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

– NEKORIGIRANA VERZIJA –

*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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija *(4 pitanja)*

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

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((2,-3),5),((2,3),-10),((-3,5),5),((5,0),20)\}$$

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 2$. Dobili smo težine $\mathbf{w} = (4.568, 0.746, -0.550)$. **Koliko iznosi L_1-regularizirana pogreška $E(\mathbf{w}|\mathcal{D})$?**

A 191.95 B 165.89 C 230.98 D 31.70

2 (T) Postupak najmanjih kvadrata (OLS) temelji se na izračunu pseudoinverza $\mathbf{X}^+$ matrice dizajna $\mathbf{X}$, što je poopćenje običnog inverza $\mathbf{X}^{-1}$. **U kojoj situaciji je rješenje dobiveno pseudoinverzom identično rješenju dobivenom običnim inverzom?**

A Kada je broj primjera jednak broju značajki plus jedan i nema multikolinearnosti

B Kada je broj primjera veći od broja značajki

C Kada je broj značajki manji od broja primjera i nema multikolinearnosti

D Kada nema multikolinearnosti i matrica dizajna je dobro kondicionirana

3 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{((0,0),1),((-1,-1),0),((1,1),0)\}$. Razmatramo sljedeće modele $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi} : \mathbb{Z}^2 \to \mathbb{Z}^2$, kojom primjere iz $\mathcal{D}$ preslikavamo u matricu dizajna $\mathbf{\Phi}$:

$$\mathcal{H}_1 : h_1(\mathbf{x}; \theta_0, \theta_1) = \mathbf{1}\{\theta_1 x_1 + \theta_0 \geqslant 0\} \qquad \boldsymbol{\phi}_1(\mathbf{x}) = (1, x_2, x_1)$$
$$\mathcal{H}_2 : h_2(\mathbf{x}; \theta_0, \theta_2) = \mathbf{1}\{\theta_2 x_2 + \theta_0 \geqslant 0\} \qquad \boldsymbol{\phi}_2(\mathbf{x}) = (1, x_1, x_1 x_2)$$
$$\mathcal{H}_3 : h_3(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\}$$
$$\mathcal{H}_4 : h_4(\mathbf{x}; \theta_0) = \mathbf{1}\{x_1^2 + x_2^2 \geqslant \theta_0\}$$

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Razmotrite sve kombinacije modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$. **Za koju kombinaciju modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$ postoji samo jedna hipoteza $h \in \mathcal{H}$ za koju $E(h|\mathcal{D}) = 0$?**

A $h_1 + \boldsymbol{\phi}_1$ B $h_3 + \boldsymbol{\phi}_2$ C $h_2 + \boldsymbol{\phi}_2$ D $h_4 + \boldsymbol{\phi}_1$

4 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\mathbf{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$. Treniranjem modela na skupu $(\mathbf{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (-0.81, 2.32, 0.66, 0.1)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x}; \mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu, \sigma^2) = (\sqrt{2\pi}\sigma)^{-1} \exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\mathbf{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln \mathcal{L}(\mathbf{w}|\mathbf{\Phi}, \mathbf{y})$?

A -12.63 B -5.69 C -10.64 D -4.73

Cjelina 2: Linearni klasifikacijski modeli *(3 pitanja)*

5 (T) Jedna od triju komponenta svakog algoritma strojnog učenja je funkcija gubitka. Razmotrite funkcije gubitka perceptrona, logističke regresije (LR) i SVM-a. **Što je specifično funkciji gubitka perceptrona u odnosu na funkcije gubitka LR-a i SVM-a?**

- A Svaki primjer nanosi gubitak, ali manji za točno klasificirane primjere nego za netočno klasificirane primjere
- B Gubitak za sve točno klasificirane primjere je nula, a za netočno klasificirane može biti manji od 1
- C Točno klasificirani primjer nanosi gubitak manji od 1 te gubitak opada što je primjer bliže granici
- D Gubitak netočno klasificiranih primjera raste linearno s udaljenošću od granice

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 $\mathbf{y} = (0, 1, 0)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi dotični primjer?**

A 4.02 B 12.02 C 6.00 D 8.00

7 (P) Na skupu od $N = 1000$ primjera sa $n = 444$ značajki rješavamo problem višeklasne klasifikacije. Imamo $K = 4$ klase, s po 400, 300, 200 i 100 primjera. Za klasifikaciju želimo koristiti binarnu logističku regresiju u shemi OVO ili u shemi OVR (ovo nije tipično, ali je moguće). Pretpostavite da ne koristimo nikakvu regularizaciju, $\lambda = 0$. Razmotrite, za obje sheme, za koliko binarnih modela će rješenje optimizacijskog postupka sigurno biti nestabilno zbog loše kondicije matrice dizajna. **Koliko modela će sigurno biti više nestabilno u shemi OVO nego u shemi OVR?**

A 2 B 3 C 4 D 5

Cjelina 3: SVM, jezgrene i neparametarske metode *(4 pitanja)*

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

$$\begin{aligned}(\mathbf{x}^{(1)}, y^{(1)}) &= ((6, 12, 6), -1)\\ (\mathbf{x}^{(2)}, y^{(2)}) &= ((-11, -26, -15), -1)\\ (\mathbf{x}^{(3)}, y^{(3)}) &= ((-1, -7, -6), +1)\end{aligned}$$

Lagrangeovi koeficijenti su $\alpha_1 = 3.496 \cdot 10^{-6}$, $\alpha_2 = 4.136 \cdot 10^{-7}$ i $\alpha_3 = 3.910 \cdot 10^{-6}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (1, 1, 25)$.**

A $+1.434$ B -0.676 C -2.330 D $+0.947$

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

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

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 veća od tvrde margine?**

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

10 (T) Stroj potpornih vektora (SVM) jedna je vrsta rijetkoga jezgrenog stroja. Jezgreni stroj za bazne funkcije koriste jezgrene funkcije izračunate u odnosu na odabrane prototipne primjere. **Po čemu je SVM specifičan u odnosu na općeniti algoritam rijetkoga jezgrenog stroja?**

A Dimenzija prostora značajki ne može biti veća od broja primjera

B Zbog L1-regularizacije, mnoge težine modela bit će pritegnute na nulu

C Broj parametara modela jednak je dimenziji prostora značajki

D Prototipni primjeri odabiru se u okviru optimizacijskog postupka

11 (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A 1/2 B 3/4 C 0 D 1/4

Cjelina 4: Procjena parametara i Bayesov klasifikator *(4 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 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 $[-4-a, -4+b]$

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

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

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

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

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

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

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

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

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

15 (T) Procjenitelj MAP kombinira funkciju izglednosti parametara s apriornom distribucijom parametara. **Što i kako maksimizira procjenitelj MAP?**

A Zajedničku vjerojatnost parametara i podataka, iterativno ako je apriorna distribucija iz eksponencijalne familije, inače u zatvorenoj formi

B Zajedničku gustoću vjerojatnosti parametara i podataka, u zatvorenoj formi, ako je apriorna distribucija konjugatna za izglednost, inače iterativno

C Aposteriornu gustoću vjerojatnosti podataka, u zatvorenoj formi ako su izglednost i apriorna distribucija konjugatne, inače iterativno

D Aposteriornu vjerojatnost podataka, u zatvorenoj formi ako je zajednička vjerojatnost parametara i podataka iz eksponencijalne familije, inače iterativno

Cjelina 5: Probabilistički grafički modeli *(3 pitanja)*

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

17 (T) Približno zaključivanje kod Bayesovih mreža može se provesti metodama uzorkovanja. Te metode generiraju slučajan uzorak primjera iz distribucije opisane Bayesovom mrežom, iz kojega se onda mogu procijeniti parametri distribucije. Međutim, da bismo mogli uzorkovati iz Bayesove mreže, već trebamo znati parametre svih uvjetnih distribucija u čvorovima. **Zašto radimo uzorkovanje iz Bayesove mreže da bismo procijenili parametre distribucije, kad te parametre već imamo?**

A Vrijednosti parametara svih distribucija su procjene koje inicijaliziramo i zatim iterativno ažuriramo u svakom koraku uzorkovanja

B Uzorkovanje koristimo da bismo procijenili parametre bilo koje distribucije, a ne samo uvjetnih distribucija u čvorovima mreže

C Ne trebamo znati parametre svih čvorova, već samo čvorova roditelja onih distribucija za koje želimo procijeniti parametre

D Uzorkovanjem procjenjujemo parametre latentnih varijabli, čije vrijednosti ne opažamo, pa nam ti parametri nisu poznati prije uzorkovanja

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

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

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

A 472 B 348 C 944 D 596

Cjelina 6: Grupiranje i vrednovanje modela *(4 pitanja)*

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

$$\begin{pmatrix} 1.0 & 0.2 & 0.8 & 0.1 & 0.4 \\ 0.2 & 1.0 & 0.9 & 0.3 & 0.7 \\ 0.8 & 0.9 & 1.0 & 0.6 & 0.5 \\ 0.1 & 0.3 & 0.6 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

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

A 2.0 B 2.4 C 1.9 D 1.8

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

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 5 & 13 & 1 \\ 6 & 15 & 4 \\ 14 & 2 & 36 \end{pmatrix} \end{array}$$

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

A 0.015 B 0.040 C 0.094 D 0.068

21 (T) Za meko grupiranje možemo koristiti model miješane gustoće s latentnim varijablama. Kod tog modela, svaki primjer $\mathbf{x}^{(i)}$ ima pridruženu latentnu varijablu $\mathbf{z}^{(i)}$. Obje ove varijable zapravo su slučajne varijable sa svojom pretpostavljenom distribucijom. **Koju distribuciju pretpostavljamo za latentnu varijablu $\mathbf{z}^{(i)}$ i zašto?**

A Gaussovu distribuciju koja opisuje vjerojatnost da je grupa generirala primjer $\mathbf{x}^{(i)}$

B Beta-distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada svakoj grupi

C Kategoričku distribuciju koja opisuje kojoj grupi primjer $\mathbf{x}^{(i)}$ zapravo pripada

D Bernoullijevu distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada grupi $\mathbf{z}^{(i)}$

22 (P) Krivuljom ROC vrednujemo binarne klasifikatore. Neka $h_1 \succ h_2$ označava da naučeni model h_1 striktno dominira nad naučenim modelom h_2 prema krivulji ROC, tj. da vrijedi

$$\forall\theta.\Big(\big(\mathrm{FPR}_\theta(h_1) = \mathrm{FPR}_\theta(h_2)\big) \Rightarrow \big(\mathrm{TPR}_\theta(h_1) > \mathrm{TPR}_\theta(h_2)\big)\Big)$$

gdje su $\mathrm{FPR}_\theta(h)$ i $\mathrm{TPR}_\theta(h)$ vrijednosti stope lažnih pozitiva odnosno stvarnih pozitiva hipoteze h s pragom θ. Nadalje, neka $h_1 > h_2$ označava da vrijedi $\mathrm{AUC}(h_1) > \mathrm{AUC}(h_2)$. **Što od sljedećeg vrijedi?**

A $h_1 \succ h_2 \Rightarrow h_1 > h_2$ B $h_1 > h_2 \Rightarrow h_1 \succ h_2$ C $h_1 > h_2 \Rightarrow h_2 \succ h_1$ D $h_1 \succ h_2 \Rightarrow h_2 > h_1$

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

– NEKORIGIRANA VERZIJA –

*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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{((0,0),1), ((-1,-1),0), ((1,1),0)\}$. Razmatramo sljedeće modele $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi} : \mathbb{Z}^2 \to \mathbb{Z}^2$, kojom primjere iz $\mathcal{D}$ preslikavamo u matricu dizajna $\boldsymbol{\Phi}$:

$$\mathcal{H}_1 : h_1(\mathbf{x}; \theta_0, \theta_1) = \mathbf{1}\{\theta_1 x_1 + \theta_0 \geqslant 0\} \qquad \boldsymbol{\phi}_1(\mathbf{x}) = (1, x_2, x_1)$$
$$\mathcal{H}_2 : h_2(\mathbf{x}; \theta_0, \theta_2) = \mathbf{1}\{\theta_2 x_2 + \theta_0 \geqslant 0\} \qquad \boldsymbol{\phi}_2(\mathbf{x}) = (1, x_1, x_1 x_2)$$
$$\mathcal{H}_3 : h_3(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\}$$
$$\mathcal{H}_4 : h_4(\mathbf{x}; \theta_0) = \mathbf{1}\{x_1^2 + x_2^2 \geqslant \theta_0\}$$

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Razmotrite sve kombinacije modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$. **Za koju kombinaciju modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$ postoji samo jedna hipoteza $h \in \mathcal{H}$ za koju $E(h|\mathcal{D}) = 0$?**

A $h_4 + \boldsymbol{\phi}_1$ B $h_2 + \boldsymbol{\phi}_2$ C $h_3 + \boldsymbol{\phi}_2$ D $h_1 + \boldsymbol{\phi}_1$

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

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((2,-3),-5), ((2,3),10), ((-3,5),5), ((5,0),20)\}$$

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 2$. Dobili smo težine $\mathbf{w} = (-0.050, 2.965, 2.482)$. **Koliko iznosi L_1-regularizirana pogreška $E(\mathbf{w}|\mathcal{D})$?**

A 191.95   B 230.98   C 31.70   D 165.89

3 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\boldsymbol{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$. Treniranjem modela na skupu $(\boldsymbol{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (-0.81, 2.32, 0.66, 0.1)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x}; \mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu, \sigma^2) = (\sqrt{2\pi}\sigma)^{-1} \exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\boldsymbol{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln \mathcal{L}(\mathbf{w}|\boldsymbol{\Phi}, \mathbf{y})$?

A -12.63 B -10.64 C -5.69 D -4.73

4 (T) Postupak najmanjih kvadrata (OLS) temelji se na izračunu pseudoinverza $\mathbf{X}^+$ matrice dizajna $\mathbf{X}$, što je poopćenje običnog inverza $\mathbf{X}^{-1}$. **U kojoj situaciji je rješenje dobiveno pseudoinverzom identično rješenju dobivenom običnim inverzom?**

A Kada je broj značajki manji od broja primjera i nema multikolinearnosti

B Kada je broj primjera veći od broja značajki

C Kada nema multikolinearnosti i matrica dizajna je dobro kondicionirana

D Kada je broj primjera jednak broju značajki plus jedan i nema multikolinearnosti

Cjelina 2: Linearni klasifikacijski modeli *(3 pitanja)*

5 (T) Jedna od triju komponenta svakog algoritma strojnog učenja je funkcija gubitka. Razmotrite funkcije gubitka perceptrona, logističke regresije (LR) i SVM-a. **Što je specifično funkciji gubitka perceptrona u odnosu na funkcije gubitka LR-a i SVM-a?**

A Gubitak netočno klasificiranih primjera raste linearno s udaljenošću od granice

B Svaki primjer nanosi gubitak, ali manji za točno klasificirane primjere nego za netočno klasificirane primjere

C Gubitak za sve točno klasificirane primjere je nula, a za netočno klasificirane može biti manji od 1

D Točno klasificirani primjer nanosi gubitak manji od 1 te gubitak opada što je primjer bliže granici

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 12.02 B 8.00 C 6.00 D 4.02

7 (P) Na skupu od $N = 1000$ primjera sa $n = 555$ značajki rješavamo problem višeklasne klasifikacije. Imamo $K = 4$ klase, s po 400, 300, 200 i 100 primjera. Za klasifikaciju želimo koristiti binarnu logističku regresiju u shemi OVO ili u shemi OVR (ovo nije tipično, ali je moguće). Pretpostavite da ne koristimo nikakvu regularizaciju, $\lambda = 0$. Razmotrite, za obje sheme, za koliko binarnih modela će rješenje optimizacijskog postupka sigurno biti nestabilno zbog loše kondicije matrice dizajna. **Koliko modela će sigurno biti više nestabilno u shemi OVO nego u shemi OVR?**

A 2 B 3 C 5 D 4

Cjelina 3: SVM, jezgrene i neparametarske metode *(4 pitanja)*

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

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

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 veća od tvrde margine?**

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

9 (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A 0 B 1/2 C 1/4 D 3/4

10 (T) Stroj potpornih vektora (SVM) jedna je vrsta rijetkoga jezgrenog stroja. Jezgreni stroj za bazne funkcije koriste jezgrene funkcije izračunate u odnosu na odabrane prototipne primjere. **Po čemu je SVM specifičan u odnosu na općeniti algoritam rijetkoga jezgrenog stroja?**

A Dimenzija prostora značajki ne može biti veća od broja primjera

B Zbog L1-regularizacije, mnoge težine modela bit će pritegnute na nulu

C Prototipni primjeri odabiru se u okviru optimizacijskog postupka

D Broj parametara modela jednak je dimenziji prostora značajki

11 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 3$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 2)^3$. Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((9, 30, 21), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-11, -26, -15), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((-1, -7, -6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 2.214 \cdot 10^{-8}$, $\alpha_2 = 3.803 \cdot 10^{-8}$ i $\alpha_3 = 6.017 \cdot 10^{-8}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3, 0, -3)$.**

A $+0.947$ B -0.676 C -2.330 D $+1.434$

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

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

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

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

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

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

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

14 (T) Procjenitelj MAP kombinira funkciju izglednosti parametara s apriornom distribucijom parametara. **Što i kako maksimizira procjenitelj MAP?**

A Zajedničku gustoću vjerojatnosti parametara i podataka, u zatvorenoj formi, ako je apriorna distribucija konjugatna za izglednost, inače iterativno

B Zajedničku vjerojatnost parametara i podataka, iterativno ako je apriorna distribucija iz eksponencijalne familije, inače u zatvorenoj formi

C Aposteriornu vjerojatnost podataka, u zatvorenoj formi ako je zajednička vjerojatnost parametara i podataka iz eksponencijalne familije, inače iterativno

D Aposteriornu gustoću vjerojatnosti podataka, u zatvorenoj formi ako su izglednost i apriorna distribucija konjugatne, inače iterativno

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

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

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

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

Cjelina 5: Probabilistički grafički modeli *(3 pitanja)*

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

17 (T) Približno zaključivanje kod Bayesovih mreža može se provesti metodama uzorkovanja. Te metode generiraju slučajan uzorak primjera iz distribucije opisane Bayesovom mrežom, iz kojega se onda mogu procijeniti parametri distribucije. Međutim, da bismo mogli uzorkovati iz Bayesove mreže, već trebamo znati parametre svih uvjetnih distribucija u čvorovima. **Zašto radimo uzorkovanje iz Bayesove mreže da bismo procijenili parametre distribucije, kad te parametre već imamo?**

A Ne trebamo znati parametre svih čvorova, već samo čvorova roditelja onih distribucija za koje želimo procijeniti parametre

B Uzorkovanje koristimo da bismo procijenili parametre bilo koje distribucije, a ne samo uvjetnih distribucija u čvorovima mreže

C Uzorkovanjem procjenjujemo parametre latentnih varijabli, čije vrijednosti ne opažamo, pa nam ti parametri nisu poznati prije uzorkovanja

D Vrijednosti parametara svih distribucija su procjene koje inicijaliziramo i zatim iterativno ažuriramo u svakom koraku uzorkovanja

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

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

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

A 944 B 596 C 472 D 348

Cjelina 6: Grupiranje i vrednovanje modela *(4 pitanja)*

19 (P) Krivuljom ROC vrednujemo binarne klasifikatore. Neka $h_1 \succ h_2$ označava da naučeni model h_1 striktno dominira nad naučenim modelom h_2 prema krivulji ROC, tj. da vrijedi

$$\forall\theta.\Big(\big(\mathrm{FPR}_\theta(h_1) = \mathrm{FPR}_\theta(h_2)\big) \Rightarrow \big(\mathrm{TPR}_\theta(h_1) > \mathrm{TPR}_\theta(h_2)\big)\Big)$$

gdje su $\mathrm{FPR}_\theta(h)$ i $\mathrm{TPR}_\theta(h)$ vrijednosti stope lažnih pozitiva odnosno stvarnih pozitiva hipoteze h s pragom θ. Nadalje, neka $h_1 > h_2$ označava da vrijedi $\mathrm{AUC}(h_1) > \mathrm{AUC}(h_2)$. **Što od sljedećeg vrijedi?**

A $h_1 > h_2 \Rightarrow h_2 \succ h_1$ B $h_1 > h_2 \Rightarrow h_1 \succ h_2$ C $h_1 \succ h_2 \Rightarrow h_1 > h_2$ D $h_1 \succ h_2 \Rightarrow h_2 > h_1$

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

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

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

A 1.8 B 2.4 C 1.9 D 2.0

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

$$\begin{array}{c|ccc} & y=1 & y=2 & y=3 \\ \hline y=1 & 5 & 13 & 1 \\ y=2 & 6 & 15 & 4 \\ y=3 & 14 & 2 & 6 \end{array}$$

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

A 0.040 B 0.094 C 0.015 D 0.068

22 (T) Za meko grupiranje možemo koristiti model miješane gustoće s latentnim varijablama. Kod tog modela, svaki primjer $\mathbf{x}^{(i)}$ ima pridruženu latentnu varijablu $\mathbf{z}^{(i)}$. Obje ove varijable zapravo su slučajne varijable sa svojom pretpostavljenom distribucijom. **Koju distribuciju pretpostavljamo za latentnu varijablu $\mathbf{z}^{(i)}$ i zašto?**

A Beta-distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada svakoj grupi

B Gaussovu distribuciju koja opisuje vjerojatnost da je grupa generirala primjer $\mathbf{x}^{(i)}$

C Kategoričku distribuciju koja opisuje kojoj grupi primjer $\mathbf{x}^{(i)}$ zapravo pripada

D Bernoullijevu distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada grupi $\mathbf{z}^{(i)}$

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

– NEKORIGIRANA VERZIJA –

*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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (T) Postupak najmanjih kvadrata (OLS) temelji se na izračunu pseudoinverza $\mathbf{X}^+$ matrice dizajna $\mathbf{X}$, što je poopćenje običnog inverza $\mathbf{X}^{-1}$. **U kojoj situaciji je rješenje dobiveno pseudoinverzom identično rješenju dobivenom običnim inverzom?**

- A Kada nema multikolinearnosti i matrica dizajna je dobro kondicionirana
- B Kada je broj primjera jednak broju značajki plus jedan i nema multikolinearnosti
- C Kada je broj primjera veći od broja značajki
- D Kada je broj značajki manji od broja primjera i nema multikolinearnosti

2 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\boldsymbol{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranjem modela na skupu $(\boldsymbol{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (-0.81, 2.32, 0.66, 0.1)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x};\mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu,\sigma^2) = (\sqrt{2\pi}\sigma)^{-1}\exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\boldsymbol{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln\mathcal{L}(\mathbf{w}|\boldsymbol{\Phi},\mathbf{y})$?

A -10.64 B -12.63 C -5.69 D -4.73

3 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{((0,0),1),((-1,-1),0),((1,1),0)\}$. Razmatramo sljedeće modele $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi} : \mathbb{Z}^2 \to \mathbb{Z}^2$, kojom primjere iz $\mathcal{D}$ preslikavamo u matricu dizajna $\boldsymbol{\Phi}$:

$$\mathcal{H}_1 : h_1(\mathbf{x};\theta_0,\theta_1) = \mathbf{1}\{\theta_1 x_1 + \theta_0 \geqslant 0\} \qquad \boldsymbol{\phi}_1(\mathbf{x}) = (1, x_2, x_1)$$
$$\mathcal{H}_2 : h_2(\mathbf{x};\theta_0,\theta_2) = \mathbf{1}\{\theta_2 x_2 + \theta_0 \geqslant 0\} \qquad \boldsymbol{\phi}_2(\mathbf{x}) = (1, x_1, x_1x_2)$$
$$\mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\}$$
$$\mathcal{H}_4 : h_4(\mathbf{x};\theta_0) = \mathbf{1}\{x_1^2 + x_2^2 \geqslant \theta_0\}$$

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Razmotrite sve kombinacije modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$. **Za koju kombinaciju modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$ postoji samo jedna hipoteza $h \in \mathcal{H}$ za koju** $E(h|\mathcal{D}) = 0$**?**

A $h_3 + \boldsymbol{\phi}_2$ B $h_1 + \boldsymbol{\phi}_1$ C $h_4 + \boldsymbol{\phi}_1$ D $h_2 + \boldsymbol{\phi}_2$

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

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((2,-3),-5),((2,3),-10),((-3,5),-5),((5,0),20)\}$$

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 2$. Dobili smo težine $\mathbf{w} = (-3.977, 2.477, 0.210)$. **Koliko iznosi L_1-regularizirana pogreška $E(\mathbf{w}|\mathcal{D})$?**

A 230.98 B 165.89 C 191.95 D 31.70

Cjelina 2: 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 12.02 B 8.00 C 6.00 D 4.02

6 (T) Jedna od triju komponenta svakog algoritma strojnog učenja je funkcija gubitka. Razmotrite funkcije gubitka perceptrona, logističke regresije (LR) i SVM-a. **Što je specifično funkciji gubitka perceptrona u odnosu na funkcije gubitka LR-a i SVM-a?**

A Gubitak netočno klasificiranih primjera raste linearno s udaljenošću od granice

B Svaki primjer nanosi gubitak, ali manji za točno klasificirane primjere nego za netočno klasificirane primjere

C Točno klasificirani primjer nanosi gubitak manji od 1 te gubitak opada što je primjer bliže granici

D Gubitak za sve točno klasificirane primjere je nula, a za netočno klasificirane može biti manji od 1

7 (P) Na skupu od $N = 1000$ primjera sa $n = 555$ značajki rješavamo problem višeklasne klasifikacije. Imamo $K = 4$ klase, s po 400, 300, 200 i 100 primjera. Za klasifikaciju želimo koristiti binarnu logističku regresiju u shemi OVO ili u shemi OVR (ovo nije tipično, ali je moguće). Pretpostavite da ne koristimo nikakvu regularizaciju, $\lambda = 0$. Razmotrite, za obje sheme, za koliko binarnih modela će rješenje optimizacijskog postupka sigurno biti nestabilno zbog loše kondicije matrice dizajna. **Koliko modela će sigurno biti više nestabilno u shemi OVO nego u shemi OVR?**

A 3 B 2 C 5 D 4

Cjelina 3: SVM, jezgrene i neparametarske metode *(4 pitanja)*

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

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

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 veća 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

9 (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A 3/4 B 0 C 1/4 D 1/2

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

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((6, 12, 6), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-11, -26, -15), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((-1, -7, -6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 3.496 \cdot 10^{-6}$, $\alpha_2 = 4.136 \cdot 10^{-7}$ i $\alpha_3 = 3.910 \cdot 10^{-6}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (1, 1, 25)$.**

A $+1.434$ B -0.676 C -2.330 D $+0.947$

11 (T) Stroj potpornih vektora (SVM) jedna je vrsta rijetkoga jezgrenog stroja. Jezgreni stroj za bazne funkcije koriste jezgrene funkcije izračunate u odnosu na odabrane prototipne primjere. **Po čemu je SVM specifičan u odnosu na općeniti algoritam rijetkoga jezgrenog stroja?**

A Dimenzija prostora značajki ne može biti veća od broja primjera

B Broj parametara modela jednak je dimenziji prostora značajki

C Zbog L1-regularizacije, mnoge težine modela bit će pritegnute na nulu

D Prototipni primjeri odabiru se u okviru optimizacijskog postupka

Cjelina 4: Procjena parametara i Bayesov klasifikator *(4 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 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 $[-4, -4 + a] \cup [5, 5 + b]$ D $[-4 - a, -4 + b]$

13 (T) Procjenitelj MAP kombinira funkciju izglednosti parametara s apriornom distribucijom parametara. **Što i kako maksimizira procjenitelj MAP?**

A Aposteriornu vjerojatnost podataka, u zatvorenoj formi ako je zajednička vjerojatnost parametara i podataka iz eksponencijalne familije, inače iterativno

B Zajedničku vjerojatnost parametara i podataka, iterativno ako je apriorna distribucija iz eksponencijalne familije, inače u zatvorenoj formi

C Zajedničku gustoću vjerojatnosti parametara i podataka, u zatvorenoj formi, ako je apriorna distribucija konjugatna za izglednost, inače iterativno

D Aposteriornu gustoću vjerojatnosti podataka, u zatvorenoj formi ako su izglednost i apriorna distribucija konjugatne, inače iterativno

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

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

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

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

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

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

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

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

Cjelina 5: Probabilistički grafički modeli *(3 pitanja)*

16 (T) Približno zaključivanje kod Bayesovih mreža može se provesti metodama uzorkovanja. Te metode generiraju slučajan uzorak primjera iz distribucije opisane Bayesovom mrežom, iz kojega se onda mogu procijeniti parametri distribucije. Međutim, da bismo mogli uzorkovati iz Bayesove mreže, već trebamo znati parametre svih uvjetnih distribucija u čvorovima. **Zašto radimo uzorkovanje iz Bayesove mreže da bismo procijenili parametre distribucije, kad te parametre već imamo?**

A Uzorkovanje koristimo da bismo procijenili parametre bilo koje distribucije, a ne samo uvjetnih distribucija u čvorovima mreže

B Uzorkovanjem procjenjujemo parametre latentnih varijabli, čije vrijednosti ne opažamo, pa nam ti parametri nisu poznati prije uzorkovanja

C Ne trebamo znati parametre svih čvorova, već samo čvorova roditelja onih distribucija za koje želimo procijeniti parametre

D Vrijednosti parametara svih distribucija su procjene koje inicijaliziramo i zatim iterativno ažuriramo u svakom koraku uzorkovanja

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

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

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

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

A 472 B 944 C 596 D 348

Cjelina 6: Grupiranje i vrednovanje modela *(4 pitanja)*

19 (T) Za meko grupiranje možemo koristiti model miješane gustoće s latentnim varijablama. Kod tog modela, svaki primjer $\mathbf{x}^{(i)}$ ima pridruženu latentnu varijablu $\mathbf{z}^{(i)}$. Obje ove varijable zapravo su slučajne varijable sa svojom pretpostavljenom distribucijom. **Koju distribuciju pretpostavljamo za latentnu varijablu $\mathbf{z}^{(i)}$ i zašto?**

A Bernoullijevu distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada grupi $\mathbf{z}^{(i)}$

B Beta-distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada svakoj grupi

C Kategoričku distribuciju koja opisuje kojoj grupi primjer $\mathbf{x}^{(i)}$ zapravo pripada

D Gaussovu distribuciju koja opisuje vjerojatnost da je grupa generirala primjer $\mathbf{x}^{(i)}$

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

$$\begin{array}{c|ccc} & y=1 & y=2 & y=3 \\ \hline y=1 & 5 & 13 & 1 \\ y=2 & 6 & 15 & 4 \\ y=3 & 14 & 2 & 36 \end{array}$$

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

A 0.040 B 0.068 C 0.094 D 0.015

21 (P) Krivuljom ROC vrednujemo binarne klasifikatore. Neka $h_1 \succ h_2$ označava da naučeni model h_1 striktno dominira nad naučenim modelom h_2 prema krivulji ROC, tj. da vrijedi

$$\forall\theta.\Big(\big(\mathrm{FPR}_\theta(h_1) = \mathrm{FPR}_\theta(h_2)\big) \Rightarrow \big(\mathrm{TPR}_\theta(h_1) > \mathrm{TPR}_\theta(h_2)\big)\Big)$$

gdje su $\mathrm{FPR}_\theta(h)$ i $\mathrm{TPR}_\theta(h)$ vrijednosti stope lažnih pozitiva odnosno stvarnih pozitiva hipoteze h s pragom θ. Nadalje, neka $h_1 > h_2$ označava da vrijedi $\mathrm{AUC}(h_1) > \mathrm{AUC}(h_2)$. **Što od sljedećeg vrijedi?**

A $h_1 > h_2 \Rightarrow h_2 \succ h_1$ B $h_1 > h_2 \Rightarrow h_1 \succ h_2$ C $h_1 \succ h_2 \Rightarrow h_2 > h_1$ D $h_1 \succ h_2 \Rightarrow h_1 > h_2$

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

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

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

A 2.4 B 2.0 C 1.8 D 1.9

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

– NEKORIGIRANA VERZIJA –

*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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija *(4 pitanja)*

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

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((2,-3),-5),((2,3),-10),((-3,5),5),((5,0),20)\}$$

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 2$. Dobili smo težine $\mathbf{w} = (-0.156, 1.330, 0.529)$. **Koliko iznosi L_1-regularizirana pogreška $E(\mathbf{w}|\mathcal{D})$?**

A 191.95 B 31.70 C 165.89 D 230.98

2 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\mathbf{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranjem modela na skupu $(\mathbf{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (-0.81, 2.32, 0.66, 0.1)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x};\mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu,\sigma^2) = (\sqrt{2\pi}\sigma)^{-1}\exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\mathbf{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln\mathcal{L}(\mathbf{w}|\mathbf{\Phi},\mathbf{y})$?

A -12.63 B -4.73 C -5.69 D -10.64

3 (T) Postupak najmanjih kvadrata (OLS) temelji se na izračunu pseudoinverza $\mathbf{X}^+$ matrice dizajna $\mathbf{X}$, što je poopćenje običnog inverza $\mathbf{X}^{-1}$. **U kojoj situaciji je rješenje dobiveno pseudoinverzom identično rješenju dobivenom običnim inverzom?**

A Kada je broj primjera jednak broju značajki plus jedan i nema multikolinearnosti

B Kada nema multikolinearnosti i matrica dizajna je dobro kondicionirana

C Kada je broj primjera veći od broja značajki

D Kada je broj značajki manji od broja primjera i nema multikolinearnosti

4 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{((0,0),1),((-1,-1),0),((1,1),0)\}$. Razmatramo sljedeće modele $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi} : \mathbb{Z}^2 \to \mathbb{Z}^2$, kojom primjere iz $\mathcal{D}$ preslikavamo u matricu dizajna $\mathbf{\Phi}$:

$$\mathcal{H}_1 : h_1(\mathbf{x};\theta_0,\theta_1) = \mathbf{1}\{\theta_1 x_1 + \theta_0 \geqslant 0\} \qquad \boldsymbol{\phi}_1(\mathbf{x}) = (1, x_2, x_1)$$
$$\mathcal{H}_2 : h_2(\mathbf{x};\theta_0,\theta_2) = \mathbf{1}\{\theta_2 x_2 + \theta_0 \geqslant 0\} \qquad \boldsymbol{\phi}_2(\mathbf{x}) = (1, x_1, x_1x_2)$$
$$\mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\}$$
$$\mathcal{H}_4 : h_4(\mathbf{x};\theta_0) = \mathbf{1}\{x_1^2 + x_2^2 \geqslant \theta_0\}$$

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Razmotrite sve kombinacije modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$. **Za koju kombinaciju modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$ postoji samo jedna hipoteza $h \in \mathcal{H}$ za koju** $E(h|\mathcal{D}) = 0$**?**

A $h_1 + \boldsymbol{\phi}_1$ B $h_4 + \boldsymbol{\phi}_1$ C $h_2 + \boldsymbol{\phi}_2$ D $h_3 + \boldsymbol{\phi}_2$

Cjelina 2: 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 8.00 C 12.02 D 4.02

6 (P) Na skupu od $N = 1000$ primjera sa $n = 444$ značajki rješavamo problem višeklasne klasifikacije. Imamo $K = 4$ klase, s po 400, 300, 200 i 100 primjera. Za klasifikaciju želimo koristiti binarnu logističku regresiju u shemi OVO ili u shemi OVR (ovo nije tipično, ali je moguće). Pretpostavite da ne koristimo nikakvu regularizaciju, $\lambda = 0$. Razmotrite, za obje sheme, za koliko binarnih modela će rješenje optimizacijskog postupka sigurno biti nestabilno zbog loše kondicije matrice dizajna. **Koliko modela će sigurno biti više nestabilno u shemi OVO nego u shemi OVR?**

A 3 B 4 C 2 D 5

7 (T) Jedna od triju komponenta svakog algoritma strojnog učenja je funkcija gubitka. Razmotrite funkcije gubitka perceptrona, logističke regresije (LR) i SVM-a. **Što je specifično funkciji gubitka perceptrona u odnosu na funkcije gubitka LR-a i SVM-a?**

A Svaki primjer nanosi gubitak, ali manji za točno klasificirane primjere nego za netočno klasificirane primjere

B Točno klasificirani primjer nanosi gubitak manji od 1 te gubitak opada što je primjer bliže granici

C Gubitak za sve točno klasificirane primjere je nula, a za netočno klasificirane može biti manji od 1

D Gubitak netočno klasificiranih primjera raste linearno s udaljenošću od granice

Cjelina 3: SVM, jezgrene i neparametarske metode *(4 pitanja)*

8 (T) Stroj potpornih vektora (SVM) jedna je vrsta rijetkoga jezgrenog stroja. Jezgreni stroj za bazne funkcije koriste jezgrene funkcije izračunate u odnosu na odabrane prototipne primjere. **Po čemu je SVM specifičan u odnosu na općeniti algoritam rijetkoga jezgrenog stroja?**

A Dimenzija prostora značajki ne može biti veća od broja primjera

B Prototipni primjeri odabiru se u okviru optimizacijskog postupka

C Broj parametara modela jednak je dimenziji prostora značajki

D Zbog L1-regularizacije, mnoge težine modela bit će pritegnute na nulu

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

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((6, 12, 6), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-11, -26, -15), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((-1, -7, -6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 3.496 \cdot 10^{-6}$, $\alpha_2 = 4.136 \cdot 10^{-7}$ i $\alpha_3 = 3.910 \cdot 10^{-6}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3, 0, -3)$.**

A -0.676 B $+1.434$ C $+0.947$ D -2.330

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

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

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 veća od tvrde margine?**

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

11 (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

[A] 1/2 [B] 0 [C] 1/4 [D] 3/4

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

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

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

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

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

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

14 (T) Procjenitelj MAP kombinira funkciju izglednosti parametara s apriornom distribucijom parametara. **Što i kako maksimizira procjenitelj MAP?**

[A] Zajedničku vjerojatnost parametara i podataka, iterativno ako je apriorna distribucija iz eksponencijalne familije, inače u zatvorenoj formi

[B] Aposteriornu vjerojatnost podataka, u zatvorenoj formi ako je zajednička vjerojatnost parametara i podataka iz eksponencijalne familije, inače iterativno

[C] Zajedničku gustoću vjerojatnosti parametara i podataka, u zatvorenoj formi, ako je apriorna distribucija konjugatna za izglednost, inače iterativno

[D] Aposteriornu gustoću vjerojatnosti podataka, u zatvorenoj formi ako su izglednost i apriorna distribucija konjugatne, inače iterativno

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

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

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

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

Cjelina 5: Probabilistički grafički modeli *(3 pitanja)*

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

17 (T) Približno zaključivanje kod Bayesovih mreža može se provesti metodama uzorkovanja. Te metode generiraju slučajan uzorak primjera iz distribucije opisane Bayesovom mrežom, iz kojega se onda mogu procijeniti parametri distribucije. Međutim, da bismo mogli uzorkovati iz Bayesove mreže, već trebamo znati parametre svih uvjetnih distribucija u čvorovima. **Zašto radimo uzorkovanje iz Bayesove mreže da bismo procijenili parametre distribucije, kad te parametre već imamo?**

A Vrijednosti parametara svih distribucija su procjene koje inicijaliziramo i zatim iterativno ažuriramo u svakom koraku uzorkovanja

B Uzorkovanje koristimo da bismo procijenili parametre bilo koje distribucije, a ne samo uvjetnih distribucija u čvorovima mreže

C Uzorkovanjem procjenjujemo parametre latentnih varijabli, čije vrijednosti ne opažamo, pa nam ti parametri nisu poznati prije uzorkovanja

D Ne trebamo znati parametre svih čvorova, već samo čvorova roditelja onih distribucija za koje želimo procijeniti parametre

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

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

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

A 596 B 472 C 944 D 348

Cjelina 6: Grupiranje i vrednovanje modela *(4 pitanja)*

19 (P) Krivuljom ROC vrednujemo binarne klasifikatore. Neka $h_1 \succ h_2$ označava da naučeni model h_1 striktno dominira nad naučenim modelom h_2 prema krivulji ROC, tj. da vrijedi

$$\forall \theta . \Big(\big(\mathrm{FPR}_\theta(h_1) = \mathrm{FPR}_\theta(h_2) \big) \Rightarrow \big(\mathrm{TPR}_\theta(h_1) > \mathrm{TPR}_\theta(h_2) \big) \Big)$$

gdje su $\mathrm{FPR}_\theta(h)$ i $\mathrm{TPR}_\theta(h)$ vrijednosti stope lažnih pozitiva odnosno stvarnih pozitiva hipoteze h s pragom θ. Nadalje, neka $h_1 > h_2$ označava da vrijedi $\mathrm{AUC}(h_1) > \mathrm{AUC}(h_2)$. **Što od sljedećeg vrijedi?**

[A] $h_1 \succ h_2 \Rightarrow h_1 > h_2$ [B] $h_1 > h_2 \Rightarrow h_2 \succ h_1$ [C] $h_1 > h_2 \Rightarrow h_1 \succ h_2$ [D] $h_1 \succ h_2 \Rightarrow h_2 > h_1$

20 (T) Za meko grupiranje možemo koristiti model miješane gustoće s latentnim varijablama. Kod tog modela, svaki primjer $\mathbf{x}^{(i)}$ ima pridruženu latentnu varijablu $\mathbf{z}^{(i)}$. Obje ove varijable zapravo su slučajne varijable sa svojom pretpostavljenom distribucijom. **Koju distribuciju pretpostavljamo za latentnu varijablu $\mathbf{z}^{(i)}$ i zašto?**

[A] Bernoullijevu distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada grupi $\mathbf{z}^{(i)}$

[B] Gaussovu distribuciju koja opisuje vjerojatnost da je grupa generirala primjer $\mathbf{x}^{(i)}$

[C] Beta-distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada svakoj grupi

[D] Kategoričku distribuciju koja opisuje kojoj grupi primjer $\mathbf{x}^{(i)}$ zapravo pripada

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

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 5 & 13 & 1 \\ 6 & 15 & 4 \\ 14 & 2 & 26 \end{pmatrix} \end{array}$$

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

[A] 0.068 [B] 0.015 [C] 0.094 [D] 0.040

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

$$\begin{pmatrix} 1.0 & 0.2 & 0.8 & 0.1 & 0.4 \\ 0.2 & 1.0 & 0.9 & 0.3 & 0.7 \\ 0.8 & 0.9 & 1.0 & 0.6 & 0.5 \\ 0.1 & 0.3 & 0.6 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

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

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

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

– NEKORIGIRANA VERZIJA –

*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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija *(4 pitanja)*

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

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((2,-3),-5),((2,3),-10),((-3,5),-5),((5,0),20)\}$$

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 2$. Dobili smo težine $\mathbf{w} = (-3.977, 2.477, 0.210)$. **Koliko iznosi L_1-regularizirana pogreška $E(\mathbf{w}|\mathcal{D})$?**

A 31.70 B 230.98 C 165.89 D 191.95

2 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \big\{((0,0),1),((-1,-1),0),((1,1),0)\big\}$. Razmatramo sljedeće modele $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi} : \mathbb{Z}^2 \to \mathbb{Z}^2$, kojom primjere iz $\mathcal{D}$ preslikavamo u matricu dizajna $\boldsymbol{\Phi}$:

$$\begin{aligned}
&\mathcal{H}_1 : h_1(\mathbf{x};\theta_0,\theta_1) = \mathbf{1}\{\theta_1 x_1 + \theta_0 \geqslant 0\} && \boldsymbol{\phi}_1(\mathbf{x}) = (1, x_2, x_1)\\
&\mathcal{H}_2 : h_2(\mathbf{x};\theta_0,\theta_2) = \mathbf{1}\{\theta_2 x_2 + \theta_0 \geqslant 0\} && \boldsymbol{\phi}_2(\mathbf{x}) = (1, x_1, x_1 x_2)\\
&\mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\} &&\\
&\mathcal{H}_4 : h_4(\mathbf{x};\theta_0) = \mathbf{1}\{x_1^2 + x_2^2 \geqslant \theta_0\} &&
\end{aligned}$$

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Razmotrite sve kombinacije modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$. **Za koju kombinaciju modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$ postoji samo jedna hipoteza $h \in \mathcal{H}$ za koju $E(h|\mathcal{D}) = 0$?**

A $h_4 + \boldsymbol{\phi}_1$ B $h_3 + \boldsymbol{\phi}_2$ C $h_2 + \boldsymbol{\phi}_2$ D $h_1 + \boldsymbol{\phi}_1$

3 (T) Postupak najmanjih kvadrata (OLS) temelji se na izračunu pseudoinverza $\mathbf{X}^+$ matrice dizajna $\mathbf{X}$, što je poopćenje običnog inverza $\mathbf{X}^{-1}$. **U kojoj situaciji je rješenje dobiveno pseudoinverzom identično rješenju dobivenom običnim inverzom?**

A Kada je broj značajki manji od broja primjera i nema multikolinearnosti

B Kada je broj primjera veći od broja značajki

C Kada nema multikolinearnosti i matrica dizajna je dobro kondicionirana

D Kada je broj primjera jednak broju značajki plus jedan i nema multikolinearnosti

4 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\boldsymbol{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$. Treniranjem modela na skupu $(\boldsymbol{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (2.2, -0.66, 0.77, -0.24)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x};\mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu,\sigma^2) = (\sqrt{2\pi}\sigma)^{-1}\exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\boldsymbol{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost $\ln \mathcal{L}(\mathbf{w}|\boldsymbol{\Phi},\mathbf{y})$?**

A -4.73 B -5.69 C -10.64 D -12.63

Cjelina 2: Linearni klasifikacijski modeli *(3 pitanja)*

5 (P) Na skupu od $N = 1000$ primjera sa $n = 555$ značajki rješavamo problem višeklasne klasifikacije. Imamo $K = 4$ klase, s po 400, 300, 200 i 100 primjera. Za klasifikaciju želimo koristiti binarnu logističku regresiju u shemi OVO ili u shemi OVR (ovo nije tipično, ali je moguće). Pretpostavite da ne koristimo nikakvu regularizaciju, $\lambda = 0$. Razmotrite, za obje sheme, za koliko binarnih modela će rješenje optimizacijskog postupka sigurno biti nestabilno zbog loše kondicije matrice dizajna. **Koliko modela će sigurno biti više nestabilno u shemi OVO nego u shemi OVR?**

A 2 B 4 C 5 D 3

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 $\mathbf{y} = (0, 1, 0)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi dotični primjer?**

A 12.02 B 4.02 C 8.00 D 6.00

7 (T) Jedna od triju komponenta svakog algoritma strojnog učenja je funkcija gubitka. Razmotrite funkcije gubitka perceptrona, logističke regresije (LR) i SVM-a. **Što je specifično funkciji gubitka perceptrona u odnosu na funkcije gubitka LR-a i SVM-a?**

A Gubitak za sve točno klasificirane primjere je nula, a za netočno klasificirane može biti manji od 1

B Svaki primjer nanosi gubitak, ali manji za točno klasificirane primjere nego za netočno klasificirane primjere

C Gubitak netočno klasificiranih primjera raste linearno s udaljenošću od granice

D Točno klasificirani primjer nanosi gubitak manji od 1 te gubitak opada što je primjer bliže granici

Cjelina 3: SVM, jezgrene i neparametarske metode *(4 pitanja)*

8 (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A 3/4 B 0 C 1/2 D 1/4

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

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

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 veća od tvrde margine?**

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

10 (T) Stroj potpornih vektora (SVM) jedna je vrsta rijetkoga jezgrenog stroja. Jezgreni stroj za bazne funkcije koriste jezgrene funkcije izračunate u odnosu na odabrane prototipne primjere. **Po čemu je SVM specifičan u odnosu na općeniti algoritam rijetkoga jezgrenog stroja?**

- A Dimenzija prostora značajki ne može biti veća od broja primjera
- B Zbog L1-regularizacije, mnoge težine modela bit će pritegnute na nulu
- C Broj parametara modela jednak je dimenziji prostora značajki
- D Prototipni primjeri odabiru se u okviru optimizacijskog postupka

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

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((6, 12, 6), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-11, -26, -15), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((-1, -7, -6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 3.496 \cdot 10^{-6}$, $\alpha_2 = 4.136 \cdot 10^{-7}$ i $\alpha_3 = 3.910 \cdot 10^{-6}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (1, 1, 25)$.**

A -2.330 B -0.676 C $+1.434$ D $+0.947$

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

12 (T) Procjenitelj MAP kombinira funkciju izglednosti parametara s apriornom distribucijom parametara. **Što i kako maksimizira procjenitelj MAP?**

- A Aposteriornu gustoću vjerojatnosti podataka, u zatvorenoj formi ako su izglednost i apriorna distribucija konjugatne, inače iterativno
- B Zajedničku gustoću vjerojatnosti parametara i podataka, u zatvorenoj formi, ako je apriorna distribucija konjugatna za izglednost, inače iterativno
- C Aposteriornu vjerojatnost podataka, u zatvorenoj formi ako je zajednička vjerojatnost parametara i podataka iz eksponencijalne familije, inače iterativno
- D Zajedničku vjerojatnost parametara i podataka, iterativno ako je apriorna distribucija iz eksponencijalne familije, inače u zatvorenoj formi

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

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

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

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

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

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

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

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

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

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

Cjelina 5: Probabilistički grafički modeli *(3 pitanja)*

16 (T) Približno zaključivanje kod Bayesovih mreža može se provesti metodama uzorkovanja. Te metode generiraju slučajan uzorak primjera iz distribucije opisane Bayesovom mrežom, iz kojega se onda mogu procijeniti parametri distribucije. Međutim, da bismo mogli uzorkovati iz Bayesove mreže, već trebamo znati parametre svih uvjetnih distribucija u čvorovima. **Zašto radimo uzorkovanje iz Bayesove mreže da bismo procijenili parametre distribucije, kad te parametre već imamo?**

A Vrijednosti parametara svih distribucija su procjene koje inicijaliziramo i zatim iterativno ažuriramo u svakom koraku uzorkovanja

B Uzorkovanje koristimo da bismo procijenili parametre bilo koje distribucije, a ne samo uvjetnih distribucija u čvorovima mreže

C Uzorkovanjem procjenjujemo parametre latentnih varijabli, čije vrijednosti ne opažamo, pa nam ti parametri nisu poznati prije uzorkovanja

D Ne trebamo znati parametre svih čvorova, već samo čvorova roditelja onih distribucija za koje želimo procijeniti parametre

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

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

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

A 472 B 348 C 596 D 944

18 (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 $w \perp y$, $z \perp x$, $x \perp w | \{z, y\}$ C $w \perp y$, $z \perp \{x, y\} | w$ D $x \perp y | z$, $z \perp w | y$

Cjelina 6: Grupiranje i vrednovanje modela *(4 pitanja)*

19 (T) Za meko grupiranje možemo koristiti model miješane gustoće s latentnim varijablama. Kod tog modela, svaki primjer $\mathbf{x}^{(i)}$ ima pridruženu latentnu varijablu $\mathbf{z}^{(i)}$. Obje ove varijable zapravo su slučajne varijable sa svojom pretpostavljenom distribucijom. **Koju distribuciju pretpostavljamo za latentnu varijablu $\mathbf{z}^{(i)}$ i zašto?**

- A Beta-distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada svakoj grupi
- B Bernoullijevu distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada grupi $\mathbf{z}^{(i)}$
- C Gaussovu distribuciju koja opisuje vjerojatnost da je grupa generirala primjer $\mathbf{x}^{(i)}$
- D Kategoričku distribuciju koja opisuje kojoj grupi primjer $\mathbf{x}^{(i)}$ zapravo pripada

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

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

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

A 1.9 B 2.0 C 1.8 D 2.4

21 (P) Krivuljom ROC vrednujemo binarne klasifikatore. Neka $h_1 \succ h_2$ označava da naučeni model h_1 striktno dominira nad naučenim modelom h_2 prema krivulji ROC, tj. da vrijedi

$$\forall\theta.\Big(\big(\mathrm{FPR}_\theta(h_1) = \mathrm{FPR}_\theta(h_2)\big) \Rightarrow \big(\mathrm{TPR}_\theta(h_1) > \mathrm{TPR}_\theta(h_2)\big)\Big)$$

gdje su $\mathrm{FPR}_\theta(h)$ i $\mathrm{TPR}_\theta(h)$ vrijednosti stope lažnih pozitiva odnosno stvarnih pozitiva hipoteze h s pragom θ. Nadalje, neka $h_1 > h_2$ označava da vrijedi $\mathrm{AUC}(h_1) > \mathrm{AUC}(h_2)$. **Što od sljedećeg vrijedi?**

A $h_1 > h_2 \Rightarrow h_2 \succ h_1$ B $h_1 > h_2 \Rightarrow h_1 \succ h_2$ C $h_1 \succ h_2 \Rightarrow h_2 > h_1$ D $h_1 \succ h_2 \Rightarrow h_1 > h_2$

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

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 5 & 13 & 1 \\ 6 & 15 & 4 \\ 14 & 2 & 46 \end{pmatrix} \end{array}$$

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

A 0.068 B 0.015 C 0.040 D 0.094

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

– NEKORIGIRANA VERZIJA –

*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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (T) Postupak najmanjih kvadrata (OLS) temelji se na izračunu pseudoinverza $\mathbf{X}^+$ matrice dizajna $\mathbf{X}$, što je poopćenje običnog inverza $\mathbf{X}^{-1}$. **U kojoj situaciji je rješenje dobiveno pseudoinverzom identično rješenju dobivenom običnim inverzom?**

A Kada je broj primjera jednak broju značajki plus jedan i nema multikolinearnosti

B Kada nema multikolinearnosti i matrica dizajna je dobro kondicionirana

C Kada je broj primjera veći od broja značajki

D Kada je broj značajki manji od broja primjera i nema multikolinearnosti

2 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\mathbf{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranjem modela na skupu $(\mathbf{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (-1.91, 0.22, -0.47, -0.06)^\mathrm{T}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x};\mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu,\sigma^2) = (\sqrt{2\pi}\sigma)^{-1}\exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\mathbf{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln\mathcal{L}(\mathbf{w}|\mathbf{\Phi},\mathbf{y})$?

A -4.73 B -5.69 C -10.64 D -12.63

3 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{((0,0),1),((-1,-1),0),((1,1),0)\}$. Razmatramo sljedeće modele $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}: \mathbb{Z}^2 \to \mathbb{Z}^2$, kojom primjere iz $\mathcal{D}$ preslikavamo u matricu dizajna $\mathbf{\Phi}$:

$$\mathcal{H}_1 : h_1(\mathbf{x};\theta_0,\theta_1) = \mathbf{1}\{\theta_1 x_1 + \theta_0 \geqslant 0\} \qquad \boldsymbol{\phi}_1(\mathbf{x}) = (1, x_2, x_1)$$
$$\mathcal{H}_2 : h_2(\mathbf{x};\theta_0,\theta_2) = \mathbf{1}\{\theta_2 x_2 + \theta_0 \geqslant 0\} \qquad \boldsymbol{\phi}_2(\mathbf{x}) = (1, x_1, x_1x_2)$$
$$\mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^\mathrm{T}\mathbf{x} \geqslant 0\}$$
$$\mathcal{H}_4 : h_4(\mathbf{x};\theta_0) = \mathbf{1}\{x_1^2 + x_2^2 \geqslant \theta_0\}$$

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Razmotrite sve kombinacije modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$. **Za koju kombinaciju modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$ postoji samo jedna hipoteza $h \in \mathcal{H}$ za koju** $E(h|\mathcal{D}) = 0$**?**

A $h_3 + \boldsymbol{\phi}_2$ B $h_2 + \boldsymbol{\phi}_2$ C $h_4 + \boldsymbol{\phi}_1$ D $h_1 + \boldsymbol{\phi}_1$

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

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((2,-3),-5),((2,3),-10),((-3,5),5),((5,0),20)\}$$

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 2$. Dobili smo težine $\mathbf{w} = (-0.156, 1.330, 0.529)$. **Koliko iznosi L_1-regularizirana pogreška** $E(\mathbf{w}|\mathcal{D})$**?**

A 31.70 B 191.95 C 230.98 D 165.89

Cjelina 2: 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 $\mathbf{y} = (0, 0, 1)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi dotični primjer?**

A 8.00 B 12.02 C 4.02 D 6.00

6 (P) Na skupu od $N = 1000$ primjera sa $n = 444$ značajki rješavamo problem višeklasne klasifikacije. Imamo $K = 4$ klase, s po 400, 300, 200 i 100 primjera. Za klasifikaciju želimo koristiti binarnu logističku regresiju u shemi OVO ili u shemi OVR (ovo nije tipično, ali je moguće). Pretpostavite da ne koristimo nikakvu regularizaciju, $\lambda = 0$. Razmotrite, za obje sheme, za koliko binarnih modela će rješenje optimizacijskog postupka sigurno biti nestabilno zbog loše kondicije matrice dizajna. **Koliko modela će sigurno biti više nestabilno u shemi OVO nego u shemi OVR?**

A 3 B 2 C 5 D 4

7 (T) Jedna od triju komponenta svakog algoritma strojnog učenja je funkcija gubitka. Razmotrite funkcije gubitka perceptrona, logističke regresije (LR) i SVM-a. **Što je specifično funkciji gubitka perceptrona u odnosu na funkcije gubitka LR-a i SVM-a?**

A Točno klasificirani primjer nanosi gubitak manji od 1 te gubitak opada što je primjer bliže granici

B Gubitak netočno klasificiranih primjera raste linearno s udaljenošću od granice

C Svaki primjer nanosi gubitak, ali manji za točno klasificirane primjere nego za netočno klasificirane primjere

D Gubitak za sve točno klasificirane primjere je nula, a za netočno klasificirane može biti manji od 1

Cjelina 3: SVM, jezgrene i neparametarske metode *(4 pitanja)*

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

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

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 veća od tvrde margine?**

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

9 (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A 3/4 B 0 C 1/2 D 1/4

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

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((6, 12, 6), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-11, -26, -15), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((-1, -7, -6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 3.496 \cdot 10^{-6}$, $\alpha_2 = 4.136 \cdot 10^{-7}$ i $\alpha_3 = 3.910 \cdot 10^{-6}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3, 0, -3)$.**

A -0.676 B $+1.434$ C -2.330 D $+0.947$

11 (T) Stroj potpornih vektora (SVM) jedna je vrsta rijetkoga jezgrenog stroja. Jezgreni stroj za bazne funkcije koriste jezgrene funkcije izračunate u odnosu na odabrane prototipne primjere. **Po čemu je SVM specifičan u odnosu na općeniti algoritam rijetkoga jezgrenog stroja?**

A Zbog L1-regularizacije, mnoge težine modela bit će pritegnute na nulu

B Prototipni primjeri odabiru se u okviru optimizacijskog postupka

C Dimenzija prostora značajki ne može biti veća od broja primjera

D Broj parametara modela jednak je dimenziji prostora značajki

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

12 (T) Procjenitelj MAP kombinira funkciju izglednosti parametara s apriornom distribucijom parametara. **Što i kako maksimizira procjenitelj MAP?**

A Zajedničku vjerojatnost parametara i podataka, iterativno ako je apriorna distribucija iz eksponencijalne familije, inače u zatvorenoj formi

B Aposteriornu gustoću vjerojatnosti podataka, u zatvorenoj formi ako su izglednost i apriorna distribucija konjugatne, inače iterativno

C Zajedničku gustoću vjerojatnosti parametara i podataka, u zatvorenoj formi, ako je apriorna distribucija konjugatna za izglednost, inače iterativno

D Aposteriornu vjerojatnost podataka, u zatvorenoj formi ako je zajednička vjerojatnost parametara i podataka iz eksponencijalne familije, inače iterativno

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 1 i da se apriorne vjerojatnosti klasa ujednače, $P(y = 1) = P(y = 2) = P(y = 3) = 1/3$. Skicirajte gustoće zajedničke vjerojatnosti naučenog i izmijenjenog modela. Neka su h_1 i h_2 MAP-hipoteze prvog i drugog modela, te neka su a i b pozitivne konstante. Razmotrite segment ulaznog prostora za koji vrijedi $-10 \leqslant x \leqslant 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

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

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

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

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

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

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

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

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

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

Cjelina 5: Probabilistički grafički modeli *(3 pitanja)*

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

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

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

A 472 B 348 C 596 D 944

17 (T) Približno zaključivanje kod Bayesovih mreža može se provesti metodama uzorkovanja. Te metode generiraju slučajan uzorak primjera iz distribucije opisane Bayesovom mrežom, iz kojega se onda mogu procijeniti parametri distribucije. Međutim, da bismo mogli uzorkovati iz Bayesove mreže, već trebamo znati parametre svih uvjetnih distribucija u čvorovima. **Zašto radimo uzorkovanje iz Bayesove mreže da bismo procijenili parametre distribucije, kad te parametre već imamo?**

A Uzorkovanjem procjenjujemo parametre latentnih varijabli, čije vrijednosti ne opažamo, pa nam ti parametri nisu poznati prije uzorkovanja

B Ne trebamo znati parametre svih čvorova, već samo čvorova roditelja onih distribucija za koje želimo procijeniti parametre

C Uzorkovanje koristimo da bismo procijenili parametre bilo koje distribucije, a ne samo uvjetnih distribucija u čvorovima mreže

D Vrijednosti parametara svih distribucija su procjene koje inicijaliziramo i zatim iterativno ažuriramo u svakom koraku uzorkovanja

18 (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 $y\perp w$, $y\perp x|\{w, z\}$ D $x\perp y|z$, $z\perp w|y$

Cjelina 6: Grupiranje i vrednovanje modela *(4 pitanja)*

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

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

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

A 2.0 B 1.9 C 1.8 D 2.4

20 (T) Za meko grupiranje možemo koristiti model miješane gustoće s latentnim varijablama. Kod tog modela, svaki primjer $\mathbf{x}^{(i)}$ ima pridruženu latentnu varijablu $\mathbf{z}^{(i)}$. Obje ove varijable zapravo su slučajne varijable sa svojom pretpostavljenom distribucijom. **Koju distribuciju pretpostavljamo za latentnu varijablu $\mathbf{z}^{(i)}$ i zašto?**

A Kategoričku distribuciju koja opisuje kojoj grupi primjer $\mathbf{x}^{(i)}$ zapravo pripada

B Beta-distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada svakoj grupi

C Gaussovu distribuciju koja opisuje vjerojatnost da je grupa generirala primjer $\mathbf{x}^{(i)}$

D Bernoullijevu distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada grupi $\mathbf{z}^{(i)}$

21 (P) Krivuljom ROC vrednujemo binarne klasifikatore. Neka $h_1 \succ h_2$ označava da naučeni model h_1 striktno dominira nad naučenim modelom h_2 prema krivulji ROC, tj. da vrijedi

$$\forall\theta.\Big(\big(\mathrm{FPR}_\theta(h_1) = \mathrm{FPR}_\theta(h_2)\big) \Rightarrow \big(\mathrm{TPR}_\theta(h_1) > \mathrm{TPR}_\theta(h_2)\big)\Big)$$

gdje su $\mathrm{FPR}_\theta(h)$ i $\mathrm{TPR}_\theta(h)$ vrijednosti stope lažnih pozitiva odnosno stvarnih pozitiva hipoteze h s pragom θ. Nadalje, neka $h_1 > h_2$ označava da vrijedi $\mathrm{AUC}(h_1) > \mathrm{AUC}(h_2)$. **Što od sljedećeg vrijedi?**

A $h_1 > h_2 \Rightarrow h_2 \succ h_1$ B $h_1 \succ h_2 \Rightarrow h_1 > h_2$ C $h_1 > h_2 \Rightarrow h_1 \succ h_2$ D $h_1 \succ h_2 \Rightarrow h_2 > h_1$

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

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array}\begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 5 & 13 & 1 \\ 6 & 15 & 4 \\ 14 & 2 & 46 \end{pmatrix} \end{array}$$

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

A 0.015 B 0.068 C 0.040 D 0.094

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

– NEKORIGIRANA VERZIJA –

*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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija *(4 pitanja)*

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

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((2,-3),5),((2,3),-10),((-3,5),5),((5,0),20)\}$$

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 2$. Dobili smo težine $\mathbf{w} = (4.568, 0.746, -0.550)$. **Koliko iznosi L_1-regularizirana pogreška $E(\mathbf{w}|\mathcal{D})$?**

[A] 165.89 [B] 191.95 [C] 31.70 [D] 230.98

2 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\mathbf{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranjem modela na skupu $(\mathbf{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (2.2, -0.66, 0.77, -0.24)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x};\mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu,\sigma^2) = (\sqrt{2\pi}\sigma)^{-1}\exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\mathbf{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln\mathcal{L}(\mathbf{w}|\mathbf{\Phi},\mathbf{y})$?

[A] -4.73 [B] -5.69 [C] -12.63 [D] -10.64

3 (T) Postupak najmanjih kvadrata (OLS) temelji se na izračunu pseudoinverza $\mathbf{X}^+$ matrice dizajna $\mathbf{X}$, što je poopćenje običnog inverza $\mathbf{X}^{-1}$. **U kojoj situaciji je rješenje dobiveno pseudoinverzom identično rješenju dobivenom običnim inverzom?**

[A] Kada je broj primjera veći od broja značajki

[B] Kada nema multikolinearnosti i matrica dizajna je dobro kondicionirana

[C] Kada je broj primjera jednak broju značajki plus jedan i nema multikolinearnosti

[D] Kada je broj značajki manji od broja primjera i nema multikolinearnosti

4 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{((0,0),1),((-1,-1),0),((1,1),0)\}$. Razmatramo sljedeće modele $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi} : \mathbb{Z}^2 \to \mathbb{Z}^2$, kojom primjere iz $\mathcal{D}$ preslikavamo u matricu dizajna $\mathbf{\Phi}$:

$$\mathcal{H}_1 : h_1(\mathbf{x};\theta_0,\theta_1) = \mathbf{1}\{\theta_1 x_1 + \theta_0 \geqslant 0\} \qquad \boldsymbol{\phi}_1(\mathbf{x}) = (1, x_2, x_1)$$
$$\mathcal{H}_2 : h_2(\mathbf{x};\theta_0,\theta_2) = \mathbf{1}\{\theta_2 x_2 + \theta_0 \geqslant 0\} \qquad \boldsymbol{\phi}_2(\mathbf{x}) = (1, x_1, x_1x_2)$$
$$\mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\}$$
$$\mathcal{H}_4 : h_4(\mathbf{x};\theta_0) = \mathbf{1}\{x_1^2 + x_2^2 \geqslant \theta_0\}$$

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Razmotrite sve kombinacije modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$. **Za koju kombinaciju modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$ postoji samo jedna hipoteza $h \in \mathcal{H}$ za koju** $E(h|\mathcal{D}) = 0$?

[A] $h_2 + \boldsymbol{\phi}_2$ [B] $h_4 + \boldsymbol{\phi}_1$ [C] $h_3 + \boldsymbol{\phi}_2$ [D] $h_1 + \boldsymbol{\phi}_1$

Cjelina 2: 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 8.00 B 4.02 C 12.02 D 6.00

6 (P) Na skupu od $N = 1000$ primjera sa $n = 555$ značajki rješavamo problem višeklasne klasifikacije. Imamo $K = 4$ klase, s po 400, 300, 200 i 100 primjera. Za klasifikaciju želimo koristiti binarnu logističku regresiju u shemi OVO ili u shemi OVR (ovo nije tipično, ali je moguće). Pretpostavite da ne koristimo nikakvu regularizaciju, $\lambda = 0$. Razmotrite, za obje sheme, za koliko binarnih modela će rješenje optimizacijskog postupka sigurno biti nestabilno zbog loše kondicije matrice dizajna. **Koliko modela će sigurno biti više nestabilno u shemi OVO nego u shemi OVR?**

A 3 B 4 C 2 D 5

7 (T) Jedna od triju komponenta svakog algoritma strojnog učenja je funkcija gubitka. Razmotrite funkcije gubitka perceptrona, logističke regresije (LR) i SVM-a. **Što je specifično funkciji gubitka perceptrona u odnosu na funkcije gubitka LR-a i SVM-a?**

A Svaki primjer nanosi gubitak, ali manji za točno klasificirane primjere nego za netočno klasificirane primjere

B Gubitak netočno klasificiranih primjera raste linearno s udaljenošću od granice

C Točno klasificirani primjer nanosi gubitak manji od 1 te gubitak opada što je primjer bliže granici

D Gubitak za sve točno klasificirane primjere je nula, a za netočno klasificirane može biti manji od 1

Cjelina 3: SVM, jezgrene i neparametarske metode *(4 pitanja)*

8 (T) Stroj potpornih vektora (SVM) jedna je vrsta rijetkoga jezgrenog stroja. Jezgreni stroj za bazne funkcije koriste jezgrene funkcije izračunate u odnosu na odabrane prototipne primjere. **Po čemu je SVM specifičan u odnosu na općeniti algoritam rijetkoga jezgrenog stroja?**

A Dimenzija prostora značajki ne može biti veća od broja primjera

B Zbog L1-regularizacije, mnoge težine modela bit će pritegnute na nulu

C Prototipni primjeri odabiru se u okviru optimizacijskog postupka

D Broj parametara modela jednak je dimenziji prostora značajki

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

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

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 veća od tvrde margine?**

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

10 (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A 3/4 B 1/2 C 1/4 D 0

11 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 3$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 2)^3$. Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((9, 30, 21), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-11, -26, -15), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((-1, -7, -6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 2.214 \cdot 10^{-8}$, $\alpha_2 = 3.803 \cdot 10^{-8}$ i $\alpha_3 = 6.017 \cdot 10^{-8}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (1, 1, 25)$.**

A -0.676 B $+0.947$ C $+1.434$ D -2.330

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

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

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

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

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

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

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

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

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

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

15 (T) Procjenitelj MAP kombinira funkciju izglednosti parametara s apriornom distribucijom parametara. **Što i kako maksimizira procjenitelj MAP?**

A Zajedničku vjerojatnost parametara i podataka, iterativno ako je apriorna distribucija iz eksponencijalne familije, inače u zatvorenoj formi

B Aposteriornu vjerojatnost podataka, u zatvorenoj formi ako je zajednička vjerojatnost parametara i podataka iz eksponencijalne familije, inače iterativno

C Aposteriornu gustoću vjerojatnosti podataka, u zatvorenoj formi ako su izglednost i apriorna distribucija konjugatne, inače iterativno

D Zajedničku gustoću vjerojatnosti parametara i podataka, u zatvorenoj formi, ako je apriorna distribucija konjugatna za izglednost, inače iterativno

Cjelina 5: Probabilistički grafički modeli *(3 pitanja)*

16 (T) Približno zaključivanje kod Bayesovih mreža može se provesti metodama uzorkovanja. Te metode generiraju slučajan uzorak primjera iz distribucije opisane Bayesovom mrežom, iz kojega se onda mogu procijeniti parametri distribucije. Međutim, da bismo mogli uzorkovati iz Bayesove mreže, već trebamo znati parametre svih uvjetnih distribucija u čvorovima. **Zašto radimo uzorkovanje iz Bayesove mreže da bismo procijenili parametre distribucije, kad te parametre već imamo?**

A Vrijednosti parametara svih distribucija su procjene koje inicijaliziramo i zatim iterativno ažuriramo u svakom koraku uzorkovanja

B Ne trebamo znati parametre svih čvorova, već samo čvorova roditelja onih distribucija za koje želimo procijeniti parametre

C Uzorkovanjem procjenjujemo parametre latentnih varijabli, čije vrijednosti ne opažamo, pa nam ti parametri nisu poznati prije uzorkovanja

D Uzorkovanje koristimo da bismo procijenili parametre bilo koje distribucije, a ne samo uvjetnih distribucija u čvorovima mreže

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

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

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

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

A 596 B 944 C 348 D 472

Cjelina 6: Grupiranje i vrednovanje modela *(4 pitanja)*

19 (T) Za meko grupiranje možemo koristiti model miješane gustoće s latentnim varijablama. Kod tog modela, svaki primjer $\mathbf{x}^{(i)}$ ima pridruženu latentnu varijablu $\mathbf{z}^{(i)}$. Obje ove varijable zapravo su slučajne varijable sa svojom pretpostavljenom distribucijom. **Koju distribuciju pretpostavljamo za latentnu varijablu $\mathbf{z}^{(i)}$ i zašto?**

A Bernoullijevu distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada grupi $\mathbf{z}^{(i)}$

B Beta-distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada svakoj grupi

C Gaussovu distribuciju koja opisuje vjerojatnost da je grupa generirala primjer $\mathbf{x}^{(i)}$

D Kategoričku distribuciju koja opisuje kojoj grupi primjer $\mathbf{x}^{(i)}$ zapravo pripada

20 (P) Krivuljom ROC vrednujemo binarne klasifikatore. Neka $h_1 \succ h_2$ označava da naučeni model h_1 striktno dominira nad naučenim modelom h_2 prema krivulji ROC, tj. da vrijedi

$$\forall\theta.\Big(\big(\mathrm{FPR}_\theta(h_1) = \mathrm{FPR}_\theta(h_2)\big) \Rightarrow \big(\mathrm{TPR}_\theta(h_1) > \mathrm{TPR}_\theta(h_2)\big)\Big)$$

gdje su $\mathrm{FPR}_\theta(h)$ i $\mathrm{TPR}_\theta(h)$ vrijednosti stope lažnih pozitiva odnosno stvarnih pozitiva hipoteze h s pragom θ. Nadalje, neka $h_1 > h_2$ označava da vrijedi $\mathrm{AUC}(h_1) > \mathrm{AUC}(h_2)$. **Što od sljedećeg vrijedi?**

A $h_1 \succ h_2 \Rightarrow h_2 > h_1$ B $h_1 > h_2 \Rightarrow h_1 \succ h_2$ C $h_1 \succ h_2 \Rightarrow h_1 > h_2$ D $h_1 > h_2 \Rightarrow h_2 \succ h_1$

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

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

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

A 1.8 B 1.9 C 2.4 D 2.0

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

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 5 & 13 & 1 \\ 6 & 15 & 4 \\ 14 & 2 & 36 \end{pmatrix} \end{array}$$

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

A 0.068 B 0.015 C 0.040 D 0.094

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

– NEKORIGIRANA VERZIJA –

*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 **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{((0,0),1),((-1,-1),0),((1,1),0)\}$. Razmatramo sljedeće modele $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi} : \mathbb{Z}^2 \to \mathbb{Z}^2$, kojom primjere iz $\mathcal{D}$ preslikavamo u matricu dizajna $\boldsymbol{\Phi}$:

$$\mathcal{H}_1 : h_1(\mathbf{x};\theta_0,\theta_1) = \mathbf{1}\{\theta_1 x_1 + \theta_0 \geqslant 0\} \qquad \boldsymbol{\phi}_1(\mathbf{x}) = (1, x_2, x_1)$$
$$\mathcal{H}_2 : h_2(\mathbf{x};\theta_0,\theta_2) = \mathbf{1}\{\theta_2 x_2 + \theta_0 \geqslant 0\} \qquad \boldsymbol{\phi}_2(\mathbf{x}) = (1, x_1, x_1 x_2)$$
$$\mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\}$$
$$\mathcal{H}_4 : h_4(\mathbf{x};\theta_0) = \mathbf{1}\{x_1^2 + x_2^2 \geqslant \theta_0\}$$

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Razmotrite sve kombinacije modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$. **Za koju kombinaciju modela $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi}$ postoji samo jedna hipoteza $h \in \mathcal{H}$ za koju $E(h|\mathcal{D}) = 0$?**

[A] $h_2 + \boldsymbol{\phi}_2$ [B] $h_1 + \boldsymbol{\phi}_1$ [C] $h_3 + \boldsymbol{\phi}_2$ [D] $h_4 + \boldsymbol{\phi}_1$

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

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((2,-3),-5),((2,3),-10),((-3,5),5),((5,0),20)\}$$

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 2$. Dobili smo težine $\mathbf{w} = (-0.156, 1.330, 0.529)$. **Koliko iznosi L_1-regularizirana pogreška $E(\mathbf{w}|\mathcal{D})$?**

[A] 31.70 [B] 230.98 [C] 191.95 [D] 165.89

3 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\boldsymbol{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$. Treniranjem modela na skupu $(\boldsymbol{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (-0.81, 2.32, 0.66, 0.1)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x};\mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu,\sigma^2) = (\sqrt{2\pi}\sigma)^{-1}\exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\boldsymbol{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln\mathcal{L}(\mathbf{w}|\boldsymbol{\Phi},\mathbf{y})$?

[A] -4.73 [B] -10.64 [C] -5.69 [D] -12.63

4 (T) Postupak najmanjih kvadrata (OLS) temelji se na izračunu pseudoinverza $\mathbf{X}^+$ matrice dizajna $\mathbf{X}$, što je poopćenje običnog inverza $\mathbf{X}^{-1}$. **U kojoj situaciji je rješenje dobiveno pseudoinverzom identično rješenju dobivenom običnim inverzom?**

[A] Kada je broj primjera veći od broja značajki

[B] Kada je broj značajki manji od broja primjera i nema multikolinearnosti

[C] Kada je broj primjera jednak broju značajki plus jedan i nema multikolinearnosti

[D] Kada nema multikolinearnosti i matrica dizajna je dobro kondicionirana

Cjelina 2: 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 12.02 B 6.00 C 8.00 D 4.02

6 (P) Na skupu od $N = 1000$ primjera sa $n = 555$ značajki rješavamo problem višeklasne klasifikacije. Imamo $K = 4$ klase, s po 400, 300, 200 i 100 primjera. Za klasifikaciju želimo koristiti binarnu logističku regresiju u shemi OVO ili u shemi OVR (ovo nije tipično, ali je moguće). Pretpostavite da ne koristimo nikakvu regularizaciju, $\lambda = 0$. Razmotrite, za obje sheme, za koliko binarnih modela će rješenje optimizacijskog postupka sigurno biti nestabilno zbog loše kondicije matrice dizajna. **Koliko modela će sigurno biti više nestabilno u shemi OVO nego u shemi OVR?**

A 5 B 3 C 2 D 4

7 (T) Jedna od triju komponenta svakog algoritma strojnog učenja je funkcija gubitka. Razmotrite funkcije gubitka perceptrona, logističke regresije (LR) i SVM-a. **Što je specifično funkciji gubitka perceptrona u odnosu na funkcije gubitka LR-a i SVM-a?**

A Točno klasificirani primjer nanosi gubitak manji od 1 te gubitak opada što je primjer bliže granici

B Svaki primjer nanosi gubitak, ali manji za točno klasificirane primjere nego za netočno klasificirane primjere

C Gubitak za sve točno klasificirane primjere je nula, a za netočno klasificirane može biti manji od 1

D Gubitak netočno klasificiranih primjera raste linearno s udaljenošću od granice

Cjelina 3: SVM, jezgrene i neparametarske metode *(4 pitanja)*

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

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

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 veća od tvrde margine?**

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

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

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((6, 12, 6), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-11, -26, -15), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((-1, -7, -6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 3.496 \cdot 10^{-6}$, $\alpha_2 = 4.136 \cdot 10^{-7}$ i $\alpha_3 = 3.910 \cdot 10^{-6}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (1, 1, 25)$.**

A $+1.434$ B $+0.947$ C -0.676 D -2.330

10 (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A 3/4 B 1/4 C 1/2 D 0

11 (T) Stroj potpornih vektora (SVM) jedna je vrsta rijetkoga jezgrenog stroja. Jezgreni stroj za bazne funkcije koriste jezgrene funkcije izračunate u odnosu na odabrane prototipne primjere. **Po čemu je SVM specifičan u odnosu na općeniti algoritam rijetkoga jezgrenog stroja?**

A Dimenzija prostora značajki ne može biti veća od broja primjera

B Prototipni primjeri odabiru se u okviru optimizacijskog postupka

C Broj parametara modela jednak je dimenziji prostora značajki

D Zbog L1-regularizacije, mnoge težine modela bit će pritegnute na nulu

Cjelina 4: Procjena parametara i Bayesov klasifikator *(4 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 $[-4-a, 5+b]$ B $[-4, -4-a] \cup [5, 5+b]$ C $[-4-a, -4+b]$ D $[-4, -4+a] \cup [5-b, 5]$

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

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

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

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

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

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

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

14 (T) Procjenitelj MAP kombinira funkciju izglednosti parametara s apriornom distribucijom parametara. **Što i kako maksimizira procjenitelj MAP?**

A Aposteriornu gustoću vjerojatnosti podataka, u zatvorenoj formi ako su izglednost i apriorna distribucija konjugatne, inače iterativno

B Aposteriornu vjerojatnost podataka, u zatvorenoj formi ako je zajednička vjerojatnost parametara i podataka iz eksponencijalne familije, inače iterativno

C Zajedničku gustoću vjerojatnosti parametara i podataka, u zatvorenoj formi, ako je apriorna distribucija konjugatna za izglednost, inače iterativno

D Zajedničku vjerojatnost parametara i podataka, iterativno ako je apriorna distribucija iz eksponencijalne familije, inače u zatvorenoj formi

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

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

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

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

Cjelina 5: Probabilistički grafički modeli *(3 pitanja)*

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

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

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

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

A 596 B 348 C 944 D 472

18 (T) Približno zaključivanje kod Bayesovih mreža može se provesti metodama uzorkovanja. Te metode generiraju slučajan uzorak primjera iz distribucije opisane Bayesovom mrežom, iz kojega se onda mogu procijeniti parametri distribucije. Međutim, da bismo mogli uzorkovati iz Bayesove mreže, već trebamo znati parametre svih uvjetnih distribucija u čvorovima. **Zašto radimo uzorkovanje iz Bayesove mreže da bismo procijenili parametre distribucije, kad te parametre već imamo?**

A Uzorkovanje koristimo da bismo procijenili parametre bilo koje distribucije, a ne samo uvjetnih distribucija u čvorovima mreže

B Ne trebamo znati parametre svih čvorova, već samo čvorova roditelja onih distribucija za koje želimo procijeniti parametre

C Uzorkovanjem procjenjujemo parametre latentnih varijabli, čije vrijednosti ne opažamo, pa nam ti parametri nisu poznati prije uzorkovanja

D Vrijednosti parametara svih distribucija su procjene koje inicijaliziramo i zatim iterativno ažuriramo u svakom koraku uzorkovanja

Cjelina 6: Grupiranje i vrednovanje modela *(4 pitanja)*

19 (P) Krivuljom ROC vrednujemo binarne klasifikatore. Neka $h_1 \succ h_2$ označava da naučeni model h_1 striktno dominira nad naučenim modelom h_2 prema krivulji ROC, tj. da vrijedi

$$\forall \theta. \Big(\big(\mathrm{FPR}_\theta(h_1) = \mathrm{FPR}_\theta(h_2)\big) \Rightarrow \big(\mathrm{TPR}_\theta(h_1) > \mathrm{TPR}_\theta(h_2)\big) \Big)$$

gdje su $\mathrm{FPR}_\theta(h)$ i $\mathrm{TPR}_\theta(h)$ vrijednosti stope lažnih pozitiva odnosno stvarnih pozitiva hipoteze h s pragom θ. Nadalje, neka $h_1 > h_2$ označava da vrijedi $\mathrm{AUC}(h_1) > \mathrm{AUC}(h_2)$. **Što od sljedećeg vrijedi?**

A $h_1 \succ h_2 \Rightarrow h_2 > h_1$ B $h_1 \succ h_2 \Rightarrow h_1 > h_2$ C $h_1 > h_2 \Rightarrow h_2 \succ h_1$ D $h_1 > h_2 \Rightarrow h_1 \succ h_2$

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

$$\begin{pmatrix} 1.0 & 0.2 & 0.8 & 0.1 & 0.4 \\ 0.2 & 1.0 & 0.9 & 0.3 & 0.7 \\ 0.8 & 0.9 & 1.0 & 0.6 & 0.5 \\ 0.1 & 0.3 & 0.6 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

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

A 2.0 B 1.9 C 2.4 D 1.8

21 (T) Za meko grupiranje možemo koristiti model miješane gustoće s latentnim varijablama. Kod tog modela, svaki primjer $\mathbf{x}^{(i)}$ ima pridruženu latentnu varijablu $\mathbf{z}^{(i)}$. Obje ove varijable zapravo su slučajne varijable sa svojom pretpostavljenom distribucijom. **Koju distribuciju pretpostavljamo za latentnu varijablu $\mathbf{z}^{(i)}$ i zašto?**

A Kategoričku distribuciju koja opisuje kojoj grupi primjer $\mathbf{x}^{(i)}$ zapravo pripada

B Gaussovu distribuciju koja opisuje vjerojatnost da je grupa generirala primjer $\mathbf{x}^{(i)}$

C Beta-distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada svakoj grupi

D Bernoullijevu distribuciju koja opisuje s kojom vjerojatnošću primjer $\mathbf{x}^{(i)}$ pripada grupi $\mathbf{z}^{(i)}$

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

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 5 & 13 & 1 \\ 6 & 15 & 4 \\ 14 & 2 & 36 \end{pmatrix} \end{array}$$

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

A 0.040 B 0.015 C 0.068 D 0.094

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