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

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (T) Optimizacija modela hrbatne regresije (L_2-regularizirane linearne regresije) ima rješenje u zatvorenoj formi. Neka je λ regularizacijski faktor, n broj značajki u ulaznom prostoru (bez "dummy" jedinice), m broj značajki u prostoru značajki (također bez "dummy" jedinice) te N broj primjera. Glavna komponenta rješenja je izračun inverza matrice izračunate na temelju matrice dizajna $\mathbf{\Phi}$. **Koliko redaka odnosno stupaca ima matrica koju invertiramo?**

A N B $m+1$ C $n+\lambda$ D $m+\lambda$

2 (N) Raspolažemo sljedećim skupom primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 1$. Dobili smo težine $\mathbf{w} = (2.12, -0.94, -0.08)$. **Koliko iznosi L_1-regularizirana pogreška $E(\mathbf{w}|\mathcal{D})$?**

A 2.69 B 7.10 C 1.58 D 0.29

3 (P) Skup označenih primjera u dvodimenzijskome ulaznom prostoru je:

$$\mathcal{D} = \{((0,0),0),((0,1),0),((1,1),1)\}$$

Koja je veličina prostora inačica, $|\mathrm{VS}_{\mathcal{H},\mathcal{D}}|$?

A Pitanje nema smisla jer nije definiran model B 14 C 16 D Beskonačno mnogo

4 (P) Koristimo regresiju za predviđanje uspjeha na studiju. Kao značajke možemo koristiti ocjene u četiri razreda srednje škole (značajke x_1–x_4), prosjek ocjena sva četiri razreda (x_5) te uspjeh iz matematike (x_6) i fizike (x_7) na državnoj maturi (ukupno 7 značajki). Ne moramo iskoristiti sve značajke, ali ih želimo iskoristiti što više. Za preslikavanje u prostor značajki koristimo preslikavanje s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki (npr. x_1x_2) i interakcije trojki (npr. $x_1x_2x_3$) između svih značajki koje koristimo. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 75 B 48 C 63 D 38

Linearni klasifikacijski modeli *(3 pitanja)*

5 (N) Raspolažemo označenim skupom primjera iz triju klasa ($K=3$) u trodimenzijskome ulaznom prostoru ($n=3$). Na tom skupu treniramo model multinomijalne logističke regresije. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 0 & -2 \\ 3 & -4 & 6 \\ -3 & 0 & 2 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (-4, 1, -3)$ s oznakom $\mathbf{y} = (0, 1, 0)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi dotični primjer?**

A 4.02 B 8.00 C 12.02 D 6.00

6 (P) Treniramo linearni diskriminativni model u dvodimenzijskome prostoru primjera. Skup za učenje čine samo dva primjera, $(\mathbf{x}_1, y_1) = ((1,0), +1)$ i $(\mathbf{x}_2, y_2) = ((0,1), -1)$. Na tom skupu treniramo model koji ima induktivnu pristranost takvu da rješenje maksimizira minimalnu udaljenost primjera od hiperravnine. Naučen model ispravno klasificira oba primjera, pri čemu za oba primjera vrijedi $y \cdot h(\mathbf{x}) = 5$. **Koliko iznosi težina w_2 tako naučenog modela?**

A -1 B -5 C 1 D 5

7 (T) Neregularizirani model logističke regresije sklon je prenaučenosti. To je pogotovo slučaj ako se model trenira na linearno odvojivim podatcima, čak i onda kada u podatcima nema nikakvoga šuma i kada se ne koristi nikakvo preslikavanje u prostor značajki. **Zbog čega dolazi do prenaučenosti modela neregularizirane logističke regresije na linearno odvojivim skupovima podataka?**

A Empirijska pogreška logističke regresije smanjuje se s porastom broja primjera

B Netočno klasificirani primjeri nanose gubitak koji je proporcionalan normi vektora težina

C Ispravno klasificirani primjeri koji su vrlo udaljeni od granice nanose malen gubitak

D S porastom norme vektora težina gubitak na točnim primjerima teži prema nuli

Jezgrene i neparametarske metode *(4 pitanja)*

8 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 4$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 1)^3$. Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((11, 32, 21), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((5, 11, 6), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((6, 12, 6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 1.063 \cdot 10^{-7}$, $\alpha_2 = 1.022 \cdot 10^{-5}$ i $\alpha_3 = 1.032 \cdot 10^{-5}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3, 0, -3)$.**

A -0.676 B -2.330 C $+1.434$ D -3.553

9 (T) Problem maksimalne margine ima svoju geometrijsku interpretaciju: maksimalna margina odgovara simetrali spojnice konveksnih ljusaka primjera iz dviju klasa. **Što je nužan i dovoljan uvjet da klasifikacijski problem bude rješiv SVM-om s tvrdom marginom?**

A Svaka spojnica između primjera iz jedne klase treba biti izvan konveksne ljuske druge klase

B Najbliži primjeri iz suprotnih klasa moraju biti u vrhovima konveksnih ljusaka

C Konveksne ljuske dviju klasa ne smiju se preklapati (trebaju biti disjunktne)

D Primjeri iz obje klase trebaju činiti konveksne skupove u ulaznom prostoru

10 (N) Algoritam k-NN koristimo za višeklasnu klasifikaciju riječi prema jeziku kojemu pripadaju. Skup za učenje sastoji se od sljedećih riječi i oznaka klasa:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{(\text{“water”}, 0), (\text{“voda”}, 1), (\text{“zrak”}, 1), (\text{“luft”}, 2), (\text{“feuer”}, 2)\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2| / |\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje je su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“voda”}) = 1/8 = 0.125$. Razmatramo dvije varijante algoritma: 3-NN i težinski k-NN. Kod potonjeg u obzir uzimamo sve primjere, tj. $k = N$. Odredite klasifikaciju primjera $\mathbf{x}$ = “fire” pomoću ova dva algoritma. U slučaju jednake sličnosti između dva primjera, kao susjed se uzima onaj koji je u skupu $\mathcal{D}$ naveden prvi. U slučaju izjednačenja glasova između klasa, prednost se daje klasi s numerički manjom oznakom y. **U koju će klasu biti klasificiran primjer x algoritmom 3-NN, a u koju algoritmom težinski k-NN?**

A $y = 0$ i $y = 2$ B $y = 0$ i $y = 1$ C $y = 0$ i $y = 0$ D $y = 1$ i $y = 1$

11 (P) Razmatramo sljedeći skup označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a, i to model s tvrdom marginom te model s mekom marginom sa $C = 1$. Kod modela s mekom marginom za dualne koeficijente vrijedi $\alpha_1 = 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, $\alpha_4 > 0$ i $\alpha_5 = 1$. Skicirajte tvrdu i meku marginu u ulaznome prostoru. **Koliko je meka margina šira od tvrde margine?**

[A] $\frac{2}{5}\sqrt{10}$ puta [B] $\frac{3}{5}\sqrt{10}$ puta [C] $\frac{1}{6}\sqrt{2}$ puta [D] $\frac{4}{5}\sqrt{10}$ puta

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

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

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

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

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

13 (T) Nepristranost je jedno od svojstava statističkih procjenitelja. Procjenitelj MLE ne mora nužno biti nepristran, tj. može biti pristran. **Za koji od sljedećih parametara distribucije procjena MLE pristrana?**

[A] Varijanca Bernoullijeve distribucije

[B] Kovarijacijska matrica Gaussove distribucije

[C] Srednja vrijednost Bernoullijeve distribucije

[D] Srednja vrijednost Gaussove distribucije

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

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

Probabilistički grafički modeli *(3 pitanja)*

15 (T) Glavna svrha probabilističkih grafičkih modela (PGM) jest provođenje probabilističkih upita. Jedna vrsta upita su MAP-upiti (upiti najvjerojatnijeg objašnjenja). Neka su $\mathbf{x}_q$, $\mathbf{x}_o$ i $\mathbf{x}_n$ skupovi varijabli upita, opaženih varijabli odnosno varijabli smetnje. **Kako je definiran rezultat MAP-upita?**

[A] $\text{argmax}_{\mathbf{x}_o} \sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$ [C] $\text{argmax}_{\mathbf{x}_q} \sum_{\mathbf{x}_o} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$

[B] $\text{argmax}_{\mathbf{x}_o} \sum_{\mathbf{x}_q} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$ [D] $\text{argmax}_{\mathbf{x}_q} \sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$

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

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

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

A 89 B 180 C 152 D 404

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

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

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

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

Grupiranje *(3 pitanja)*

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

$\mathcal{H}_1$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina
$\mathcal{H}_2$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom
$\mathcal{H}_3$: Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom
$\mathcal{H}_4$: Model sa $K = 10$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

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

A $LL^0_3 \geqslant LL^0_4,\ LL^*_1 \geqslant LL^*_3 \geqslant LL^*_4$ C $LL^0_2 \geqslant LL^0_4 \geqslant LL^0_3,\ LL^*_1 \geqslant LL^*_2$
B $LL^0_2 \geqslant LL^0_4,\ LL^*_1 \geqslant LL^*_2 \geqslant LL^*_3$ D $LL^0_2 \leqslant LL^0_4,\ LL^*_2 \leqslant LL^*_1 \geqslant LL^*_3$

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

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

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

A 0.292 B 0.535 C 0.583 D 0.354

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

A 6.66 B 1.85 C 4.25 D 3.00

Vrednovanje modela *(2 pitanja)*

21 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \ldots, 2^{15}\}$ i $\gamma \in \{2^{-15}, 2^{-14}, \ldots, 2^{15}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 10 ponavljanja u vanjskoj petlji i 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 49600 B 35721 C 44640 D 69201

22 (T) Ugniježđena k-struka unakrsna provjera često se koristi za procjenu točnosti modela. **Što je prednost ugniježđene k-struke unakrsne provjere u odnosu na običnu unakrsnu provjeru?**

A Procjenjujemo očekivanu ispitnu pogrešku modela optimalne složenosti

B Procjenjujemo pogrešku generalizacije, a ne pogrešku učenja

C Model ispitujemo na cijelom raspoloživom skupu primjera

D Procjenjujemo ispitnu pogrešku modela s najmanjom pogreškom učenja

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

Ispit sadrži ***22 pitanja*** *i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti* ***20 pitanja****, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je* ***150 minuta****. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (T) Optimizacija modela hrbatne regresije (L_2-regularizirane linearne regresije) ima rješenje u zatvorenoj formi. Neka je λ regularizacijski faktor, n broj značajki u ulaznom prostoru (bez "dummy" jedinice), m broj značajki u prostoru značajki (također bez "dummy" jedinice) te N broj primjera. Glavna komponenta rješenja je izračun inverza matrice izračunate na temelju matrice dizajna $\mathbf{\Phi}$. **Koliko redaka odnosno stupaca ima matrica koju invertiramo?**

A N B $n+\lambda$ C $m+1$ D $m+\lambda$

2 (P) Koristimo regresiju za predviđanje uspjeha na studiju. Kao značajke možemo koristiti ocjene u četiri razreda srednje škole (značajke x_1–x_4), prosjek ocjena sva četiri razreda (x_5) te uspjeh iz matematike (x_6) i fizike (x_7) na državnoj maturi (ukupno 7 značajki). Ne moramo iskoristiti sve značajke, ali ih želimo iskoristiti što više. Za preslikavanje u prostor značajki koristimo preslikavanje s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki (npr. x_1x_2) i interakcije trojki (npr. $x_1x_2x_3$) između svih značajki koje koristimo. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 63 B 75 C 38 D 48

3 (N) Raspolažemo sljedećim skupom primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 1$. Dobili smo težine $\mathbf{w} = (2.12, -0.94, -0.08)$. **Koliko iznosi L_1-regularizirana pogreška $E(\mathbf{w}|\mathcal{D})$?**

A 1.58 B 7.10 C 0.29 D 2.69

4 (P) Skup označenih primjera u dvodimenzijskome ulaznom prostoru je:

$$\mathcal{D} = \{((0,0),0),((0,1),0),((1,1),1)\}$$

Koja je veličina prostora inačica, $|\mathrm{VS}_{\mathcal{H},\mathcal{D}}|$?

A Beskonačno mnogo B 14 C Pitanje nema smisla jer nije definiran model D 16

Linearni klasifikacijski modeli *(3 pitanja)*

5 (T) Neregularizirani model logističke regresije sklon je prenaučenosti. To je pogotovo slučaj ako se model trenira na linearno odvojivim podatcima, čak i onda kada u podatcima nema nikakvoga šuma i kada se ne koristi nikakvo preslikavanje u prostor značajki. **Zbog čega dolazi do prenaučenosti modela neregularizirane logističke regresije na linearno odvojivim skupovima podataka?**

A Netočno klasificirani primjeri nanose gubitak koji je proporcionalan normi vektora težina

B S porastom norme vektora težina gubitak na točnim primjerima teži prema nuli

C Empirijska pogreška logističke regresije smanjuje se s porastom broja primjera

D Ispravno klasificirani primjeri koji su vrlo udaljeni od granice nanose malen gubitak

6 (N) Raspolažemo označenim skupom primjera iz triju klasa ($K = 3$) u trodimenzijskome ulaznom prostoru ($n = 3$). Na tom skupu treniramo model multinomijalne logističke regresije. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 0 & -2 \\ 3 & -4 & 6 \\ -3 & 0 & 2 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (-4, -1, -3)$ s oznakom $\mathbf{y} = (0, 0, 1)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi dotični primjer?**

A 8.00 B 6.00 C 4.02 D 12.02

7 (P) Treniramo linearni diskriminativni model u dvodimenzijskome prostoru primjera. Skup za učenje čine samo dva primjera, $(\mathbf{x}_1, y_1) = ((1,0), +1)$ i $(\mathbf{x}_2, y_2) = ((0,1), -1)$. Na tom skupu treniramo model koji ima induktivnu pristranost takvu da rješenje maksimizira minimalnu udaljenost primjera od hiperravnine. Naučen model ispravno klasificira oba primjera, pri čemu za oba primjera vrijedi $y \cdot h(\mathbf{x}) = 5$. **Koliko iznosi težina w_2 tako naučenog modela?**

A -1 B -5 C 5 D 1

Jezgrene i neparametarske metode *(4 pitanja)*

8 (N) Algoritam k-NN koristimo za višeklasnu klasifikaciju riječi prema jeziku kojemu pripadaju. Skup za učenje sastoji se od sljedećih riječi i oznaka klasa:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{(\text{“water”}, 0), (\text{“voda”}, 1), (\text{“zrak”}, 1), (\text{“luft”}, 2), (\text{“feuer”}, 2)\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2| / |\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje je su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“voda”}) = 1/8 = 0.125$. Razmatramo dvije varijante algoritma: 3-NN i težinski k-NN. Kod potonjeg u obzir uzimamo sve primjere, tj. $k = N$. Odredite klasifikaciju primjera $\mathbf{x}$ = “fire” pomoću ova dva algoritma. U slučaju jednake sličnosti između dva primjera, kao susjed se uzima onaj koji je u skupu $\mathcal{D}$ naveden prvi. U slučaju izjednačenja glasova između klasa, prednost se daje klasi s numerički manjom oznakom y. **U koju će klasu biti klasificiran primjer x algoritmom 3-NN, a u koju algoritmom težinski k-NN?**

A $y = 0$ i $y = 0$ B $y = 0$ i $y = 1$ C $y = 0$ i $y = 2$ D $y = 1$ i $y = 1$

9 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 4$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 1)^3$. Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((11, 32, 21), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((5, 11, 6), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((6, 12, 6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 1.063 \cdot 10^{-7}$, $\alpha_2 = 1.022 \cdot 10^{-5}$ i $\alpha_3 = 1.032 \cdot 10^{-5}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (-3, -5, 3)$.**

A -3.553 B -2.330 C $+1.434$ D -4.093

10 (P) Razmatramo sljedeći skup označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a, i to model s tvrdom marginom te model s mekom marginom sa $C = 1$. Kod modela s mekom marginom za dualne koeficijente vrijedi $\alpha_1 = 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, $\alpha_4 > 0$ i $\alpha_5 = 1$. Skicirajte tvrdu i meku marginu u ulaznome prostoru. **Koliko je meka margina šira od tvrde margine?**

A $\frac{1}{6}\sqrt{2}$ puta B $\frac{2}{5}\sqrt{10}$ puta C $\frac{4}{5}\sqrt{10}$ puta D $\frac{3}{5}\sqrt{10}$ puta

11 (T) Problem maksimalne margine ima svoju geometrijsku interpretaciju: maksimalna margina odgovara simetrali spojnice konveksnih ljusaka primjera iz dviju klasa. **Što je nužan i dovoljan uvjet da klasifikacijski problem bude rješiv SVM-om s tvrdom marginom?**

- A Konveksne ljuske dviju klasa ne smiju se preklapati (trebaju biti disjunktne)
- B Primjeri iz obje klase trebaju činiti konveksne skupove u ulaznom prostoru
- C Svaka spojnica između primjera iz jedne klase treba biti izvan konveksne ljuske druge klase
- D Najbliži primjeri iz suprotnih klasa moraju biti u vrhovima konveksnih ljusaka

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (T) Nepristranost je jedno od svojstava statističkih procjenitelja. Procjenitelj MLE ne mora nužno biti nepristran, tj. može biti pristran. **Za koji od sljedećih parametara distribucije procjena MLE pristrana?**

- A Kovarijacijska matrica Gaussove distribucije
- B Varijanca Bernoullijeve distribucije
- C Srednja vrijednost Bernoullijeve distribucije
- D Srednja vrijednost Gaussove distribucije

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

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

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

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

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

A 0.488 B 0.588 C 0.741 D 0.322

Probabilistički grafički modeli *(3 pitanja)*

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

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

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

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

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

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

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

A 89 B 152 C 404 D 180

17 (T) Glavna svrha probabilističkih grafičkih modela (PGM) jest provođenje probabilističkih upita. Jedna vrsta upita su MAP-upiti (upiti najvjerojatnijeg objašnjenja). Neka su $\mathbf{x}_q$, $\mathbf{x}_o$ i $\mathbf{x}_n$ skupovi varijabli upita, opaženih varijabli odnosno varijabli smetnje. **Kako je definiran rezultat MAP-upita?**

A $\text{argmax}_{\mathbf{x}_q} \sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$ C $\text{argmax}_{\mathbf{x}_o} \sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$

B $\text{argmax}_{\mathbf{x}_o} \sum_{\mathbf{x}_q} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$ D $\text{argmax}_{\mathbf{x}_q} \sum_{\mathbf{x}_o} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$

Grupiranje *(3 pitanja)*

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

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

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

A 0.583 B 0.478 C 0.292 D 0.535

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

A 3.00 B 4.25 C 1.85 D 6.66

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

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

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

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

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

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

A $LL_2^0 \geqslant LL_4^0,\ LL_1^* \geqslant LL_2^* \geqslant LL_3^*$ C $LL_2^0 \geqslant LL_4^0 \geqslant LL_3^0,\ LL_1^* \geqslant LL_2^*$

B $LL_2^0 \leqslant LL_4^0,\ LL_2^* \leqslant LL_1^* \geqslant LL_3^*$ D $LL_3^0 \geqslant LL_4^0,\ LL_1^* \geqslant LL_3^* \geqslant LL_4^*$

Vrednovanje modela *(2 pitanja)*

21 (T) Ugniježđena k-struka unakrsna provjera često se koristi za procjenu točnosti modela. **Što je prednost ugniježđene k-struke unakrsne provjere u odnosu na običnu unakrsnu provjeru?**

A Procjenjujemo pogrešku generalizacije, a ne pogrešku učenja

B Procjenjujemo ispitnu pogrešku modela s najmanjom pogreškom učenja

C Procjenjujemo očekivanu ispitnu pogrešku modela optimalne složenosti

D Model ispitujemo na cijelom raspoloživom skupu primjera

22 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \ldots, 2^{15}\}$ i $\gamma \in \{2^{-15}, 2^{-14}, \ldots, 2^{15}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 10 ponavljanja u vanjskoj petlji i 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 49600 B 35721 C 69201 D 44640

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

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) Koristimo regresiju za predviđanje uspjeha na studiju. Kao značajke možemo koristiti ocjene u četiri razreda srednje škole (značajke x_1–x_4), prosjek ocjena sva četiri razreda (x_5) te uspjeh iz matematike (x_6) i fizike (x_7) na državnoj maturi (ukupno 7 značajki). Ne moramo iskoristiti sve značajke, ali ih želimo iskoristiti što više. Za preslikavanje u prostor značajki koristimo preslikavanje s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki (npr. x_1x_2) i interakcije trojki (npr. $x_1x_2x_3$) između svih značajki koje koristimo. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 63 B 48 C 38 D 75

2 (P) Skup označenih primjera u dvodimenzijskome ulaznom prostoru je:

$$\mathcal{D} = \{((0,0),0),((0,1),0),((1,1),1)\}$$

Koja je veličina prostora inačica, $|\mathrm{VS}_{\mathcal{H},\mathcal{D}}|$?

A 16 B 14 C Pitanje nema smisla jer nije definiran model D Beskonačno mnogo

3 (T) Optimizacija modela hrbatne regresije (L_2-regularizirane linearne regresije) ima rješenje u zatvorenoj formi. Neka je λ regularizacijski faktor, n broj značajki u ulaznom prostoru (bez "dummy" jedinice), m broj značajki u prostoru značajki (također bez "dummy" jedinice) te N broj primjera. Glavna komponenta rješenja je izračun inverza matrice izračunate na temelju matrice dizajna $\boldsymbol{\Phi}$. **Koliko redaka odnosno stupaca ima matrica koju invertiramo?**

A $m+\lambda$ B N C $m+1$ D $n+\lambda$

4 (N) Raspolažemo sljedećim skupom primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 1$. Dobili smo težine $\mathbf{w} = (2.12, -0.94, -0.08)$. **Koliko iznosi L_1-regularizirana pogreška $E(\mathbf{w}|\mathcal{D})$?**

A 1.58 B 7.10 C 2.69 D 0.29

Linearni klasifikacijski modeli *(3 pitanja)*

5 (P) Treniramo linearni diskriminativni model u dvodimenzijskome prostoru primjera. Skup za učenje čine samo dva primjera, $(\mathbf{x}_1, y_1) = ((1,0),+1)$ i $(\mathbf{x}_2, y_2) = ((0,1),-1)$. Na tom skupu treniramo model koji ima induktivnu pristranost takvu da rješenje maksimizira minimalnu udaljenost primjera od hiperravnine. Naučen model ispravno klasificira oba primjera, pri čemu za oba primjera vrijedi $y \cdot h(\mathbf{x}) = 5$. **Koliko iznosi težina w_2 tako naučenog modela?**

A -1 B 1 C 5 D -5

6 (T) Neregularizirani model logističke regresije sklon je prenaučenosti. To je pogotovo slučaj ako se model trenira na linearno odvojivim podatcima, čak i onda kada u podatcima nema nikakvoga šuma i kada se ne koristi nikakvo preslikavanje u prostor značajki. **Zbog čega dolazi do prenaučenosti modela neregularizirane logističke regresije na linearno odvojivim skupovima podataka?**

A S porastom norme vektora težina gubitak na točnim primjerima teži prema nuli

B Empirijska pogreška logističke regresije smanjuje se s porastom broja primjera

C Netočno klasificirani primjeri nanose gubitak koji je proporcionalan normi vektora težina

D Ispravno klasificirani primjeri koji su vrlo udaljeni od granice nanose malen gubitak

7 (N) Raspolažemo označenim skupom primjera iz triju klasa ($K = 3$) u trodimenzijskome ulaznom prostoru ($n = 3$). Na tom skupu treniramo model multinomijalne logističke regresije. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 0 & -2 \\ 3 & -4 & 6 \\ -3 & 0 & 2 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (-4, -1, -3)$ s oznakom $\mathbf{y} = (0, 0, 1)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi dotični primjer?**

A 12.02 B 4.02 C 6.00 D 8.00

Jezgrene i neparametarske metode *(4 pitanja)*

8 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 4$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 1)^3$. Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((11, 32, 21), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((5, 11, 6), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((6, 12, 6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 1.063 \cdot 10^{-7}$, $\alpha_2 = 1.022 \cdot 10^{-5}$ i $\alpha_3 = 1.032 \cdot 10^{-5}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (5, 15, 1)$.**

A -4.093 B $+5.454$ C -2.330 D -3.553

9 (T) Problem maksimalne margine ima svoju geometrijsku interpretaciju: maksimalna margina odgovara simetrali spojnice konveksnih ljusaka primjera iz dviju klasa. **Što je nužan i dovoljan uvjet da klasifikacijski problem bude rješiv SVM-om s tvrdom marginom?**

A Svaka spojnica između primjera iz jedne klase treba biti izvan konveksne ljuske druge klase

B Konveksne ljuske dviju klasa ne smiju se preklapati (trebaju biti disjunktne)

C Primjeri iz obje klase trebaju činiti konveksne skupove u ulaznom prostoru

D Najbliži primjeri iz suprotnih klasa moraju biti u vrhovima konveksnih ljusaka

10 (P) Razmatramo sljedeći skup označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a, i to model s tvrdom marginom te model s mekom marginom sa $C = 1$. Kod modela s mekom marginom za dualne koeficijente vrijedi $\alpha_1 = 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, $\alpha_4 > 0$ i $\alpha_5 = 1$. Skicirajte tvrdu i meku marginu u ulaznome prostoru. **Koliko je meka margina šira od tvrde margine?**

A $\frac{4}{5}\sqrt{10}$ puta B $\frac{1}{6}\sqrt{2}$ puta C $\frac{2}{5}\sqrt{10}$ puta D $\frac{3}{5}\sqrt{10}$ puta

11 (N) Algoritam k-NN koristimo za višeklasnu klasifikaciju riječi prema jeziku kojemu pripadaju. Skup za učenje sastoji se od sljedećih riječi i oznaka klasa:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{(\text{“water”}, 0), (\text{“voda”}, 1), (\text{“zrak”}, 1), (\text{“luft”}, 2), (\text{“feuer”}, 2)\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje je su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“voda”}) = 1/8 = 0.125$. Razmatramo dvije varijante algoritma: 3-NN i težinski k-NN. Kod potonjeg u obzir uzimamo sve primjere, tj. $k = N$. Odredite klasifikaciju primjera $\mathbf{x} =$ “wasser” pomoću ova dva algoritma. U slučaju jednake sličnosti između dva primjera, kao susjed se uzima onaj koji je u skupu $\mathcal{D}$ naveden prvi. U slučaju izjednačenja glasova između klasa, prednost se daje klasi s numerički manjom oznakom y. **U koju će klasu biti klasificiran primjer x algoritmom 3-NN, a u koju algoritmom težinski k-NN?**

A $y = 0$ i $y = 2$ B $y = 0$ i $y = 1$ C $y = 0$ i $y = 0$ D $y = 1$ i $y = 1$

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

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

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

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

A 0.588 B 0.741 C 0.488 D 0.322

13 (T) Nepristranost je jedno od svojstava statističkih procjenitelja. Procjenitelj MLE ne mora nužno biti nepristran, tj. može biti pristran. **Za koji od sljedećih parametara distribucije procjena MLE pristrana?**

A Srednja vrijednost Gaussove distribucije

B Varijanca Bernoullijeve distribucije

C Srednja vrijednost Bernoullijeve distribucije

D Kovarijacijska matrica Gaussove distribucije

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

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

Probabilistički grafički modeli *(3 pitanja)*

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

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

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

A 404 B 152 C 180 D 89

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

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

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

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

17 (T) Glavna svrha probabilističkih grafičkih modela (PGM) jest provođenje probabilističkih upita. Jedna vrsta upita su MAP-upiti (upiti najvjerojatnijeg objašnjenja). Neka su $\mathbf{x}_q$, $\mathbf{x}_o$ i $\mathbf{x}_n$ skupovi varijabli upita, opaženih varijabli odnosno varijabli smetnje. **Kako je definiran rezultat MAP-upita?**

A $\mathrm{argmax}_{\mathbf{x}_q} \sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$ C $\mathrm{argmax}_{\mathbf{x}_q} \sum_{\mathbf{x}_o} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$

B $\mathrm{argmax}_{\mathbf{x}_o} \sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$ D $\mathrm{argmax}_{\mathbf{x}_o} \sum_{\mathbf{x}_q} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$

Grupiranje *(3 pitanja)*

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

A 4.25 B 6.66 C 3.00 D 1.85

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

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

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

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

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

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

A $LL^0_2 \geqslant LL^0_4$, $LL^*_1 \geqslant LL^*_2 \geqslant LL^*_3$ C $LL^0_2 \geqslant LL^0_4 \geqslant LL^0_3$, $LL^*_1 \geqslant LL^*_2$

B $LL^0_3 \geqslant LL^0_4$, $LL^*_1 \geqslant LL^*_3 \geqslant LL^*_4$ D $LL^0_2 \leqslant LL^0_4$, $LL^*_2 \leqslant LL^*_1 \geqslant LL^*_3$

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

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

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

A 0.354 B 0.292 C 0.535 D 0.583

Vrednovanje modela *(2 pitanja)*

21 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$ i $\gamma \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 10 ponavljanja u vanjskoj petlji i 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 35721 B 44640 C 69201 D 49600

22 (T) Ugniježđena k-struka unakrsna provjera često se koristi za procjenu točnosti modela. **Što je prednost ugniježđene k-struke unakrsne provjere u odnosu na običnu unakrsnu provjeru?**

A Procjenjujemo pogrešku generalizacije, a ne pogrešku učenja

B Procjenjujemo očekivanu ispitnu pogrešku modela optimalne složenosti

C Model ispitujemo na cijelom raspoloživom skupu primjera

D Procjenjujemo ispitnu pogrešku modela s najmanjom pogreškom učenja

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

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (T) Optimizacija modela hrbatne regresije (L_2-regularizirane linearne regresije) ima rješenje u zatvorenoj formi. Neka je λ regularizacijski faktor, n broj značajki u ulaznom prostoru (bez "dummy" jedinice), m broj značajki u prostoru značajki (također bez "dummy" jedinice) te N broj primjera. Glavna komponenta rješenja je izračun inverza matrice izračunate na temelju matrice dizajna $\mathbf{\Phi}$. **Koliko redaka odnosno stupaca ima matrica koju invertiramo?**

A $m+1$ B N C $m+\lambda$ D $n+\lambda$

2 (N) Raspolažemo sljedećim skupom primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 1$. Dobili smo težine $\mathbf{w} = (2.12, -0.94, -0.08)$. **Koliko iznosi L_1-regularizirana pogreška $E(\mathbf{w}|\mathcal{D})$?**

A 0.29 B 7.10 C 1.58 D 2.69

3 (P) Koristimo regresiju za predviđanje uspjeha na studiju. Kao značajke možemo koristiti ocjene u četiri razreda srednje škole (značajke x_1–x_4), prosjek ocjena sva četiri razreda (x_5) te uspjeh iz matematike (x_6) i fizike (x_7) na državnoj maturi (ukupno 7 značajki). Ne moramo iskoristiti sve značajke, ali ih želimo iskoristiti što više. Za preslikavanje u prostor značajki koristimo preslikavanje s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki (npr. x_1x_2) i interakcije trojki (npr. $x_1x_2x_3$) između svih značajki koje koristimo. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 38 B 63 C 75 D 48

4 (P) Skup označenih primjera u dvodimenzijskome ulaznom prostoru je:

$$\mathcal{D} = \{((0,0),0),((0,1),0),((1,1),1)\}$$

Koja je veličina prostora inačica, $|\mathrm{VS}_{\mathcal{H},\mathcal{D}}|$?

A Beskonačno mnogo B Pitanje nema smisla jer nije definiran model C 14 D 16

Linearni klasifikacijski modeli *(3 pitanja)*

5 (N) Raspolažemo označenim skupom primjera iz triju klasa ($K = 3$) u trodimenzijskome ulaznom prostoru ($n = 3$). Na tom skupu treniramo model multinomijalne logističke regresije. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 0 & -2 \\ 3 & -4 & 6 \\ -3 & 0 & 2 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (-4, -1, -3)$ s oznakom $\mathbf{y} = (1, 0, 0)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi dotični primjer?**

A 4.02 B 8.00 C 6.00 D 12.02

6 (P) Treniramo linearni diskriminativni model u dvodimenzijskome prostoru primjera. Skup za učenje čine samo dva primjera, $(\mathbf{x}_1, y_1) = ((1,0), +1)$ i $(\mathbf{x}_2, y_2) = ((0,1), -1)$. Na tom skupu treniramo model koji ima induktivnu pristranost takvu da rješenje maksimizira minimalnu udaljenost primjera od hiperravnine. Naučen model ispravno klasificira oba primjera, pri čemu za oba primjera vrijedi $y \cdot h(\mathbf{x}) = 5$. **Koliko iznosi težina w_2 tako naučenog modela?**

A -1 B 1 C 5 D -5

7 (T) Neregularizirani model logističke regresije sklon je prenaučenosti. To je pogotovo slučaj ako se model trenira na linearno odvojivim podatcima, čak i onda kada u podatcima nema nikakvoga šuma i kada se ne koristi nikakvo preslikavanje u prostor značajki. **Zbog čega dolazi do prenaučenosti modela neregularizirane logističke regresije na linearno odvojivim skupovima podataka?**

A Netočno klasificirani primjeri nanose gubitak koji je proporcionalan normi vektora težina

B Empirijska pogreška logističke regresije smanjuje se s porastom broja primjera

C Ispravno klasificirani primjeri koji su vrlo udaljeni od granice nanose malen gubitak

D S porastom norme vektora težina gubitak na točnim primjerima teži prema nuli

Jezgrene i neparametarske metode *(4 pitanja)*

8 (N) Algoritam k-NN koristimo za višeklasnu klasifikaciju riječi prema jeziku kojemu pripadaju. Skup za učenje sastoji se od sljedećih riječi i oznaka klasa:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{(\text{“water”}, 0), (\text{“voda”}, 1), (\text{“zrak”}, 1), (\text{“luft”}, 2), (\text{“feuer”}, 2)\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2| / |\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje je su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“voda”}) = 1/8 = 0.125$. Razmatramo dvije varijante algoritma: 3-NN i težinski k-NN. Kod potonjeg u obzir uzimamo sve primjere, tj. $k = N$. Odredite klasifikaciju primjera $\mathbf{x} = \text{“wasser”}$ pomoću ova dva algoritma. U slučaju jednake sličnosti između dva primjera, kao susjed se uzima onaj koji je u skupu $\mathcal{D}$ naveden prvi. U slučaju izjednačenja glasova između klasa, prednost se daje klasi s numerički manjom oznakom y. **U koju će klasu biti klasificiran primjer x algoritmom 3-NN, a u koju algoritmom težinski k-NN?**

A $y = 0$ i $y = 0$ B $y = 0$ i $y = 2$ C $y = 1$ i $y = 1$ D $y = 0$ i $y = 1$

9 (P) Razmatramo sljedeći skup označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a, i to model s tvrdom marginom te model s mekom marginom sa $C = 1$. Kod modela s mekom marginom za dualne koeficijente vrijedi $\alpha_1 = 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, $\alpha_4 > 0$ i $\alpha_5 = 1$. Skicirajte tvrdu i meku marginu u ulaznome prostoru. **Koliko je meka margina šira od tvrde margine?**

A $\frac{3}{5}\sqrt{10}$ puta B $\frac{4}{5}\sqrt{10}$ puta C $\frac{2}{5}\sqrt{10}$ puta D $\frac{1}{6}\sqrt{2}$ puta

10 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 4$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 1)^3$. Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((11, 32, 21), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((5, 11, 6), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((6, 12, 6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 1.063 \cdot 10^{-7}$, $\alpha_2 = 1.022 \cdot 10^{-5}$ i $\alpha_3 = 1.032 \cdot 10^{-5}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (-3, -5, 3)$.**

A -4.093 B -2.330 C -3.553 D $+1.434$

11 (T) Problem maksimalne margine ima svoju geometrijsku interpretaciju: maksimalna margina odgovara simetrali spojnice konveksnih ljusaka primjera iz dviju klasa. **Što je nužan i dovoljan uvjet da klasifikacijski problem bude rješiv SVM-om s tvrdom marginom?**

- A Primjeri iz obje klase trebaju činiti konveksne skupove u ulaznom prostoru
- B Konveksne ljuske dviju klasa ne smiju se preklapati (trebaju biti disjunktne)
- C Svaka spojnica između primjera iz jedne klase treba biti izvan konveksne ljuske druge klase
- D Najbliži primjeri iz suprotnih klasa moraju biti u vrhovima konveksnih ljusaka

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (T) Nepristranost je jedno od svojstava statističkih procjenitelja. Procjenitelj MLE ne mora nužno biti nepristran, tj. može biti pristran. **Za koji od sljedećih parametara distribucije procjena MLE pristrana?**

- A Kovarijacijska matrica Gaussove distribucije
- B Varijanca Bernoullijeve distribucije
- C Srednja vrijednost Gaussove distribucije
- D Srednja vrijednost Bernoullijeve distribucije

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

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

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

A 0.488 B 0.322 C 0.588 D 0.741

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

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

Probabilistički grafički modeli *(3 pitanja)*

15 (T) Glavna svrha probabilističkih grafičkih modela (PGM) jest provođenje probabilističkih upita. Jedna vrsta upita su MAP-upiti (upiti najvjerojatnijeg objašnjenja). Neka su $\mathbf{x}_q$, $\mathbf{x}_o$ i $\mathbf{x}_n$ skupovi varijabli upita, opaženih varijabli odnosno varijabli smetnje. **Kako je definiran rezultat MAP-upita?**

A $\text{argmax}_{\mathbf{x}_o} \sum_{\mathbf{x}_q} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$ C $\text{argmax}_{\mathbf{x}_q} \sum_{\mathbf{x}_o} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$

B $\text{argmax}_{\mathbf{x}_q} \sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$ D $\text{argmax}_{\mathbf{x}_o} \sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$

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

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

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

A 180 B 152 C 89 D 404

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

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

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

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

Grupiranje *(3 pitanja)*

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

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

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

A 0.292 B 0.478 C 0.535 D 0.583

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

$\mathcal{H}_1$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina
$\mathcal{H}_2$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom
$\mathcal{H}_3$: Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom
$\mathcal{H}_4$: Model sa $K = 10$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

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

A $LL^0_2 \geqslant LL^0_4 \geqslant LL^0_3$, $LL^*_1 \geqslant LL^*_2$ C $LL^0_2 \geqslant LL^0_4$, $LL^*_1 \geqslant LL^*_2 \geqslant LL^*_3$
B $LL^0_2 \leqslant LL^0_4$, $LL^*_2 \leqslant LL^*_1 \geqslant LL^*_3$ D $LL^0_3 \geqslant LL^0_4$, $LL^*_1 \geqslant LL^*_3 \geqslant LL^*_4$

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

A 3.00 B 6.66 C 4.25 D 1.85

Vrednovanje modela *(2 pitanja)*

21 (T) Ugniježđena k-struka unakrsna provjera često se koristi za procjenu točnosti modela. **Što je prednost ugniježđene k-struke unakrsne provjere u odnosu na običnu unakrsnu provjeru?**

- A Procjenjujemo očekivanu ispitnu pogrešku modela optimalne složenosti
- B Procjenjujemo ispitnu pogrešku modela s najmanjom pogreškom učenja
- C Procjenjujemo pogrešku generalizacije, a ne pogrešku učenja
- D Model ispitujemo na cijelom raspoloživom skupu primjera

22 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$ i $\gamma \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 10 ponavljanja u vanjskoj petlji i 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 44640 B 35721 C 49600 D 69201

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

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) Skup označenih primjera u dvodimenzijskome ulaznom prostoru je:

$$\mathcal{D} = \{((0,0),0),((0,1),0),((1,1),1)\}$$

Koja je veličina prostora inačica, $|\mathrm{VS}_{\mathcal{H},\mathcal{D}}|$?

A Pitanje nema smisla jer nije definiran model B 16 C 14 D Beskonačno mnogo

2 (T) Optimizacija modela hrbatne regresije (L_2-regularizirane linearne regresije) ima rješenje u zatvorenoj formi. Neka je λ regularizacijski faktor, n broj značajki u ulaznom prostoru (bez "dummy" jedinice), m broj značajki u prostoru značajki (također bez "dummy" jedinice) te N broj primjera. Glavna komponenta rješenja je izračun inverza matrice izračunate na temelju matrice dizajna $\boldsymbol{\Phi}$. **Koliko redaka odnosno stupaca ima matrica koju invertiramo?**

A $m + \lambda$ B N C $n + \lambda$ D $m + 1$

3 (P) Koristimo regresiju za predviđanje uspjeha na studiju. Kao značajke možemo koristiti ocjene u četiri razreda srednje škole (značajke x_1–x_4), prosjek ocjena sva četiri razreda (x_5) te uspjeh iz matematike (x_6) i fizike (x_7) na državnoj maturi (ukupno 7 značajki). Ne moramo iskoristiti sve značajke, ali ih želimo iskoristiti što više. Za preslikavanje u prostor značajki koristimo preslikavanje s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki (npr. x_1x_2) i interakcije trojki (npr. $x_1x_2x_3$) između svih značajki koje koristimo. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 75 B 38 C 63 D 48

4 (N) Raspolažemo sljedećim skupom primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 1$. Dobili smo težine $\mathbf{w} = (2.12, -0.94, -0.08)$. **Koliko iznosi L_1-regularizirana pogreška $E(\mathbf{w}|\mathcal{D})$?**

A 2.69 B 7.10 C 1.58 D 0.29

Linearni klasifikacijski modeli *(3 pitanja)*

5 (T) Neregularizirani model logističke regresije sklon je prenaučenosti. To je pogotovo slučaj ako se model trenira na linearno odvojivim podatcima, čak i onda kada u podatcima nema nikakvoga šuma i kada se ne koristi nikakvo preslikavanje u prostor značajki. **Zbog čega dolazi do prenaučenosti modela neregularizirane logističke regresije na linearno odvojivim skupovima podataka?**

A S porastom norme vektora težina gubitak na točnim primjerima teži prema nuli

B Ispravno klasificirani primjeri koji su vrlo udaljeni od granice nanose malen gubitak

C Netočno klasificirani primjeri nanose gubitak koji je proporcionalan normi vektora težina

D Empirijska pogreška logističke regresije smanjuje se s porastom broja primjera

6 (P) Treniramo linearni diskriminativni model u dvodimenzijskome prostoru primjera. Skup za učenje čine samo dva primjera, $(\mathbf{x}_1, y_1) = ((1,0), +1)$ i $(\mathbf{x}_2, y_2) = ((0,1), -1)$. Na tom skupu treniramo model koji ima induktivnu pristranost takvu da rješenje maksimizira minimalnu udaljenost primjera od hiperravnine. Naučen model ispravno klasificira oba primjera, pri čemu za oba primjera vrijedi $y \cdot h(\mathbf{x}) = 5$. **Koliko iznosi težina w_2 tako naučenog modela?**

[A] 5 [B] 1 [C] -1 [D] -5

7 (N) Raspolažemo označenim skupom primjera iz triju klasa ($K = 3$) u trodimenzijskome ulaznom prostoru ($n = 3$). Na tom skupu treniramo model multinomijalne logističke regresije. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 0 & -2 \\ 3 & -4 & 6 \\ -3 & 0 & 2 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (-4, 1, -3)$ s oznakom $\mathbf{y} = (0, 1, 0)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi dotični primjer?**

[A] 8.00 [B] 6.00 [C] 4.02 [D] 12.02

Jezgrene i neparametarske metode *(4 pitanja)*

8 (T) Problem maksimalne margine ima svoju geometrijsku interpretaciju: maksimalna margina odgovara simetrali spojnice konveksnih ljusaka primjera iz dviju klasa. **Što je nužan i dovoljan uvjet da klasifikacijski problem bude rješiv SVM-om s tvrdom marginom?**

[A] Najbliži primjeri iz suprotnih klasa moraju biti u vrhovima konveksnih ljusaka

[B] Primjeri iz obje klase trebaju činiti konveksne skupove u ulaznom prostoru

[C] Svaka spojnica između primjera iz jedne klase treba biti izvan konveksne ljuske druge klase

[D] Konveksne ljuske dviju klasa ne smiju se preklapati (trebaju biti disjunktne)

9 (P) Razmatramo sljedeći skup označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a, i to model s tvrdom marginom te model s mekom marginom sa $C = 1$. Kod modela s mekom marginom za dualne koeficijente vrijedi $\alpha_1 = 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, $\alpha_4 > 0$ i $\alpha_5 = 1$. Skicirajte tvrdu i meku marginu u ulaznome prostoru. **Koliko je meka margina šira od tvrde margine?**

[A] $\frac{3}{5}\sqrt{10}$ puta [B] $\frac{2}{5}\sqrt{10}$ puta [C] $\frac{1}{6}\sqrt{2}$ puta [D] $\frac{4}{5}\sqrt{10}$ puta

10 (N) Algoritam k-NN koristimo za višeklasnu klasifikaciju riječi prema jeziku kojemu pripadaju. Skup za učenje sastoji se od sljedećih riječi i oznaka klasa:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{(\text{“water”}, 0), (\text{“voda”}, 1), (\text{“zrak”}, 1), (\text{“luft”}, 2), (\text{“feuer”}, 2)\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2| / |\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“voda”}) = 1/8 = 0.125$. Razmatramo dvije varijante algoritma: 3-NN i težinski k-NN. Kod potonjeg u obzir uzimamo sve primjere, tj. $k = N$. Odredite klasifikaciju primjera $\mathbf{x} = \text{“zemlja”}$ pomoću ova dva algoritma. U slučaju jednake sličnosti između dva primjera, kao susjed se uzima onaj koji je u skupu $\mathcal{D}$ naveden prvi. U slučaju izjednačenja glasova između klasa, prednost se daje klasi s numerički manjom oznakom y. **U koju će klasu biti klasificiran primjer x algoritmom 3-NN, a u koju algoritmom težinski k-NN?**

[A] $y = 0$ i $y = 2$ [B] $y = 1$ i $y = 1$ [C] $y = 0$ i $y = 0$ [D] $y = 0$ i $y = 1$

11 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 4$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 1)^3$. Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((11, 32, 21), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((5, 11, 6), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((6, 12, 6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 1.063 \cdot 10^{-7}$, $\alpha_2 = 1.022 \cdot 10^{-5}$ i $\alpha_3 = 1.032 \cdot 10^{-5}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (5, -15, -1)$.**

A -1.486 B -4.093 C $+5.454$ D -3.553

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

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

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

13 (T) Nepristranost je jedno od svojstava statističkih procjenitelja. Procjenitelj MLE ne mora nužno biti nepristran, tj. može biti pristran. **Za koji od sljedećih parametara distribucije procjena MLE pristrana?**

A Srednja vrijednost Gaussove distribucije

B Varijanca Bernoullijeve distribucije

C Srednja vrijednost Bernoullijeve distribucije

D Kovarijacijska matrica Gaussove distribucije

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

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

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

A 0.741 B 0.588 C 0.488 D 0.322

Probabilistički grafički modeli *(3 pitanja)*

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

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

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

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

16 (T) Glavna svrha probabilističkih grafičkih modela (PGM) jest provođenje probabilističkih upita. Jedna vrsta upita su MAP-upiti (upiti najvjerojatnijeg objašnjenja). Neka su $\mathbf{x}_q$, $\mathbf{x}_o$ i $\mathbf{x}_n$ skupovi varijabli upita, opaženih varijabli odnosno varijabli smetnje. **Kako je definiran rezultat MAP-upita?**

A $\mathrm{argmax}_{\mathbf{x}_q} \sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$ C $\mathrm{argmax}_{\mathbf{x}_o} \sum_{\mathbf{x}_q} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$
B $\mathrm{argmax}_{\mathbf{x}_o} \sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$ D $\mathrm{argmax}_{\mathbf{x}_q} \sum_{\mathbf{x}_o} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$

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

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

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

A 180 B 89 C 404 D 152

Grupiranje *(3 pitanja)*

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

A 3.00 B 4.25 C 6.66 D 1.85

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

$\mathcal{H}_1$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina
$\mathcal{H}_2$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom
$\mathcal{H}_3$: Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom
$\mathcal{H}_4$: Model sa $K = 10$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

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

A $LL^0_2 \geqslant LL^0_4,\ LL^*_1 \geqslant LL^*_2 \geqslant LL^*_3$ C $LL^0_2 \geqslant LL^0_4 \geqslant LL^0_3,\ LL^*_1 \geqslant LL^*_2$
B $LL^0_3 \geqslant LL^0_4,\ LL^*_1 \geqslant LL^*_3 \geqslant LL^*_4$ D $LL^0_2 \leqslant LL^0_4,\ LL^*_2 \leqslant LL^*_1 \geqslant LL^*_3$

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

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

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

A 0.354 B 0.292 C 0.535 D 0.583

Vrednovanje modela *(2 pitanja)*

21 (T) Ugniježđena k-struka unakrsna provjera često se koristi za procjenu točnosti modela. **Što je prednost ugniježđene k-struke unakrsne provjere u odnosu na običnu unakrsnu provjeru?**

A Procjenjujemo ispitnu pogrešku modela s najmanjom pogreškom učenja

B Model ispitujemo na cijelom raspoloživom skupu primjera

C Procjenjujemo očekivanu ispitnu pogrešku modela optimalne složenosti

D Procjenjujemo pogrešku generalizacije, a ne pogrešku učenja

22 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \ldots, 2^{15}\}$ i $\gamma \in \{2^{-15}, 2^{-14}, \ldots, 2^{15}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 10 ponavljanja u vanjskoj petlji i 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 35721 B 49600 C 69201 D 44640

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

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (N) Raspolažemo sljedećim skupom primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 1$. Dobili smo težine $\mathbf{w} = (2.12, -0.94, -0.08)$. **Koliko iznosi L_1-regularizirana pogreška $E(\mathbf{w}|\mathcal{D})$?**

A 1.58 B 2.69 C 7.10 D 0.29

2 (T) Optimizacija modela hrbatne regresije (L_2-regularizirane linearne regresije) ima rješenje u zatvorenoj formi. Neka je λ regularizacijski faktor, n broj značajki u ulaznom prostoru (bez "dummy" jedinice), m broj značajki u prostoru značajki (također bez "dummy" jedinice) te N broj primjera. Glavna komponenta rješenja je izračun inverza matrice izračunate na temelju matrice dizajna $\mathbf{\Phi}$. **Koliko redaka odnosno stupaca ima matrica koju invertiramo?**

A $n + \lambda$ B $m + 1$ C N D $m + \lambda$

3 (P) Koristimo regresiju za predviđanje uspjeha na studiju. Kao značajke možemo koristiti ocjene u četiri razreda srednje škole (značajke x_1–x_4), prosjek ocjena sva četiri razreda (x_5) te uspjeh iz matematike (x_6) i fizike (x_7) na državnoj maturi (ukupno 7 značajki). Ne moramo iskoristiti sve značajke, ali ih želimo iskoristiti što više. Za preslikavanje u prostor značajki koristimo preslikavanje s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki (npr. x_1x_2) i interakcije trojki (npr. $x_1x_2x_3$) između svih značajki koje koristimo. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 48 B 38 C 63 D 75

4 (P) Skup označenih primjera u dvodimenzijskome ulaznom prostoru je:

$$\mathcal{D} = \{((0,0),0),((0,1),0),((1,1),1)\}$$

Koja je veličina prostora inačica, $|\mathrm{VS}_{\mathcal{H},\mathcal{D}}|$?

A Beskonačno mnogo B Pitanje nema smisla jer nije definiran model C 14 D 16

Linearni klasifikacijski modeli *(3 pitanja)*

5 (N) Raspolažemo označenim skupom primjera iz triju klasa ($K = 3$) u trodimenzijskome ulaznom prostoru ($n = 3$). Na tom skupu treniramo model multinomijalne logističke regresije. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 0 & -2 \\ 3 & -4 & 6 \\ -3 & 0 & 2 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (-4, -1, -3)$ s oznakom $\mathbf{y} = (1, 0, 0)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi dotični primjer?**

A 6.00 B 4.02 C 8.00 D 12.02

6 (P) Treniramo linearni diskriminativni model u dvodimenzijskome prostoru primjera. Skup za učenje čine samo dva primjera, $(\mathbf{x}_1, y_1) = ((1,0), +1)$ i $(\mathbf{x}_2, y_2) = ((0,1), -1)$. Na tom skupu treniramo model koji ima induktivnu pristranost takvu da rješenje maksimizira minimalnu udaljenost primjera od hiperravnine. Naučen model ispravno klasificira oba primjera, pri čemu za oba primjera vrijedi $y \cdot h(\mathbf{x}) = 5$. **Koliko iznosi težina w_2 tako naučenog modela?**

A -5 B 1 C -1 D 5

7 (T) Neregularizirani model logističke regresije sklon je prenaučenosti. To je pogotovo slučaj ako se model trenira na linearno odvojivim podatcima, čak i onda kada u podatcima nema nikakvoga šuma i kada se ne koristi nikakvo preslikavanje u prostor značajki. **Zbog čega dolazi do prenaučenosti modela neregularizirane logističke regresije na linearno odvojivim skupovima podataka?**

A S porastom norme vektora težina gubitak na točnim primjerima teži prema nuli

B Empirijska pogreška logističke regresije smanjuje se s porastom broja primjera

C Netočno klasificirani primjeri nanose gubitak koji je proporcionalan normi vektora težina

D Ispravno klasificirani primjeri koji su vrlo udaljeni od granice nanose malen gubitak

Jezgrene i neparametarske metode *(4 pitanja)*

8 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 4$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 1)^3$. Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((11, 32, 21), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((5, 11, 6), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((6, 12, 6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 1.063 \cdot 10^{-7}$, $\alpha_2 = 1.022 \cdot 10^{-5}$ i $\alpha_3 = 1.032 \cdot 10^{-5}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3, 0, -3)$.**

A -2.330 B $+1.434$ C -0.676 D -3.553

9 (N) Algoritam k-NN koristimo za višeklasnu klasifikaciju riječi prema jeziku kojemu pripadaju. Skup za učenje sastoji se od sljedećih riječi i oznaka klasa:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{(\text{“water”}, 0), (\text{“voda”}, 1), (\text{“zrak”}, 1), (\text{“luft”}, 2), (\text{“feuer”}, 2)\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2| / |\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje je su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“voda”}) = 1/8 = 0.125$. Razmatramo dvije varijante algoritma: 3-NN i težinski k-NN. Kod potonjeg u obzir uzimamo sve primjere, tj. $k = N$. Odredite klasifikaciju primjera $\mathbf{x} =$ “love” pomoću ova dva algoritma. U slučaju jednake sličnosti između dva primjera, kao susjed se uzima onaj koji je u skupu $\mathcal{D}$ naveden prvi. U slučaju izjednačenja glasova između klasa, prednost se daje klasi s numerički manjom oznakom y. **U koju će klasu biti klasificiran primjer x algoritmom 3-NN, a u koju algoritmom težinski k-NN?**

A $y = 2$ i $y = 1$ B $y = 0$ i $y = 0$ C $y = 0$ i $y = 2$ D $y = 1$ i $y = 1$

10 (P) Razmatramo sljedeći skup označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a, i to model s tvrdom marginom te model s mekom marginom sa $C = 1$. Kod modela s mekom marginom za dualne koeficijente vrijedi $\alpha_1 = 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, $\alpha_4 > 0$ i $\alpha_5 = 1$. Skicirajte tvrdu i meku marginu u ulaznome prostoru. **Koliko je meka margina šira od tvrde margine?**

A $\frac{1}{6}\sqrt{2}$ puta B $\frac{3}{5}\sqrt{10}$ puta C $\frac{4}{5}\sqrt{10}$ puta D $\frac{2}{5}\sqrt{10}$ puta

11 (T) Problem maksimalne margine ima svoju geometrijsku interpretaciju: maksimalna margina odgovara simetrali spojnice konveksnih ljusaka primjera iz dviju klasa. **Što je nužan i dovoljan uvjet da klasifikacijski problem bude rješiv SVM-om s tvrdom marginom?**

- A Konveksne ljuske dviju klasa ne smiju se preklapati (trebaju biti disjunktne)
- B Primjeri iz obje klase trebaju činiti konveksne skupove u ulaznom prostoru
- C Najbliži primjeri iz suprotnih klasa moraju biti u vrhovima konveksnih ljusaka
- D Svaka spojnica između primjera iz jedne klase treba biti izvan konveksne ljuske druge klase

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

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

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

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

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

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

A 0.488 B 0.322 C 0.588 D 0.741

14 (T) Nepristranost je jedno od svojstava statističkih procjenitelja. Procjenitelj MLE ne mora nužno biti nepristran, tj. može biti pristran. **Za koji od sljedećih parametara distribucije procjena MLE pristrana?**

- A Varijanca Bernoullijeve distribucije
- B Srednja vrijednost Gaussove distribucije
- C Kovarijacijska matrica Gaussove distribucije
- D Srednja vrijednost Bernoullijeve distribucije

Probabilistički grafički modeli *(3 pitanja)*

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

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

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

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

16 (T) Glavna svrha probabilističkih grafičkih modela (PGM) jest provođenje probabilističkih upita. Jedna vrsta upita su MAP-upiti (upiti najvjerojatnijeg objašnjenja). Neka su $\mathbf{x}_q$, $\mathbf{x}_o$ i $\mathbf{x}_n$ skupovi varijabli upita, opaženih varijabli odnosno varijabli smetnje. **Kako je definiran rezultat MAP-upita?**

A $\text{argmax}_{\mathbf{x}_o} \sum_{\mathbf{x}_q} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$ C $\text{argmax}_{\mathbf{x}_q} \sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$

B $\text{argmax}_{\mathbf{x}_q} \sum_{\mathbf{x}_o} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$ D $\text{argmax}_{\mathbf{x}_o} \sum_{\mathbf{x}_n} p(\mathbf{x}_q, \mathbf{x}_o, \mathbf{x}_n)$

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

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

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

A 152 B 180 C 89 D 404

Grupiranje *(3 pitanja)*

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

A 1.85 B 4.25 C 6.66 D 3.00

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

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

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

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

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

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

A $LL^0_2 \leqslant LL^0_4,\ LL^*_2 \leqslant LL^*_1 \geqslant LL^*_3$ C $LL^0_3 \geqslant LL^0_4,\ LL^*_1 \geqslant LL^*_3 \geqslant LL^*_4$

B $LL^0_2 \geqslant LL^0_4,\ LL^*_1 \geqslant LL^*_2 \geqslant LL^*_3$ D $LL^0_2 \geqslant LL^0_4 \geqslant LL^0_3,\ LL^*_1 \geqslant LL^*_2$

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

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

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

A 0.535 B 0.292 C 0.478 D 0.583

Vredňovanje modela *(2 pitanja)*

21 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$ i $\gamma \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 10 ponavljanja u vanjskoj petlji i 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

[A] 35721 [B] 49600 [C] 69201 [D] 44640

22 (T) Ugniježđena k-struka unakrsna provjera često se koristi za procjenu točnosti modela. **Što je prednost ugniježđene k-struke unakrsne provjere u odnosu na običnu unakrsnu provjeru?**

[A] Procjenjujemo pogrešku generalizacije, a ne pogrešku učenja

[B] Procjenjujemo ispitnu pogrešku modela s najmanjom pogreškom učenja

[C] Model ispitujemo na cijelom raspoloživom skupu primjera

[D] Procjenjujemo očekivanu ispitnu pogrešku modela optimalne složenosti

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