

Završni ispit iz Strojnog učenja 1 (ak. god. 2022./2023.) – NEKORIGIRANA VERZIJA –

Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.

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

- 1** (T) Za Bayesov klasifikator procjenjujemo parametar μ Bernoullijeve distribucije. Skup primjera za učenje je razmjerno malen. Naš procjenitelj $\hat{\mu}$ parametra μ može biti pristran ili nepristran, dok model čiji je parametar procijenjen s $\hat{\mu}$ može biti prenaučan ili može dobro generalizirati. Razmotrite procjenitelje MLE i MAP (konkretno, Laplaceovo zaglađivanje). **Što od sljedećega općenito vrijedi u ovom slučaju?**

- ☐ A MAP procjenitelj je nepristran i očekujemo da će model dobro generalizirati
☐ B MLE procjenitelj je nepristran i očekujemo da će model dobro generalizirati
☐ C MLE procjenitelj je pristran i očekujemo da će model loše generalizirati
☐ D MAP procjenitelj je pristran, ali očekujemo da će model dobro generalizirati

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

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

Na temelju beta-Bernoullijevog modela na skupu \mathcal{D} računamo MAP procjenu parametra μ Bernoullijeve distribucije. MLE procjena za isti parametar na skupu \mathcal{D} iznosi 0.3. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

- ☐ A $\alpha = 4, \beta = 8$ ☐ B $\alpha = 2, \beta = 10$ ☐ C $\alpha = 5, \beta = 7$ ☐ D $\alpha = 2, \beta = 5$

- 3** (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y = 1$ i $y = 2$) u dvodimenzijaskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klase su jednake, dok su izglednosti klase modelirane bivarijantnim Gaussovim distribucijama sa sljedećim parametrima:

$$\mu_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \Sigma_1 = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \quad \mu_2 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \Sigma_2 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klase definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

- ☐ A (1, 6), (3, 3), (5, 0) ☐ B (1, 2), (3, 4), (6, 7) ☐ C (3, 2), (4, 5), (9, 7) ☐ D (1, 3), (5, 4), (7, 9)

- 4** (T) Gaussov Bayesov klasifikator s dijeljenom kovarijacijskom matricom (GBC) i logistička regresija (LR) su generativno-diskriminativni par modela. **Što to znači?**

- ☐ A Nevisno o optimizacijskom postupku, GBC i LR ostvaruju istu pogrešku na skupu za učenje, ali uz moguće različit broj parametara
☐ B Aposteriorna vjerojatnost klase za GBC može se izraziti kao poopćeni linearni model sa sigmoidnom aktivacijskom funkcijom
☐ C GBC i neregularizirana LR modeliraju identične distribucije zajedničke vjerojatnosti primjera i oznaka, ali s različitim brojem parametara
☐ D Izlaz modela LR jednak je zajedničkoj vjerojatnosti modela GBC, ali model LR iziskuje manje parametara

- 5 (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju 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 $[-4-a, 5+b]$ ☐ B $[-4-a, -4] \cup [5-b, 5]$ ☐ C $[-4, -4+a] \cup [5, 5+b]$ ☐ D $[-4-a, -4+b]$

- 6 (N) Raspolažemo sljedećim skupom označenih primjera:

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

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

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

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

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

- 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:

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

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. Za procjenu kovarijacijske matrice koristite dva procjenitelja: MLE i nepristrani procjenitelj. Izlaz modela za klasu $y=j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y=j)$. Neka su h_1^{MLE} i h_1^{UB} hipoteze za prvu klasu modela s pristranom odnosno nepristranom procjenom kovarijacijske matrice. Zanima nas predikcija za primjer $\mathbf{x} = (1, 1)$. **Koliko iznosi apsolutna razlika u predikciji za \mathbf{x} za ova dva modela, $|h_1^{\text{MLE}}(\mathbf{x}) - h_1^{\text{UB}}(\mathbf{x})|$?**

- ☐ A 0.2276 ☐ B 0.2537 ☐ C 0.1638 ☐ D 0.0337

- 8 (N) Treniramo polunaivan Bayesov klasifikator sa tri binarne značajke, x_1 , x_2 i x_3 . Skup primjera za učenje \mathcal{D} sastoji se od sljedećih deset primjera:

x_1	x_2	x_3	y	x_1	x_2	x_3	y
1	1	0	1	1	1	0	0
0	1	0	1	1	0	1	1
1	1	0	0	1	0	0	0
0	1	1	1	0	1	1	1
0	1	0	1	0	0	1	0

Prije treniranja koristimo uzajamnu informaciju kako bismo procijenili koje su varijable najviše statistički zavisne, jer se te varijable isplati združiti u zajednički faktor. Izračun provodimo tako da za svaki par varijabli x_i i x_j procjenjujemo parametre zajedničke distribucije $P(x_i, x_j)$, a zatim iz zajedničke distribucije računamo marginalne vjerojatnosti i uzajamnu informaciju $I(x_i, x_j)$. Budući da je skup \mathcal{D} malen, za procjenu parametara distribucije $P(x_i, x_j)$ koristimo Laplaceov procjenitelj. **Koliko iznosi na taj način izračunata uzajamna informacija između varijabli x_1 i x_2 ?**

- ☐ A 0.0112 ☐ B 0.0334 ☐ C 0.0078 ☐ D 0.0423

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

- 9 (N) Razmotrite Bayesovu mrežu koja odgovara faktorizaciji $P(w, x, y, z) = P(w)P(x)P(y|w, x)P(z|x)$. Sve varijable su binarne. Vrijedi $P(w = 1) = 0.1$, $P(x = 1) = 0.2$, $P(z = 1|x = 0) = 0.9$ i $P(z = 1|x = 1) = 0.7$. Tablica uvjetnih vjerojatnosti za čvor y je sljedeća:

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

Postupkom uzorkovanja s odbijanjem želimo procijeniti parametar μ uvjetne distribucije $P(x = 0|y = 1, z = 0)$. Uzorkovanje smo ponovili ukupno $N = 100$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N . Na temelju dobivenog uzorka parametar μ procjenjujemo MAP procjeniteljem uz $\alpha = \beta = 2$. **Koliko iznosi očekivana MAP procjena parametra μ ?**

- ☐ A 0.0786 ☐ B 0.0490 ☐ C 0.1877 ☐ D 0.1274

- 10 (P) Razmotrite Bayesovu mrežu sa šest čvorova koja odgovara sljedećoj faktorizaciji:

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

Upotrijebite metodu d-odvajanja da biste ispitati uvjetnu nezavisnost varijabli x i z u ovisnosti o preostale četiri varijable. Ispitajte koje varijable od preostalih četiri varijabli trebaju biti opažene a koje neopažene, a da bi varijable x i z bile d-odvojene. **Za koju od preostale četiri varijable je svejedno je li opažena, ako su varijable x i z d-odvojene?**

- ☐ A w ☐ B u ☐ C y ☐ D v

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

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

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

- ☐ A 23 ☐ B 24 ☐ C 21 ☐ D 18

- 12 (T) Parametre probabilističkih grafičkih modela, uključivo Bayesove mreže, možemo procjenjivati iz potpunih podataka ili nepotpunih podataka. **Zašto i kako Bayesovu mrežu učimo nad nepotpunim podacima?**

- ☐ A Jer mreža ima manje čvorova nego što je opaženih varijabli, pa koristimo eliminaciju varijabli
☐ B Jer MLE nema rješenje u zatvorenoj formi, pa umjesto MLE koristimo gradijentni uspon ili EM-algoritam
☐ C Jer mreža ima skrivene varijable koje ne opažamo, pa moramo koristiti iterativne metode za MAP ili MLE
☐ D Jer se log-izglednost dekomponira po čvorovima mreže, pa MLE ili MAP procjenjujemo u zatvorenoj formi

- 13 (N) Bayesovu mrežu koristimo za medicinsku dijagnostiku te modeliramo sljedeće kauzalne odnose. Upala grla ($U = 1$) može biti uzrokovana virusom ($V = 1$) ili bakterijom ($B = 1$). Povišena temperatura ($T = 1$) može biti uzrokovana upalom grla ili sunčanicom ($S = 1$). Sve varijable su binarne. Na temelju podataka o pacijentima procijenili smo parametre mreže: $P(V = 1) = 0.3$, $P(B = 1) = 0.1$ i $P(S = 1) = 0.05$. Uvjetne vjerojatnosti za čvorove U i T su:

V	B	$P(U = 1 V, B)$	U	S	$P(T = 1 U, S)$
0	0	0.2	0	0	0
0	1	0.5	0	1	0.2
1	0	0.4	1	0	0.4
1	1	0.7	1	1	0.4

Zanima nas koje je najvjerojatnije objašnjenje izravnog uzroka povišene temperature u pacijenata kod kojih nije dokazano prisustvo virusa. U tu svrhu računamo MAP-upit za par varijabli upita U i S uz opažene varijable $V = 0$ i $T = 1$, tj. računamo $\operatorname{argmax}_{U, S} P(U, S|V = 0, T = 1)$. Neka je p_1 vjerojatnost najvjerojatnijeg (MAP)

objašnjenja za varijable U i S , a p_2 vjerojatnost drugog po redu najvjerojatnijeg objašnjenja za te varijable. **Koliko je puta najvjerojatnije objašnjenje vjerojatnije od drugog najvjerojatnijeg objašnjenja, tj. koliko iznosi p_1/p_2 ?**

- ☐ A 13.35 ☐ B 11.35 ☐ C 15.52 ☐ D 17.88

Cjelina 6: Grupiranje i vrednovanje modela (9 pitanja)

- 14** (P) Skup neoznačenih primjera u dvodimenzijaskome ulaznom prostoru neka je sljedeći:

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

Primjere grupiramo algoritmom K-sredina sa $K = 2$ grupe. Grupiranje možemo shvatiti kao pretraživanje prostora stanja, gdje svako pojedino stanje odgovara jednom pridjeljivanju primjera grupama. Pritom dva stanja smatramo identičnima ako su grupiranja identična, neovisno o identitetu grupa (npr., grupiranje kod kojega prva grupa sadrži samo primjer $\mathbf{x}^{(1)}$ je identično kao i grupiranje kod kojega druga grupa sadrži samo primjer $\mathbf{x}^{(1)}$). Neka je A_1 algoritam K-sredina s potpuno slučajno inicijaliziranim središtima, a A_2 algoritam K-sredina gdje su središta inicijalizirana algoritmom K-means++. Neka je $S(A_1)$ skup stanja koje pretražuje algoritam A_1 , a $S(A_2)$ skup stanja koje pretražuje algoritam A_2 . Izračunajte veličine ovih skupova, uzevši u obzir mogućnost da pojedina grupa bude prazna te da dođe do izjednačenja udaljenosti primjera do centroida, što se razrješava slučajnim mehanizmom. **Koliko algoritam A_2 pretražuje manje stanja od algoritma A_1 , tj. koliko iznosi $|S(A_1)| - |S(A_2)|$?**

- ☐ A 6 ☐ B 4 ☐ C 10 ☐ D 12

- 15** (P) Raspoložemo 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 144 ☐ C 60 ☐ D 120

- 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 2.0 ☐ B 1.9 ☐ C 1.8 ☐ D 2.4

- 17** (N) Na ispitnome skupu sa $K = 3$ klase vrednujemo algoritam multinomijalne logističke regresije. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci predikcije klasifikatora):

$$\begin{array}{ccc} & y = 1 & y = 2 & y = 3 \\ \begin{array}{l} y = 1 \\ y = 2 \\ y = 3 \end{array} & \begin{pmatrix} 15 & 3 & 5 \\ 9 & 5 & 4 \\ 4 & 2 & 3 \end{pmatrix} \end{array}$$

Kao referentni model koristimo klasifikator koji nasumično pogađa oznaku y , i to s jednakom vjerojatnošću za svaku od tri klase. Za vrednovanje koristimo mjeru F_1 -makro. Izračunajte očekivanu vrijednost mjere F_1 -makro za logističku regresiju i za referentni model. **Koliko iznosi očekivana razlika u mjeri F_1 -makro između logističke regresije i referentnog modela?**

- ☐ A 0.1420 ☐ B 0.1756 ☐ C 0.0947 ☐ D 0.2238

18 (T) Algoritam K-medoida općenitiji je od algoritma K-sredina budući da se može koristiti za primjere koji nisu prikazani kao vektori. Međutim, razmotrite slučaj kada primjeri jesu prikazani kao vektori, ali ih želimo grupirati na temelju mjere udaljenosti koja nije euklidska. **Koji bismo algoritam koristili u tom slučaju i zašto?**

- ☐ A Algoritam K-sredina, jer za vektorizirane primjere možemo izračunati centroide grupa
- ☐ B Algoritam K-medoida, jer vektorizirani primjeri također mogu biti medoidi
- ☐ C Algoritam K-medoida, jer kriterijska funkcija algoritma K-sredina koristi euklidsku udaljenost
- ☐ D Algoritam K-sredina, jer koristi mjeru udaljenosti, dok algoritam K-medoida koristi mjeru sličnosti

19 (T) Mjera točnosti nije prikladna za vrednovanje klasifikatora na skupovima podataka s neuravnoteženim brojem primjera po klasama. Jedna alternativa mjeri točnosti je F_1 -mjera, međutim ni ta mjera nije uvijek prikladna. Pretpostavite da vrijednost F_1 -mjere postavljamo na nulu u slučajevima kada je harmonijska sredina preciznosti i odziva nedefinirana. **U kojem slučaju F_1 -mjera ne bi bila prikladna mjera za vrednovanje klasifikatora jer bi bila previše optimistična?**

- ☐ A Ako je većina primjera pozitivna, a klasifikator sve primjere klasificira negativno
- ☐ B Ako je većina primjera negativna i klasifikator sve primjere klasificira negativno
- ☐ C Ako je većina primjera negativna, a klasifikator sve primjere klasificira pozitivno
- ☐ D Ako je većina primjera pozitivna i klasifikator sve primjere klasificira pozitivno

20 (T) Za grupiranje primjera u K grupa koristimo model Gaussove mješavine (GMM) s dijeljenom kovarijacijskom matricom. Nakon grupiranja, odgovornosti zaokružujemo na cijeli broj, čime dobivamo tvrdo grupiranje. Iste podatke grupiramo algoritmom K-medoida (KM). **Uz koje parametre ovih algoritama očekujemo dobiti najbližnje rezultate grupiranja?**

- ☐ A GMM: $\Sigma_k = \sigma^2 \mathbf{I}$; KM: euklidska udaljenost
- ☐ B GMM: $\Sigma_k = \sigma^2 \mathbf{I}$ i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost
- ☐ C GMM: puna Σ_k i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost
- ☐ D GMM: $\Sigma_k = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ i $\pi_k = 1/K$; KM: euklidska udaljenost

21 (P) Skup neoznačenih primjera \mathcal{D} grupiramo modelom GMM treniranim EM-algoritmom. Koristimo nekoliko varijanti tog modela:

\mathcal{H}_1 : Model sa $K = 25$ središta inicijaliziranim algoritmom K-means++

\mathcal{H}_2 : Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom

\mathcal{H}_3 : Model sa $K = 50$ središta inicijaliziranim algoritmom K-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL_α^0 prosječna log-izglednost za model \mathcal{H}_α na početku izvođenja EM-algoritma, LL_α^* prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma te neka je k_α broj iteracija EM-algoritma za taj model. **Što možemo zaključiti o očekivanim odnosima između ovih vrijednosti?**

- ☐ A $LL_3^0 \geq LL_2^0$, $LL_3^* \geq LL_2^*$, $k_2 \geq k_3$
- ☐ B $LL_1^0 \geq LL_3^0$, $LL_1^* \geq LL_3^*$, $k_2 \geq k_1$
- ☐ C $LL_1^0 \geq LL_2^0$, $LL_2^* \geq LL_3^*$, $k_1 \geq k_2$
- ☐ D $LL_2^0 \geq LL_3^0$, $LL_3^* \geq LL_2^*$, $k_3 \geq k_2$

22 (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1, 2, 3, 4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{aligned} K = 3 & \quad \{\{1, 2, 4, 4\}, \{2, 2, 3\}, \{1, 1, 3, 4\}\} \\ K = 4 & \quad \{\{1, 2, 2, 4, 4\}, \{2, 3, 3\}, \{1, 1\}, \{4\}\} \end{aligned}$$

gdje podskupovi odgovaraju grupama, a brojeke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

- ☐ A 0.0727
- ☐ B 0.0545
- ☐ C 0.0364
- ☐ D 0.0182

Završni ispit iz Strojnog učenja 1 (ak. god. 2022./2023.)

– NEKORIGIRANA VERZIJA –

Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.

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

- 1** (T) Gaussov Bayesov klasifikator s dijeljenom kovarijacijskom matricom (GBC) i logistička regresija (LR) su generativno-diskriminativni par modela. **Što to znači?**
- ☐ A Izlaz modela LR jednak je zajedničkoj vjerojatnosti modela GBC, ali model LR iziskuje manje parametara
- ☐ B Aposteriorna vjerojatnost klase za GBC može se izraziti kao poopćeni linearni model sa sigmoidnom aktivacijskom funkcijom
- ☐ C GBC i neregularizirana LR modeliraju identične distribucije zajedničke vjerojatnosti primjera i oznaka, ali s različitim brojem parametara
- ☐ D Neovisno o optimizacijskom postupku, GBC i LR ostvaruju istu pogrešku na skupu za učenje, ali uz moguće različit broj parametara
- 2** (T) Za Bayesov klasifikator procjenjujemo parametar μ Bernoullijeve distribucije. Skup primjera za učenje je razmjerno malen. Naš procjenitelj $\hat{\mu}$ parametra μ može biti pristran ili nepristran, dok model čiji je parametar procijenjen s $\hat{\mu}$ može biti prenaučan ili može dobro generalizirati. Razmotrite procjenitelje MLE i MAP (konkretno, Laplaceovo zaglađivanje). **Što od sljedećega općenito vrijedi u ovom slučaju?**
- ☐ A MAP procjenitelj je nepristran i očekujemo da će model dobro generalizirati
- ☐ B MAP procjenitelj je pristran, ali očekujemo da će model dobro generalizirati
- ☐ C MLE procjenitelj je pristran i očekujemo da će model loše generalizirati
- ☐ D MLE procjenitelj je nepristran i očekujemo da će model dobro generalizirati
- 3** (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:

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

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

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

Procijenite parametre modela na ovom skupu primjera. Za procjenu kovarijacijske matrice koristite dva procjenitelja: MLE i nepristrani procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. Neka su h_1^{MLE} i h_1^{UB} hipoteze za prvu klasu modela s pristranom odnosno nepristranom procjenom kovarijacijske matrice. Zanima nas predikcija za primjer $\mathbf{x} = (1, 1)$. **Koliko iznosi apsolutna razlika u predikciji za \mathbf{x} za ova dva modela, $|h_1^{\text{MLE}}(\mathbf{x}) - h_1^{\text{UB}}(\mathbf{x})|$?**

- ☐ A 0.0337 ☐ B 0.2537 ☐ C 0.1638 ☐ D 0.2276

- 4** (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju jednodimenzijskih podataka u tri klase. Procijenjene izglednosti klase 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 klase 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 $[-4 - a, -4] \cup [5 - b, 5]$ ☐ D $[-4 - a, -4 + b]$

- 5** (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y = 1$ i $y = 2$) u dvodimenzijaskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klasa su jednake, dok su izglednosti klasa modelirane bivarijantnim Gaussovim distribucijama sa sljedećim parametrima:

$$\mu_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \Sigma_1 = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \quad \mu_2 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \Sigma_2 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klasa definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

- ☐ A $(1, 2), (3, 4), (6, 7)$ ☐ B $(1, 3), (5, 4), (7, 9)$ ☐ C $(3, 2), (4, 5), (9, 7)$ ☐ D $(1, 6), (3, 3), (5, 0)$

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

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

Na temelju beta-Bernoullijevog modela na skupu \mathcal{D} računamo MAP procjenu parametra μ Bernoullijeve distribucije. MLE procjena za isti parametar na skupu \mathcal{D} iznosi 0.2. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

- ☐ A $\alpha = 2, \beta = 10$ ☐ B $\alpha = 2, \beta = 5$ ☐ C $\alpha = 5, \beta = 7$ ☐ D $\alpha = 4, \beta = 8$

- 7** (N) Raspolažemo sljedećim skupom označenih primjera:

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

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

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

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

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

- 8** (N) Treniramo polunaivan Bayesov klasifikator sa tri binarne značajke, x_1 , x_2 i x_3 . Skup primjera za učenje \mathcal{D} sastoji se od sljedećih deset primjera:

x_1	x_2	x_3	y	x_1	x_2	x_3	y
1	0	1	1	1	1	0	0
0	1	0	1	1	0	1	1
1	1	0	0	1	0	0	0
0	1	1	1	0	1	1	1
0	1	0	1	0	0	1	0

Prije treniranja koristimo uzajamnu informaciju kako bismo procijenili koje su varijable najviše statistički zavisne, jer se te varijable isplati združiti u zajednički faktor. Izračun provodimo tako da za svaki par varijabli x_i i x_j procjenjujemo parametre zajedničke distribucije $P(x_i, x_j)$, a zatim iz zajedničke distribucije računamo marginalne vjerojatnosti i uzajamnu informaciju $I(x_i, x_j)$. Budući da je skup \mathcal{D} malen, za procjenu parametara distribucije $P(x_i, x_j)$ koristimo Laplaceov procjenitelj. **Koliko iznosi na taj način izračunata uzajamna informacija između varijabli x_1 i x_2 ?**

- ☐ A 0.0423 ☐ B 0.0078 ☐ C 0.0112 ☐ D 0.0334

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

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

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

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

- ☐ A 18 ☐ B 23 ☐ C 21 ☐ D 24

- 10 (N) Bayesovu mrežu koristimo za medicinsku dijagnostiku te modeliramo sljedeće kauzalne odnose. Upala grla ($U = 1$) može biti uzrokovana virusom ($V = 1$) ili bakterijom ($B = 1$). Povišena temperatura ($T = 1$) može biti uzrokovana upalom grla ili sunčanicom ($S = 1$). Sve varijable su binarne. Na temelju podataka o pacijentima procijenili smo parametre mreže: $P(V = 1) = 0.3$, $P(B = 1) = 0.3$ i $P(S = 1) = 0.05$. Uvjetne vjerojatnosti za čvorove U i T su:

V	B	$P(U = 1 V, B)$	U	S	$P(T = 1 U, S)$
0	0	0.2	0	0	0
0	1	0.5	0	1	0.2
1	0	0.4	1	0	0.4
1	1	0.7	1	1	0.4

Zanima nas koje je najvjerojatnije objašnjenje izravnog uzroka povišene temperature u pacijenata kod kojih nije dokazano prisustvo virusa. U tu svrhu računamo MAP-upit za par varijabli upita U i S uz opažene varijable $V = 0$ i $T = 1$, tj. računamo $\operatorname{argmax}_{U, S} P(U, S|V = 0, T = 1)$. Neka je p_1 vjerojatnost najvjerojatnijeg (MAP) objašnjenja za varijable U i S , a p_2 vjerojatnost drugog po redu najvjerojatnijeg objašnjenja za te varijable. **Koliko je puta najvjerojatnije objašnjenje vjerojatnije od drugog najvjerojatnijeg objašnjenja, tj. koliko iznosi p_1/p_2 ?**

- ☐ A 15.52 ☐ B 17.88 ☐ C 13.35 ☐ D 11.35

- 11 (T) Parametre probabilističkih grafičkih modela, uključivo Bayesove mreže, možemo procjenjivati iz potpunih podataka ili nepotpunih podataka. **Zašto i kako Bayesovu mrežu učimo nad nepotpunim podacima?**

- ☐ A Jer MLE nema rješenje u zatvorenoj formi, pa umjesto MLE koristimo gradijentni uspon ili EM-algoritam
☐ B Jer se log-izglednost dekomponira po čvorovima mreže, pa MLE ili MAP procjenjujemo u zatvorenoj formi
☐ C Jer mreža ima manje čvorova nego što je opaženih varijabli, pa koristimo eliminaciju varijabli
☐ D Jer mreža ima skrivene varijable koje ne opažamo, pa moramo koristiti iterativne metode za MAP ili MLE

- 12 (P) Razmotrite Bayesovu mrežu sa šest čvorova koja odgovara sljedećoj faktorizaciji:

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

Upotrijebite metodu d-odvajanja da biste ispitali uvjetnu nezavisnost varijabli x i z u ovisnosti o preostale četiri varijable. Ispitajte koje varijable od preostalih četiri varijabli trebaju biti opažene a koje neopažene, a da bi varijable x i z bile d-odvojene. **Za koju od preostale četiri varijable je svejedno je li opažena, ako su varijable x i z d-odvojene?**

- ☐ A u ☐ B w ☐ C v ☐ D y

- 13 (N) Razmotrite Bayesovu mrežu koja odgovara faktorizaciji $P(w, x, y, z) = P(w)P(x)P(y|w, x)P(z|x)$. Sve varijable su binarne. Vrijedi $P(w = 1) = 0.1$, $P(x = 1) = 0.2$, $P(z = 1|x = 0) = 0.9$ i $P(z = 1|x = 1) = 0.7$. Tablica uvjetnih vjerojatnosti za čvor y je sljedeća:

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

Postupkom uzorkovanja s odbijanjem želimo procijeniti parametar μ uvjetne distribucije $P(x = 0|y = 1, z = 0)$. Uzorkovanje smo ponovili ukupno $N = 100$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N . Na temelju dobivenog uzorka parametar μ procjenjujemo MAP procjeniteljem uz $\alpha = \beta = 2$. **Koliko iznosi očekivana MAP procjena parametra μ ?**

- ☐ A 0.1877 ☐ B 0.0786 ☐ C 0.1274 ☐ D 0.0490

Cjelina 6: Grupiranje i vrednovanje modela (9 pitanja)

- 14** (N) Na ispitnome skupu sa $K = 3$ klase vrednujemo algoritam multinomijalne logističke regresije. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci predikcije klasifikatora):

$$\begin{array}{c} y = 1 \quad y = 2 \quad y = 3 \\ \begin{array}{c} y = 1 \\ y = 2 \\ y = 3 \end{array} \begin{pmatrix} 15 & 3 & 1 \\ 9 & 5 & 4 \\ 4 & 2 & 3 \end{pmatrix} \end{array}$$

Kao referentni model koristimo klasifikator koji nasumično pogađa oznaku y , i to s jednakom vjerojatnošću za svaku od tri klase. Za vrednovanje koristimo mjeru F_1 -makro. Izračunajte očekivanu vrijednost mjere F_1 -makro za logističku regresiju i za referentni model. **Koliko iznosi očekivana razlika u mjeri F_1 -makro između logističke regresije i referentnog modela?**

- ☐ A 0.1420 ☐ B 0.1756 ☐ C 0.2238 ☐ D 0.0947

- 15** (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

- 16** (P) Skup neoznačenih primjera u dvodimenzijaskome ulaznom prostoru neka je sljedeći:

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

Primjere grupiramo algoritmom K-sredina sa $K = 2$ grupe. Grupiranje možemo shvatiti kao pretraživanje prostora stanja, gdje svako pojedino stanje odgovara jednom pridjeljivanju primjera grupama. Pritom dva stanja smatramo identičnima ako su grupiranja identična, neovisno o identitetu grupa (npr., grupiranje kod kojega prva grupa sadrži samo primjer $\mathbf{x}^{(1)}$ je identično kao i grupiranje kod kojega drugo grupa sadrži samo primjer $\mathbf{x}^{(1)}$). Neka je A_1 algoritam K-sredina s potpuno slučajno inicijaliziranim središtima, a A_2 algoritam K-sredina gdje su središta inicijalizirana algoritmom K-means++. Neka je $S(A_1)$ skup stanja koje pretražuje algoritam A_1 , a $S(A_2)$ skup stanja koje pretražuje algoritam A_2 . Izračunajte veličine ovih skupova, uzevši u obzir mogućnost da pojedina grupa bude prazna te da dođe do izjednačenja udaljenosti primjera do centroida, što se razrješava slučajnim mehanizmom. **Koliko algoritam A_2 pretražuje manje stanja od algoritma A_1 , tj. koliko iznosi $|S(A_1)| - |S(A_2)|$?**

- ☐ A 6 ☐ B 12 ☐ C 10 ☐ D 4

- 17** (P) Raspoložemo sa 1500 označenih primjera. Na tom skupu treniramo i vrednujemo algoritam SVM, optimizirajući hiperparametre C i γ . Za vrednovanje koristimo ugniježdenu 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 ugniježdene unakrsne provjere. **Za koliko bi taj broj narastao kada bismo broj preklopa u unutarnjoj petlji povećali na 10?**

- ☐ A 150 ☐ B 288 ☐ C 120 ☐ D 240

- 18 (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1, 2, 3, 4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{aligned} K = 3 & \quad \{\{1, 2, 4, 4\}, \{2, 2, 3\}, \{1, 1, 3, 4\}\} \\ K = 4 & \quad \{\{1, 2, 2, 4, 4\}, \{2, 3, 3\}, \{1, 1\}, \{4\}\} \end{aligned}$$

gdje podskupovi odgovaraju grupama, a brojke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

- ☐ A 0.0727 ☐ B 0.0182 ☐ C 0.0364 ☐ D 0.0545

- 19 (T) Za grupiranje primjera u K grupa koristimo model Gaussove mješavine (GMM) s dijeljenom kovarijacijskom matricom. Nakon grupiranja, odgovornosti zaokružujemo na cijeli broj, čime dobivamo tvrdo grupiranje. Iste podatke grupiramo algoritmom K-medoida (KM). **Uz koje parametre ovih algoritama očekujemo dobiti najbližnje rezultate grupiranja?**

- ☐ A GMM: $\Sigma_k = \sigma^2 \mathbf{I}$ i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost
☐ B GMM: puna Σ_k i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost
☐ C GMM: $\Sigma_k = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ i $\pi_k = 1/K$; KM: euklidska udaljenost
☐ D GMM: $\Sigma_k = \sigma^2 \mathbf{I}$; KM: euklidska udaljenost

- 20 (P) Skup neoznačenih primjera \mathcal{D} grupiramo modelom GMM treniranim EM-algoritmom. Koristimo nekoliko varijanti tog modela:

\mathcal{H}_1 : Model sa $K = 25$ središta inicijaliziranim algoritmom K-means++

\mathcal{H}_2 : Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom

\mathcal{H}_3 : Model sa $K = 50$ središta inicijaliziranim algoritmom K-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL_α^0 prosječna log-izglednost za model \mathcal{H}_α na početku izvođenja EM-algoritma, LL_α^* prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma te neka je k_α broj iteracija EM-algoritma za taj model. **Što možemo zaključiti o očekivanim odnosima između ovih vrijednosti?**

- ☐ A $LL_2^0 \geq LL_3^0$, $LL_3^* \geq LL_2^*$, $k_3 \geq k_2$ ☐ C $LL_1^0 \geq LL_2^0$, $LL_2^* \geq LL_3^*$, $k_1 \geq k_2$
☐ B $LL_3^0 \geq LL_2^0$, $LL_3^* \geq LL_2^*$, $k_2 \geq k_3$ ☐ D $LL_1^0 \geq LL_3^0$, $LL_1^* \geq LL_3^*$, $k_2 \geq k_1$

- 21 (T) Mjera točnosti nije prikladna za vrednovanje klasifikatora na skupovima podataka s neuravnoteženim brojem primjera po klasama. Jedna alternativa mjeri točnosti je F_1 -mjera, međutim ni ta mjera nije uvijek prikladna. Pretpostavite da vrijednost F_1 -mjere postavljamo na nulu u slučajevima kada je harmonijska sredina preciznosti i odziva nedefinirana. **U kojem slučaju F_1 -mjera ne bi bila prikladna mjera za vrednovanje klasifikatora jer bi bila previše optimistična?**

- ☐ A Ako je većina primjera pozitivna i klasifikator sve primjere klasificira pozitivno
☐ B Ako je većina primjera pozitivna, a klasifikator sve primjere klasificira negativno
☐ C Ako je većina primjera negativna i klasifikator sve primjere klasificira negativno
☐ D Ako je većina primjera negativna, a klasifikator sve primjere klasificira pozitivno

- 22 (T) Algoritam K-medoida općenitiji je od algoritma K-sredina budući da se može koristiti za primjere koji nisu prikazani kao vektori. Međutim, razmotrite slučaj kada primjeri jesu prikazani kao vektori, ali ih želimo grupirati na temelju mjere udaljenosti koja nije euklidska. **Koji bismo algoritam koristili u tom slučaju i zašto?**

- ☐ A Algoritam K-medoida, jer vektorizirani primjeri također mogu biti medoidi
☐ B Algoritam K-sredina, jer za vektorizirane primjere možemo izračunati centroide grupa
☐ C Algoritam K-sredina, jer koristi mjeru udaljenosti, dok algoritam K-medoida koristi mjeru sličnosti
☐ D Algoritam K-medoida, jer kriterijska funkcija algoritma K-sredina koristi euklidsku udaljenost

Završni ispit iz Strojnog učenja 1 (ak. god. 2022./2023.) – NEKORIGIRANA VERZIJA –

Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **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 klasifikaciju u dvije klase ($y = 1$ i $y = 2$) u dvodimenzijaskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klase su jednake, dok su izglednosti klase modelirane bivarijatnim Gaussovim distribucijama sa sljedećim parametrima:

$$\mu_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \Sigma_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \mu_2 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \Sigma_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klase definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

- ☐ A (3, 2), (4, 5), (9, 7) ☐ B (1, 2), (3, 4), (6, 7) ☐ C (1, 6), (3, 3), (5, 0) ☐ D (1, 3), (5, 4), (7, 9)

- 2** (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju jednodimenzijaskih podataka u tri klase. Procijenjene izglednosti klase 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 klase 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 klase 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 $[-4, -4 + a] \cup [5, 5 + b]$ ☐ C $[-4 - a, 5 + b]$ ☐ D $[-4 - a, -4 + b]$

- 3** (N) Treniramo polunaivan Bayesov klasifikator sa tri binarne značajke, x_1 , x_2 i x_3 . Skup primjera za učenje \mathcal{D} sastoji se od sljedećih deset primjera:

x_1	x_2	x_3	y	x_1	x_2	x_3	y
1	1	0	1	1	1	0	0
0	1	0	1	1	0	1	1
1	1	0	0	1	0	0	0
0	1	1	1	0	1	1	1
0	1	0	1	0	0	1	0

Prije treniranja koristimo uzajamnu informaciju kako bismo procijenili koje su varijable najviše statistički zavisne, jer se te varijable isplati združiti u zajednički faktor. Izračun provodimo tako da za svaki par varijabli x_i i x_j procjenjujemo parametre zajedničke distribucije $P(x_i, x_j)$, a zatim iz zajedničke distribucije računamo marginalne vjerojatnosti i uzajamnu informaciju $I(x_i, x_j)$. Budući da je skup \mathcal{D} malen, za procjenu parametara distribucije $P(x_i, x_j)$ koristimo Laplaceov procjenitelj. **Koliko iznosi na taj način izračunata uzajamna informacija između varijabli x_1 i x_2 ?**

- ☐ A 0.0423 ☐ B 0.0112 ☐ C 0.0334 ☐ D 0.0078

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

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

Na temelju beta-Bernoullijevog modela na skupu \mathcal{D} računamo MAP procjenu parametra μ Bernoullijeve distribucije. MLE procjena za isti parametar na skupu \mathcal{D} iznosi 0.3. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

- ☐ A $\alpha = 2, \beta = 10$ ☐ B $\alpha = 5, \beta = 7$ ☐ C $\alpha = 2, \beta = 5$ ☐ D $\alpha = 4, \beta = 8$

- 5 (T) Za Bayesov klasifikator procjenjujemo parametar μ Bernoullijeve distribucije. Skup primjera za učenje je razmjerno malen. Naš procjenitelj $\hat{\mu}$ parametra μ može biti pristran ili nepristran, dok model čiji je parametar procijenjen s $\hat{\mu}$ može biti prenaučan ili može dobro generalizirati. Razmotrite procjenitelje MLE i MAP (konkretno, Laplaceovo zaglađivanje). **Što od sljedećega općenito vrijedi u ovom slučaju?**

- ☐ A MLE procjenitelj je nepristran i očekujemo da će model dobro generalizirati
☐ B MAP procjenitelj je nepristran i očekujemo da će model dobro generalizirati
☐ C MLE procjenitelj je pristran i očekujemo da će model loše generalizirati
☐ D MAP procjenitelj je pristran, ali očekujemo da će model dobro generalizirati

- 6 (T) Gaussov Bayesov klasifikator s dijeljenom kovarijacijskom matricom (GBC) i logistička regresija (LR) su generativno-diskriminativni par modela. **Što to znači?**

- ☐ A GBC i neregularizirana LR modeliraju identične distribucije zajedničke vjerojatnosti primjera i oznaka, ali s različitim brojem parametara
☐ B Izlaz modela LR jednak je zajedničkoj vjerojatnosti modela GBC, ali model LR iziskuje manje parametara
☐ C Aposteriorna vjerojatnost klase za GBC može se izraziti kao poopceni linearni model sa sigmoidnom aktivacijskom funkcijom
☐ D Neovisno o optimizacijskom postupku, GBC i LR ostvaruju istu pogrešku na skupu za učenje, ali uz moguće različit broj parametara

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

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

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. Za procjenu kovarijacijske matrice koristite dva procjenitelja: MLE i nepristrani procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. Neka su h_1^{MLE} i h_1^{UB} hipoteze za prvu klasu modela s pristranom odnosno nepristranom procjenom kovarijacijske matrice. Zanima nas predikcija za primjer $\mathbf{x} = (1, 1)$. **Koliko iznosi apsolutna razlika u predikciji za \mathbf{x} za ova dva modela, $|h_1^{\text{MLE}}(\mathbf{x}) - h_1^{\text{UB}}(\mathbf{x})|$?**

- ☐ A 0.1638 ☐ B 0.2276 ☐ C 0.2537 ☐ D 0.0337

- 8 (N) Raspoložemo sljedećim skupom označenih primjera:

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

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

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

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

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

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

- 9 (P) Razmotrite Bayesovu mrežu sa šest čvorova koja odgovara sljedećoj faktorizaciji:

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

Upotrijebite metodu d-odvajanja da biste ispitali uvjetnu nezavisnost varijabli x i z u ovisnosti o preostale četiri varijable. Ispitajte koje varijable od preostalih četiri varijabli trebaju biti opažene a koje neopažene, a da bi varijable x i z bile d-odvojene. **Za koju od preostale četiri varijable je svejedno je li opažena, ako su varijable x i z d-odvojene?**

☐ A w ☐ B u ☐ C y ☐ D v

- 10 (N) Bayesovu mrežu koristimo za medicinsku dijagnostiku te modeliramo sljedeće kauzalne odnose. Upala grla ($U = 1$) može biti uzrokovana virusom ($V = 1$) ili bakterijom ($B = 1$). Povišena temperatura ($T = 1$) može biti uzrokovana upalom grla ili sunčanicom ($S = 1$). Sve varijable su binarne. Na temelju podataka o pacijentima procijenili smo parametre mreže: $P(V = 1) = 0.3$, $P(B = 1) = 0.3$ i $P(S = 1) = 0.05$. Uvjetne vjerojatnosti za čvorove U i T su:

V	B	$P(U = 1 V, B)$	U	S	$P(T = 1 U, S)$
0	0	0.2	0	0	0
0	1	0.5	0	1	0.2
1	0	0.4	1	0	0.4
1	1	0.7	1	1	0.4

Zanima nas koje je najvjerojatnije objašnjenje izravnog uzroka povišene temperature u pacijenata kod kojih nije dokazano prisustvo virusa. U tu svrhu računamo MAP-upit za par varijabli upita U i S uz opažene varijable $V = 0$ i $T = 1$, tj. računamo $\arg\max_{U, S} P(U, S|V = 0, T = 1)$. Neka je p_1 vjerojatnost najvjerojatnijeg (MAP) objašnjenja za varijable U i S , a p_2 vjerojatnost drugog po redu najvjerojatnijeg objašnjenja za te varijable. **Koliko je puta najvjerojatnije objašnjenje vjerojatnije od drugog najvjerojatnijeg objašnjenja, tj. koliko iznosi p_1/p_2 ?**

☐ A 17.88 ☐ B 11.35 ☐ C 15.52 ☐ D 13.35

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

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

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

☐ A 23 ☐ B 21 ☐ C 18 ☐ D 24

- 12 (N) Razmotrite Bayesovu mrežu koja odgovara faktorizaciji $P(w, x, y, z) = P(w)P(x)P(y|w, x)P(z|x)$. Sve varijable su binarne. Vrijedi $P(w = 1) = 0.1$, $P(x = 1) = 0.2$, $P(z = 1|x = 0) = 0.9$ i $P(z = 1|x = 1) = 0.7$. Tablica uvjetnih vjerojatnosti za čvor y je sljedeća:

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

Postupkom uzorkovanja s odbijanjem želimo procijeniti parametar μ uvjetne distribucije $P(x = 0|y = 1, z = 0)$. Uzorkovanje smo ponovili ukupno $N = 200$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N . Na temelju dobivenog uzorka parametar μ procjenjujemo MAP procjeniteljem uz $\alpha = \beta = 2$. **Koliko iznosi očekivana MAP procjena parametra μ ?**

☐ A 0.0786 ☐ B 0.0490 ☐ C 0.1877 ☐ D 0.1274

- 13 (T) Parametre probabilističkih grafičkih modela, uključivo Bayesove mreže, možemo procjenjivati iz potpunih podataka ili nepotpunih podataka. **Zašto i kako Bayesovu mrežu učimo nad nepotpunim podacima?**
- ☐ A Jer mreža ima manje čvorova nego što je opaženih varijabli, pa koristimo eliminaciju varijabli
- ☐ B Jer mreža ima skrivene varijable koje ne opažamo, pa moramo koristiti iterativne metode za MAP ili MLE
- ☐ C Jer MLE nema rješenje u zatvorenoj formi, pa umjesto MLE koristimo gradijentni uspon ili EM-algoritam
- ☐ D Jer se log-izglednost dekomponira po čvorovima mreže, pa MLE ili MAP procjenjujemo u zatvorenoj formi

Cjelina 6: Grupiranje i vrednovanje modela (9 pitanja)

- 14 (N) Na ispitnome skupu sa $K = 3$ klase vrednujemo algoritam multinomijalne logističke regresije. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci predikcije klasifikatora):

$$\begin{matrix} & y = 1 & y = 2 & y = 3 \\ \begin{matrix} y = 1 \\ y = 2 \\ y = 3 \end{matrix} & \begin{pmatrix} 15 & 3 & 5 \\ 9 & 5 & 4 \\ 4 & 2 & 3 \end{pmatrix} \end{matrix}$$

Kao referentni model koristimo klasifikator koji nasumično pogađa oznaku y , i to s jednakom vjerojatnošću za svaku od tri klase. Za vrednovanje koristimo mjeru F_1 -makro. Izračunajte očekivanu vrijednost mjere F_1 -makro za logističku regresiju i za referentni model. **Koliko iznosi očekivana razlika u mjeri F_1 -makro između logističke regresije i referentnog modela?**

- ☐ A 0.0947 ☐ B 0.2238 ☐ C 0.1756 ☐ D 0.1420
- 15 (P) Skup neoznačenih primjera u dvodimenzijaskome ulaznom prostoru neka je sljedeći:

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

Primjere grupiramo algoritmom K-sredina sa $K = 2$ grupe. Grupiranje možemo shvatiti kao pretraživanje prostora stanja, gdje svako pojedino stanje odgovara jednom pridjeljivanju primjera grupama. Pritom dva stanja smatramo identičnima ako su grupiranja identična, neovisno o identitetu grupa (npr., grupiranje kod kojega prva grupa sadrži samo primjer $\mathbf{x}^{(1)}$ je identično kao i grupiranje kod kojega druga grupa sadrži samo primjer $\mathbf{x}^{(1)}$). Neka je A_1 algoritam K-sredina s potpuno slučajno inicijaliziranim središtima, a A_2 algoritam K-sredina gdje su središta inicijalizirana algoritmom K-means++. Neka je $S(A_1)$ skup stanja koje pretražuje algoritam A_1 , a $S(A_2)$ skup stanja koje pretražuje algoritam A_2 . Izračunajte veličine ovih skupova, uzevši u obzir mogućnost da pojedina grupa bude prazna te da dođe do izjednačenja udaljenosti primjera do centroida, što se razrješava slučajnim mehanizmom. **Koliko algoritam A_2 pretražuje manje stanja od algoritma A_1 , tj. koliko iznosi $|S(A_1)| - |S(A_2)|$?**

- ☐ A 12 ☐ B 10 ☐ C 4 ☐ D 6
- 16 (P) Skup neoznačenih primjera \mathcal{D} grupiramo modelom GMM treniranim EM-algoritmom. Koristimo nekoliko varijanti tog modela:

\mathcal{H}_1 : Model sa $K = 25$ središta inicijaliziranim algoritmom K-means++

\mathcal{H}_2 : Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom

\mathcal{H}_3 : Model sa $K = 50$ središta inicijaliziranim algoritmom K-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL_α^0 prosječna log-izglednost za model \mathcal{H}_α na početku izvođenja EM-algoritma, LL_α^* prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma te neka je k_α broj iteracija EM-algoritma za taj model. **Što možemo zaključiti o očekivanim odnosima između ovih vrijednosti?**

- ☐ A $LL_2^0 \geq LL_3^0$, $LL_3^* \geq LL_2^*$, $k_3 \geq k_2$ ☐ C $LL_1^0 \geq LL_3^0$, $LL_1^* \geq LL_3^*$, $k_2 \geq k_1$
- ☐ B $LL_1^0 \geq LL_2^0$, $LL_2^* \geq LL_3^*$, $k_1 \geq k_2$ ☐ D $LL_3^0 \geq LL_2^0$, $LL_3^* \geq LL_2^*$, $k_2 \geq k_3$

- 17 (P) Raspoložemo sa 1250 označenih primjera. Na tom skupu treniramo i vrednujemo algoritam SVM, optimizirajući hiperparametre C i γ . Za vrednovanje koristimo ugniježdenu 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 ugniježdene unakrsne provjere. **Za koliko bi taj broj narastao kada bismo broj preklopa u unutarnjoj petlji povećali na 10?**

☐ A 125 ☐ B 200 ☐ C 100 ☐ D 240

- 18 (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1, 2, 3, 4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{aligned} K = 3 : & \quad \{\{1, 2, 4, 4\}, \{2, 3, 3\}, \{1, 1, 3, 4\}\} \\ K = 4 : & \quad \{\{1, 2, 2, 4, 4\}, \{3, 3\}, \{1, 1\}, \{3, 4\}\} \end{aligned}$$

gdje podskupovi odgovaraju grupama, a brojke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

☐ A 0.0727 ☐ B 0.0545 ☐ C 0.0364 ☐ D 0.0182

- 19 (T) Za grupiranje primjera u K grupa koristimo model Gaussove mješavine (GMM) s dijeljenom kovarijacijskom matricom. Nakon grupiranja, odgovornosti zaokružujemo na cijeli broj, čime dobivamo tvrdo grupiranje. Iste podatke grupiramo algoritmom K-medoida (KM). **Uz koje parametre ovih algoritama očekujemo dobiti najbližnje rezultate grupiranja?**

- ☐ A GMM: $\Sigma_k = \sigma^2 \mathbf{I}$; KM: euklidska udaljenost
☐ B GMM: puna Σ_k i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost
☐ C GMM: $\Sigma_k = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ i $\pi_k = 1/K$; KM: euklidska udaljenost
☐ D GMM: $\Sigma_k = \sigma^2 \mathbf{I}$ i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost

- 20 (T) Mjera točnosti nije prikladna za vrednovanje klasifikatora na skupovima podataka s neuravnoteženim brojem primjera po klasama. Jedna alternativa mjeri točnosti je F_1 -mjera, međutim ni ta mjera nije uvijek prikladna. Pretpostavite da vrijednost F_1 -mjere postavljamo na nulu u slučajevima kada je harmonijska sredina preciznosti i odziva nedefinirana. **U kojem slučaju F_1 -mjera ne bi bila prikladna mjera za vrednovanje klasifikatora jer bi bila previše optimistična?**

- ☐ A Ako je većina primjera pozitivna, a klasifikator sve primjere klasificira negativno
☐ B Ako je većina primjera negativna, a klasifikator sve primjere klasificira pozitivno
☐ C Ako je većina primjera pozitivna i klasifikator sve primjere klasificira pozitivno
☐ D Ako je većina primjera negativna i klasifikator sve primjere klasificira negativno

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

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

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 1.8 ☐ B 2.4 ☐ C 1.9 ☐ D 2.0

- 22 (T) Algoritam K-medoida općenitiji je od algoritma K-sredina budući da se može koristiti za primjere koji nisu prikazani kao vektori. Međutim, razmotrite slučaj kada primjeri jesu prikazani kao vektori, ali ih želimo grupirati na temelju mjere udaljenosti koja nije euklidska. **Koji bismo algoritam koristili u tom slučaju i zašto?**
- ☐ A Algoritam K-sredina, jer koristi mjeru udaljenosti, dok algoritam K-medoida koristi mjeru sličnosti
 - ☐ B Algoritam K-sredina, jer za vektorizirane primjere možemo izračunati centroide grupa
 - ☐ C Algoritam K-medoida, jer kriterijska funkcija algoritma K-sredina koristi euklidsku udaljenost
 - ☐ D Algoritam K-medoida, jer vektorizirani primjeri također mogu biti medoidi

Završni ispit iz Strojnog učenja 1 (ak. god. 2022./2023.) – NEKORIGIRANA VERZIJA –

Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **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 klasifikaciju jednodimenzijskih podataka u tri klase. Procijenjene izglednosti klase 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 klase 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 klase 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 $[-4, -4+a] \cup [5, 5+b]$ ☐ D $[-4-a, -4+b]$

- 2** (N) Raspoložemo sljedećim skupom označenih primjera:

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

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

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

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

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

- 3** (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y=1$ i $y=2$) u dvodimenzijskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klase su jednake, dok su izglednosti klase modelirane bivarijatnim Gausovim distribucijama sa sljedećim parametrima:

$$\mu_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \Sigma_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \mu_2 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \Sigma_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klase definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

- ☐ A $(1, 3), (5, 4), (7, 9)$ ☐ B $(3, 2), (4, 5), (9, 7)$ ☐ C $(1, 2), (3, 4), (6, 7)$ ☐ D $(1, 6), (3, 3), (5, 0)$

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

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

Na temelju beta-Bernoullijevog modela na skupu \mathcal{D} računamo MAP procjenu parametra μ Bernoullijeve distribucije. MLE procjena za isti parametar na skupu \mathcal{D} iznosi 0.3. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

- ☐ A $\alpha = 2, \beta = 10$ ☐ B $\alpha = 4, \beta = 8$ ☐ C $\alpha = 2, \beta = 5$ ☐ D $\alpha = 5, \beta = 7$

- 5 (N) Treniramo polunaivan Bayesov klasifikator sa tri binarne značajke, x_1 , x_2 i x_3 . Skup primjera za učenje \mathcal{D} sastoji se od sljedećih deset primjera:

x_1	x_2	x_3	y	x_1	x_2	x_3	y
0	1	0	1	1	1	0	0
0	1	0	1	1	0	1	1
1	1	0	0	1	0	0	0
0	1	1	1	0	1	1	1
0	1	0	1	0	0	1	0

Prije treniranja koristimo uzajamnu informaciju kako bismo procijenili koje su varijable najviše statistički zavisne, jer se te varijable isplati združiti u zajednički faktor. Izračun provodimo tako da za svaki par varijabli x_i i x_j procjenjujemo parametre zajedničke distribucije $P(x_i, x_j)$, a zatim iz zajedničke distribucije računamo marginalne vjerojatnosti i uzajamnu informaciju $I(x_i, x_j)$. Budući da je skup \mathcal{D} malen, za procjenu parametara distribucije $P(x_i, x_j)$ koristimo Laplaceov procjenitelj. **Koliko iznosi na taj način izračunata uzajamna informacija između varijabli x_1 i x_2 ?**

- A 0.0423 B 0.0112 C 0.0334 D 0.0078

- 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:

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

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. Za procjenu kovarijacijske matrice koristite dva procjenitelja: MLE i nepristrani procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. Neka su h_1^{MLE} i h_1^{UB} hipoteze za prvu klasu modela s pristranom odnosno nepristranom procjenom kovarijacijske matrice. Zanima nas predikcija za primjer $\mathbf{x} = (1, 1)$. **Koliko iznosi apsolutna razlika u predikciji za \mathbf{x} za ova dva modela, $|h_1^{\text{MLE}}(\mathbf{x}) - h_1^{\text{UB}}(\mathbf{x})|$?**

- A 0.2537 B 0.1638 C 0.2276 D 0.0337

- 7 (T) Za Bayesov klasifikator procjenjujemo parametar μ Bernoullijeve distribucije. Skup primjera za učenje je razmjerno malen. Naš procjenitelj $\hat{\mu}$ parametra μ može biti pristran ili nepristran, dok model čiji je parametar procijenjen s $\hat{\mu}$ može biti prenaučan ili može dobro generalizirati. Razmotrite procjenitelje MLE i MAP (konkretno, Laplaceovo zaglađivanje). **Što od sljedećega općenito vrijedi u ovom slučaju?**

- A MLE procjenitelj je nepristran i očekujemo da će model dobro generalizirati
 B MAP procjenitelj je pristran, ali očekujemo da će model dobro generalizirati
 C MAP procjenitelj je nepristran i očekujemo da će model dobro generalizirati
 D MLE procjenitelj je pristran i očekujemo da će model loše generalizirati

- 8 (T) Gaussov Bayesov klasifikator s dijeljenom kovarijacijskom matricom (GBC) i logistička regresija (LR) su generativno-diskriminativni par modela. **Što to znači?**

- A Neovisno o optimizacijskom postupku, GBC i LR ostvaruju istu pogrešku na skupu za učenje, ali uz moguće različit broj parametara
 B Izlaz modela LR jednak je zajedničkoj vjerojatnosti modela GBC, ali model LR iziskuje manje parametara
 C Aposteriorna vjerojatnost klase za GBC može se izraziti kao poopceni linearni model sa sigmoidnom aktivacijskom funkcijom
 D GBC i neregularizirana LR modeliraju identične distribucije zajedničke vjerojatnosti primjera i oznaka, ali s različitim brojem parametara

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

9 (T) Parametre probabilističkih grafičkih modela, uključivo Bayesove mreže, možemo procjenjivati iz potpunih podataka ili nepotpunih podataka. **Zašto i kako Bayesovu mrežu učimo nad nepotpunim podatcima?**

- ☐ A Jer mreža ima skrivene varijable koje ne opažamo, pa moramo koristiti iterativne metode za MAP ili MLE
- ☐ B Jer mreža ima manje čvorova nego što je opaženih varijabli, pa koristimo eliminaciju varijabli
- ☐ C Jer se log-izglednost dekomponira po čvorovima mreže, pa MLE ili MAP procjenjujemo u zatvorenoj formi
- ☐ D Jer MLE nema rješenje u zatvorenoj formi, pa umjesto MLE koristimo gradijentni uspon ili EM-algoritam

10 (P) Razmotrite Bayesovu mrežu sa šest čvorova koja odgovara sljedećoj faktorizaciji:

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

Upotrijebite metodu d-odvajanja da biste ispitati uvjetnu nezavisnost varijabli x i z u ovisnosti o preostale četiri varijable. Ispitajte koje varijable od preostalih četiri varijabli trebaju biti opažene a koje neopažene, a da bi varijable x i z bile d-odvojene. **Za koju od preostale četiri varijable je svejedno je li opažena, ako su varijable x i z d-odvojene?**

- ☐ A v ☐ B y ☐ C w ☐ D u

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

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

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

- ☐ A 24 ☐ B 21 ☐ C 23 ☐ D 18

12 (N) Razmotrite Bayesovu mrežu koja odgovara faktorizaciji $P(w, x, y, z) = P(w)P(x)P(y|w, x)P(z|x)$. Sve varijable su binarne. Vrijedi $P(w = 1) = 0.1$, $P(x = 1) = 0.2$, $P(z = 1|x = 0) = 0.9$ i $P(z = 1|x = 1) = 0.7$. Tablica uvjetnih vjerojatnosti za čvor y je sljedeća:

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

Postupkom uzorkovanja s odbijanjem želimo procijeniti parametar μ uvjetne distribucije $P(x = 0|y = 1, z = 0)$. Uzorkovanje smo ponovili ukupno $N = 200$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N . Na temelju dobivenog uzorka parametar μ procjenjujemo MAP procjeniteljem uz $\alpha = \beta = 2$. **Koliko iznosi očekivana MAP procjena parametra μ ?**

- ☐ A 0.0490 ☐ B 0.1274 ☐ C 0.0786 ☐ D 0.1877

13 (N) Bayesovu mrežu koristimo za medicinsku dijagnostiku te modeliramo sljedeće kauzalne odnose. Upala grla ($U = 1$) može biti uzrokovana virusom ($V = 1$) ili bakterijom ($B = 1$). Povišena temperatura ($T = 1$) može biti uzrokovana upalom grla ili sunčanicom ($S = 1$). Sve varijable su binarne. Na temelju podataka o pacijentima procijenili smo parametre mreže: $P(V = 1) = 0.3$, $P(B = 1) = 0.1$ i $P(S = 1) = 0.05$. Uvjetne vjerojatnosti za čvorove U i T su:

V	B	$P(U = 1 V, B)$	U	S	$P(T = 1 U, S)$
0	0	0.2	0	0	0
0	1	0.5	0	1	0.2
1	0	0.4	1	0	0.4
1	1	0.7	1	1	0.4

Zanima nas koje je najvjerojatnije objašnjenje izravnog uzroka povišene temperature u pacijenata kod kojih nije dokazano prisustvo virusa. U tu svrhu računamo MAP-upit za par varijabli upita U i S uz opažene varijable $V = 0$ i $T = 1$, tj. računamo $\operatorname{argmax}_{U, S} P(U, S|V = 0, T = 1)$. Neka je p_1 vjerojatnost najvjerojatnijeg (MAP)

objašnjenja za varijable U i S , a p_2 vjerojatnost drugog po redu najvjerojatnijeg objašnjenja za te varijable. **Koliko je puta najvjerojatnije objašnjenje vjerojatnije od drugog najvjerojatnijeg objašnjenja, tj. koliko iznosi p_1/p_2 ?**

- ☐ A 17.88 ☐ B 13.35 ☐ C 15.52 ☐ D 11.35

Cjelina 6: Grupiranje i vrednovanje modela (9 pitanja)

- 14** (P) Raspoložemo sa 1500 označenih primjera. Na tom skupu treniramo i vrednujemo algoritam SVM, optimizirajući hiperparametre C i γ . Za vrednovanje koristimo ugniježdenu 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 disjunktne: svaki par treniranih modela dijele određeni broj primjera za učenje. Izračunajte koliko primjera za učenje dijele svaki par modela koje treniramo u unutarnjoj petlji ugniježdene unakrsne provjere. **Za koliko bi taj broj narastao kada bismo broj preklopa u unutarnjoj petlji povećali na 10?**

- ☐ A 240 ☐ B 150 ☐ C 288 ☐ D 120

- 15** (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1, 2, 3, 4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{aligned} K = 3 : & \quad \{\{1, 2, 4, 4\}, \{2, 3, 3\}, \{1, 1, 3, 4\}\} \\ K = 4 : & \quad \{\{1, 2, 2, 4, 4\}, \{3, 3\}, \{1, 1\}, \{3, 4\}\} \end{aligned}$$

gdje podskupovi odgovaraju grupama, a brojeke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

- ☐ A 0.0364 ☐ B 0.0545 ☐ C 0.0727 ☐ D 0.0182

- 16** (T) Algoritam K-medoida općenitiji je od algoritma K-sredina budući da se može koristiti za primjere koji nisu prikazani kao vektori. Međutim, razmotrite slučaj kada primjeri jesu prikazani kao vektori, ali ih želimo grupirati na temelju mjere udaljenosti koja nije euklidska. **Koji bismo algoritam koristili u tom slučaju i zašto?**

- ☐ A Algoritam K-medoida, jer vektorizirani primjeri također mogu biti medoidi
☐ B Algoritam K-medoida, jer kriterijska funkcija algoritma K-sredina koristi euklidsku udaljenost
☐ C Algoritam K-sredina, jer koristi mjeru udaljenosti, dok algoritam K-medoida koristi mjeru sličnosti
☐ D Algoritam K-sredina, jer za vektorizirane primjere možemo izračunati centroide grupa

- 17** (T) Mjera točnosti nije prikladna za vrednovanje klasifikatora na skupovima podataka s neuravnoteženim brojem primjera po klasama. Jedna alternativa mjeri točnosti je F_1 -mjera, međutim ni ta mjera nije uvijek prikladna. Pretpostavite da vrijednost F_1 -mjere postavljamo na nulu u slučajevima kada je harmonijska sredina preciznosti i odziva nedefinirana. **U kojem slučaju F_1 -mjera ne bi bila prikladna mjera za vrednovanje klasifikatora jer bi bila previše optimistična?**

- ☐ A Ako je većina primjera negativna i klasifikator sve primjere klasificira negativno
☐ B Ako je većina primjera negativna, a klasifikator sve primjere klasificira pozitivno
☐ C Ako je većina primjera pozitivna, a klasifikator sve primjere klasificira negativno
☐ D Ako je većina primjera pozitivna i klasifikator sve primjere klasificira pozitivno

- 18** (P) Skup neoznačenih primjera \mathcal{D} grupiramo modelom GMM treniranim EM-algoritmom. Koristimo nekoliko varijanti tog modela:

\mathcal{H}_1 : Model sa $K = 25$ središta inicijaliziranim algoritmom K-means++

\mathcal{H}_2 : Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom

\mathcal{H}_3 : Model sa $K = 50$ središta inicijaliziranim algoritmom K-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL_α^0 prosječna log-izglednost za model \mathcal{H}_α na početku izvođenja EM-algoritma, LL_α^* prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma te neka je k_α broj iteracija EM-algoritma za taj model. **Što možemo zaključiti o očekivanim odnosima između ovih vrijednosti?**

- ☐ A $LL_3^0 \geq LL_2^0, LL_3^* \geq LL_2^*, k_2 \geq k_3$ ☐ C $LL_2^0 \geq LL_3^0, LL_3^* \geq LL_2^*, k_3 \geq k_2$
☐ B $LL_1^0 \geq LL_3^0, LL_1^* \geq LL_3^*, k_2 \geq k_1$ ☐ D $LL_1^0 \geq LL_2^0, LL_2^* \geq LL_3^*, k_1 \geq k_2$

- 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 1.9 ☐ B 2.0 ☐ C 2.4 ☐ D 1.8

- 20 (T) Za grupiranje primjera u K grupa koristimo model Gaussove mješavine (GMM) s dijeljenom kovarijacijskom matricom. Nakon grupiranja, odgovornosti zaokružujemo na cijeli broj, čime dobivamo tvrdo grupiranje. Iste podatke grupiramo algoritmom K-medoida (KM). **Uz koje parametre ovih algoritama očekujemo dobiti najsličnije rezultate grupiranja?**

- ☐ A GMM: $\Sigma_k = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ i $\pi_k = 1/K$; KM: euklidska udaljenost
☐ B GMM: puna Σ_k i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost
☐ C GMM: $\Sigma_k = \sigma^2 \mathbf{I}$ i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost
☐ D GMM: $\Sigma_k = \sigma^2 \mathbf{I}$; KM: euklidska udaljenost

- 21 (N) Na ispitnome skupu sa $K = 3$ klase vrednujemo algoritam multinomijalne logističke regresije. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci predikcije klasifikatora):

$$\begin{array}{ccc} & y = 1 & y = 2 & y = 3 \\ \begin{array}{c} y = 1 \\ y = 2 \\ y = 3 \end{array} & \begin{pmatrix} 15 & 3 & 5 \\ 9 & 15 & 4 \\ 4 & 2 & 3 \end{pmatrix} \end{array}$$

Kao referentni model koristimo klasifikator koji nasumično pogađa oznaku y , i to s jednakom vjerojatnošću za svaku od tri klase. Za vrednovanje koristimo mjeru F_1 -makro. Izračunajte očekivanu vrijednost mjere F_1 -makro za logističku regresiju i za referentni model. **Koliko iznosi očekivana razlika u mjeri F_1 -makro između logističke regresije i referentnog modela?**

- ☐ A 0.1420 ☐ B 0.0947 ☐ C 0.2238 ☐ D 0.1756

- 22 (P) Skup neoznačenih primjera u dvodimenzijaskome ulaznom prostoru neka je sljedeći:

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

Primjere grupiramo algoritmom K-sredina sa $K = 2$ grupe. Grupiranje možemo shvatiti kao pretraživanje prostora stanja, gdje svako pojedino stanje odgovara jednom pridjeljivanju primjera grupama. Pritom dva stanja smatramo identičnima ako su grupiranja identična, neovisno o identitetu grupa (npr., grupiranje kod kojega prva grupa sadrži samo primjer $\mathbf{x}^{(1)}$ je identično kao i grupiranje kod kojega drugo grupa sadrži samo primjer $\mathbf{x}^{(1)}$). Neka je A_1 algoritam K-sredina s potpuno slučajno inicijaliziranim središtima, a A_2 algoritam K-sredina gdje su središta inicijalizirana algoritmom K-means++. Neka je $S(A_1)$ skup stanja koje pretražuje algoritam A_1 , a $S(A_2)$ skup stanja koje pretražuje algoritam A_2 . Izračunajte veličine ovih skupova, uzevši u obzir mogućnost da pojedina grupa bude prazna te da dođe do izjednačenja udaljenosti primjera do centroida, što se razrješava slučajnim mehanizmom. **Koliko algoritam A_2 pretražuje manje stanja od algoritma A_1 , tj. koliko iznosi $|S(A_1)| - |S(A_2)|$?**

- ☐ A 4 ☐ B 6 ☐ C 10 ☐ D 12

Završni ispit iz Strojnog učenja 1 (ak. god. 2022./2023.)

– NEKORIGIRANA VERZIJA –

Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.

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

1 (T) Gaussov Bayesov klasifikator s dijeljenom kovarijacijskom matricom (GBC) i logistička regresija (LR) su generativno-diskriminativni par modela. **Što to znači?**

- ☐ A Izlaz modela LR jednak je zajedničkoj vjerojatnosti modela GBC, ali model LR iziskuje manje parametara
- ☐ B GBC i neregularizirana LR modeliraju identične distribucije zajedničke vjerojatnosti primjera i oznaka, ali s različitim brojem parametara
- ☐ C Neovisno o optimizacijskom postupku, GBC i LR ostvaruju istu pogrešku na skupu za učenje, ali uz moguće različit broj parametara
- ☐ D Aposteriorna vjerojatnost klase za GBC može se izraziti kao poopćeni linearni model sa sigmoidnom aktivacijskom funkcijom

2 (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju jednodimenzijskih podataka u tri klase. Procijenjene izglednosti klase 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 klase 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 klase 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 $[-4, -4 + a] \cup [5, 5 + b]$
- ☐ C $[-4 - a, -4 + b]$
- ☐ D $[-4 - a, -4] \cup [5 - b, 5]$

3 (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y=1$ i $y=2$) u dvodimenzijskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klase su jednake, dok su izglednosti klase modelirane bivarijantnim Gaussovim distribucijama sa sljedećim parametrima:

$$\mu_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \Sigma_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \mu_2 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \Sigma_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klase definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

- ☐ A $(3, 2), (4, 5), (9, 7)$
- ☐ B $(1, 3), (5, 4), (7, 9)$
- ☐ C $(1, 6), (3, 3), (5, 0)$
- ☐ D $(1, 2), (3, 4), (6, 7)$

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:

$$p(\mathbf{x}|y=j) = \frac{1}{(2\pi)^{n/2} |\Sigma_j|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu_j)^T \Sigma_j^{-1} (\mathbf{x} - \mu_j) \right\}$$

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

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

Procijenite parametre modela na ovom skupu primjera. Za procjenu kovarijacijske matrice koristite dva procjenitelja: MLE i nepristrani procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. Neka su h_1^{MLE} i h_1^{UB} hipoteze za prvu klasu modela s pristranom odnosno nepristranom procjenom kovarijacijske matrice. Zanima nas predikcija za primjer $\mathbf{x} = (1, 1)$. **Koliko iznosi apsolutna razlika u predikciji za \mathbf{x} za ova dva modela, $|h_1^{\text{MLE}}(\mathbf{x}) - h_1^{\text{UB}}(\mathbf{x})|$?**

- ☐ A 0.2537 ☐ B 0.2276 ☐ C 0.1638 ☐ D 0.0337

- 5** (N) Treniramo polunaivan Bayesov klasifikator sa tri binarne značajke, x_1 , x_2 i x_3 . Skup primjera za učenje \mathcal{D} sastoji se od sljedećih deset primjera:

x_1	x_2	x_3	y	x_1	x_2	x_3	y
1	1	0	1	1	1	0	0
0	1	0	1	1	0	1	1
1	1	0	0	1	0	0	0
0	1	1	1	0	1	1	1
0	1	0	1	0	0	1	0

Prije treniranja koristimo uzajamnu informaciju kako bismo procijenili koje su varijable najviše statistički zavisne, jer se te varijable isplati združiti u zajednički faktor. Izračun provodimo tako da za svaki par varijabli x_i i x_j procjenjujemo parametre zajedničke distribucije $P(x_i, x_j)$, a zatim iz zajedničke distribucije računamo marginalne vjerojatnosti i uzajamnu informaciju $I(x_i, x_j)$. Budući da je skup \mathcal{D} malen, za procjenu parametara distribucije $P(x_i, x_j)$ koristimo Laplaceov procjenitelj. **Koliko iznosi na taj način izračunata uzajamna informacija između varijabli x_1 i x_2 ?**

- ☐ A 0.0112 ☐ B 0.0423 ☐ C 0.0334 ☐ D 0.0078

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

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

Na temelju beta-Bernoullijevog modela na skupu \mathcal{D} računamo MAP procjenu parametra μ Bernoullijeve distribucije. MLE procjena za isti parametar na skupu \mathcal{D} iznosi 0.3. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

- ☐ A $\alpha = 5, \beta = 7$ ☐ B $\alpha = 4, \beta = 8$ ☐ C $\alpha = 2, \beta = 5$ ☐ D $\alpha = 2, \beta = 10$

- 7** (N) Raspoložemo sljedećim skupom označenih primjera:

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

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

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

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

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

- 8** (T) Za Bayesov klasifikator procjenjujemo parametar μ Bernoullijeve distribucije. Skup primjera za učenje je razmjerno malen. Naš procjenitelj $\hat{\mu}$ parametra μ može biti pristran ili nepristran, dok model čiji je parametar procijenjen s $\hat{\mu}$ može biti prenaučan ili može dobro generalizirati. Razmotrite procjenitelje MLE i MAP (konkretno, Laplaceovo zaglađivanje). **Što od sljedećega općenito vrijedi u ovom slučaju?**

- ☐ A MLE procjenitelj je nepristran i očekujemo da će model dobro generalizirati
☐ B MLE procjenitelj je pristran i očekujemo da će model loše generalizirati
☐ C MAP procjenitelj je pristran, ali očekujemo da će model dobro generalizirati
☐ D MAP procjenitelj je nepristran i očekujemo da će model dobro generalizirati

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

9 (T) Parametre probabilističkih grafičkih modela, uključivo Bayesove mreže, možemo procjenjivati iz potpunih podataka ili nepotpunih podataka. **Zašto i kako Bayesovu mrežu učimo nad nepotpunim podatcima?**

- ☐ A Jer mreža ima skrivene varijable koje ne opažamo, pa moramo koristiti iterativne metode za MAP ili MLE
- ☐ B Jer mreža ima manje čvorova nego što je opaženih varijabli, pa koristimo eliminaciju varijabli
- ☐ C Jer se log-izglednost dekomponira po čvorovima mreže, pa MLE ili MAP procjenjujemo u zatvorenoj formi
- ☐ D Jer MLE nema rješenje u zatvorenoj formi, pa umjesto MLE koristimo gradijentni uspon ili EM-algoritam

10 (N) Bayesovu mrežu koristimo za medicinsku dijagnostiku te modeliramo sljedeće kauzalne odnose. Upala grla ($U = 1$) može biti uzrokovana virusom ($V = 1$) ili bakterijom ($B = 1$). Povišena temperatura ($T = 1$) može biti uzrokovana upalom grla ili sunčanicom ($S = 1$). Sve varijable su binarne. Na temelju podataka o pacijentima procijenili smo parametre mreže: $P(V = 1) = 0.3$, $P(B = 1) = 0.4$ i $P(S = 1) = 0.05$. Uvjetne vjerojatnosti za čvorove U i T su:

V	B	$P(U = 1 V, B)$	U	S	$P(T = 1 U, S)$
0	0	0.2	0	0	0
0	1	0.5	0	1	0.2
1	0	0.4	1	0	0.4
1	1	0.7	1	1	0.4

Zanima nas koje je najvjerojatnije objašnjenje izravnog uzroka povišene temperature u pacijenata kod kojih nije dokazano prisustvo virusa. U tu svrhu računamo MAP-upit za par varijabli upita U i S uz opažene varijable $V = 0$ i $T = 1$, tj. računamo $\arg\max_{U, S} P(U, S|V = 0, T = 1)$. Neka je p_1 vjerojatnost najvjerojatnijeg (MAP) objašnjenja za varijable U i S , a p_2 vjerojatnost drugog po redu najvjerojatnijeg objašnjenja za te varijable. **Koliko je puta najvjerojatnije objašnjenje vjerojatnije od drugog najvjerojatnijeg objašnjenja, tj. koliko iznosi p_1/p_2 ?**

- ☐ A 13.35 ☐ B 17.88 ☐ C 11.35 ☐ D 15.52

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

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

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

- ☐ A 18 ☐ B 24 ☐ C 23 ☐ D 21

12 (N) Razmotrite Bayesovu mrežu koja odgovara faktorizaciji $P(w, x, y, z) = P(w)P(x)P(y|w, x)P(z|x)$. Sve varijable su binarne. Vrijedi $P(w = 1) = 0.1$, $P(x = 1) = 0.2$, $P(z = 1|x = 0) = 0.9$ i $P(z = 1|x = 1) = 0.7$. Tablica uvjetnih vjerojatnosti za čvor y je sljedeća:

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

Postupkom uzorkovanja s odbijanjem želimo procijeniti parametar μ uvjetne distribucije $P(x = 0|y = 1, z = 0)$. Uzorkovanje smo ponovili ukupno $N = 200$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N . Na temelju dobivenog uzorka parametar μ procjenjujemo MAP procjeniteljem uz $\alpha = \beta = 2$. **Koliko iznosi očekivana MAP procjena parametra μ ?**

- ☐ A 0.0786 ☐ B 0.1274 ☐ C 0.1877 ☐ D 0.0490

13 (P) Razmotrite Bayesovu mrežu sa šest čvorova koja odgovara sljedećoj faktorizaciji:

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

Upotrijebite metodu d-odvajanja da biste ispitali uvjetnu nezavisnost varijabli x i z u ovisnosti o preostale četiri varijable. Ispitajte koje varijable od preostalih četiri varijabli trebaju biti opažene a koje neopažene, a da bi varijable x i z bile d-odvojene. **Za koju od preostale četiri varijable je svejedno je li opažena, ako su varijable x i z d-odvojene?**

- ☐ A u ☐ B w ☐ C v ☐ D y

Cjelina 6: Grupiranje i vrednovanje modela (9 pitanja)

- 14** (T) Algoritam K-medoida općenitiji je od algoritma K-sredina budući da se može koristiti za primjere koji nisu prikazani kao vektori. Međutim, razmotrite slučaj kada primjeri jesu prikazani kao vektori, ali ih želimo grupirati na temelju mjere udaljenosti koja nije euklidska. **Koji bismo algoritam koristili u tom slučaju i zašto?**

- ☐ A Algoritam K-sredina, jer koristi mjeru udaljenosti, dok algoritam K-medoida koristi mjeru sličnosti
☐ B Algoritam K-sredina, jer za vektorizirane primjere možemo izračunati centroide grupa
☐ C Algoritam K-medoida, jer kriterijska funkcija algoritma K-sredina koristi euklidsku udaljenost
☐ D Algoritam K-medoida, jer vektorizirani primjeri također mogu biti medoidi

- 15** (T) Za grupiranje primjera u K grupa koristimo model Gaussove mješavine (GMM) s dijeljenom kovarijacijskom matricom. Nakon grupiranja, odgovornosti zaokružujemo na cijeli broj, čime dobivamo tvrdo grupiranje. Iste podatke grupiramo algoritmom K-medoida (KM). **Uz koje parametre ovih algoritama očekujemo dobiti najbližnje rezultate grupiranja?**

- ☐ A GMM: $\Sigma_k = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ i $\pi_k = 1/K$; KM: euklidska udaljenost
☐ B GMM: puna Σ_k i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost
☐ C GMM: $\Sigma_k = \sigma^2 \mathbf{I}$; KM: euklidska udaljenost
☐ D GMM: $\Sigma_k = \sigma^2 \mathbf{I}$ i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost

- 16** (P) Skup neoznačenih primjera \mathcal{D} grupiramo modelom GMM treniranim EM-algoritmom. Koristimo nekoliko varijanti tog modela:

\mathcal{H}_1 : Model sa $K = 25$ središta inicijaliziranim algoritmom K-means++

\mathcal{H}_2 : Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom

\mathcal{H}_3 : Model sa $K = 50$ središta inicijaliziranim algoritmom K-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL_α^0 prosječna log-izglednost za model \mathcal{H}_α na početku izvođenja EM-algoritma, LL_α^* prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma te neka je k_α broj iteracija EM-algoritma za taj model. **Što možemo zaključiti o očekivanim odnosima između ovih vrijednosti?**

- ☐ A $LL_1^0 \geq LL_2^0, LL_2^* \geq LL_3^*, k_1 \geq k_2$ ☐ C $LL_3^0 \geq LL_2^0, LL_3^* \geq LL_2^*, k_2 \geq k_3$
☐ B $LL_2^0 \geq LL_3^0, LL_3^* \geq LL_2^*, k_3 \geq k_2$ ☐ D $LL_1^0 \geq LL_3^0, LL_1^* \geq LL_3^*, k_2 \geq k_1$

- 17** (T) Mjera točnosti nije prikladna za vrednovanje klasifikatora na skupovima podataka s neuravnoteženim brojem primjera po klasama. Jedna alternativa mjeri točnosti je F_1 -mjera, međutim ni ta mjera nije uvijek prikladna. Pretpostavite da vrijednost F_1 -mjere postavljamo na nulu u slučajevima kada je harmonijska sredina preciznosti i odziva nedefinirana. **U kojem slučaju F_1 -mjera ne bi bila prikladna mjera za vrednovanje klasifikatora jer bi bila previše optimistična?**

- ☐ A Ako je većina primjera pozitivna, a klasifikator sve primjere klasificira negativno
☐ B Ako je većina primjera negativna, a klasifikator sve primjere klasificira pozitivno
☐ C Ako je većina primjera pozitivna i klasifikator sve primjere klasificira pozitivno
☐ D Ako je većina primjera negativna i klasifikator sve primjere klasificira negativno

- 18 (N) Na ispitnome skupu sa $K = 3$ klase vrednujemo algoritam multinomijalne logističke regresije. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci predikcije klasifikatora):

$$\begin{matrix} & y = 1 & y = 2 & y = 3 \\ \begin{matrix} y = 1 \\ y = 2 \\ y = 3 \end{matrix} & \begin{pmatrix} 15 & 3 & 1 \\ 9 & 15 & 4 \\ 4 & 2 & 3 \end{pmatrix} \end{matrix}$$

Kao referentni model koristimo klasifikator koji nasumično pogađa oznaku y , i to s jednakom vjerojatnošću za svaku od tri klase. Za vrednovanje koristimo mjeru F_1 -makro. Izračunajte očekivanu vrijednost mjere F_1 -makro za logističku regresiju i za referentni model. **Koliko iznosi očekivana razlika u mjeri F_1 -makro između logističke regresije i referentnog modela?**

- ☐ A 0.1420 ☐ B 0.2238 ☐ C 0.1756 ☐ D 0.0947

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

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

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 2.0 ☐ B 1.8 ☐ C 1.9 ☐ D 2.4

- 20 (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1, 2, 3, 4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{aligned} K = 3 & \quad \{\{1, 2, 4, 4\}, \{2, 2, 3\}, \{1, 1, 3, 4\}\} \\ K = 4 & \quad \{\{1, 2, 2, 4, 4\}, \{2, 3, 3\}, \{1, 1\}, \{4\}\} \end{aligned}$$

gdje podskupovi odgovaraju grupama, a brojke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

- ☐ A 0.0727 ☐ B 0.0182 ☐ C 0.0545 ☐ D 0.0364

- 21 (P) Raspoložemo 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 disjunktne: 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 125 ☐ B 200 ☐ C 240 ☐ D 100

- 22 (P) Skup neoznačenih primjera u dvodimenzijaskome ulaznom prostoru neka je sljedeći:

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

Primjere grupiramo algoritmom K-sredina sa $K = 2$ grupe. Grupiranje možemo shvatiti kao pretraživanje prostora stanja, gdje svako pojedino stanje odgovara jednom pridjeljivanju primjera grupama. Pritom dva stanja smatramo identičnima ako su grupiranja identična, neovisno o identitetu grupa (npr., grupiranje kod kojega prva grupa sadrži samo primjer $\mathbf{x}^{(1)}$ je identično kao i grupiranje kod kojega druga grupa sadrži samo primjer $\mathbf{x}^{(1)}$). Neka je A_1 algoritam K-sredina s potpuno slučajno inicijaliziranim središtima, a A_2 algoritam K-sredina gdje su središta inicijalizirana algoritmom K-means++. Neka je $S(A_1)$ skup stanja koje pretražuje algoritam A_1 , a $S(A_2)$ skup stanja koje pretražuje algoritam A_2 . Izračunajte veličine ovih skupova, uzevši u obzir mogućnost da pojedina grupa

bude prazna te da dođe do izjednačenja udaljenosti primjera do centroida, što se razrješava slučajnim mehanizmom.

Koliko algoritam A_2 pretražuje manje stanja od algoritma A_1 , tj. koliko iznosi $|S(A_1)| - |S(A_2)|$?

- ☐ A 4 ☐ B 10 ☐ C 6 ☐ D 12

Završni ispit iz Strojnog učenja 1 (ak. god. 2022./2023.)

– NEKORIGIRANA VERZIJA –

Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.

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

- 1** (T) Gaussov Bayesov klasifikator s dijeljenom kovarijacijskom matricom (GBC) i logistička regresija (LR) su generativno-diskriminativni par modela. **Što to znači?**

- ☐ A) Aposteriorna vjerojatnost klase za GBC može se izraziti kao poopceni linearni model sa sigmoidnom aktivacijskom funkcijom
- ☐ B) GBC i neregularizirana LR modeliraju identične distribucije zajedničke vjerojatnosti primjera i oznaka, ali s različitim brojem parametara
- ☐ C) Neovisno o optimizacijskom postupku, GBC i LR ostvaruju istu pogrešku na skupu za učenje, ali uz moguće različit broj parametara
- ☐ D) Izlaz modela LR jednak je zajedničkoj vjerojatnosti modela GBC, ali model LR iziskuje manje parametara

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

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

Na temelju beta-Bernoullijevog modela na skupu \mathcal{D} računamo MAP procjenu parametra μ Bernoullijeve distribucije. MLE procjena za isti parametar na skupu \mathcal{D} iznosi 0.2. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

- ☐ A) $\alpha = 5, \beta = 7$ ☐ B) $\alpha = 4, \beta = 8$ ☐ C) $\alpha = 2, \beta = 10$ ☐ D) $\alpha = 2, \beta = 5$

- 3** (N) Raspoložemo sljedećim skupom označenih primjera:

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

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

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

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

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

- 4** (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju 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 $[-4 - a, -4 + b]$ ☐ B $[-4 - a, -4] \cup [5, 5 + b]$ ☐ C $[-4, -4 + a] \cup [5 - b, 5]$ ☐ D $[-4 - a, 5 + b]$

- 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:

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

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. Za procjenu kovarijacijske matrice koristite dva procjenitelja: MLE i nepristrani procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. Neka su h_1^{MLE} i h_1^{UB} hipoteze za prvu klasu modela s pristranom odnosno nepristranom procjenom kovarijacijske matrice. Zanima nas predikcija za primjer $\mathbf{x} = (1, 1)$. **Koliko iznosi apsolutna razlika u predikciji za \mathbf{x} za ova dva modela, $|h_1^{\text{MLE}}(\mathbf{x}) - h_1^{\text{UB}}(\mathbf{x})|$?**

- ☐ A 0.2537 ☐ B 0.1638 ☐ C 0.0337 ☐ D 0.2276

- 6** (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y = 1$ i $y = 2$) u dvodimenzijaskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klase su jednake, dok su izglednosti klase modelirane bivarijantnim Gausovim distribucijama sa sljedećim parametrima:

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \Sigma_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \boldsymbol{\mu}_2 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \Sigma_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klase definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

- ☐ A $(1, 2), (3, 4), (6, 7)$ ☐ B $(3, 2), (4, 5), (9, 7)$ ☐ C $(1, 3), (5, 4), (7, 9)$ ☐ D $(1, 6), (3, 3), (5, 0)$

- 7** (T) Za Bayesov klasifikator procjenjujemo parametar μ Bernoullijeve distribucije. Skup primjera za učenje je razmjerno malen. Naš procjenitelj $\hat{\mu}$ parametra μ može biti pristran ili nepristran, dok model čiji je parametar procijenjen s $\hat{\mu}$ može biti prenaučan ili može dobro generalizirati. Razmotrite procjenitelje MLE i MAP (konkretno, Laplaceovo zaglađivanje). **Što od sljedećega općenito vrijedi u ovom slučaju?**

- ☐ A MAP procjenitelj je nepristran i očekujemo da će model dobro generalizirati
☐ B MAP procjenitelj je pristran, ali očekujemo da će model dobro generalizirati
☐ C MLE procjenitelj je pristran i očekujemo da će model loše generalizirati
☐ D MLE procjenitelj je nepristran i očekujemo da će model dobro generalizirati

- 8** (N) Treniramo polunaivan Bayesov klasifikator sa tri binarne značajke, x_1 , x_2 i x_3 . Skup primjera za učenje \mathcal{D} sastoji se od sljedećih deset primjera:

x_1	x_2	x_3	y	x_1	x_2	x_3	y
0	1	0	1	1	1	0	0
0	1	0	1	1	0	1	1
1	1	0	0	1	0	0	0
0	1	1	1	0	1	1	1
0	1	0	1	0	0	1	0

Prije treniranja koristimo uzajamnu informaciju kako bismo procijenili koje su varijable najviše statistički zavisne, jer se te varijable isplati združiti u zajednički faktor. Izračun provodimo tako da za svaki par varijabli x_i i x_j procjenjujemo parametre zajedničke distribucije $P(x_i, x_j)$, a zatim iz zajedničke distribucije računamo marginalne vjerojatnosti i uzajamnu informaciju $I(x_i, x_j)$. Budući da je skup \mathcal{D} malen, za procjenu parametara distribucije $P(x_i, x_j)$ koristimo Laplaceov procjenitelj. **Koliko iznosi na taj način izračunata uzajamna informacija između varijabli x_1 i x_2 ?**

- ☐ A 0.0423 ☐ B 0.0078 ☐ C 0.0112 ☐ D 0.0334

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

- 9 (N) Bayesovu mrežu koristimo za medicinsku dijagnostiku te modeliramo sljedeće kauzalne odnose. Upala grla ($U = 1$) može biti uzrokovana virusom ($V = 1$) ili bakterijom ($B = 1$). Povišena temperatura ($T = 1$) može biti uzrokovana upalom grla ili sunčanicom ($S = 1$). Sve varijable su binarne. Na temelju podataka o pacijentima procijenili smo parametre mreže: $P(V = 1) = 0.3$, $P(B = 1) = 0.1$ i $P(S = 1) = 0.05$. Uvjetne vjerojatnosti za čvorove U i T su:

V	B	$P(U = 1 V, B)$	U	S	$P(T = 1 U, S)$
0	0	0.2	0	0	0
0	1	0.5	0	1	0.2
1	0	0.4	1	0	0.4
1	1	0.7	1	1	0.4

Zanima nas koje je najvjerojatnije objašnjenje izravnog uzroka povišene temperature u pacijenata kod kojih nije dokazano prisustvo virusa. U tu svrhu računamo MAP-upit za par varijabli upita U i S uz opažene varijable $V = 0$ i $T = 1$, tj. računamo $\operatorname{argmax}_{U, S} P(U, S|V = 0, T = 1)$. Neka je p_1 vjerojatnost najvjerojatnijeg (MAP) objašnjenja za varijable U i S , a p_2 vjerojatnost drugog po redu najvjerojatnijeg objašnjenja za te varijable. **Koliko je puta najvjerojatnije objašnjenje vjerojatnije od drugog najvjerojatnijeg objašnjenja, tj. koliko iznosi p_1/p_2 ?**

- ☐ A 13.35 ☐ B 17.88 ☐ C 11.35 ☐ D 15.52

- 10 (T) Parametre probabilističkih grafičkih modela, uključivo Bayesove mreže, možemo procjenjivati iz potpunih podataka ili nepotpunih podataka. **Zašto i kako Bayesovu mrežu učimo nad nepotpunim podacima?**

- ☐ A Jer mreža ima skrivene varijable koje ne opažamo, pa moramo koristiti iterativne metode za MAP ili MLE
☐ B Jer mreža ima manje čvorova nego što je opaženih varijabli, pa koristimo eliminaciju varijabli
☐ C Jer MLE nema rješenje u zatvorenoj formi, pa umjesto MLE koristimo gradijentni uspon ili EM-algoritam
☐ D Jer se log-izglednost dekomponira po čvorovima mreže, pa MLE ili MAP procjenjujemo u zatvorenoj formi

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

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

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

- ☐ A 24 ☐ B 23 ☐ C 18 ☐ D 21

- 12 (N) Razmotrite Bayesovu mrežu koja odgovara faktorizaciji $P(w, x, y, z) = P(w)P(x)P(y|w, x)P(z|x)$. Sve varijable su binarne. Vrijedi $P(w = 1) = 0.1$, $P(x = 1) = 0.2$, $P(z = 1|x = 0) = 0.9$ i $P(z = 1|x = 1) = 0.7$. Tablica uvjetnih vjerojatnosti za čvor y je sljedeća:

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

Postupkom uzorkovanja s odbijanjem želimo procijeniti parametar μ uvjetne distribucije $P(x = 0|y = 1, z = 0)$. Uzorkovanje smo ponovili ukupno $N = 500$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N . Na temelju dobivenog uzorka parametar μ procjenjujemo MAP procjeniteljem uz $\alpha = \beta = 2$. **Koliko iznosi očekivana MAP procjena parametra μ ?**

- ☐ A 0.1274 ☐ B 0.0490 ☐ C 0.1877 ☐ D 0.0786

- 13 (P) Razmotrite Bayesovu mrežu sa šest čvorova koja odgovara sljedećoj faktorizaciji:

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

Upotrijebite metodu d-odvajanja da biste ispitali uvjetnu nezavisnost varijabli x i z u ovisnosti o preostale četiri varijable. Ispitajte koje varijable od preostalih četiri varijabli trebaju biti opažene a koje neopažene, a da bi varijable x i z bile d-odvojene. **Za koju od preostale četiri varijable je svejedno je li opažena, ako su varijable x i z d-odvojene?**

- ☐ A v ☐ B y ☐ C u ☐ D w

Cjelina 6: Grupiranje i vrednovanje modela (9 pitanja)

- 14** (P) Raspoložemo 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 ugniježdene unakrsne provjere. **Za koliko bi taj broj narastao kada bismo broj preklopa u unutarnjoj petlji povećali na 10?**

- ☐ A 144 ☐ B 120 ☐ C 75 ☐ D 60

- 15** (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1, 2, 3, 4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{aligned} K = 3 & \quad \{\{1, 2, 4, 4\}, \{2, 2, 3\}, \{1, 1, 3, 4\}\} \\ K = 4 & \quad \{\{1, 2, 2, 4, 4\}, \{2, 3\}, \{1, 1\}, \{3, 4\}\} \end{aligned}$$

gdje podskupovi odgovaraju grupama, a brojeke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

- ☐ A 0.0182 ☐ B 0.0545 ☐ C 0.0364 ☐ D 0.0727

- 16** (T) Algoritam K-medoida općenitiji je od algoritma K-sredina budući da se može koristiti za primjere koji nisu prikazani kao vektori. Međutim, razmotrite slučaj kada primjeri jesu prikazani kao vektori, ali ih želimo grupirati na temelju mjere udaljenosti koja nije euklidska. **Koji bismo algoritam koristili u tom slučaju i zašto?**

- ☐ A Algoritam K-sredina, jer koristi mjeru udaljenosti, dok algoritam K-medoida koristi mjeru sličnosti
☐ B Algoritam K-sredina, jer za vektorizirane primjere možemo izračunati centroide grupa
☐ C Algoritam K-medoida, jer vektorizirani primjeri također mogu biti medoidi
☐ D Algoritam K-medoida, jer kriterijska funkcija algoritma K-sredina koristi euklidsku udaljenost

- 17** (N) Na ispitnome skupu sa $K = 3$ klase vrednujemo algoritam multinomijalne logističke regresije. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci predikcije klasifikatora):

$$\begin{array}{c} y = 1 \quad y = 2 \quad y = 3 \\ \begin{array}{c} y = 1 \\ y = 2 \\ y = 3 \end{array} \left(\begin{array}{ccc} 15 & 3 & 1 \\ 9 & 5 & 4 \\ 4 & 2 & 3 \end{array} \right) \end{array}$$

Kao referentni model koristimo klasifikator koji nasumično pogađa oznaku y , i to s jednakom vjerojatnošću za svaku od tri klase. Za vrednovanje koristimo mjeru F_1 -makro. Izračunajte očekivanu vrijednost mjere F_1 -makro za logističku regresiju i za referentni model. **Koliko iznosi očekivana razlika u mjeri F_1 -makro između logističke regresije i referentnog modela?**

- ☐ A 0.0947 ☐ B 0.2238 ☐ C 0.1420 ☐ D 0.1756

- 18** (T) Za grupiranje primjera u K grupa koristimo model Gaussove mješavine (GMM) s dijeljenom kovarijacijskom matricom. Nakon grupiranja, odgovornosti zaokružujemo na cijeli broj, čime dobivamo tvrdo grupiranje. Iste

podatke grupiramo algoritmom K-medoida (KM). Uz koje parametre ovih algoritama očekujemo dobiti najslabije rezultate grupiranja?

- ☐ A GMM: $\Sigma_k = \sigma^2 \mathbf{I}$; KM: euklidska udaljenost
- ☐ B GMM: $\Sigma_k = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ i $\pi_k = 1/K$; KM: euklidska udaljenost
- ☐ C GMM: $\Sigma_k = \sigma^2 \mathbf{I}$ i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost
- ☐ D GMM: puna Σ_k i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost

19 (T) Mjera točnosti nije prikladna za vrednovanje klasifikatora na skupovima podataka s neuravnoteženim brojem primjera po klasama. Jedna alternativa mjeri točnosti je F_1 -mjera, međutim ni ta mjera nije uvijek prikladna. Pretpostavite da vrijednost F_1 -mjere postavljamo na nulu u slučajevima kada je harmonijska sredina preciznosti i odziva nedefinirana. U kojem slučaju F_1 -mjera ne bi bila prikladna mjera za vrednovanje klasifikatora jer bi bila previše optimistična?

- ☐ A Ako je većina primjera pozitivna, a klasifikator sve primjere klasificira negativno
- ☐ B Ako je većina primjera pozitivna i klasifikator sve primjere klasificira pozitivno
- ☐ C Ako je većina primjera negativna, a klasifikator sve primjere klasificira pozitivno
- ☐ D Ako je većina primjera negativna i klasifikator sve primjere klasificira negativno

20 (P) Skup neoznačenih primjera u dvodimenzijaskome ulaznom prostoru neka je sljedeći:

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

Primjere grupiramo algoritmom K-sredina sa $K = 2$ grupe. Grupiranje možemo shvatiti kao pretraživanje prostora stanja, gdje svako pojedino stanje odgovara jednom pridjeljivanju primjera grupama. Pritom dva stanja smatramo identičnima ako su grupiranja identična, neovisno o identitetu grupa (npr., grupiranje kod kojega prva grupa sadrži samo primjer $\mathbf{x}^{(1)}$ je identično kao i grupiranje kod kojega druga grupa sadrži samo primjer $\mathbf{x}^{(1)}$). Neka je A_1 algoritam K-sredina s potpuno slučajno inicijaliziranim središtima, a A_2 algoritam K-sredina gdje su središta inicijalizirana algoritmom K-means++. Neka je $S(A_1)$ skup stanja koje pretražuje algoritam A_1 , a $S(A_2)$ skup stanja koje pretražuje algoritam A_2 . Izračunajte veličine ovih skupova, uzevši u obzir mogućnost da pojedina grupa bude prazna te da dođe do izjednačenja udaljenosti primjera do centroida, što se razrješava slučajnim mehanizmom. **Koliko algoritam A_2 pretražuje manje stanja od algoritma A_1 , tj. koliko iznosi $|S(A_1)| - |S(A_2)|$?**

- ☐ A 4 ☐ B 12 ☐ C 6 ☐ D 10

21 (P) Skup neoznačenih primjera \mathcal{D} grupiramo modelom GMM treniranim EM-algoritmom. Koristimo nekoliko varijanti tog modela:

- \mathcal{H}_1 : Model sa $K = 25$ središta inicijaliziranim algoritmom K-means++
- \mathcal{H}_2 : Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom
- \mathcal{H}_3 : Model sa $K = 50$ središta inicijaliziranim algoritmom K-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL_α^0 prosječna log-izglednost za model \mathcal{H}_α na početku izvođenja EM-algoritma, LL_α^* prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma te neka je k_α broj iteracija EM-algoritma za taj model. **Što možemo zaključiti o očekivanim odnosima između ovih vrijednosti?**

- ☐ A $LL_1^0 \geq LL_2^0$, $LL_2^* \geq LL_3^*$, $k_1 \geq k_2$ ☐ C $LL_3^0 \geq LL_2^0$, $LL_3^* \geq LL_2^*$, $k_2 \geq k_3$
- ☐ B $LL_1^0 \geq LL_3^0$, $LL_1^* \geq LL_3^*$, $k_2 \geq k_1$ ☐ D $LL_2^0 \geq LL_3^0$, $LL_3^* \geq LL_2^*$, $k_3 \geq k_2$

22 (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 1.9 ☐ C 1.8 ☐ D 2.0

Završni ispit iz Strojnog učenja 1 (ak. god. 2022./2023.)

– NEKORIGIRANA VERZIJA –

Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.

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

- 1** (T) Gaussov Bayesov klasifikator s dijeljenom kovarijacijskom matricom (GBC) i logistička regresija (LR) su generativno-diskriminativni par modela. **Što to znači?**
- ☐ A Izlaz modela LR jednak je zajedničkoj vjerojatnosti modela GBC, ali model LR iziskuje manje parametara
- ☐ B Aposteriorna vjerojatnost klase za GBC može se izraziti kao poopćeni linearni model sa sigmoidnom aktivacijskom funkcijom
- ☐ C Neovisno o optimizacijskom postupku, GBC i LR ostvaruju istu pogrešku na skupu za učenje, ali uz moguće različit broj parametara
- ☐ D GBC i neregularizirana LR modeliraju identične distribucije zajedničke vjerojatnosti primjera i oznaka, ali s različitim brojem parametara
- 2** (T) Za Bayesov klasifikator procjenjujemo parametar μ Bernoullijeve distribucije. Skup primjera za učenje je razmjerno malen. Naš procjenitelj $\hat{\mu}$ parametra μ može biti pristran ili nepristran, dok model čiji je parametar procijenjen s $\hat{\mu}$ može biti prenaučan ili može dobro generalizirati. Razmotrite procjenitelje MLE i MAP (konkretno, Laplaceovo zaglađivanje). **Što od sljedećega općenito vrijedi u ovom slučaju?**
- ☐ A MAP procjenitelj je pristran, ali očekujemo da će model dobro generalizirati
- ☐ B MLE procjenitelj je nepristran i očekujemo da će model dobro generalizirati
- ☐ C MLE procjenitelj je pristran i očekujemo da će model loše generalizirati
- ☐ D MAP procjenitelj je nepristran i očekujemo da će model dobro generalizirati
- 3** (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:

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

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. Za procjenu kovarijacijske matrice koristite dva procjenitelja: MLE i nepristrani procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. Neka su h_1^{MLE} i h_1^{UB} hipoteze za prvu klasu modela s pristranom odnosno nepristranom procjenom kovarijacijske matrice. Zanima nas predikcija za primjer $\mathbf{x} = (1, 1)$. **Koliko iznosi apsolutna razlika u predikciji za \mathbf{x} za ova dva modela, $|h_1^{\text{MLE}}(\mathbf{x}) - h_1^{\text{UB}}(\mathbf{x})|$?**

- ☐ A 0.2537 ☐ B 0.1638 ☐ C 0.2276 ☐ D 0.0337

- 4** (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y = 1$ i $y = 2$) u dvodimenzijaskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klase su jednake, dok su izglednosti klase modelirane bivarijatnim Gaussovim distribucijama sa sljedećim parametrima:

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \Sigma_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \boldsymbol{\mu}_2 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \Sigma_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klasa definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

- A** (1, 6), (3, 3), (5, 0) **B** (1, 3), (5, 4), (7, 9) **C** (3, 2), (4, 5), (9, 7) **D** (1, 2), (3, 4), (6, 7)

- 5** (N) Treniramo polunaivan Bayesov klasifikator sa tri binarne značajke, x_1 , x_2 i x_3 . Skup primjera za učenje \mathcal{D} sastoji se od sljedećih deset primjera:

x_1	x_2	x_3	y	x_1	x_2	x_3	y
0	1	0	1	1	1	0	0
0	1	0	1	1	0	1	1
1	1	0	0	1	0	0	0
0	1	1	1	0	1	1	1
0	1	0	1	0	0	1	0

Prije treniranja koristimo uzajamnu informaciju kako bismo procijenili koje su varijable najviše statistički zavisne, jer se te varijable isplati združiti u zajednički faktor. Izračun provodimo tako da za svaki par varijabli x_i i x_j procjenjujemo parametre zajedničke distribucije $P(x_i, x_j)$, a zatim iz zajedničke distribucije računamo marginalne vjerojatnosti i uzajamnu informaciju $I(x_i, x_j)$. Budući da je skup \mathcal{D} malen, za procjenu parametara distribucije $P(x_i, x_j)$ koristimo Laplaceov procjenitelj. **Koliko iznosi na taj način izračunata uzajamna informacija između varijabli x_1 i x_2 ?**

- A** 0.0334 **B** 0.0423 **C** 0.0112 **D** 0.0078

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

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

Na temelju beta-Bernoullijevog modela na skupu \mathcal{D} računamo MAP procjenu parametra μ Bernoullijeve distribucije. MLE procjena za isti parametar na skupu \mathcal{D} iznosi 0.1. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

- A** $\alpha = 2, \beta = 10$ **B** $\alpha = 5, \beta = 7$ **C** $\alpha = 2, \beta = 5$ **D** $\alpha = 4, \beta = 8$

- 7** (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju 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** $[-4-a, -4] \cup [5, 5+b]$

- 8** (N) Raspoložemo sljedećim skupom označenih primjera:

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

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

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

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

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

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

- 9 (N) Razmotrite Bayesovu mrežu koja odgovara faktorizaciji $P(w, x, y, z) = P(w)P(x)P(y|w, x)P(z|x)$. Sve varijable su binarne. Vrijedi $P(w = 1) = 0.1$, $P(x = 1) = 0.2$, $P(z = 1|x = 0) = 0.9$ i $P(z = 1|x = 1) = 0.7$. Tablica uvjetnih vjerojatnosti za čvor y je sljedeća:

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

Postupkom uzorkovanja s odbijanjem želimo procijeniti parametar μ uvjetne distribucije $P(x = 0|y = 1, z = 0)$. Uzorkovanje smo ponovili ukupno $N = 2000$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N . Na temelju dobivenog uzorka parametar μ procjenjujemo MAP procjeniteljem uz $\alpha = \beta = 2$. **Koliko iznosi očekivana MAP procjena parametra μ ?**

- ☐ A 0.1877 ☐ B 0.0490 ☐ C 0.1274 ☐ D 0.0786

- 10 (T) Parametre probabilističkih grafičkih modela, uključivo Bayesove mreže, možemo procjenjivati iz potpunih podataka ili nepotpunih podataka. **Zašto i kako Bayesovu mrežu učimo nad nepotpunim podacima?**

- ☐ A Jer mreža ima skrivene varijable koje ne opažamo, pa moramo koristiti iterativne metode za MAP ili MLE
☐ B Jer mreža ima manje čvorova nego što je opaženih varijabli, pa koristimo eliminaciju varijabli
☐ C Jer se log-izglednost dekomponira po čvorovima mreže, pa MLE ili MAP procjenjujemo u zatvorenoj formi
☐ D Jer MLE nema rješenje u zatvorenoj formi, pa umjesto MLE koristimo gradijentni uspon ili EM-algoritam

- 11 (P) Razmotrite Bayesovu mrežu sa šest čvorova koja odgovara sljedećoj faktorizaciji:

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

Upotrijebite metodu d-odvajanja da biste ispitati uvjetnu nezavisnost varijabli x i z u ovisnosti o preostale četiri varijable. Ispitajte koje varijable od preostalih četiri varijabli trebaju biti opažene a koje neopažene, a da bi varijable x i z bile d-odvojene. **Za koju od preostale četiri varijable je svejedno je li opažena, ako su varijable x i z d-odvojene?**

- ☐ A w ☐ B v ☐ C y ☐ D u

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

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

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

- ☐ A 18 ☐ B 23 ☐ C 21 ☐ D 24

- 13 (N) Bayesovu mrežu koristimo za medicinsku dijagnostiku te modeliramo sljedeće kauzalne odnose. Upala grla ($U = 1$) može biti uzrokovana virusom ($V = 1$) ili bakterijom ($B = 1$). Povišena temperatura ($T = 1$) može biti uzrokovana upalom grla ili sunčanicom ($S = 1$). Sve varijable su binarne. Na temelju podataka o pacijentima procijenili smo parametre mreže: $P(V = 1) = 0.3$, $P(B = 1) = 0.3$ i $P(S = 1) = 0.05$. Uvjetne vjerojatnosti za čvorove U i T su:

V	B	$P(U = 1 V, B)$	U	S	$P(T = 1 U, S)$
0	0	0.2	0	0	0
0	1	0.5	0	1	0.2
1	0	0.4	1	0	0.4
1	1	0.7	1	1	0.4

Zanima nas koje je najvjerojatnije objašnjenje izravnog uzroka povišene temperature u pacijenata kod kojih nije dokazano prisustvo virusa. U tu svrhu računamo MAP-upit za par varijabli upita U i S uz opažene varijable $V = 0$ i $T = 1$, tj. računamo $\operatorname{argmax}_{U, S} P(U, S|V = 0, T = 1)$. Neka je p_1 vjerojatnost najvjerojatnijeg (MAP)

objašnjenja za varijable U i S , a p_2 vjerojatnost drugog po redu najvjerojatnijeg objašnjenja za te varijable. **Koliko je puta najvjerojatnije objašnjenje vjerojatnije od drugog najvjerojatnijeg objašnjenja, tj. koliko iznosi p_1/p_2 ?**

- ☐ A 17.88 ☐ B 15.52 ☐ C 11.35 ☐ D 13.35

Cjelina 6: Grupiranje i vrednovanje modela (9 pitanja)

- 14** (N) Na ispitnome skupu sa $K = 3$ klase vrednujemo algoritam multinomijalne logističke regresije. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci predikcije klasifikatora):

$$\begin{array}{c} y = 1 \quad y = 2 \quad y = 3 \\ \begin{array}{c} y = 1 \\ y = 2 \\ y = 3 \end{array} \begin{pmatrix} 15 & 3 & 1 \\ 9 & 15 & 4 \\ 4 & 2 & 3 \end{pmatrix} \end{array}$$

Kao referentni model koristimo klasifikator koji nasumično pogađa oznaku y , i to s jednakom vjerojatnošću za svaku od tri klase. Za vrednovanje koristimo mjeru F_1 -makro. Izračunajte očekivanu vrijednost mjere F_1 -makro za logističku regresiju i za referentni model. **Koliko iznosi očekivana razlika u mjeri F_1 -makro između logističke regresije i referentnog modela?**

- ☐ A 0.1756 ☐ B 0.0947 ☐ C 0.2238 ☐ D 0.1420

- 15** (P) Skup neoznačenih primjera \mathcal{D} grupiramo modelom GMM treniranim EM-algoritmom. Koristimo nekoliko varijanti tog modela:

\mathcal{H}_1 : Model sa $K = 25$ središta inicijaliziranim algoritmom K-means++

\mathcal{H}_2 : Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom

\mathcal{H}_3 : Model sa $K = 50$ središta inicijaliziranim algoritmom K-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL_α^0 prosječna log-izglednost za model \mathcal{H}_α na početku izvođenja EM-algoritma, LL_α^* prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma te neka je k_α broj iteracija EM-algoritma za taj model. **Što možemo zaključiti o očekivanim odnosima između ovih vrijednosti?**

- ☐ A $LL_2^0 \geq LL_3^0$, $LL_3^* \geq LL_2^*$, $k_3 \geq k_2$ ☐ C $LL_1^0 \geq LL_2^0$, $LL_2^* \geq LL_3^*$, $k_1 \geq k_2$
☐ B $LL_3^0 \geq LL_2^0$, $LL_3^* \geq LL_2^*$, $k_2 \geq k_3$ ☐ D $LL_1^0 \geq LL_3^0$, $LL_1^* \geq LL_3^*$, $k_2 \geq k_1$

- 16** (P) Raspoložemo 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 ugniježdene unakrsne provjere. **Za koliko bi taj broj narastao kada bismo broj preklopa u unutarnjoj petlji povećali na 10?**

- ☐ A 144 ☐ B 60 ☐ C 120 ☐ D 75

- 17** (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1, 2, 3, 4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{array}{l} K = 3 \quad \{\{1, 2, 4, 4\}, \{2, 2, 3\}, \{1, 1, 3, 4\}\} \\ K = 4 \quad \{\{1, 2, 2, 4, 4\}, \{2, 3\}, \{1, 1\}, \{3, 4\}\} \end{array}$$

gdje podskupovi odgovaraju grupama, a brojke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

- ☐ A 0.0364 ☐ B 0.0545 ☐ C 0.0182 ☐ D 0.0727

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

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

Primjere grupiramo algoritmom K-sredina sa $K = 2$ grupe. Grupiranje možemo shvatiti kao pretraživanje prostora stanja, gdje svako pojedino stanje odgovara jednom pridjeljivanju primjera grupama. Pritom dva stanja smatramo identičnima ako su grupiranja identična, neovisno o identitetu grupa (npr., grupiranje kod kojega prva grupa sadrži samo primjer $\mathbf{x}^{(1)}$ je identično kao i grupiranje kod kojega druga grupa sadrži samo primjer $\mathbf{x}^{(1)}$). Neka je A_1 algoritam K-sredina s potpuno slučajno inicijaliziranim središtima, a A_2 algoritam K-sredina gdje su središta inicijalizirana algoritmom K-means++. Neka je $S(A_1)$ skup stanja koje pretražuje algoritam A_1 , a $S(A_2)$ skup stanja koje pretražuje algoritam A_2 . Izračunajte veličine ovih skupova, uzevši u obzir mogućnost da pojedina grupa bude prazna te da dođe do izjednačenja udaljenosti primjera do centroida, što se razrješava slučajnim mehanizmom. **Koliko algoritam A_2 pretražuje manje stanja od algoritma A_1 , tj. koliko iznosi $|S(A_1)| - |S(A_2)|$?**

- ☐ A 10 ☐ B 12 ☐ C 4 ☐ D 6

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

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

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 1.9 ☐ B 2.0 ☐ C 2.4 ☐ D 1.8

- 20 (T) Za grupiranje primjera u K grupa koristimo model Gaussove mješavine (GMM) s dijeljenom kovarijacijskom matricom. Nakon grupiranja, odgovornosti zaokružujemo na cijeli broj, čime dobivamo tvrdo grupiranje. Iste podatke grupiramo algoritmom K-medoida (KM). **Uz koje parametre ovih algoritama očekujemo dobiti najbližnje rezultate grupiranja?**

- ☐ A GMM: $\Sigma_k = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ i $\pi_k = 1/K$; KM: euklidska udaljenost
☐ B GMM: $\Sigma_k = \sigma^2 \mathbf{I}$ i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost
☐ C GMM: $\Sigma_k = \sigma^2 \mathbf{I}$; KM: euklidska udaljenost
☐ D GMM: puna Σ_k i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost

- 21 (T) Algoritam K-medoida općenitiji je od algoritma K-sredina budući da se može koristiti za primjere koji nisu prikazani kao vektori. Međutim, razmotrite slučaj kada primjeri jesu prikazani kao vektori, ali ih želimo grupirati na temelju mjere udaljenosti koja nije euklidska. **Koji bismo algoritam koristili u tom slučaju i zašto?**

- ☐ A Algoritam K-medoida, jer vektorizirani primjeri također mogu biti medoidi
☐ B Algoritam K-sredina, jer za vektorizirane primjere možemo izračunati centroide grupa
☐ C Algoritam K-sredina, jer koristi mjeru udaljenosti, dok algoritam K-medoida koristi mjeru sličnosti
☐ D Algoritam K-medoida, jer kriterijska funkcija algoritma K-sredina koristi euklidsku udaljenost

- 22 (T) Mjera točnosti nije prikladna za vrednovanje klasifikatora na skupovima podataka s neuravnoteženim brojem primjera po klasama. Jedna alternativa mjeri točnosti je F_1 -mjera, međutim ni ta mjera nije uvijek prikladna. Pretpostavite da vrijednost F_1 -mjere postavljamo na nulu u slučajevima kada je harmonijska sredina preciznosti i odziva nedefinirana. **U kojem slučaju F_1 -mjera ne bi bila prikladna mjera za vrednovanje klasifikatora jer bi bila previše optimistična?**

- ☐ A Ako je većina primjera negativna i klasifikator sve primjere klasificira negativno
☐ B Ako je većina primjera pozitivna i klasifikator sve primjere klasificira pozitivno
☐ C Ako je većina primjera pozitivna, a klasifikator sve primjere klasificira negativno
☐ D Ako je većina primjera negativna, a klasifikator sve primjere klasificira pozitivno

Završni ispit iz Strojnog učenja 1 (ak. god. 2022./2023.)

– NEKORIGIRANA VERZIJA –

Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a $1/3$ boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.

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

- 1** (N) Treniramo polunaivan Bayesov klasifikator sa tri binarne značajke, x_1 , x_2 i x_3 . Skup primjera za učenje \mathcal{D} sastoji se od sljedećih deset primjera:

x_1	x_2	x_3	y	x_1	x_2	x_3	y
0	1	0	1	1	1	0	0
0	1	0	1	1	0	1	1
1	1	0	0	1	0	0	0
0	1	1	1	0	1	1	1
0	1	0	1	0	0	1	0

Prije treniranja koristimo uzajamnu informaciju kako bismo procijenili koje su varijable najviše statistički zavisne, jer se te varijable isplati združiti u zajednički faktor. Izračun provodimo tako da za svaki par varijabli x_i i x_j procjenjujemo parametre zajedničke distribucije $P(x_i, x_j)$, a zatim iz zajedničke distribucije računamo marginalne vjerojatnosti i uzajamnu informaciju $I(x_i, x_j)$. Budući da je skup \mathcal{D} malen, za procjenu parametara distribucije $P(x_i, x_j)$ koristimo Laplaceov procjenitelj. **Koliko iznosi na taj način izračunata uzajamna informacija između varijabli x_1 i x_2 ?**

- ☐ A 0.0078 ☐ B 0.0334 ☐ C 0.0423 ☐ D 0.0112

- 2** (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju jednodimenzijskih podataka u tri klase. Procijenjene izglednosti klase 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 klase 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 klase 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 $[-4 - a, -4 + b]$ ☐ C $[-4, -4 - a] \cup [5, 5 + b]$ ☐ D $[-4, -4 + a] \cup [5 - b, 5]$

- 3** (T) Za Bayesov klasifikator procjenjujemo parametar μ Bernoullijeve distribucije. Skup primjera za učenje je razmjerno malen. Naš procjenitelj $\hat{\mu}$ parametra μ može biti pristran ili nepristran, dok model čiji je parametar procijenjen s $\hat{\mu}$ može biti prenaučan ili može dobro generalizirati. Razmotrite procjenitelje MLE i MAP (konkretno, Laplaceovo zaglađivanje). **Što od sljedećega općenito vrijedi u ovom slučaju?**

- ☐ A MAP procjenitelj je pristran, ali očekujemo da će model dobro generalizirati
☐ B MAP procjenitelj je nepristran i očekujemo da će model dobro generalizirati
☐ C MLE procjenitelj je pristran i očekujemo da će model loše generalizirati
☐ D MLE procjenitelj je nepristran i očekujemo da će model dobro generalizirati

- 4 (N) Raspoložemo sljedećim skupom označenih primjera:

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

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

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

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

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

- 5 (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y=1$ i $y=2$) u dvodimenzijaskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klasa su jednake, dok su izglednosti klasa modelirane bivarijatnim Gaussovim distribucijama sa sljedećim parametrima:

$$\mu_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \Sigma_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \mu_2 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \Sigma_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klasa definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

- ☐ A (3, 2), (4, 5), (9, 7) ☐ B (1, 6), (3, 3), (5, 0) ☐ C (1, 3), (5, 4), (7, 9) ☐ D (1, 2), (3, 4), (6, 7)

- 6 (T) Gaussov Bayesov klasifikator s dijeljenom kovarijacijskom matricom (GBC) i logistička regresija (LR) su generativno-diskriminativni par modela. **Što to znači?**

- ☐ A Izlaz modela LR jednak je zajedničkoj vjerojatnosti modela GBC, ali model LR iziskuje manje parametara
☐ B GBC i neregularizirana LR modeliraju identične distribucije zajedničke vjerojatnosti primjera i oznaka, ali s različitim brojem parametara
☐ C Aposteriorna vjerojatnost klase za GBC može se izraziti kao poopćeni linearni model sa sigmoidnom aktivacijskom funkcijom
☐ D Neovisno o optimizacijskom postupku, GBC i LR ostvaruju istu pogrešku na skupu za učenje, ali uz moguće različit broj parametara

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

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

Na temelju beta-Bernoullijevog modela na skupu \mathcal{D} računamo MAP procjenu parametra μ Bernoullijeve distribucije. MLE procjena za isti parametar na skupu \mathcal{D} iznosi 0.4. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

- ☐ A $\alpha = 5, \beta = 7$ ☐ B $\alpha = 2, \beta = 5$ ☐ C $\alpha = 2, \beta = 10$ ☐ D $\alpha = 4, \beta = 8$

- 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:

$$p(\mathbf{x}|y=j) = \frac{1}{(2\pi)^{n/2} |\Sigma_j|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu_j)^T \Sigma_j^{-1}(\mathbf{x}-\mu_j)\right\}$$

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. Za procjenu kovarijacijske matrice koristite dva procjenitelja: MLE i nepristrani procjenitelj. Izlaz modela za klasu $y=j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y=j)$. Neka su h_1^{MLE} i h_1^{UB} hipoteze za prvu klasu modela s pristranom odnosno nepristranom procjenom kovarijacijske matrice. Zanima nas predikcija za primjer $\mathbf{x} = (1, 1)$. **Koliko iznosi apsolutna razlika u predikciji za \mathbf{x} za ova dva modela, $|h_1^{\text{MLE}}(\mathbf{x}) - h_1^{\text{UB}}(\mathbf{x})|$?**

- ☐ A 0.2537 ☐ B 0.1638 ☐ C 0.0337 ☐ D 0.2276

Cjelina 5: Probabilistički grafički modeli (5 pitanja)

- 9 (P) Razmotrite Bayesovu mrežu sa šest čvorova koja odgovara sljedećoj faktorizaciji:

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

Upotrijebite metodu d-odvajanja da biste ispitali uvjetnu nezavisnost varijabli x i z u ovisnosti o preostale četiri varijable. Ispitajte koje varijable od preostalih četiri varijabli trebaju biti opažene a koje neopažene, a da bi varijable x i z bile d-odvojene. **Za koju od preostale četiri varijable je svejedno je li opažena, ako su varijable x i z d-odvojene?**

☐ A w ☐ B v ☐ C u ☐ D y

- 10 (N) Razmotrite Bayesovu mrežu koja odgovara faktorizaciji $P(w, x, y, z) = P(w)P(x)P(y|w, x)P(z|x)$. Sve varijable su binarne. Vrijedi $P(w = 1) = 0.1$, $P(x = 1) = 0.2$, $P(z = 1|x = 0) = 0.9$ i $P(z = 1|x = 1) = 0.7$. Tablica uvjetnih vjerojatnosti za čvor y je sljedeća:

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

Postupkom uzorkovanja s odbijanjem želimo procijeniti parametar μ uvjetne distribucije $P(x = 0|y = 1, z = 0)$. Uzorkovanje smo ponovili ukupno $N = 500$ puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od N . Na temelju dobivenog uzorka parametar μ procjenjujemo MAP procjeniteljem uz $\alpha = \beta = 2$. **Koliko iznosi očekivana MAP procjena parametra μ ?**

☐ A 0.0786 ☐ B 0.1877 ☐ C 0.1274 ☐ D 0.0490

- 11 (N) Bayesovu mrežu koristimo za medicinsku dijagnostiku te modeliramo sljedeće kauzalne odnose. Upala grla ($U = 1$) može biti uzrokovana virusom ($V = 1$) ili bakterijom ($B = 1$). Povišena temperatura ($T = 1$) može biti uzrokovana upalom grla ili sunčanicom ($S = 1$). Sve varijable su binarne. Na temelju podataka o pacijentima procijenili smo parametre mreže: $P(V = 1) = 0.3$, $P(B = 1) = 0.4$ i $P(S = 1) = 0.05$. Uvjetne vjerojatnosti za čvorove U i T su:

V	B	$P(U = 1 V, B)$	U	S	$P(T = 1 U, S)$
0	0	0.2	0	0	0
0	1	0.5	0	1	0.2
1	0	0.4	1	0	0.4
1	1	0.7	1	1	0.4

Zanima nas koje je najvjerojatnije objašnjenje izravnog uzroka povišene temperature u pacijenata kod kojih nije dokazano prisustvo virusa. U tu svrhu računamo MAP-upit za par varijabli upita U i S uz opažene varijable $V = 0$ i $T = 1$, tj. računamo $\arg\max_{U, S} P(U, S|V = 0, T = 1)$. Neka je p_1 vjerojatnost najvjerojatnijeg (MAP) objašnjenja za varijable U i S , a p_2 vjerojatnost drugog po redu najvjerojatnijeg objašnjenja za te varijable. **Koliko je puta najvjerojatnije objašnjenje vjerojatnije od drugog najvjerojatnijeg objašnjenja, tj. koliko iznosi p_1/p_2 ?**

☐ A 15.52 ☐ B 13.35 ☐ C 17.88 ☐ D 11.35

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

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

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

☐ A 23 ☐ B 24 ☐ C 21 ☐ D 18

- 13** (T) Parametre probabilističkih grafičkih modela, uključivo Bayesove mreže, možemo procjenjivati iz potpunih podataka ili nepotpunih podataka. **Zašto i kako Bayesovu mrežu učimo nad nepotpunim podacima?**
- ☐ A Jer mreža ima manje čvorova nego što je opaženih varijabli, pa koristimo eliminaciju varijabli
- ☐ B Jer MLE nema rješenje u zatvorenoj formi, pa umjesto MLE koristimo gradijentni uspon ili EM-algoritam
- ☐ C Jer mreža ima skrivene varijable koje ne opažamo, pa moramo koristiti iterativne metode za MAP ili MLE
- ☐ D Jer se log-izglednost dekomponira po čvorovima mreže, pa MLE ili MAP procjenjujemo u zatvorenoj formi

Cjelina 6: Grupiranje i vrednovanje modela (9 pitanja)

- 14** (T) Algoritam K-medoida općenitiji je od algoritma K-sredina budući da se može koristiti za primjere koji nisu prikazani kao vektori. Međutim, razmotrite slučaj kada primjeri jesu prikazani kao vektori, ali ih želimo grupirati na temelju mjere udaljenosti koja nije euklidska. **Koji bismo algoritam koristili u tom slučaju i zašto?**
- ☐ A Algoritam K-medoida, jer kriterijska funkcija algoritma K-sredina koristi euklidsku udaljenost
- ☐ B Algoritam K-sredina, jer koristi mjeru udaljenosti, dok algoritam K-medoida koristi mjeru sličnosti
- ☐ C Algoritam K-medoida, jer vektorizirani primjeri također mogu biti medoidi
- ☐ D Algoritam K-sredina, jer za vektorizirane primjere možemo izračunati centroide grupa
- 15** (N) Na ispitnome skupu sa $K = 3$ klase vrednujemo algoritam multinomijalne logističke regresije. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci predikcije klasifikatora):

$$\begin{array}{c} y = 1 \quad y = 2 \quad y = 3 \\ \begin{array}{c} y = 1 \\ y = 2 \\ y = 3 \end{array} \begin{pmatrix} 15 & 3 & 5 \\ 9 & 5 & 4 \\ 4 & 2 & 3 \end{pmatrix} \end{array}$$

Kao referentni model koristimo klasifikator koji nasumično pogađa oznaku y , i to s jednakom vjerojatnošću za svaku od tri klase. Za vrednovanje koristimo mjeru F_1 -makro. Izračunajte očekivanu vrijednost mjere F_1 -makro za logističku regresiju i za referentni model. **Koliko iznosi očekivana razlika u mjeri F_1 -makro između logističke regresije i referentnog modela?**

- ☐ A 0.2238 ☐ B 0.1756 ☐ C 0.0947 ☐ D 0.1420
- 16** (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

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

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 1.9 ☐ D 2.4
- 17** (T) Za grupiranje primjera u K grupa koristimo model Gaussove mješavine (GMM) s dijeljenom kovarijacijskom matricom. Nakon grupiranja, odgovornosti zaokružujemo na cijeli broj, čime dobivamo tvrdo grupiranje. Iste podatke grupiramo algoritmom K-medoida (KM). **Uz koje parametre ovih algoritama očekujemo dobiti najslabije rezultate grupiranja?**
- ☐ A GMM: $\Sigma_k = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ i $\pi_k = 1/K$; KM: euklidska udaljenost
- ☐ B GMM: puna Σ_k i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost
- ☐ C GMM: $\Sigma_k = \sigma^2 \mathbf{I}$; KM: euklidska udaljenost
- ☐ D GMM: $\Sigma_k = \sigma^2 \mathbf{I}$ i $\pi_k = 1/K$; KM: Mahalanobisova udaljenost

- 18 (P) Raspoložemo sa 1000 označenih primjera. Na tom skupu treniramo i vrednujemo algoritam SVM, optimizirajući hiperparametre C i γ . Za vrednovanje koristimo ugniježdenu 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 ugniježdene unakrsne provjere. **Za koliko bi taj broj narastao kada bismo broj preklopa u unutarnjoj petlji povećali na 10?**

☐ A 100 ☐ B 160 ☐ C 80 ☐ D 192

- 19 (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1, 2, 3, 4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{aligned} K = 3 & \quad \{\{1, 2, 4, 4\}, \{2, 2, 3\}, \{1, 1, 3, 4\}\} \\ K = 4 & \quad \{\{1, 2, 2, 4, 4\}, \{2, 3\}, \{1, 1\}, \{3, 4\}\} \end{aligned}$$

gdje podskupovi odgovaraju grupama, a brojke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

☐ A 0.0545 ☐ B 0.0182 ☐ C 0.0727 ☐ D 0.0364

- 20 (P) Skup neoznačenih primjera u dvodimenzijaskome ulaznom prostoru neka je sljedeći:

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

Primjere grupiramo algoritmom K-sredina sa $K = 2$ grupe. Grupiranje možemo shvatiti kao pretraživanje prostora stanja, gdje svako pojedino stanje odgovara jednom pridjeljivanju primjera grupama. Pritom dva stanja smatramo identičnima ako su grupiranja identična, neovisno o identitetu grupa (npr., grupiranje kod kojega prva grupa sadrži samo primjer $\mathbf{x}^{(1)}$ je identično kao i grupiranje kod kojega druga grupa sadrži samo primjer $\mathbf{x}^{(1)}$). Neka je A_1 algoritam K-sredina s potpuno slučajno inicijaliziranim središtima, a A_2 algoritam K-sredina gdje su središta inicijalizirana algoritmom K-means++. Neka je $S(A_1)$ skup stanja koje pretražuje algoritam A_1 , a $S(A_2)$ skup stanja koje pretražuje algoritam A_2 . Izračunajte veličine ovih skupova, uzevši u obzir mogućnost da pojedina grupa bude prazna te da dođe do izjednačenja udaljenosti primjera do centroida, što se razrješava slučajnim mehanizmom. **Koliko algoritam A_2 pretražuje manje stanja od algoritma A_1 , tj. koliko iznosi $|S(A_1)| - |S(A_2)|$?**

☐ A 10 ☐ B 12 ☐ C 4 ☐ D 6

- 21 (P) Skup neoznačenih primjera \mathcal{D} grupiramo modelom GMM treniranim EM-algoritmom. Koristimo nekoliko varijanti tog modela:

\mathcal{H}_1 : Model sa $K = 25$ središta inicijaliziranim algoritmom K-means++

\mathcal{H}_2 : Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom

\mathcal{H}_3 : Model sa $K = 50$ središta inicijaliziranim algoritmom K-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL_α^0 prosječna log-izglednost za model \mathcal{H}_α na početku izvođenja EM-algoritma, LL_α^* prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma te neka je k_α broj iteracija EM-algoritma za taj model. **Što možemo zaključiti o očekivanim odnosima između ovih vrijednosti?**

☐ A $LL_3^0 \geq LL_2^0$, $LL_3^* \geq LL_2^*$, $k_2 \geq k_3$ ☐ C $LL_2^0 \geq LL_3^0$, $LL_3^* \geq LL_2^*$, $k_3 \geq k_2$
☐ B $LL_1^0 \geq LL_2^0$, $LL_2^* \geq LL_3^*$, $k_1 \geq k_2$ ☐ D $LL_1^0 \geq LL_3^0$, $LL_1^* \geq LL_3^*$, $k_2 \geq k_1$

- 22 (T) Mjera točnosti nije prikladna za vrednovanje klasifikatora na skupovima podataka s neuravnoteženim brojem primjera po klasama. Jedna alternativa mjeri točnosti je F_1 -mjera, međutim ni ta mjera nije uvijek prikladna. Pretpostavite da vrijednost F_1 -mjere postavljamo na nulu u slučajevima kada je harmonijska sredina preciznosti i odziva nedefinirana. **U kojem slučaju F_1 -mjera ne bi bila prikladna mjera za vrednovanje klasifikatora jer bi bila previše optimistična?**

☐ A Ako je većina primjera pozitivna, a klasifikator sve primjere klasificira negativno
☐ B Ako je većina primjera pozitivna i klasifikator sve primjere klasificira pozitivno
☐ C Ako je većina primjera negativna, a klasifikator sve primjere klasificira pozitivno
☐ D Ako je većina primjera negativna i klasifikator sve primjere klasificira negativno

		0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	2	2	2
		1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2
		-----+-----																					
Grupa A		D	A	D	B	C	A	B	A	C	B	B	C	B	B	D	D	C	C	D	C	A	B
Grupa B		B	B	C	C	B	B	A	A	D	A	D	C	A	A	D	D	D	D	B	B	A	D
Grupa C		A	A	B	D	D	C	C	D	B	C	D	D	B	A	C	D	B	A	B	C	C	C
Grupa D		A	B	B	B	C	D	B	C	A	A	B	B	D	A	C	B	D	A	B	B	D	A
Grupa E		D	B	A	B	A	B	D	C	A	B	D	B	C	C	B	C	C	B	C	C	B	A
Grupa F		A	D	A	B	C	B	B	D	C	A	D	D	C	B	C	D	C	D	B	A	C	A
Grupa G		B	A	D	C	A	A	D	A	B	A	D	C	B	C	B	C	A	C	A	D	D	B
Grupa H		B	D	A	D	A	C	A	A	C	A	C	B	C	A	C	A	B	B	D	C	A	B