}$$

Razmatramo dva modela: $\mathcal{H}_1$ i $\mathcal{H}_2$. Oba modela koriste punu kovarijacijsku matricu, međutim kod modela $\mathcal{H}_1$ kovarijacijska matrica nije dijeljena, dok je kod modela $\mathcal{H}_2$ kovarijacijska matrica dijeljena. Za svaki model razmotrite ono grupiranje koje maksimizira log-izglednost na skupu $\mathcal{D}$. Neka je $p_j(\mathbf{x}^{(i)}|y = k)$ izglednost grupe

$y = k$ za primjer $\mathbf{x}^{(i)}$ prema modelu $\mathcal{H}_j$. Oznake grupa su $y = 1$ i $y = 2$, te pretpostavite da je grupiranje takvo da vrijedi $\|\boldsymbol{\mu}_1\| > \|\boldsymbol{\mu}_2\|$. **Što od sljedećega vrijedi?**

A $p_2(\mathbf{x}^{(1)}|y=2) > p_1(\mathbf{x}^{(2)}|y=2)$ C $p_2(\mathbf{x}^{(4)}|y=2) > p_1(\mathbf{x}^{(3)}|y=2)$

B $p_2(\mathbf{x}^{(6)}|y=1) < p_1(\mathbf{x}^{(7)}|y=1)$ D $p_2(\mathbf{x}^{(8)}|y=1) < p_1(\mathbf{x}^{(5)}|y=1)$

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

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_i = \big\{(1,1),(1,3),(1,4),(3,4)\big\}$$

Primjere grupiramo u $K = 2$ grupe algoritmom K-means++. Konačna vrijednost funkcije pogreške J bit će različita za različita pokretanja algoritma. Razmotrite sve moguće konačne vrijednosti funkcije J, tj. vrijednosti funkcije J nakon što se algoritam K-means++ zaustavi. **Koliko iznosi maksimalna moguća konačna vrijednost funkcije J?**

A 4 B 14/3 C 11/2 D 10/3

20 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

$$\begin{pmatrix} 1.0 & 0.2 & 0.6 & 0.7 & 0.4 \\ 0.2 & 1.0 & 0.3 & 0.3 & 0.7 \\ 0.6 & 0.3 & 1.0 & 0.9 & 0.5 \\ 0.7 & 0.3 & 0.9 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

Provedite grupiranje algoritmom HAC s potpunim povezivanjem. Pritom u svakoj iteraciji bilježite na kojoj razini sličnosti se odvija stapanje dviju grupa. **Koliko iznosi zbroj po svim razinama sličnosti na kojima se odvija stapanje grupa?**

A 2.4 B 1.8 C 1.9 D 2.0

21 (N) Vrednujemo višeklasni klasifikator koji koristimo kao referentni model za vrednovanje drugih klasifikatora. Klasifikator nasumično dodjeljuje oznake za tri klase ($K = 3$), i to prema distribuciji oznaka na skupu za učenje. Skup za učenje sadrži 100 primjera sa 25, 70 i 5 primjera u prvoj, drugoj odnosno trećoj klasi. S druge strane, ispitni skup sadrži 50 primjera, od kojih je po 10 u prvoj i drugoj klasi, a 30 u trećoj klasi. **Koliko iznosi očekivana makro-F1 vrijednost nasumičnog klasifikatora na ispitnome skupu?**

A 0.3111 B 0.2989 C 0.2156 D 0.2085

22 (T) Algoritam K-sredina iterativno smanjuje vrijednost funkcije pogreške J. Neovisno o broju grupa K, algoritam uvijek konvergira (završava s izvođenjem). Međutim, konačna vrijednost funkcije pogreške J može se razlikovati između dva pokretanja algoritma s istim brojem grupa K. **Zašto?**

A Zbog razlike u početnim središtima $\boldsymbol{\mu}_k$

B Zbog loše odabrane vrijednosti za broj grupa K

C Zbog izotropnosti kovarijacijskih matrica Σ_k

D Zbog nelinearnosti granice između K grupa

Završni ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% 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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 4: Procjena parametara i Bayesov klasifikator (8 pitanja)

1 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{10, 5, 6, 20\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Za procjenu parametra μ koristimo procjenitelj $\hat{\mu}_{\text{MLE}}$, a za procjenu parametra σ^2 procjenitelj $\hat{\sigma}^2_{\text{UB}}$, koji je nepristran procjenitelj izveden odgovarajućom korekcijom iz procjenitelja $\hat{\sigma}^2_{\text{MLE}}$. Izračunajte log-izglednost procijenjenih parametara na skupu $\mathcal{D}$, tj. $\ln \mathcal{L}(\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{UB}}|\mathcal{D})$. **Koliko iznosi ta log-izglednost?**

A -15.200 B -12.873 C -13.163 D -10.832

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 skupu 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 -2.75 C -3.03 D -4.13

3 (P) Uz određene restrikcije na parametre Gaussovog Bayesovog klasifikatora, Gaussov Bayesov klasifikator i logistička regresija su generativno-diskriminativni par modela. U pravilu, logistička regresija imat će manje parametara od njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. Razmotrite slučaj binarne klasifikacije u ulaznome prostoru dimenzije $n = 200$ pomoću modela logističke regresije i njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. **Koliko će model Gaussovog Bayesovog klasifikatora imati više parametara od modela logističke regresije?**

A 45450 B 40400 C 90600 D 20300

4 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{da}, \text{zimski}, \text{preddipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.588   B 0.488   C 0.741   D 0.322

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

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

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

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

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

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

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

6 (T) Bayesov klasifikator temelji se na hipotezi *maximum aposteriori* (MAP). Za treniranje Bayesovog klasifikatora možemo koristiti procjenitelj MLE. **Koja je razlika između hipoteze MAP kod Bayesovog klasifikatora i procjenitelja MLE?**

A Hipoteza MAP maksimizira aposterionu vjerojatnost parametara, a procjenitelj MLE određuje aposteriornu vjerojatnost klase za zadani primjer

B Procjenitelj MLE određuje apriornu vjerojatnost klase, a hipoteza MAP maksimizira aposterionu vjerojatnost parametara

C Hipoteza MAP određuje koju klasu odabrati za primjer, a procjenitelj MLE određuje koju vrijednost odabrati za parametar

D Procjenitelj MLE primjer klasificira u klasu koja je za njega navjerojatnija, a hipoteza MAP tu odluku kombinira s apriornim znanjem

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 $P(y = 1) = P(y = 2) = 1/5$ 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 $-10 \leq x \leq 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

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

8 (T) Polunaivan Bayesov klasifikator može ostvariti veću klasifikacijsku točnost od naivnog Bayesovog klasifikatora u situacijama kada između značajki postoji zavisnost. Jedan način određivanja zavisnosti između značajki temelji se na izračunu Kullback-Leiblerove divergencije (KL-divergencije) između distribucija dviju varijabli. **Koji je princip uporabe KL-divergencije u izgradnji polunaivnog Bayesovog klasifikatora?**

A Što je manja KL-divergencija između $P(x_1)$ i $P(x_2)$, to su varijable x_1 i x_2 više zavisne, pa je to veća korist od njihovog združivanja u jedan faktor

B Ako je KL-divergencija između $P(x_1|x_2)$ i $P(x_2|x_1)$ vrlo mala, onda su varijable x_1 i x_2 vrlo zavisne i treba ih odvojiti u zasebne faktore

C Ako je KL-divergencija između $P(x_1|x_2)$ i $P(x_2|x_1)$ vrlo velika, onda su varijable x_1 i x_2 vrlo zavisne i treba ih združiti u jedan faktor

D Što je veća KL-divergencija između $P(x_1)P(x_2)$ i $P(x_1, x_2)$, to su varijable x_1 i x_2 manje nezavisne, pa je to veća korist od njihovog združivanja u jedan faktor

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

9 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

Primjenom algoritma d-odvajanja ispitujemo zavisnosti između parova varijabli. **Koje od sljedećih tvrdnji o nezavisnosti vrijede u ovoj Bayesovoj mreži?**

A $v \perp z | x$ B $v \perp y | x$ C $x \perp z | w$ D $v \perp z | y$

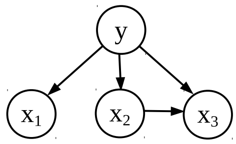
10 (N) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x, y, z, w) = P(z)P(w)P(y|z, w)P(x|y)$. Sve varijable su binarne. Vrijedi $P(z = 1) = 0.1$, $P(w = 1) = 0.3$, $P(x = 1|y = 0) = 0.2$ i $P(x = 1|y = 1) = 0.6$. Tablica uvjetne vjerojatnosti za čvor y je sljedeća:

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 konačni 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

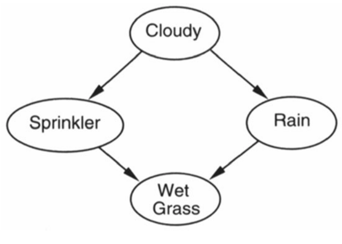
11 (P) Na slici ispod prikazana je Bayesova mreža koja odgovara polunaivnom Bayesovom klasifikatoru. Pretpostavite da su značajke x_1, x_2 i x_3 ternarne varijable te da je oznaka klase y također ternarna varijabla. Označimo ovaj model sa $\mathcal{H}_2$. Model $\mathcal{H}_2$ može se pojednostaviti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



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

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

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



C	$P(S = 1\|C)$
0	0.5
1	0.3

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

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

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

A 0.223 B 0.069 C 0.309 D 0.144

13 (T) Probablistički grafički modeli na sažet način prikazuju zajedničku distribuciju n slučajnih varijabli. Ta je distribucija prikazana usmjerenim grafom. Neka graf sadrži čvor označen sa x. **Čemu odgovara taj čvor?**

A Distribuciji varijable x uvjetovane vrijednostima varijabli roditeljskih čvorova od x

B Zajedničkoj vjerojatnosti varijable x i varijabli roditeljskih čvorova od x

C Distribuciji varijable x uz opažene sve varijable koje varijabli x prethode u topološkom uređaju

D Marginalnoj vjerojatnosti $P(x)$ ili marginalnoj gustoći vjerojatnosti $p(x)$

Cjelina 6: Grupiranje i vrednovanje modela (9 pitanja)

14 (P) Podatke u dvodimenzijskome ulaznom prostoru ($n = 2$) grupiramo u dvije grupe ($K = 2$) algoritmom Gaussove mješavine (GMM). Skup podataka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{8} = \big\{(-2,-2),(-2,-1),(-1,-2),(-1,-1),(1,2),(3,1),(3,3),(5,2)\big\}$$

Razmatramo dva modela: $\mathcal{H}_1$ i $\mathcal{H}_2$. Oba modela koriste punu kovarijacijsku matricu, međutim kod modela $\mathcal{H}_1$ kovarijacijska matrica nije dijeljena, dok je kod modela $\mathcal{H}_2$ kovarijacijska matrica dijeljena. Za svaki model razmotrite ono grupiranje koje maksimizira log-izglednost na skupu $\mathcal{D}$. Neka je $p_j(\mathbf{x}^{(i)}|y = k)$ izglednost grupe $y = k$ za primjer $\mathbf{x}^{(i)}$ prema modelu $\mathcal{H}_j$. Oznake grupa su $y = 1$ i $y = 2$, te pretpostavite da je grupiranje takvo da vrijedi $\|\boldsymbol{\mu}_1\| > \|\boldsymbol{\mu}_2\|$. **Što od sljedećega vrijedi?**

A $p_2(\mathbf{x}^{(1)}|y = 2) > p_1(\mathbf{x}^{(2)}|y = 2)$ C $p_2(\mathbf{x}^{(6)}|y = 1) < p_1(\mathbf{x}^{(7)}|y = 1)$

B $p_2(\mathbf{x}^{(8)}|y = 1) < p_1(\mathbf{x}^{(5)}|y = 1)$ D $p_2(\mathbf{x}^{(4)}|y = 2) > p_1(\mathbf{x}^{(3)}|y = 2)$

15 (T) Za procjenu parametara modela GMM (središta grupa $\boldsymbol{\mu}_k$, kovarijacijske matrice Σ_k i koeficijenti mješavine π_k) tipično se koristi algoritam maksimizacije očekivanja (EM-algoritam). Prije pokretanja EM-algoritma, parametre je potrebno inicijalizirati. U praksi se za inicijalizaciju parametara $\boldsymbol{\mu}_k$ koristi algoritam K-sredina. **Koji je očekivani efekt uporabe algoritma K-sredina za inicijalizaciju EM-algoritma?**

A Manja log-izglednost početnih i konačnih parametara te optimalan broj grupa K

B Veća log-izglednost početnih i konačnih parametara te manje iteracija do konvergencije

C Manja log-izglednost početnih parametara, ali veća log-izglednost konačnih parametara

D Veća potpuna log-izglednost konačnih parametara i veće vrijednosti gustoća komponenti

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

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_i = \big\{(1,1),(1,3),(1,4),(3,4)\big\}$$

Primjere grupiramo u $K = 2$ grupe algoritmom K-means++. Konačna vrijednost funkcije pogreške J bit će različita za različita pokretanja algoritma. Razmotrite sve moguće konačne vrijednosti funkcije J, tj. vrijednosti funkcije J nakon što se algoritam K-means++ zaustavi. **Koliko iznosi maksimalna moguća konačna vrijednost funkcije J?**

A 14/3 B 10/3 C 4 D 11/2

17 (T) Kod vrednovanja algoritma strojnog učenja važno je da vrednujemo model optimalne složenosti. Zbog toga označeni skup podataka dijelimo na tri disjunktna podskupa: skup za učenje, skup za provjeru i ispitni skup. Za odabir modela (tj. optimizaciju hiperparametara) koristimo skup za učenje i skup za provjeru, dok točnost optimalnog modela ispitujemo na ispitnome skupu. Zamislite da smo zaboravili upotrijebiti skup za provjeru: umjesto toga hiperparametre modela optimirali smo na skupu za učenje, a model smo potom ispitali na ispitnom skupu. **U odnosu na to da smo pravilno proveli vrednovanje algoritma, što možemo očekivati u ovom slučaju?**

A Model će biti prenaučen, ali imat će veću točnost na ispitnome skupu

B Model će biti optimalne složenosti i imat će veću točnost na ispitnome skupu

C Model će biti prenaučen i imat će manju točnost na ispitnome skupu

D Model će biti podnaučen, ali imat će veću točnost na ispitnome skupu

18 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

$$\begin{pmatrix} 1.0 & 0.2 & 0.8 & 0.2 & 0.4 \\ 0.2 & 1.0 & 0.3 & 0.3 & 0.7 \\ 0.8 & 0.3 & 1.0 & 0.6 & 0.5 \\ 0.2 & 0.3 & 0.6 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

Provedite grupiranje algoritmom HAC s potpunim povezivanjem. Pritom u svakoj iteraciji bilježite na kojoj razini sličnosti se odvija stapanje dviju grupa. **Koliko iznosi zbroj po svim razinama sličnosti na kojima se odvija stapanje grupa?**

A 1.8 B 1.9 C 2.4 D 2.0

19 (T) Algoritam K-sredina iterativno smanjuje vrijednost funkcije pogreške J. Neovisno o broju grupa K, algoritam uvijek konvergira (završava s izvođenjem). Međutim, konačna vrijednost funkcije pogreške J može se razlikovati između dva pokretanja algoritma s istim brojem grupa K. **Zašto?**

A Zbog razlike u početnim središtima $\boldsymbol{\mu}_k$

B Zbog loše odabrane vrijednosti za broj grupa K

C Zbog nelinearnosti granice između K grupa

D Zbog izotropnosti kovarijacijskih matrica Σ_k

20 (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 0.58 C 0.53 D 0.62

21 (N) Vrednujemo višeklasni klasifikator koji koristimo kao referentni model za vrednovanje drugih klasifikatora. Klasifikator nasumično dodjeljuje oznake za tri klase ($K = 3$), i to prema distribuciji oznaka na skupu za učenje. Skup za učenje sadrži 100 primjera sa 35, 60 i 5 primjera u prvoj, drugoj odnosno trećoj klasi. S druge strane, ispitni skup sadrži 50 primjera, od kojih je po 20 u prvoj i drugoj klasi, a 10 u trećoj klasi. **Koliko iznosi očekivana makro-F1 vrijednost nasumičnog klasifikatora na ispitnome skupu?**

A 0.2085 B 0.2156 C 0.2989 D 0.3111

22 (P) Raspolažemo sa 750 označenih primjera. Na tom skupu treniramo i vrednujemo algoritam SVM, optimizirajući hiperparametre C i γ. Za vrednovanje koristimo ugniježđenu unakrsnu provjeru sa 5 preklopa u vanjskoj petlji i 5 preklopa u unutarnjoj petlji. To znači da za svaku kombinaciju vrijednosti hiperparametara C i γ treniramo pet modela. Ti modeli trenirani su na skupovima za učenje koji nisu disjunktni: svaki par treniranih modela dijele određeni broj primjera za učenje. Izračunajte koliko primjera za učenje dijele svaki par modela koje treniramo u unutarnjoj petlji ugniježđene unakrsne provjere. **Za koliko bi taj broj narastao kada bismo broj preklopa u unutarnjoj petlji povećali na 10?**

A 60 B 144 C 75 D 120

Završni ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% 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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 4: Procjena parametara i Bayesov klasifikator (8 pitanja)

1 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{da}, \text{zimski}, \text{preddipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.741 B 0.488 C 0.588 D 0.322

2 (T) Polunaivan Bayesov klasifikator može ostvariti veću klasifikacijsku točnost od naivnog Bayesovog klasifikatora u situacijama kada između značajki postoji zavisnost. Jedan način određivanja zavisnosti između značajki temelji se na izračunu Kullback-Leiblerove divergencije (KL-divergencije) između distribucija dviju varijabli. **Koji je princip uporabe KL-divergencije u izgradnji polunaivnog Bayesovog klasifikatora?**

A Što je manja KL-divergencija između $P(x_1)$ i $P(x_2)$, to su varijable x_1 i x_2 više zavisne, pa je to veća korist od njihovog združivanja u jedan faktor

B Što je veća KL-divergencija između $P(x_1)P(x_2)$ i $P(x_1, x_2)$, to su varijable x_1 i x_2 manje nezavisne, pa je to veća korist od njihovog združivanja u jedan faktor

C Ako je KL-divergencija između $P(x_1)P(x_2)$ i $P(x_1, x_2)$ jednaka nuli, onda su varijable linearno zavisne i treba ih odvojiti u zasebne faktore

D Ako je KL-divergencija između $P(x_1|x_2)$ i $P(x_2|x_1)$ vrlo mala, onda su varijable x_1 i x_2 vrlo zavisne i treba ih odvojiti u zasebne faktore

3 (T) Bayesov klasifikator temelji se na hipotezi *maximum aposteriori* (MAP). Za treniranje Bayesovog klasifikatora možemo koristiti procjenitelj MLE. **Koja je razlika između hipoteze MAP kod Bayesovog klasifikatora i procjenitelja MLE?**

A Procjenitelj MLE primjer klasificira u klasu koja je za njega navjerojatnija, a hipoteza MAP tu odluku kombinira s apriornim znanjem

B Procjenitelj MLE određuje apriornu vjerojatnost klase, a hipoteza MAP maksimizira aposterionu vjerojatnost parametara

C Hipoteza MAP maksimizira aposterionu vjerojatnost parametara, a procjenitelj MLE određuje aposteriornu vjerojatnost klase za zadani primjer

D Hipoteza MAP određuje koju klasu odabrati za primjer, a procjenitelj MLE određuje koju vrijednost odabrati za parametar

4 (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 skupu 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 -2.75 C -3.03 D -4.13

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 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 $-10 \leq x \leq 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

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

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

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

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

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

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

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

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

7 (P) Uz određene restrikcije na parametre Gaussovog Bayesovog klasifikatora, Gaussov Bayesov klasifikator i logistička regresija su generativno-diskriminativni par modela. U pravilu, logistička regresija imat će manje parametara od njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. Razmotrite slučaj binarne klasifikacije u ulaznome prostoru dimenzije $n = 300$ pomoću modela logističke regresije i njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. **Koliko će model Gaussovog Bayesovog klasifikatora imati više parametara od modela logističke regresije?**

A 20300 B 40400 C 45450 D 90600

8 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{-10, 5, 6, 20\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

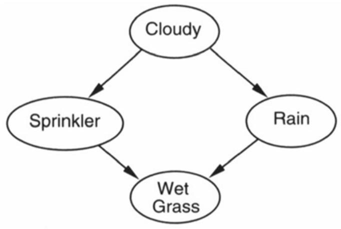
$$p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{ -\frac{(x-\mu)^2}{2\sigma^2}\Big\}$$

Za procjenu parametra μ koristimo procjenitelj $\hat{\mu}_{\mathrm{MLE}}$, a za procjenu parametra σ^2 procjenitelj $\hat{\sigma}^2_{\mathrm{UB}}$, koji je nepristran procjenitelj izveden odgovarajućom korekcijom iz procjenitelja $\hat{\sigma}^2_{\mathrm{MLE}}$. Izračunajte log-izglednost procijenjenih parametara na skupu $\mathcal{D}$, tj. $\ln \mathcal{L}(\hat{\mu}_{\mathrm{MLE}}, \hat{\sigma}^2_{\mathrm{UB}}|\mathcal{D})$. **Koliko iznosi ta log-izglednost?**

A -15.200 B -10.832 C -13.163 D -12.873

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

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



C	$P(S=1\|C)$
0	0.5
1	0.4

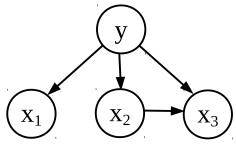
C	$P(R=1\|C)$
0	0.2
1	0.8

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

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

A 0.069 B 0.223 C 0.309 D 0.144

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



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

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

11 (T) Probablistički grafički modeli na sažet način prikazuju zajedničku distribuciju n slučajnih varijabli. Ta je distribucija prikazana usmjerenim grafom. Neka graf sadrži čvor označen sa x. **Čemu odgovara taj čvor?**

A Distribuciji varijable x uvjetovane vrijednostima varijabli roditeljskih čvorova od x

B Distribuciji varijable x uz opažene sve varijable koje varijabli x prethode u topološkom uređaju

C Marginalnoj vjerojatnosti $P(x)$ ili marginalnoj gustoći vjerojatnosti $p(x)$

D Zajedničkoj vjerojatnosti varijable x i varijabli roditeljskih čvorova od x

12 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

Primjenom algoritma d-odvajanja ispitujemo zavisnosti između parova varijabli. **Koje od sljedećih tvrdnji o nezavisnosti vrijede u ovoj Bayesovoj mreži?**

A $v\perp z|w$ B $x\perp z|w$ C $w\perp y|z$ D $w\perp y|x$

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 $P(z=1) = 0.1$, $P(w=1) = 0.2$, $P(x=1|y=0) = 0.2$ 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 konačni 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 596 D 348

Cjelina 6: Grupiranje i vrednovanje modela (9 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
$y^{(i)}_{pred}$	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 0.58 B 0.62 C 0.53 D 0.49

15 (P) Podatke u dvodimenzijskome ulaznom prostoru ($n = 2$) grupiramo u dvije grupe ($K = 2$) algoritmom Gaussove mješavine (GMM). Skup podataka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{8} = \big\{(-2,-2),(-2,-1),(-1,-2),(-1,-1),(1,2),(3,1),(3,3),(5,2)\big\}$$

Razmatramo dva modela: $\mathcal{H}_1$ i $\mathcal{H}_2$. Oba modela koriste punu kovarijacijsku matricu, međutim kod modela $\mathcal{H}_1$ kovarijacijska matrica nije dijeljena, dok je kod modela $\mathcal{H}_2$ kovarijacijska matrica dijeljena. Za svaki model razmotrite ono grupiranje koje maksimizira log-izglednost na skupu $\mathcal{D}$. Neka je $p_j(\mathbf{x}^{(i)}|y=k)$ izglednost grupe $y = k$ za primjer $\mathbf{x}^{(i)}$ prema modelu $\mathcal{H}_j$. Oznake grupa su $y = 1$ i $y = 2$, te pretpostavite da je grupiranje takvo da vrijedi $\|\boldsymbol{\mu}_1\| > \|\boldsymbol{\mu}_2\|$. **Što od sljedećega vrijedi?**

A $p_2(\mathbf{x}^{(4)}|y=2) > p_1(\mathbf{x}^{(3)}|y=2)$ C $p_2(\mathbf{x}^{(8)}|y=1) < p_1(\mathbf{x}^{(5)}|y=1)$

B $p_2(\mathbf{x}^{(6)}|y=1) < p_1(\mathbf{x}^{(7)}|y=1)$ D $p_2(\mathbf{x}^{(1)}|y=2) > p_1(\mathbf{x}^{(2)}|y=2)$

16 (T) Za procjenu parametara modela GMM (središta grupa $\boldsymbol{\mu}_k$, kovarijacijske matrice Σ_k i koeficijenti mješavine π_k) tipično se koristi algoritam maksimizacije očekivanja (EM-algoritam). Prije pokretanja EM-algoritma, parametre je potrebno inicijalizirati. U praksi se za inicijalizaciju parametara $\boldsymbol{\mu}_k$ koristi algoritam K-sredina. **Koji je očekivani efekt uporabe algoritma K-sredina za inicijalizaciju EM-algoritma?**

A Manja log-izglednost početnih parametara, ali veća log-izglednost konačnih parametara

B Veća potpuna log-izglednost konačnih parametara i veće vrijednosti gustoća komponenti

C Veća log-izglednost početnih i konačnih parametara te manje iteracija do konvergencije

D Manja log-izglednost početnih i konačnih parametara te optimalan broj grupa K

17 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

$$\begin{pmatrix} 1.0 & 0.2 & 0.8 & 0.2 & 0.4 \\ 0.2 & 1.0 & 0.3 & 0.3 & 0.7 \\ 0.8 & 0.3 & 1.0 & 0.6 & 0.5 \\ 0.2 & 0.3 & 0.6 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

Provedite grupiranje algoritmom HAC s potpunim povezivanjem. Pritom u svakoj iteraciji bilježite na kojoj razini sličnosti se odvija stapanje dviju grupa. **Koliko iznosi zbroj po svim razinama sličnosti na kojima se odvija stapanje grupa?**

A 1.9 B 1.8 C 2.4 D 2.0

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

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_i = \big\{(1,1),(1,3),(1,4),(3,4)\big\}$$

Primjere grupiramo u $K = 2$ grupe algoritmom K-means++. Konačna vrijednost funkcije pogreške J bit će različita za različita pokretanja algoritma. Razmotrite sve moguće konačne vrijednosti funkcije J, tj. vrijednosti funkcije J nakon što se algoritam K-means++ zaustavi. **Koliko iznosi maksimalna moguća konačna vrijednost funkcije J?**

A 4 B 10/3 C 14/3 D 11/2

19 (T) Algoritam K-sredina iterativno smanjuje vrijednost funkcije pogreške J. Neovisno o broju grupa K, algoritam uvijek konvergira (završava s izvođenjem). Međutim, konačna vrijednost funkcije pogreške J može se razlikovati između dva pokretanja algoritma s istim brojem grupa K. **Zašto?**

A Zbog nelinearnosti granice između K grupa

B Zbog razlike u početnim središtima $\boldsymbol{\mu}_k$

C Zbog loše odabrane vrijednosti za broj grupa K

D Zbog izotropnosti kovarijacijskih matrica Σ_k

20 (P) Raspolažemo sa 750 označenih primjera. Na tom skupu treniramo i vrednujemo algoritam SVM, optimizirajući hiperparametre C i γ. Za vrednovanje koristimo ugniježđenu unakrsnu provjeru sa 5 preklopa u vanjskoj petlji i 5 preklopa u unutarnjoj petlji. To znači da za svaku kombinaciju vrijednosti hiperparametara C i γ treniramo pet modela. Ti modeli trenirani su na skupovima za učenje koji nisu disjunktni: svaki par treniranih modela dijele određeni broj primjera za učenje. Izračunajte koliko primjera za učenje dijele svaki par modela koje treniramo u unutarnjoj petlji ugniježđene unakrsne provjere. **Za koliko bi taj broj narastao kada bismo broj preklopa u unutarnjoj petlji povećali na 10?**

A 75 B 120 C 60 D 144

21 (N) Vrednujemo višeklasni klasifikator koji koristimo kao referentni model za vrednovanje drugih klasifikatora. Klasifikator nasumično dodjeljuje oznake za tri klase ($K = 3$), i to prema distribuciji oznaka na skupu za učenje. Skup za učenje sadrži 100 primjera sa 35, 60 i 5 primjera u prvoj, drugoj odnosno trećoj klasi. S druge strane, ispitni skup sadrži 50 primjera, od kojih je po 20 u prvoj i drugoj klasi, a 10 u trećoj klasi. **Koliko iznosi očekivana makro-F1 vrijednost nasumičnog klasifikatora na ispitnome skupu?**

A 0.3111 B 0.2085 C 0.2989 D 0.2156

22 (T) Kod vrednovanja algoritma strojnog učenja važno je da vrednujemo model optimalne složenosti. Zbog toga označeni skup podataka dijelimo na tri disjunktna podskupa: skup za učenje, skup za provjeru i ispitni skup. Za odabir modela (tj. optimizaciju hiperparametara) koristimo skup za učenje i skup za provjeru, dok točnost optimalnog modela ispitujemo na ispitnome skupu. Zamislite da smo zaboravili upotrijebiti skup za provjeru: umjesto toga hiperparametre modela optimirali smo na skupu za učenje, a model smo potom ispitali na ispitnom skupu. **U odnosu na to da smo pravilno proveli vrednovanje algoritma, što možemo očekivati u ovom slučaju?**

A Model će biti podnaučen, ali imat će veću točnost na ispitnome skupu

B Model će biti prenaučen, ali imat će veću točnost na ispitnome skupu

C Model će biti prenaučen i imat će manju točnost na ispitnome skupu

D Model će biti optimalne složenosti i imat će veću točnost na ispitnome skupu

Završni ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% 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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 4: Procjena parametara i Bayesov klasifikator (8 pitanja)

1 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{10, 5, 6, 20\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Za procjenu parametra μ koristimo procjenitelj $\hat{\mu}_{\mathrm{MLE}}$, a za procjenu parametra σ^2 procjenitelj $\hat{\sigma}^2_{\mathrm{UB}}$, koji je nepristran procjenitelj izveden odgovarajućom korekcijom iz procjenitelja $\hat{\sigma}^2_{\mathrm{MLE}}$. Izračunajte log-izglednost procijenjenih parametara na skupu $\mathcal{D}$, tj. $\ln \mathcal{L}(\hat{\mu}_{\mathrm{MLE}}, \hat{\sigma}^2_{\mathrm{UB}}|\mathcal{D})$. **Koliko iznosi ta log-izglednost?**

A -13.163 B -12.873 C -10.832 D -15.200

2 (T) Polunaivan Bayesov klasifikator može ostvariti veću klasifikacijsku točnost od naivnog Bayesovog klasifikatora u situacijama kada između značajki postoji zavisnost. Jedan način određivanja zavisnosti između značajki temelji se na izračunu Kullback-Leiblerove divergencije (KL-divergencije) između distribucija dviju varijabli. **Koji je princip uporabe KL-divergencije u izgradnji polunaivnog Bayesovog klasifikatora?**

A Ako je KL-divergencija između $P(x_1)P(x_2)$ i $P(x_1, x_2)$ jednaka nuli, onda su varijable linearno zavisne i treba ih odvojiti u zasebne faktore

B Što je veća KL-divergencija između $P(x_1)P(x_2)$ i $P(x_1, x_2)$, to su varijable x_1 i x_2 manje nezavisne, pa je to veća korist od njihovog združivanja u jedan faktor

C Ako je KL-divergencija između $P(x_1|x_2)$ i $P(x_2|x_1)$ vrlo velika, onda su varijable x_1 i x_2 vrlo zavisne i treba ih združiti u jedan faktor

D Što je manja KL-divergencija između $P(x_1)$ i $P(x_2)$, to su varijable x_1 i x_2 više zavisne, pa je to veća korist od njihovog združivanja u jedan faktor

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

$\mathcal{H}_1$: Značajke imaju različite varijance, ali iste za sve klase, te nisu korelirane
$\mathcal{H}_2$: Značajke nisu korelirane, imaju jednaku varijancu unutar svake klase, no različitu za svaku klasu
$\mathcal{H}_3$: Između značajki postoje korelacije, ali se one ne razlikuju između klasa

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

A $\mathcal{H}_1 > \mathcal{H}_2 > \mathcal{H}_3$, $\mathcal{H}_1 \supset \mathcal{H}_2$
C $\mathcal{H}_1 > \mathcal{H}_3 > \mathcal{H}_2$, $\mathcal{H}_1 \supset \mathcal{H}_2 \supset \mathcal{H}_3$
B $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$, $\mathcal{H}_3 \supset \mathcal{H}_1$
D $\mathcal{H}_3 > \mathcal{H}_2 > \mathcal{H}_1$, $\mathcal{H}_3 \supset \mathcal{H}_1 \supset \mathcal{H}_2$

4 (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 $-10 \leq x \leq 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

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

5 (T) Bayesov klasifikator temelji se na hipotezi *maximum aposteriori* (MAP). Za treniranje Bayesovog klasifikatora možemo koristiti procjenitelj MLE. **Koja je razlika između hipoteze MAP kod Bayesovog klasifikatora i procjenitelja MLE?**

A Hipoteza MAP maksimizira aposterionu vjerojatnost parametara, a procjenitelj MLE određuje aposteriornu vjerojatnost klase za zadani primjer

B Procjenitelj MLE određuje apriornu vjerojatnost klase, a hipoteza MAP maksimizira aposterionu vjerojatnost parametara

C Procjenitelj MLE primjer klasificira u klasu koja je za njega navjerojatnija, a hipoteza MAP tu odluku kombinira s apriornim znanjem

D Hipoteza MAP određuje koju klasu odabrati za primjer, a procjenitelj MLE određuje koju vrijednost odabrati za parametar

6 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{da}, \text{zimski}, \text{dipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.488 B 0.322 C 0.741 D 0.588

7 (P) Uz određene restrikcije na parametre Gaussovog Bayesovog klasifikatora, Gaussov Bayesov klasifikator i logistička regresija su generativno-diskriminativni par modela. U pravilu, logistička regresija imat će manje parametara od njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. Razmotrite slučaj binarne klasifikacije u ulaznome prostoru dimenzije $n = 200$ pomoću modela logističke regresije i njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. **Koliko će model Gaussovog Bayesovog klasifikatora imati više parametara od modela logističke regresije?**

A 90600 B 40400 C 20300 D 45450

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 skupu podataka od $N = 7$ primjera:

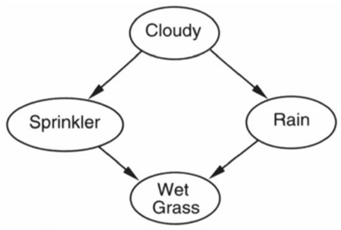
$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((-4, -2), 0), ((0, 0), 0), ((4, 2), 0), ((3, -1), 1), ((4, -1), 1), ((4, 1), 1), ((5, 1), 1)\}$$

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 -2.75 B -3.03 C -4.13 D -3.84

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

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



C	$P(S=1\|C)$
0	0.5
1	0.4

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

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

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

A 0.144   B 0.223   C 0.069   D 0.309

10 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

Primjenom algoritma d-odvajanja ispitujemo zavisnosti između parova varijabli. **Koje od sljedećih tvrdnji o nezavisnosti vrijede u ovoj Bayesovoj mreži?**

A $v\perp z|y$ B $w\perp y|z$ C $w\perp y|x$ D $v\perp y|z$

11 (T) Probablistički grafički modeli na sažet način prikazuju zajedničku distribuciju n slučajnih varijabli. Ta je distribucija prikazana usmjerenim grafom. Neka graf sadrži čvor označen sa x. **Čemu odgovara taj čvor?**

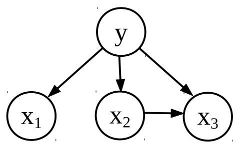
A Distribuciji varijable x uz opažene sve varijable koje varijabli x prethode u topološkom uređaju

B Distribuciji varijable x uvjetovane vrijednostima varijabli roditeljskih čvorova od x

C Marginalnoj vjerojatnosti $P(x)$ ili marginalnoj gustoći vjerojatnosti $p(x)$

D Zajedničkoj vjerojatnosti varijable x i varijabli roditeljskih čvorova od x

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



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

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

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 $P(z=1)=0.1$, $P(w=1)=0.2$, $P(x=1|y=0)=0.2$ 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 konačni uzorak manji od N. **U uzorku neodbačenih vektora, koliko je očekivano više vektora sa $w=0$ od vektora sa $w=1$?**

A 596   B 944   C 472   D 348

Cjelina 6: Grupiranje i vrednovanje modela (9 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	1
$y_{true}^{(i)}$	1	0	0	2	1	0	2	1	0	1

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.49 B 0.58 C 0.62 D 0.53

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

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_i = \big\{(1,1),(1,3),(1,4),(3,4)\big\}$$

Primjere grupiramo u $K = 2$ grupe algoritmom K-means++. Konačna vrijednost funkcije pogreške J bit će različita za različita pokretanja algoritma. Razmotrite sve moguće konačne vrijednosti funkcije J, tj. vrijednosti funkcije J nakon što se algoritam K-means++ zaustavi. **Koliko iznosi maksimalna moguća konačna vrijednost funkcije J?**

A 14/3 B 4 C 11/2 D 10/3

16 (T) Kod vrednovanja algoritma strojnog učenja važno je da vrednujemo model optimalne složenosti. Zbog toga označeni skup podataka dijelimo na tri disjunktna podskupa: skup za učenje, skup za provjeru i ispitni skup. Za odabir modela (tj. optimizaciju hiperparametara) koristimo skup za učenje i skup za provjeru, dok točnost optimalnog modela ispitujemo na ispitnome skupu. Zamislite da smo zaboravili upotrijebiti skup za provjeru: umjesto toga hiperparametre modela optimirali smo na skupu za učenje, a model smo potom ispitali na ispitnom skupu. **U odnosu na to da smo pravilno proveli vrednovanje algoritma, što možemo očekivati u ovom slučaju?**

A Model će biti prenaučen, ali imat će veću točnost na ispitnome skupu

B Model će biti prenaučen i imat će manju točnost na ispitnome skupu

C Model će biti podnaučen, ali imat će veću točnost na ispitnome skupu

D Model će biti optimalne složenosti i imat će veću točnost na ispitnome skupu

17 (T) Za procjenu parametara modela GMM (središta grupa $\boldsymbol{\mu}_k$, kovarijacijske matrice Σ_k i koeficijenti mješavine π_k) tipično se koristi algoritam maksimizacije očekivanja (EM-algoritam). Prije pokretanja EM-algoritma, parametre je potrebno inicijalizirati. U praksi se za inicijalizaciju parametara $\boldsymbol{\mu}_k$ koristi algoritam K-sredina. **Koji je očekivani efekt uporabe algoritma K-sredina za inicijalizaciju EM-algoritma?**

A Veća potpuna log-izglednost konačnih parametara i veće vrijednosti gustoća komponenti

B Manja log-izglednost početnih parametara, ali veća log-izglednost konačnih parametara

C Veća log-izglednost početnih i konačnih parametara te manje iteracija do konvergencije

D Manja log-izglednost početnih i konačnih parametara te optimalan broj grupa K

18 (P) Raspolažemo sa 750 označenih primjera. Na tom skupu treniramo i vrednujemo algoritam SVM, optimizirajući hiperparametre C i γ. Za vrednovanje koristimo ugniježđenu unakrsnu provjeru sa 5 preklopa u vanjskoj petlji i 5 preklopa u unutarnjoj petlji. To znači da za svaku kombinaciju vrijednosti hiperparametara C i γ treniramo pet modela. Ti modeli trenirani su na skupovima za učenje koji nisu disjunktni: svaki par treniranih modela dijele određeni broj primjera za učenje. Izračunajte koliko primjera za učenje dijele svaki par modela koje treniramo u unutarnjoj petlji ugniježđene unakrsne provjere. **Za koliko bi taj broj narastao kada bismo broj preklopa u unutarnjoj petlji povećali na 10?**

A 60 B 120 C 75 D 144

19 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

$$\begin{pmatrix} 1.0 & 0.2 & 0.8 & 0.2 & 0.4 \\ 0.2 & 1.0 & 0.3 & 0.3 & 0.7 \\ 0.8 & 0.3 & 1.0 & 0.6 & 0.5 \\ 0.2 & 0.3 & 0.6 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

Provedite grupiranje algoritmom HAC s potpunim povezivanjem. Pritom u svakoj iteraciji bilježite na kojoj razini sličnosti se odvija stapanje dviju grupa. **Koliko iznosi zbroj po svim razinama sličnosti na kojima se odvija stapanje grupa?**

A 2.0   B 1.8   C 2.4   D 1.9

20 (T) Algoritam K-sredina iterativno smanjuje vrijednost funkcije pogreške J. Neovisno o broju grupa K, algoritam uvijek konvergira (završava s izvođenjem). Međutim, konačna vrijednost funkcije pogreške J može se razlikovati između dva pokretanja algoritma s istim brojem grupa K. **Zašto?**

A Zbog nelinearnosti granice između K grupa

B Zbog izotropnosti kovarijacijskih matrica Σ_k

C Zbog razlike u početnim središtima $\boldsymbol{\mu}_k$

D Zbog loše odabrane vrijednosti za broj grupa K

21 (P) Podatke u dvodimenzijskome ulaznom prostoru ($n = 2$) grupiramo u dvije grupe ($K = 2$) algoritmom Gaussove mješavine (GMM). Skup podataka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{8} = \big\{(-2,-2), (-2,-1), (-1,-2), (-1,-1), (1,2), (3,1), (3,3), (5,2)\big\}$$

Razmatramo dva modela: $\mathcal{H}_1$ i $\mathcal{H}_2$. Oba modela koriste punu kovarijacijsku matricu, međutim kod modela $\mathcal{H}_1$ kovarijacijska matrica nije dijeljena, dok je kod modela $\mathcal{H}_2$ kovarijacijska matrica dijeljena. Za svaki model razmotrite ono grupiranje koje maksimizira log-izglednost na skupu $\mathcal{D}$. Neka je $p_j(\mathbf{x}^{(i)}|y = k)$ izglednost grupe $y = k$ za primjer $\mathbf{x}^{(i)}$ prema modelu $\mathcal{H}_j$. Oznake grupa su $y = 1$ i $y = 2$, te pretpostavite da je grupiranje takvo da vrijedi $\|\boldsymbol{\mu}_1\| > \|\boldsymbol{\mu}_2\|$. **Što od sljedećega vrijedi?**

A $p_2(\mathbf{x}^{(1)}|y = 2) > p_1(\mathbf{x}^{(2)}|y = 2)$ C $p_2(\mathbf{x}^{(8)}|y = 1) < p_1(\mathbf{x}^{(5)}|y = 1)$

B $p_2(\mathbf{x}^{(4)}|y = 2) > p_1(\mathbf{x}^{(3)}|y = 2)$ D $p_2(\mathbf{x}^{(6)}|y = 1) < p_1(\mathbf{x}^{(7)}|y = 1)$

22 (N) Vrednujemo višeklasni klasifikator koji koristimo kao referentni model za vrednovanje drugih klasifikatora. Klasifikator nasumično dodjeljuje oznake za tri klase ($K = 3$), i to prema distribuciji oznaka na skupu za učenje. Skup za učenje sadrži 100 primjera sa 60, 35 i 5 primjera u prvoj, drugoj odnosno trećoj klasi. S druge strane, ispitni skup sadrži 50 primjera, od kojih je po 10 u prvoj i drugoj klasi, a 30 u trećoj klasi. **Koliko iznosi očekivana makro-F1 vrijednost nasumičnog klasifikatora na ispitnome skupu?**

A 0.3111   B 0.2989   C 0.2085   D 0.2156

Završni ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% 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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 4: Procjena parametara i Bayesov klasifikator (8 pitanja)

1 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{10, 5, 6, 20\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Za procjenu parametra μ koristimo procjenitelj $\hat{\mu}_{\text{MLE}}$, a za procjenu parametra σ^2 procjenitelj $\hat{\sigma}^2_{\text{UB}}$, koji je nepristran procjenitelj izveden odgovarajućom korekcijom iz procjenitelja $\hat{\sigma}^2_{\text{MLE}}$. Izračunajte log-izglednost procijenjenih parametara na skupu $\mathcal{D}$, tj. $\ln \mathcal{L}(\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{UB}}|\mathcal{D})$. **Koliko iznosi ta log-izglednost?**

[A] -15.200 [B] -13.163 [C] -12.873 [D] -10.832

2 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{da}, \text{zimski}, \text{preddipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

[A] 0.322 [B] 0.741 [C] 0.588 [D] 0.488

3 (T) Bayesov klasifikator temelji se na hipotezi *maximum aposteriori* (MAP). Za treniranje Bayesovog klasifikatora možemo koristiti procjenitelj MLE. **Koja je razlika između hipoteze MAP kod Bayesovog klasifikatora i procjenitelja MLE?**

[A] Hipoteza MAP maksimizira aposterionu vjerojatnost parametara, a procjenitelj MLE određuje aposteriornu vjerojatnost klase za zadani primjer

[B] Procjenitelj MLE primjer klasificira u klasu koja je za njega navjerojatnija, a hipoteza MAP tu odluku kombinira s apriornim znanjem

[C] Hipoteza MAP određuje koju klasu odabrati za primjer, a procjenitelj MLE određuje koju vrijednost odabrati za parametar

[D] Procjenitelj MLE određuje apriornu vjerojatnost klase, a hipoteza MAP maksimizira aposterionu vjerojatnost parametara

4 (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 skupu 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] -3.03 [B] -3.84 [C] -2.75 [D] -4.13

5 (P) Uz određene restrikcije na parametre Gaussovog Bayesovog klasifikatora, Gaussov Bayesov klasifikator i logistička regresija su generativno-diskriminativni par modela. U pravilu, logistička regresija imat će manje parametara od njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. Razmotrite slučaj binarne klasifikacije u ulaznome prostoru dimenzije $n = 200$ pomoću modela logističke regresije i njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. **Koliko će model Gaussovog Bayesovog klasifikatora imati više parametara od modela logističke regresije?**

[A] 20300 [B] 40400 [C] 90600 [D] 45450

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 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 $-10 \leq x \leq 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

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

7 (T) Polunaivan Bayesov klasifikator može ostvariti veću klasifikacijsku točnost od naivnog Bayesovog klasifikatora u situacijama kada između značajki postoji zavisnost. Jedan način određivanja zavisnosti između značajki temelji se na izračunu Kullback-Leiblerove divergencije (KL-divergencije) između distribucija dviju varijabli. **Koji je princip uporabe KL-divergencije u izgradnji polunaivnog Bayesovog klasifikatora?**

[A] Što je veća KL-divergencija između $P(x_1)P(x_2)$ i $P(x_1, x_2)$, to su varijable x_1 i x_2 manje nezavisne, pa je to veća korist od njihovog združivanja u jedan faktor

[B] Ako je KL-divergencija između $P(x_1|x_2)$ i $P(x_2|x_1)$ vrlo mala, onda su varijable x_1 i x_2 vrlo zavisne i treba ih odvojiti u zasebne faktore

[C] Što je manja KL-divergencija između $P(x_1)$ i $P(x_2)$, to su varijable x_1 i x_2 više zavisne, pa je to veća korist od njihovog združivanja u jedan faktor

[D] Ako je KL-divergencija između $P(x_1|x_2)$ i $P(x_2|x_1)$ vrlo velika, onda su varijable x_1 i x_2 vrlo zavisne i treba ih združiti u jedan faktor

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

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

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

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

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

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

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

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

9 (T) Probablistički grafički modeli na sažet način prikazuju zajedničku distribuciju n slučajnih varijabli. Ta je distribucija prikazana usmjerenim grafom. Neka graf sadrži čvor označen sa x. **Čemu odgovara taj čvor?**

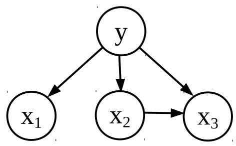
A Marginalnoj vjerojatnosti $P(x)$ ili marginalnoj gustoći vjerojatnosti $p(x)$

B Distribuciji varijable x uvjetovane vrijednostima varijabli roditeljskih čvorova od x

C Distribuciji varijable x uz opažene sve varijable koje varijabli x prethode u topološkom uređaju

D Zajedničkoj vjerojatnosti varijable x i varijabli roditeljskih čvorova od x

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



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

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

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 $P(z = 1) = 0.1$, $P(w = 1) = 0.1$, $P(x = 1|y = 0) = 0.2$ 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 konačni 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 596 D 348

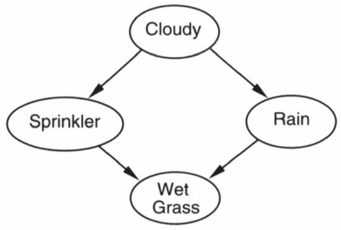
12 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

Primjenom algoritma d-odvajanja ispitujemo zavisnosti između parova varijabli. **Koje od sljedećih tvrdnji o nezavisnosti vrijede u ovoj Bayesovoj mreži?**

A $w \perp y | z$ B $v \perp z | y$ C $v \perp y | z$ D $v \perp y | x$

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



C	$P(S=1\|C)$
0	0.5
1	0.2

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

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

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

[A] 0.144 [B] 0.309 [C] 0.223 [D] 0.069

Cjelina 6: Grupiranje i vrednovanje modela (9 pitanja)

14 (N) Vrednujemo višeklasni klasifikator koji koristimo kao referentni model za vrednovanje drugih klasifikatora. Klasifikator nasumično dodjeljuje oznake za tri klase ($K = 3$), i to prema distribuciji oznaka na skupu za učenje. Skup za učenje sadrži 100 primjera sa 35, 60 i 5 primjera u prvoj, drugoj odnosno trećoj klasi. S druge strane, ispitni skup sadrži 50 primjera, od kojih je po 20 u prvoj i drugoj klasi, a 10 u trećoj klasi. **Koliko iznosi očekivana makro-F1 vrijednost nasumičnog klasifikatora na ispitnome skupu?**

[A] 0.2989 [B] 0.3111 [C] 0.2085 [D] 0.2156

15 (T) Kod vrednovanja algoritma strojnog učenja važno je da vrednujemo model optimalne složenosti. Zbog toga označeni skup podataka dijelimo na tri disjunktna podskupa: skup za učenje, skup za provjeru i ispitni skup. Za odabir modela (tj. optimizaciju hiperparametara) koristimo skup za učenje i skup za provjeru, dok točnost optimalnog modela ispitujemo na ispitnome skupu. Zamislite da smo zaboravili upotrijebiti skup za provjeru: umjesto toga hiperparametre modela optimirali smo na skupu za učenje, a model smo potom ispitali na ispitnom skupu. **U odnosu na to da smo pravilno proveli vrednovanje algoritma, što možemo očekivati u ovom slučaju?**

[A] Model će biti prenaučen i imat će manju točnost na ispitnome skupu

[B] Model će biti podnaučen, ali imat će veću točnost na ispitnome skupu

[C] Model će biti prenaučen, ali imat će veću točnost na ispitnome skupu

[D] Model će biti optimalne složenosti i imat će veću točnost na ispitnome skupu

16 (P) Podatke u dvodimenzijskome ulaznom prostoru ($n = 2$) grupiramo u dvije grupe ($K = 2$) algoritmom Gaussove mješavine (GMM). Skup podataka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{8} = \big\{(-2,-2),(-2,-1),(-1,-2),(-1,-1),(1,2),(3,1),(3,3),(5,2)\big\}$$

Razmatramo dva modela: $\mathcal{H}_1$ i $\mathcal{H}_2$. Oba modela koriste punu kovarijacijsku matricu, međutim kod modela $\mathcal{H}_1$ kovarijacijska matrica nije dijeljena, dok je kod modela $\mathcal{H}_2$ kovarijacijska matrica dijeljena. Za svaki model razmotrite ono grupiranje koje maksimizira log-izglednost na skupu $\mathcal{D}$. Neka je $p_j(\mathbf{x}^{(i)}|y=k)$ izglednost grupe $y = k$ za primjer $\mathbf{x}^{(i)}$ prema modelu $\mathcal{H}_j$. Oznake grupa su $y = 1$ i $y = 2$, te pretpostavite da je grupiranje takvo da vrijedi $\|\boldsymbol{\mu}_1\| > \|\boldsymbol{\mu}_2\|$. **Što od sljedećega vrijedi?**

[A] $p_2(\mathbf{x}^{(8)}|y=1) < p_1(\mathbf{x}^{(5)}|y=1)$ [C] $p_2(\mathbf{x}^{(1)}|y=2) > p_1(\mathbf{x}^{(2)}|y=2)$

[B] $p_2(\mathbf{x}^{(6)}|y=1) < p_1(\mathbf{x}^{(7)}|y=1)$ [D] $p_2(\mathbf{x}^{(4)}|y=2) > p_1(\mathbf{x}^{(3)}|y=2)$

17 (T) Algoritam K-sredina iterativno smanjuje vrijednost funkcije pogreške J. Neovisno o broju grupa K, algoritam uvijek konvergira (završava s izvođenjem). Međutim, konačna vrijednost funkcije pogreške J može se razlikovati između dva pokretanja algoritma s istim brojem grupa K. **Zašto?**

[A] Zbog izotropnosti kovarijacijskih matrica Σ_k

[B] Zbog loše odabrane vrijednosti za broj grupa K

[C] Zbog nelinearnosti granice između K grupa

[D] Zbog razlike u početnim središtima $\boldsymbol{\mu}_k$

18 (T) Za procjenu parametara modela GMM (središta grupa $\boldsymbol{\mu}_k$, kovarijacijske matrice Σ_k i koeficijenti mješavine π_k) tipično se koristi algoritam maksimizacije očekivanja (EM-algoritam). Prije pokretanja EM-algoritma, parametre je potrebno inicijalizirati. U praksi se za inicijalizaciju parametara $\boldsymbol{\mu}_k$ koristi algoritam K-sredina. **Koji je očekivani efekt uporabe algoritma K-sredina za inicijalizaciju EM-algoritma?**

A Manja log-izglednost početnih i konačnih parametara te optimalan broj grupa K

B Veća potpuna log-izglednost konačnih parametara i veće vrijednosti gustoća komponenti

C Manja log-izglednost početnih parametara, ali veća log-izglednost konačnih parametara

D Veća log-izglednost početnih i konačnih parametara te manje iteracija do konvergencije

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

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_i = \big\{(1,1),(1,3),(1,4),(3,4)\big\}$$

Primjere grupiramo u $K = 2$ grupe algoritmom K-means++. Konačna vrijednost funkcije pogreške J bit će različita za različita pokretanja algoritma. Razmotrite sve moguće konačne vrijednosti funkcije J, tj. vrijednosti funkcije J nakon što se algoritam K-means++ zaustavi. **Koliko iznosi maksimalna moguća konačna vrijednost funkcije J?**

A 4 B 14/3 C 10/3 D 11/2

20 (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 0.58 B 0.49 C 0.62 D 0.53

21 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

$$\begin{pmatrix} 1.0 & 0.2 & 0.6 & 0.7 & 0.4 \\ 0.2 & 1.0 & 0.3 & 0.3 & 0.7 \\ 0.6 & 0.3 & 1.0 & 0.9 & 0.5 \\ 0.7 & 0.3 & 0.9 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

Provedite grupiranje algoritmom HAC s potpunim povezivanjem. Pritom u svakoj iteraciji bilježite na kojoj razini sličnosti se odvija stapanje dviju grupa. **Koliko iznosi zbroj po svim razinama sličnosti na kojima se odvija stapanje grupa?**

A 1.9 B 2.4 C 1.8 D 2.0

22 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i vrednujemo algoritam SVM, optimizirajući hiperparametre C i γ. Za vrednovanje koristimo ugniježđenu unakrsnu provjeru sa 5 preklopa u vanjskoj petlji i 5 preklopa u unutarnjoj petlji. To znači da za svaku kombinaciju vrijednosti hiperparametara C i γ treniramo pet modela. Ti modeli trenirani su na skupovima za učenje koji nisu disjunktni: svaki par treniranih modela dijele određeni broj primjera za učenje. Izračunajte koliko primjera za učenje dijele svaki par modela koje treniramo u unutarnjoj petlji ugniježđene unakrsne provjere. **Za koliko bi taj broj narastao kada bismo broj preklopa u unutarnjoj petlji povećali na 10?**

A 192 B 160 C 100 D 80

Završni ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% 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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 4: Procjena parametara i Bayesov klasifikator (8 pitanja)

1 (P) Uz određene restrikcije na parametre Gaussovog Bayesovog klasifikatora, Gaussov Bayesov klasifikator i logistička regresija su generativno-diskriminativni par modela. U pravilu, logistička regresija imat će manje parametara od njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. Razmotrite slučaj binarne klasifikacije u ulaznome prostoru dimenzije $n = 300$ pomoću modela logističke regresije i njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. **Koliko će model Gaussovog Bayesovog klasifikatora imati više parametara od modela logističke regresije?**

A 40400 B 45450 C 20300 D 90600

2 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{10, 5, 6, 0\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Za procjenu parametra μ koristimo procjenitelj $\hat{\mu}_{\mathrm{MLE}}$, a za procjenu parametra σ^2 procjenitelj $\hat{\sigma}^2_{\mathrm{UB}}$, koji je nepristran procjenitelj izveden odgovarajućom korekcijom iz procjenitelja $\hat{\sigma}^2_{\mathrm{MLE}}$. Izračunajte log-izglednost procijenjenih parametara na skupu $\mathcal{D}$, tj. $\ln \mathcal{L}(\hat{\mu}_{\mathrm{MLE}}, \hat{\sigma}^2_{\mathrm{UB}}|\mathcal{D})$. **Koliko iznosi ta log-izglednost?**

A -13.163 B -10.832 C -12.873 D -15.200

3 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{ne}, \text{zimski}, \text{dipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.588 B 0.488 C 0.741 D 0.322

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

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

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

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

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

- A $\mathcal{H}_1 > \mathcal{H}_2 > \mathcal{H}_3$, $\mathcal{H}_1 \supset \mathcal{H}_2$
- C $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$, $\mathcal{H}_3 \supset \mathcal{H}_1$
- B $\mathcal{H}_3 > \mathcal{H}_2 > \mathcal{H}_1$, $\mathcal{H}_3 \supset \mathcal{H}_1 \supset \mathcal{H}_2$
- D $\mathcal{H}_1 > \mathcal{H}_3 > \mathcal{H}_2$, $\mathcal{H}_1 \supset \mathcal{H}_2 \supset \mathcal{H}_3$

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 $P(y=1) = P(y=2) = 1/5$ 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 $-10 \leq x \leq 10$. **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 $[-4-a, -4+b]$
- D $[-4-a, -4] \cup [5-b, 5]$

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:

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

Model treniramo na skupu 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 -4.13
- B -2.75
- C -3.03
- D -3.84

7 (T) Bayesov klasifikator temelji se na hipotezi *maximum aposteriori* (MAP). Za treniranje Bayesovog klasifikatora možemo koristiti procjenitelj MLE. **Koja je razlika između hipoteze MAP kod Bayesovog klasifikatora i procjenitelja MLE?**

- A Hipoteza MAP maksimizira aposterionu vjerojatnost parametara, a procjenitelj MLE određuje aposteriornu vjerojatnost klase za zadani primjer
- B Procjenitelj MLE primjer klasificira u klasu koja je za njega navjerojatnija, a hipoteza MAP tu odluku kombinira s apriornim znanjem
- C Procjenitelj MLE određuje apriornu vjerojatnost klase, a hipoteza MAP maksimizira aposterionu vjerojatnost parametara
- D Hipoteza MAP određuje koju klasu odabrati za primjer, a procjenitelj MLE određuje koju vrijednost odabrati za parametar

8 (T) Polunaivan Bayesov klasifikator može ostvariti veću klasifikacijsku točnost od naivnog Bayesovog klasifikatora u situacijama kada između značajki postoji zavisnost. Jedan način određivanja zavisnosti između značajki temelji se na izračunu Kullback-Leiblerove divergencije (KL-divergencije) između distribucija dviju varijabli. **Koji je princip uporabe KL-divergencije u izgradnji polunaivnog Bayesovog klasifikatora?**

- A Što je veća KL-divergencija između $P(x_1)P(x_2)$ i $P(x_1, x_2)$, to su varijable x_1 i x_2 manje nezavisne, pa je to veća korist od njihovog združivanja u jedan faktor
- B Ako je KL-divergencija između $P(x_1|x_2)$ i $P(x_2|x_1)$ vrlo velika, onda su varijable x_1 i x_2 vrlo zavisne i treba ih združiti u jedan faktor
- C Ako je KL-divergencija između $P(x_1|x_2)$ i $P(x_2|x_1)$ vrlo mala, onda su varijable x_1 i x_2 vrlo zavisne i treba ih odvojiti u zasebne faktore
- D Što je manja KL-divergencija između $P(x_1)$ i $P(x_2)$, to su varijable x_1 i x_2 više zavisne, pa je to veća korist od njihovog združivanja u jedan faktor

Cjelina 5: 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 $P(z=1) = 0.1$, $P(w=1) = 0.2$, $P(x=1|y=0) = 0.2$ 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 konačni 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 596 D 472

10 (T) Probablistički grafički modeli na sažet način prikazuju zajedničku distribuciju n slučajnih varijabli. Ta je distribucija prikazana usmjerenim grafom. Neka graf sadrži čvor označen sa x. **Čemu odgovara taj čvor?**

A Marginalnoj vjerojatnosti $P(x)$ ili marginalnoj gustoći vjerojatnosti $p(x)$

B Distribuciji varijable x uvjetovane vrijednostima varijabli roditeljskih čvorova od x

C Distribuciji varijable x uz opažene sve varijable koje varijabli x prethode u topološkom uređaju

D Zajedničkoj vjerojatnosti varijable x i varijabli roditeljskih čvorova od x

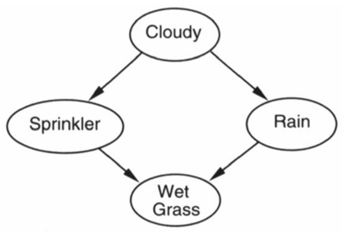
11 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

Primjenom algoritma d-odvajanja ispitujemo zavisnosti između parova varijabli. **Koje od sljedećih tvrdnji o nezavisnosti vrijede u ovoj Bayesovoj mreži?**

A $v\perp z|y$ B $v\perp y|w$ C $w\perp y|x$ D $v\perp z|w$

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



C	$P(S=1\|C)$
0	0.5
1	0.4

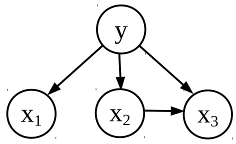
C	$P(R=1\|C)$
0	0.2
1	0.8

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

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

A 0.144 B 0.069 C 0.309 D 0.223

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



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

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

Cjelina 6: Grupiranje i vrednovanje modela (9 pitanja)

14 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

$$\begin{pmatrix} 1.0 & 0.2 & 0.6 & 0.7 & 0.4 \\ 0.2 & 1.0 & 0.3 & 0.3 & 0.7 \\ 0.6 & 0.3 & 1.0 & 0.9 & 0.5 \\ 0.7 & 0.3 & 0.9 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

Provedite grupiranje algoritmom HAC s potpunim povezivanjem. Pritom u svakoj iteraciji bilježite na kojoj razini sličnosti se odvija stapanje dviju grupa. **Koliko iznosi zbroj po svim razinama sličnosti na kojima se odvija stapanje grupa?**

A 1.9 B 2.0 C 2.4 D 1.8

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	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 0.62 C 0.53 D 0.58

16 (N) Vrednujemo višeklasni klasifikator koji koristimo kao referentni model za vrednovanje drugih klasifikatora. Klasifikator nasumično dodjeljuje oznake za tri klase ($K = 3$), i to prema distribuciji oznaka na skupu za učenje. Skup za učenje sadrži 100 primjera sa 35, 60 i 5 primjera u prvoj, drugoj odnosno trećoj klasi. S druge strane, ispitni skup sadrži 50 primjera, od kojih je po 20 u prvoj i drugoj klasi, a 10 u trećoj klasi. **Koliko iznosi očekivana makro-F1 vrijednost nasumičnog klasifikatora na ispitnome skupu?**

A 0.2156 B 0.2085 C 0.2989 D 0.3111

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

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_i = \big\{(1,1), (1,3), (1,4), (3,4)\big\}$$

Primjere grupiramo u $K = 2$ grupe algoritmom K-means++. Konačna vrijednost funkcije pogreške J bit će različita za različita pokretanja algoritma. Razmotrite sve moguće konačne vrijednosti funkcije J, tj. vrijednosti funkcije J nakon što se algoritam K-means++ zaustavi. **Koliko iznosi maksimalna moguća konačna vrijednost funkcije J?**

A 14/3 B 10/3 C 11/2 D 4

18 (T) Algoritam K-sredina iterativno smanjuje vrijednost funkcije pogreške J. Neovisno o broju grupa K, algoritam uvijek konvergira (završava s izvođenjem). Međutim, konačna vrijednost funkcije pogreške J može se razlikovati između dva pokretanja algoritma s istim brojem grupa K. **Zašto?**

A Zbog loše odabrane vrijednosti za broj grupa K

B Zbog razlike u početnim središtima $\boldsymbol{\mu}_k$

C Zbog izotropnosti kovarijacijskih matrica Σ_k

D Zbog nelinearnosti granice između K grupa

19 (P) Raspolažemo sa 750 označenih primjera. Na tom skupu treniramo i vrednujemo algoritam SVM, optimizirajući hiperparametre C i γ. Za vrednovanje koristimo ugniježđenu unakrsnu provjeru sa 5 preklopa u vanjskoj petlji i 5 preklopa u unutarnjoj petlji. To znači da za svaku kombinaciju vrijednosti hiperparametara C i γ treniramo pet modela. Ti modeli trenirani su na skupovima za učenje koji nisu disjunktni: svaki par treniranih modela dijele određeni broj primjera za učenje. Izračunajte koliko primjera za učenje dijele svaki par modela koje treniramo u unutarnjoj petlji ugniježđene unakrsne provjere. **Za koliko bi taj broj narastao kada bismo broj preklopa u unutarnjoj petlji povećali na 10?**

A 120 B 60 C 75 D 144

20 (P) Podatke u dvodimenzijskome ulaznom prostoru ($n = 2$) grupiramo u dvije grupe ($K = 2$) algoritmom Gaussove mješavine (GMM). Skup podataka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{8} = \big\{(-2,-2),(-2,-1),(-1,-2),(-1,-1),(1,2),(3,1),(3,3),(5,2)\big\}$$

Razmatramo dva modela: $\mathcal{H}_1$ i $\mathcal{H}_2$. Oba modela koriste punu kovarijacijsku matricu, međutim kod modela $\mathcal{H}_1$ kovarijacijska matrica nije dijeljena, dok je kod modela $\mathcal{H}_2$ kovarijacijska matrica dijeljena. Za svaki model razmotrite ono grupiranje koje maksimizira log-izglednost na skupu $\mathcal{D}$. Neka je $p_j(\mathbf{x}^{(i)}|y = k)$ izglednost grupe $y = k$ za primjer $\mathbf{x}^{(i)}$ prema modelu $\mathcal{H}_j$. Oznake grupa su $y = 1$ i $y = 2$, te pretpostavite da je grupiranje takvo da vrijedi $\|\boldsymbol{\mu}_1\| > \|\boldsymbol{\mu}_2\|$. **Što od sljedećega vrijedi?**

A $p_2(\mathbf{x}^{(1)}|y = 2) > p_1(\mathbf{x}^{(2)}|y = 2)$

C $p_2(\mathbf{x}^{(8)}|y = 1) < p_1(\mathbf{x}^{(5)}|y = 1)$

B $p_2(\mathbf{x}^{(4)}|y = 2) > p_1(\mathbf{x}^{(3)}|y = 2)$

D $p_2(\mathbf{x}^{(6)}|y = 1) < p_1(\mathbf{x}^{(7)}|y = 1)$

21 (T) Kod vrednovanja algoritma strojnog učenja važno je da vrednujemo model optimalne složenosti. Zbog toga označeni skup podataka dijelimo na tri disjunktna podskupa: skup za učenje, skup za provjeru i ispitni skup. Za odabir modela (tj. optimizaciju hiperparametara) koristimo skup za učenje i skup za provjeru, dok točnost optimalnog modela ispitujemo na ispitnome skupu. Zamislite da smo zaboravili upotrijebiti skup za provjeru: umjesto toga hiperparametre modela optimirali smo na skupu za učenje, a model smo potom ispitali na ispitnom skupu. **U odnosu na to da smo pravilno proveli vrednovanje algoritma, što možemo očekivati u ovom slučaju?**

A Model će biti prenaučen, ali imat će veću točnost na ispitnome skupu

B Model će biti optimalne složenosti i imat će veću točnost na ispitnome skupu

C Model će biti prenaučen i imat će manju točnost na ispitnome skupu

D Model će biti podnaučen, ali imat će veću točnost na ispitnome skupu

22 (T) Za procjenu parametara modela GMM (središta grupa $\boldsymbol{\mu}_k$, kovarijacijske matrice Σ_k i koeficijenti mješavine π_k) tipično se koristi algoritam maksimizacije očekivanja (EM-algoritam). Prije pokretanja EM-algoritma, parametre je potrebno inicijalizirati. U praksi se za inicijalizaciju parametara $\boldsymbol{\mu}_k$ koristi algoritam K-sredina. **Koji je očekivani efekt uporabe algoritma K-sredina za inicijalizaciju EM-algoritma?**

A Manja log-izglednost početnih i konačnih parametara te optimalan broj grupa K

B Veća log-izglednost početnih i konačnih parametara te manje iteracija do konvergencije

C Manja log-izglednost početnih parametara, ali veća log-izglednost konačnih parametara

D Veća potpuna log-izglednost konačnih parametara i veće vrijednosti gustoća komponenti

Završni ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% 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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 4: Procjena parametara i Bayesov klasifikator (8 pitanja)

1 (P) Uz određene restrikcije na parametre Gaussovog Bayesovog klasifikatora, Gaussov Bayesov klasifikator i logistička regresija su generativno-diskriminativni par modela. U pravilu, logistička regresija imat će manje parametara od njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. Razmotrite slučaj binarne klasifikacije u ulaznome prostoru dimenzije $n = 200$ pomoću modela logističke regresije i njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. **Koliko će model Gaussovog Bayesovog klasifikatora imati više parametara od modela logističke regresije?**

[A] 45450 [B] 90600 [C] 20300 [D] 40400

2 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{ne}, \text{zimski}, \text{preddipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

[A] 0.588 [B] 0.488 [C] 0.741 [D] 0.322

3 (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 $-10 \leq x \leq 10$. **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] $[-4, -4 + a] \cup [5, 5 + b]$

4 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{-10, 5, 6, 2\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

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

Za procjenu parametra μ koristimo procjenitelj $\hat{\mu}_{\text{MLE}}$, a za procjenu parametra σ^2 procjenitelj $\hat{\sigma}^2_{\text{UB}}$, koji je nepristran procjenitelj izveden odgovarajućom korekcijom iz procjenitelja $\hat{\sigma}^2_{\text{MLE}}$. Izračunajte log-izglednost procijenjenih parametara na skupu $\mathcal{D}$, tj. $\ln \mathcal{L}(\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{UB}}|\mathcal{D})$. **Koliko iznosi ta log-izglednost?**

A -10.832 B -15.200 C -12.873 D -13.163

5 (T) Bayesov klasifikator temelji se na hipotezi *maximum aposteriori* (MAP). Za treniranje Bayesovog klasifikatora možemo koristiti procjenitelj MLE. **Koja je razlika između hipoteze MAP kod Bayesovog klasifikatora i procjenitelja MLE?**

A Procjenitelj MLE određuje apriornu vjerojatnost klase, a hipoteza MAP maksimizira aposterionu vjerojatnost parametara

B Hipoteza MAP određuje koju klasu odabrati za primjer, a procjenitelj MLE određuje koju vrijednost odabrati za parametar

C Procjenitelj MLE primjer klasificira u klasu koja je za njega navjerojatnija, a hipoteza MAP tu odluku kombinira s apriornim znanjem

D Hipoteza MAP maksimizira aposterionu vjerojatnost parametara, a procjenitelj MLE određuje aposteriornu vjerojatnost klase za zadani primjer

6 (T) Polunaivan Bayesov klasifikator može ostvariti veću klasifikacijsku točnost od naivnog Bayesovog klasifikatora u situacijama kada između značajki postoji zavisnost. Jedan način određivanja zavisnosti između značajki temelji se na izračunu Kullback-Leiblerove divergencije (KL-divergencije) između distribucija dviju varijabli. **Koji je princip uporabe KL-divergencije u izgradnji polunaivnog Bayesovog klasifikatora?**

A Što je manja KL-divergencija između $P(x_1)$ i $P(x_2)$, to su varijable x_1 i x_2 više zavisne, pa je to veća korist od njihovog združivanja u jedan faktor

B Ako je KL-divergencija između $P(x_1|x_2)$ i $P(x_2|x_1)$ vrlo mala, onda su varijable x_1 i x_2 vrlo zavisne i treba ih odvojiti u zasebne faktore

C Što je veća KL-divergencija između $P(x_1)P(x_2)$ i $P(x_1, x_2)$, to su varijable x_1 i x_2 manje nezavisne, pa je to veća korist od njihovog združivanja u jedan faktor

D Ako je KL-divergencija između $P(x_1)P(x_2)$ i $P(x_1, x_2)$ jednaka nuli, onda su varijable linearno zavisne i treba ih odvojiti u zasebne faktore

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:

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

Model treniramo na skupu 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 -2.75 D -4.13

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

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

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

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

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

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

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

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

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

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

9 (T) Probablistički grafički modeli na sažet način prikazuju zajedničku distribuciju n slučajnih varijabli. Ta je distribucija prikazana usmjerenim grafom. Neka graf sadrži čvor označen sa x. **Čemu odgovara taj čvor?**

- A Distribuciji varijable x uvjetovane vrijednostima varijabli roditeljskih čvorova od x
- B Marginalnoj vjerojatnosti $P(x)$ ili marginalnoj gustoći vjerojatnosti $p(x)$
- C Zajedničkoj vjerojatnosti varijable x i varijabli roditeljskih čvorova od x
- D Distribuciji varijable x uz opažene sve varijable koje varijabli x prethode u topološkom uređaju

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 $P(z = 1) = 0.1$, $P(w = 1) = 0.2$, $P(x = 1|y = 0) = 0.2$ 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 konačni uzorak manji od N. **U uzorku neodbačenih vektora, koliko je očekivano više vektora sa $w = 0$ od vektora sa $w = 1$?**

A 472 B 348 C 944 D 596

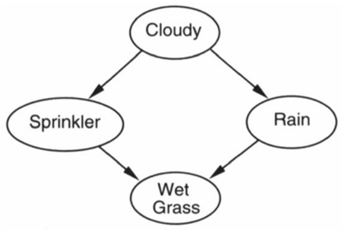
11 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

Primjenom algoritma d-odvajanja ispitujemo zavisnosti između parova varijabli. **Koje od sljedećih tvrdnji o nezavisnosti vrijede u ovoj Bayesovoj mreži?**

A $v\perp z|x$ B $w\perp y|z$ C $x\perp z|y$ D $w\perp y|x$

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



C	$P(S = 1\|C)$
0	0.5
1	0.3

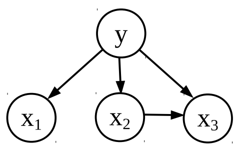
C	$P(R = 1\|C)$
0	0.2
1	0.8

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

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

A 0.144 B 0.309 C 0.069 D 0.223

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



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

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

Cjelina 6: Grupiranje i vrednovanje modela (9 pitanja)

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

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_i = \big\{(1,1),(1,3),(1,4),(3,4)\big\}$$

Primjere grupiramo u $K = 2$ grupe algoritmom K-means++. Konačna vrijednost funkcije pogreške J bit će različita za različita pokretanja algoritma. Razmotrite sve moguće konačne vrijednosti funkcije J, tj. vrijednosti funkcije J nakon što se algoritam K-means++ zaustavi. **Koliko iznosi maksimalna moguća konačna vrijednost funkcije J?**

[A] 11/2 [B] 10/3 [C] 14/3 [D] 4

15 (P) Raspolažemo sa 750 označenih primjera. Na tom skupu treniramo i vrednujemo algoritam SVM, optimizirajući hiperparametre C i γ. Za vrednovanje koristimo ugniježđenu unakrsnu provjeru sa 5 preklopa u vanjskoj petlji i 5 preklopa u unutarnjoj petlji. To znači da za svaku kombinaciju vrijednosti hiperparametara C i γ treniramo pet modela. Ti modeli trenirani su na skupovima za učenje koji nisu disjunktni: svaki par treniranih modela dijele određeni broj primjera za učenje. Izračunajte koliko primjera za učenje dijele svaki par modela koje treniramo u unutarnjoj petlji ugniježđene unakrsne provjere. **Za koliko bi taj broj narastao kada bismo broj preklopa u unutarnjoj petlji povećali na 10?**

[A] 120 [B] 60 [C] 144 [D] 75

16 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

$$\begin{pmatrix} 1.0 & 0.2 & 0.6 & 0.7 & 0.4 \\ 0.2 & 1.0 & 0.3 & 0.3 & 0.7 \\ 0.6 & 0.3 & 1.0 & 0.9 & 0.5 \\ 0.7 & 0.3 & 0.9 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

Provedite grupiranje algoritmom HAC s potpunim povezivanjem. Pritom u svakoj iteraciji bilježite na kojoj razini sličnosti se odvija stapanje dviju grupa. **Koliko iznosi zbroj po svim razinama sličnosti na kojima se odvija stapanje grupa?**

[A] 1.9 [B] 2.4 [C] 2.0 [D] 1.8

17 (P) Podatke u dvodimenzijskome ulaznom prostoru ($n = 2$) grupiramo u dvije grupe ($K = 2$) algoritmom Gaussove mješavine (GMM). Skup podataka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{8} = \big\{(-2,-2),(-2,-1),(-1,-2),(-1,-1),(1,2),(3,1),(3,3),(5,2)\big\}$$

Razmatramo dva modela: $\mathcal{H}_1$ i $\mathcal{H}_2$. Oba modela koriste punu kovarijacijsku matricu, međutim kod modela $\mathcal{H}_1$ kovarijacijska matrica nije dijeljena, dok je kod modela $\mathcal{H}_2$ kovarijacijska matrica dijeljena. Za svaki model razmotrite ono grupiranje koje maksimizira log-izglednost na skupu $\mathcal{D}$. Neka je $p_j(\mathbf{x}^{(i)}|y = k)$ izglednost grupe $y = k$ za primjer $\mathbf{x}^{(i)}$ prema modelu $\mathcal{H}_j$. Oznake grupa su $y = 1$ i $y = 2$, te pretpostavite da je grupiranje takvo da vrijedi $\|\boldsymbol{\mu}_1\| > \|\boldsymbol{\mu}_2\|$. **Što od sljedećega vrijedi?**

[A] $p_2(\mathbf{x}^{(1)}|y = 2) > p_1(\mathbf{x}^{(2)}|y = 2)$ [C] $p_2(\mathbf{x}^{(4)}|y = 2) > p_1(\mathbf{x}^{(3)}|y = 2)$

[B] $p_2(\mathbf{x}^{(6)}|y = 1) < p_1(\mathbf{x}^{(7)}|y = 1)$ [D] $p_2(\mathbf{x}^{(8)}|y = 1) < p_1(\mathbf{x}^{(5)}|y = 1)$

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

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

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

19 (T) Kod vrednovanja algoritma strojnog učenja važno je da vrednujemo model optimalne složenosti. Zbog toga označeni skup podataka dijelimo na tri disjunktna podskupa: skup za učenje, skup za provjeru i ispitni skup. Za odabir modela (tj. optimizaciju hiperparametara) koristimo skup za učenje i skup za provjeru, dok točnost optimalnog modela ispitujemo na ispitnome skupu. Zamislite da smo zaboravili upotrijebiti skup za provjeru: umjesto toga hiperparametre modela optimirali smo na skupu za učenje, a model smo potom ispitali na ispitnom skupu. **U odnosu na to da smo pravilno proveli vrednovanje algoritma, što možemo očekivati u ovom slučaju?**

- A Model će biti prenaučen i imat će manju točnost na ispitnome skupu
- B Model će biti prenaučen, ali imat će veću točnost na ispitnome skupu
- C Model će biti podnaučen, ali imat će veću točnost na ispitnome skupu
- D Model će biti optimalne složenosti i imat će veću točnost na ispitnome skupu

20 (T) Za procjenu parametara modela GMM (središta grupa $\boldsymbol{\mu}_k$, kovarijacijske matrice Σ_k i koeficijenti mješavine π_k) tipično se koristi algoritam maksimizacije očekivanja (EM-algoritam). Prije pokretanja EM-algoritma, parametre je potrebno inicijalizirati. U praksi se za inicijalizaciju parametara $\boldsymbol{\mu}_k$ koristi algoritam K-sredina. **Koji je očekivani efekt uporabe algoritma K-sredina za inicijalizaciju EM-algoritma?**

- A Veća log-izglednost početnih i konačnih parametara te manje iteracija do konvergencije
- B Manja log-izglednost početnih parametara, ali veća log-izglednost konačnih parametara
- C Veća potpuna log-izglednost konačnih parametara i veće vrijednosti gustoća komponenti
- D Manja log-izglednost početnih i konačnih parametara te optimalan broj grupa K

21 (N) Vrednujemo višeklasni klasifikator koji koristimo kao referentni model za vrednovanje drugih klasifikatora. Klasifikator nasumično dodjeljuje oznake za tri klase ($K = 3$), i to prema distribuciji oznaka na skupu za učenje. Skup za učenje sadrži 100 primjera sa 35, 60 i 5 primjera u prvoj, drugoj odnosno trećoj klasi. S druge strane, ispitni skup sadrži 50 primjera, od kojih je po 20 u prvoj i drugoj klasi, a 10 u trećoj klasi. **Koliko iznosi očekivana makro-F1 vrijednost nasumičnog klasifikatora na ispitnome skupu?**

A 0.3111 B 0.2156 C 0.2085 D 0.2989

22 (T) Algoritam K-sredina iterativno smanjuje vrijednost funkcije pogreške J. Neovisno o broju grupa K, algoritam uvijek konvergira (završava s izvođenjem). Međutim, konačna vrijednost funkcije pogreške J može se razlikovati između dva pokretanja algoritma s istim brojem grupa K. **Zašto?**

- A Zbog izotropnosti kovarijacijskih matrica Σ_k
- B Zbog nelinearnosti granice između K grupa
- C Zbog razlike u početnim središtima $\boldsymbol{\mu}_k$
- D Zbog loše odabrane vrijednosti za broj grupa K

Grupa	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
A	D	C	C	D	B	D	B	A	A	B	B	B	C	B	A	D	B	B	A	B	A	A
B	A	D	D	B	C	B	A	C	A	C	A	C	B	C	C	A	B	D	B	A	D	A
C	B	D	D	B	B	C	B	D	B	A	D	A	A	B	B	A	C	D	A	B	D	D
D	B	B	D	D	A	B	C	A	C	B	A	D	A	A	C	C	D	C	B	B	A	C
E	B	B	B	B	D	C	C	D	D	C	B	D	C	A	A	B	C	B	A	C	C	D
F	C	D	C	B	A	D	A	B	B	A	C	D	A	B	A	A	D	D	B	B	B	B
G	B	B	A	C	D	C	D	A	D	B	C	C	D	C	D	D	A	B	A	C	C	B
H	C	D	D	D	B	C	D	C	A	A	D	D	A	C	A	B	D	A	A	A	A	C