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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija (6 pitanja)

1 (P) Koristimo regresiju za predviđanje uspjeha studija na temelju prosjeka ocjena u četiri razreda srednje škole te uspjeha iz matematike, fizike i informatike na državnoj maturi (ukupno 7 značajki). Za preslikavanje u prostor značajki koristimo polinom drugog stupnja s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki. Pretpostavite da nema multikolinearnosti. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 22 B 56 C 36 D 28

2 (P) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka su $\mathcal{H}_1$ i $\mathcal{H}_2$ dva modela različite složenosti te neka su $h_1^* \in \mathcal{H}_1$ i $h_2^* \in \mathcal{H}_2$ dvije hipoteze koje minimiziraju pogrešku $E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h)$ razlika između pogreške na ispitnom skupu i pogreške na skupu za učenje za hipotezu h, tj. $\Delta(h) = E(h|\mathcal{D}_{\text{test}}) - E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h_2^*) < \Delta(h_1^*)$. **Što možemo zaključiti o modelima $\mathcal{H}_1$ i $\mathcal{H}_2$?**

A Ako $E(h_1^*|\mathcal{D}_{\text{test}}) > E(h_2^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_1$ prenaučen

B Ako $E(h_2^*|\mathcal{D}_{\text{train}}) < E(h_1^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_2$ podnaučen

C Ako $E(h_2^*|\mathcal{D}_{\text{test}}) > E(h_1^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_2$ prenaučen

D Ako $E(h_1^*|\mathcal{D}_{\text{train}}) < E(h_2^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_1$ podnaučen

3 (T) Algoritam linearne regresije koristi kvadratni gubitak. **Koje je probabilističko opravdanje za korištenje kvadratnog gubitka?**

A Vjerojatnost parametara za fiksni označeni skup podataka raste kvadratno s očekivanjem gubitka

B Primjeri za koje je predviđanje hipoteze pogrešno pokoravaju se normalnoj razdiobi sa središtem u nuli

C Minimizacija kvadratnog gubitka istovjetna je maksimizaciji negativnog logaritma vjerojatnosti parametara

D Vjerojatnost skupa primjera najveća je za parametre koji minimiziraju očekivanje kvadratnog gubitka

4 (N) U dvodimenzijskome ulaznom prostoru postupkom najmanjih kvadrata (OLS) treniramo L2-regularizirani model linearne regresije. Koristimo funkciju preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_2^2)$. Označen skup podataka je:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{cc|c} 1 & 5 & 2 \\ 2 & -2 & -5 \\ 3 & 1 & 3 \\ -5 & 6 & -3 \end{array}\right) \qquad \mathbf{G}^+ = \begin{pmatrix} 0.722 & -0.035 & 0.005 & -0.018 & -0.018 \\ -0.035 & 0.01 & 0 & 0.001 & 0.001 \\ 0.005 & 0 & 0.009 & 0 & -0.002 \\ -0.018 & 0.001 & 0 & 0.003 & -0.001 \\ -0.018 & 0.001 & -0.002 & -0.001 & 0.002 \end{pmatrix}$$

gdje je $\mathbf{G}^+$ pseudoinverz odgovarajuće Gramove matrice u prostoru značajki modificirane za regularizaciju uz $\lambda = 100$. **Koliko iznosi L_2-regularizirana pogreška naučenog modela?** (Napomena: matrica $\mathbf{G}^+$ dana je sa zaokruženim vrijednostima i koristite tu matricu.)

A 4.38 B 13.74 C 2.92 D 19.37

5 (T) Premda je linearna regresija linearan model, ona može modelirati nelinearne hipoteze. Nelinearnost možemo ostvariti uporabom prikladne funkcije preslikavanja $\boldsymbol{\phi}$ koja primjere iz ulaznog prostora preslikava u prostor značajki. **Koja pretpostavka stoji iza takvog postupka?**

- A Primjeri koji se u prostoru značajki nalaze blizu nelinearne hiperpovršine u ulaznome prostoru leže na hiperravnini
- B Primjeri koji su linearno odvojivi u ulaznome prostoru postaju linearno neodvojivi u prostoru značajki dovoljno visoke dimenzije
- C Primjeri koji u ulaznome prostoru leže na nelinearnoj hiperpovršini u prostoru značajki leže na hiperravnini
- D Hiperravnina koja minimizira pogrešku u ulaznome prostoru odgovara nelinearnoj hiperpovršini u prostoru značajki

6 (P) Razmatramo binarni klasifikacijski problem sa $\mathcal{Y} = \{0, 1\}$ u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((0,0),1),((2,2),1),((0,1),0),((0,-1),0),((1,1),0)\big\}$. Razmatramo sljedeće modele, parametrizirane sa $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$:

$$\mathcal{H}_1 : h_1(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geq 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = h_1(\mathbf{x}; \boldsymbol{\theta}_1) \cdot h_2(\mathbf{x}; \boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{1}\{(x_1 - \theta_1)^2 + (x_2 - \theta_2)^2 \geq \theta_0^2\} \qquad \mathcal{H}_4 : h_4(\mathbf{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \min\big(1, h_1(\mathbf{x}; \boldsymbol{\theta}_1) + h_2(\mathbf{x}; \boldsymbol{\theta}_2)\big)$$

Neka je E_α minimalna empirijska pogreška modela $\mathcal{H}_\alpha$ na skupu $\mathcal{D}$, tj. $E_\alpha = \mathrm{argmin}_{h \in \mathcal{H}_\alpha} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

A $E_2 > E_1 = E_3 > E_4$ B $E_1 > E_4 > E_2 = E_3$ C $E_1 = E_2 = E_4 > E_3$ D $E_1 = E_3 > E_2 = E_4$

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

7 (P) Algoritmom perceptrona rješavamo klasifikacijski problem sa $K = 10$ klasa. Koristimo pet značajki, $n = 5$. Za preslikavanje $\boldsymbol{\phi}$ koristimo polinom drugog stupnja sa svim interakcijskim parovima značajki. Koristimo shemu jedan-naspram-jedan (OVO). **Koliko iznosi ukupan broj parametara takvog višeklasnog klasifikatora?**

A 495 B 210 C 945 D 900

8 (N) Na skupu označenih primjera $\mathcal{D}$ trenirali smo model logističke regresije. Dobili smo neki vektor težina $\mathbf{w}$ i pomak $w_0 = 0.15$. Tako naučenom modelu neki primjer $\mathbf{x}$, čija je oznaka u skupu primjera $y = 0$, nanosi gubitak unakrsne entropije od $L(0, h(\mathbf{x})) = 0.274$. **Koliki gubitak unakrsne entropije bi nanosio primjer x kada bismo njegove značajke pomnožili sa dva i promijenili mu oznaku**?

A 4.03 B 2.54 C 7.11 D 1.19

9 (N) Model regularizirane logističke regresije treniramo stohastičkim gradijentnim spustom. Koristimo faktor regularizacije $\lambda = 1000$ i stopu učenja $\eta = 0.01$. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.2, 0.5, -1.1, 2.7)$. **Koliko u toj iteraciji iznosi promjena težine w_1 za primjer $(\mathbf{x}, y) = ((-1, 2), 1)$?**

A -12 B -2 C $+22$ D -5

10 (T) Jedan od nedostataka algoritma perceptrona jest taj što se za linearno neodovojive probleme algoritam ne zaustavlja. **Zašto je tomu tako?**

- A Sve točke minimuma funkcije pogreške u prostoru parametara leže na istoj ravnini
- B Gradijent funkcije gubitka uvijek je različit od nul-vektora za barem jedan označeni primjer
- C Površina funkcije pogreške u prostoru parametara po dijelovima je linearna i stoga nederivabilna
- D Svaki minimum funkcije pogreške u prostoru parametara veći je od nule

11 (T) Multinomijalna logistička regresija primjere klasificira u jednu od K klasa. Neka je $\mathbf{y}^{(i)} = (y_1^i, \ldots, y_k^i, \ldots, y_K^i)$ vektor indikatorskih varijabli koji odgovara oznaci primjera $\mathbf{x}^{(i)}$. Neka je $h_k(\mathbf{x}; \mathbf{W})$ vrijednost funkcije *softmax* za klasu k. Optimizacija parametara $\mathbf{W}$ odgovara maksimizaciji logaritma vjerojatnost oznaka. **Čemu je jednaka ta vjerojatnost?**

A $\sum_{i=1}^N \sum_{k=1}^K \ln y_k^i h_k(\mathbf{x}^{(i)}; \mathbf{W})$ C $\sum_{i=1}^N \ln \prod_{k=1}^K h_k(\mathbf{x}^{(i)}; \mathbf{W})^{y_k^i}$

B $\prod_{i=1}^N \sum_{k=1}^K \ln h_k(\mathbf{x}^{(i)}; \mathbf{W})^{y_k^i}$ D $\ln \prod_{i=1}^N \prod_{k=1}^K y_k^i h_k(\mathbf{x}^{(i)}; \mathbf{W})$

12 (N) Raspolažemo sljedećim skupom za učenje u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijski prostor značajki, definiranu kao $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$. Početne težine perceptrona su $\mathbf{w} = (1,0,-1,2,-3,0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 16   B 6   C 32   D 22

13 (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 6.00   C 4.02   D 8.00

14 (P) Algoritmom perceptrona rješavamo klasifikacijski problem. Oznake su iz skupa $\{-1,+1\}$. Za stopu učenja koristimo $\eta = \frac{1}{2}$. U jednoj iteraciji algoritma, kod ažuriranja težina na temelju primjera $\mathbf{x}$, vektor težina $\mathbf{w}$ mijenja se tako da mu se duljina prepolavlja, ali se zadržavaju smjer i orijentacija. Razmotite da, umjesto perceptrona, za klasifikaciju koristimo linearnu regresiju treniranu na oznakama $\{-1,+1\}$. **Koliko bi iznosio gubitak za isti primjer x kada bismo koristili linearnu regresiju s težinama w?**

A $1 - y\|\mathbf{w}\|^2$ B $1 - \|\mathbf{x}\|^2$ C $(y - \|\mathbf{x}\|^2)^2$ D $(\|\mathbf{w}\|^2 + 1)^2$

Cjelina 3: SVM, jezgrene i neparametarske metode (8 pitanja)

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

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

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

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

16 (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 1/2   D 1/4

17 (P) Razmatramo troklasni ($K = 3$) klasifikacijski problem sa $N = 100$ primjera u ulaznom prostoru dimenzije $n = 5$ (ne računajući značajku x_0). Razmatramo dva modela. Model $\mathcal{H}_1$ je rijetki jezgreni stroj koji za bazne funkcije koristi Gaussove jezgre. Model $\mathcal{H}_2$ je multinomijalna regresija s adaptivnim baznim funkcijama. Adaptivnih baznih funkcija ima ukupno deset, uključujući baznu funkciju $\phi_0(\mathbf{x}) = 1$. Za oba modela izračunajte njihov ukupan broj parametara prije učenja. **Koliko iznosi razlika u broju parametara između ova dva modela?**

A 719 B 219 C 725 D 819

18 (T) Razmotrite problem maksimalne margine kao standardni problem optimizacije kriterijske funkcije $f(\mathbf{x})$ uz ograničenja nejednakosti $g_i(\mathbf{x}) \leq 0$ i jednakosti $h_i(\mathbf{x}) = 0$. Razmotrite taj problem za slučaj tvrde margine i za slučaj meke margine, oba u dualu. **Po čemu se dualni optimizacijski problem meke margine razlikuje od dualnog optimizacijskog problema tvrde margine?**

A Ima dodatnih n ograničenja jednakosti

B Ima dodatnih N optimizacijskih varijabli i dodatnih $2N$ ograničenja nejednakosti

C Ima dodatnih n optimizacijskih varijabli i dodatnih N ograničenja jednakosti

D Ima dodatnih N ograničenja nejednakosti

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

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

Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(2, 0)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina veća od stare?**

A $\frac{5}{4}\sqrt{2}$ puta B $\frac{1}{2}\sqrt{17}$ puta C $\frac{1}{2}\sqrt{10}$ puta D $\frac{1}{2}\sqrt{13}$ puta

20 (N) Za klasifikacijski problem koristimo stroj potpornih vektora s linearnom jezgrom i $C = 2$. Koristimo optimizaciju u dualu. Označen skup podataka i vektor dualnih koeficijenata jesu sljedeći:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} 25 & 70 & 45 & +1 \\ 6 & 12 & 6 & +1 \\ 11 & 32 & 21 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right) \qquad \boldsymbol{\alpha} = \begin{pmatrix} 0 \\ 2 \\ 0.077 \\ 1.923 \end{pmatrix}$$

Izračunajte parametre primarnog modela. **Koliko je primjer $\mathbf{x}^{(1)}$ udaljen od margine?**

A 0.44 B 5.25 C 4.09 D 3.23

21 (N) Rješavamo problem određivanja podrijetla pojedinih riječi u jeziku: za svaku riječ trebamo odrediti je li engleskog ($y = 1$) ili francuskog ($y = 0$) podrijetla. Problem rješavamo logističkom regresijom izvedenom kao rijetki jezgreni stroj, gdje za bazne funkcije koristimo jezgru κ nad znakovnim nizovima. Funkcija κ definirana je kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Na primjer, $\kappa(\text{water}, \text{eau}) = 2/6 = 0.33$. Skup za učenje je sljedeći:

$$\mathcal{D} = \{(\mathbf{x}, y)\}_i = \big\{(\text{water}, 1), (\text{eau}, 0), (\text{dog}, 1), (\text{chien}, 0), (\text{paperclip}, 1), (\text{trombone}), 0), (\text{chance}, 1), (\text{hasard}, 0)\big\}$$

Treniranjem rijetkoga jezgrenog stroja dobili smo vektor težina $\mathbf{w} = (-0.5, 0, 0, -3.5, 0, -1, 0, 0, 1)$. Razmotrite primjer $(\mathbf{x}, y) = (\text{nounours}, 0)$. **Koliko iznosi gubitak modela na primjeru** $(\mathbf{x}, y)$**?**

A 0.552 B 0.795 C 0.359 D 0.456

22 (T) Koristimo algoritam SVM s jezgrenom funkcijom κ. Optimizaciju provodimo u dualu. Ako je κ Mercerova jezgra, onda takva jezgra implicitno definira preslikavanje $\boldsymbol{\phi}$. Međutim, u nekim slučajevima trebamo eksplicitno izraziti funkciju $\boldsymbol{\phi}$ na temelju korištene jezgre κ. **Kada je potrebno eksplicitno izraziti $\boldsymbol{\phi}$?**

A Za izračun funkcije gubitka

B Za izračun dualnih koeficijenata

C Za izračun težine w_0

D Za izračun širine margine

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija (6 pitanja)

1 (P) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka su $\mathcal{H}_1$ i $\mathcal{H}_2$ dva modela različite složenosti te neka su $h_1^* \in \mathcal{H}_1$ i $h_2^* \in \mathcal{H}_2$ dvije hipoteze koje minimiziraju pogrešku $E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h)$ razlika između pogreške na ispitnom skupu i pogreške na skupu za učenje za hipotezu h, tj. $\Delta(h) = E(h|\mathcal{D}_{\text{test}}) - E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h_2^*) > \Delta(h_1^*)$. **Što možemo zaključiti o modelima $\mathcal{H}_1$ i $\mathcal{H}_2$?**

- A Ako $E(h_2^*|\mathcal{D}_{\text{test}}) > E(h_1^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_2$ prenaučen
- B Ako $E(h_1^*|\mathcal{D}_{\text{test}}) > E(h_2^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_1$ prenaučen
- C Ako $E(h_2^*|\mathcal{D}_{\text{train}}) < E(h_1^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_2$ podnaučen
- D Ako $E(h_1^*|\mathcal{D}_{\text{train}}) < E(h_2^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_1$ podnaučen

2 (T) Algoritam linearne regresije koristi kvadratni gubitak. **Koje je probabilističko opravdanje za korištenje kvadratnog gubitka?**

- A Primjeri za koje je predviđanje hipoteze pogrešno pokoravaju se normalnoj razdiobi sa središtem u nuli
- B Vjerojatnost parametara za fiksni označeni skup podataka raste kvadratno s očekivanjem gubitka
- C Minimizacija kvadratnog gubitka istovjetna je maksimizaciji negativnog logaritma vjerojatnosti parametara
- D Vjerojatnost skupa primjera najveća je za parametre koji minimiziraju očekivanje kvadratnog gubitka

3 (T) Premda je linearna regresija linearan model, ona može modelirati nelinearne hipoteze. Nelinearnost možemo ostvariti uporabom prikladne funkcije preslikavanja $\boldsymbol{\phi}$ koja primjere iz ulaznog prostora preslikava u prostor značajki. **Koja pretpostavka stoji iza takvog postupka?**

- A Hiperravnina koja minimizira pogrešku u ulaznome prostoru odgovara nelinearnoj hiperpovršini u prostoru značajki
- B Primjeri koji u ulaznome prostoru leže na nelinearnoj hiperpovršini u prostoru značajki leže na hiperravnini
- C Primjeri koji se u prostoru značajki nalaze blizu nelinearne hiperpovršine u ulaznome prostoru leže na hiperavnini
- D Primjeri koji su linearno odvojivi u ulaznome prostoru postaju linearno neodvojivi u prostoru značajki dovoljno visoke dimenzije

4 (N) U dvodimenzijskome ulaznom prostoru postupkom najmanjih kvadrata (OLS) treniramo L2-regularizirani model linearne regresije. Koristimo funkciju preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_2^2)$. Označen skup podataka je:

$$
\mathbf{X}|\mathbf{y} = \left(\begin{array}{rr|r} 1 & 5 & 1 \\ 2 & -2 & 5 \\ 3 & 1 & 2 \\ -5 & 6 & -2 \end{array}\right) \qquad
\mathbf{G}^+ = \begin{pmatrix} 0.722 & -0.035 & 0.005 & -0.018 & -0.018 \\ -0.035 & 0.01 & 0 & 0.001 & 0.001 \\ 0.005 & 0 & 0.009 & 0 & -0.002 \\ -0.018 & 0.001 & 0 & 0.003 & -0.001 \\ -0.018 & 0.001 & -0.002 & -0.001 & 0.002 \end{pmatrix}
$$

gdje je $\mathbf{G}^+$ pseudoinverz odgovarajuće Gramove matrice u prostoru značajki modificirane za regularizaciju uz $\lambda = 100$. **Koliko iznosi L_2-regularizirana pogreška naučenog modela?** (Napomena: matrica $\mathbf{G}^+$ dana je sa zaokruženim vrijednostima i koristite tu matricu.)

A 2.92 B 19.37 C 13.74 D 4.38

5 (P) Koristimo regresiju za predviđanje uspjeha studija na temelju prosjeka ocjena u četiri razreda srednje škole te uspjeha iz matematike, fizike i informatike na državnoj maturi (ukupno 7 značajki). Za preslikavanje u prostor značajki koristimo polinom drugog stupnja s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki. Pretpostavite da nema multikolinearnosti. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 22 B 36 C 56 D 28

6 (P) Razmatramo binarni klasifikacijski problem sa $\mathcal{Y} = \{0, 1\}$ u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((0,0),1),((2,2),1),((0,1),0),((0,-1),0),((1,1),0)\big\}$. Razmatramo sljedeće modele, parametrizirane sa $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$:

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geq 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = h_1(\mathbf{x};\boldsymbol{\theta}_1)\cdot h_2(\mathbf{x};\boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \geq \theta_0^2\big\} \qquad \mathcal{H}_4 : h_4(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \min\big(1, h_1(\mathbf{x};\boldsymbol{\theta}_1) + h_2(\mathbf{x};\boldsymbol{\theta}_2)\big)$$

Neka je E_α minimalna empirijska pogreška modela $\mathcal{H}_\alpha$ na skupu $\mathcal{D}$, tj. $E_\alpha = \mathrm{argmin}_{h\in\mathcal{H}_\alpha} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

A $E_2 > E_1 = E_3 > E_4$ B $E_1 = E_3 > E_2 = E_4$ C $E_1 = E_2 = E_4 > E_3$ D $E_1 > E_4 > E_2 = E_3$

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

7 (N) Model regularizirane logističke regresije treniramo stohastičkim gradijentnim spustom. Koristimo faktor regularizacije $\lambda = 1000$ i stopu učenja $\eta = 0.01$. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.2, 0.5, -1.1, 2.7)$. **Koliko u toj iteraciji iznosi promjena težine w_1 za primjer $(\mathbf{x}, y) = ((-1, 2), 1)$?**

A +22 B −5 C −2 D −12

8 (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 8.00 B 6.00 C 12.02 D 4.02

9 (T) Multinomijalna logistička regresija primjere klasificira u jednu od K klasa. Neka je $\mathbf{y}^{(i)} = (y_1^i, \ldots, y_k^i, \ldots, y_K^i)$ vektor indikatorskih varijabli koji odgovara oznaci primjera $\mathbf{x}^{(i)}$. Neka je $h_k(\mathbf{x};\mathbf{W})$ vrijednost funkcije *softmax* za klasu k. Optimizacija parametara $\mathbf{W}$ odgovara maksimizaciji logaritma vjerojatnost oznaka. **Čemu je jednaka ta vjerojatnost?**

A $\sum_{i=1}^N \ln \prod_{k=1}^K h_k(\mathbf{x}^{(i)};\mathbf{W})^{y_k^i}$ C $\sum_{i=1}^N \sum_{k=1}^K \ln y_k^i h_k(\mathbf{x}^{(i)};\mathbf{W})$

B $\prod_{i=1}^N \sum_{k=1}^K \ln h_k(\mathbf{x}^{(i)};\mathbf{W})^{y_k^i}$ D $\ln \prod_{i=1}^N \prod_{k=1}^K y_k^i h_k(\mathbf{x}^{(i)};\mathbf{W})$

10 (P) Algoritmom perceptrona rješavamo klasifikacijski problem. Oznake su iz skupa $\{-1, +1\}$. Za stopu učenja koristimo $\eta = \frac{1}{2}$. U jednoj iteraciji algoritma, kod ažuriranja težina na temelju primjera $\mathbf{x}$, vektor težina $\mathbf{w}$ mijenja se tako da mu se duljina prepolavlja, ali se zadržavaju smjer i orijentacija. Razmotite da, umjesto perceptrona, za klasifikaciju koristimo linearnu regresiju treniranu na oznakama $\{-1, +1\}$. **Koliko bi iznosio gubitak za isti primjer x kada bismo koristili linearnu regresiju s težinama w?**

A $(\|\mathbf{w}\|^2 + 1)^2$ B $(y - \|\mathbf{x}\|^2)^2$ C $1 - y\|\mathbf{w}\|^2$ D $1 - \|\mathbf{x}\|^2$

11 (P) Algoritmom perceptrona rješavamo klasifikacijski problem sa $K = 10$ klasa. Koristimo pet značajki, $n = 5$. Za preslikavanje $\boldsymbol{\phi}$ koristimo polinom drugog stupnja sa svim interakcijskim parovima značajki. Koristimo shemu jedan-naspram-jedan (OVO). **Koliko iznosi ukupan broj parametara takvog višeklasnog klasifikatora?**

A 900 B 945 C 495 D 210

12 (T) Jedan od nedostataka algoritma perceptrona jest taj što se za linearno neodovojive probleme algoritam ne zaustavlja. **Zašto je tomu tako?**

A Gradijent funkcije gubitka uvijek je različit od nul-vektora za barem jedan označeni primjer

B Sve točke minimuma funkcije pogreške u prostoru parametara leže na istoj ravnini

C Svaki minimum funkcije pogreške u prostoru parametara veći je od nule

D Površina funkcije pogreške u prostoru parametara po dijelovima je linearna i stoga nederivabilna

13 (N) Raspolažemo sljedećim skupom za učenje u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijski prostor značajki, definiranu kao $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$. Početne težine perceptrona su $\mathbf{w} = (1, 0, -1, 2, -3, 0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 32 B 6 C 16 D 22

14 (N) Na skupu označenih primjera $\mathcal{D}$ trenirali smo model logističke regresije. Dobili smo neki vektor težina $\mathbf{w}$ i pomak $w_0 = 3.15$. Tako naučenom modelu neki primjer $\mathbf{x}$, čija je oznaka u skupu primjera $y = 0$, nanosi gubitak unakrsne entropije od $L(0, h(\mathbf{x})) = 0.5$. **Koliki gubitak unakrsne entropije bi nanosio primjer x kada bismo njegove značajke pomnožili sa dva i promijenili mu oznaku**?

A 1.19 B 7.11 C 2.54 D 4.03

Cjelina 3: SVM, jezgrene i neparametarske metode (8 pitanja)

15 (T) Koristimo algoritam SVM s jezgrenom funkcijom κ. Optimizaciju provodimo u dualu. Ako je κ Mercerova jezgra, onda takva jezgra implicitno definira preslikavanje $\boldsymbol{\phi}$. Međutim, u nekim slučajevima trebamo eksplicitno izraziti funkciju $\boldsymbol{\phi}$ na temelju korištene jezgre κ. **Kada je potrebno eksplicitno izraziti $\boldsymbol{\phi}$?**

A Za izračun dualnih koeficijenata

B Za izračun težine w_0

C Za izračun širine margine

D Za izračun funkcije gubitka

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

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

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

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

17 (N) Rješavamo problem određivanja podrijetla pojedinih riječi u jeziku: za svaku riječ trebamo odrediti je li engleskog ($y = 1$) ili francuskog ($y = 0$) podrijetla. Problem rješavamo logističkom regresijom izvedenom kao rijetki jezgreni stroj, gdje za bazne funkcije koristimo jezgru κ nad znakovnim nizovima. Funkcija κ definirana je kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Na primjer, $\kappa(\text{water}, \text{eau}) = 2/6 = 0.33$. Skup za učenje je sljedeći:

$$\mathcal{D} = \{(\mathbf{x}, y)\}_i = \big\{(\text{water}, 1), (\text{eau}, 0), (\text{dog}, 1), (\text{chien}, 0), (\text{paperclip}, 1), (\text{trombone}), 0), (\text{chance}, 1), (\text{hasard}, 0)\big\}$$

Treniranjem rijetkoga jezgrenog stroja dobili smo vektor težina $\mathbf{w} = (-0.5, 0, 0, 0, -3.5, 1, 0, 0, 1)$. Razmotrite primjer $(\mathbf{x}, y) = (\text{nounours}, 0)$. **Koliko iznosi gubitak modela na primjeru $(\mathbf{x}, y)$?**

A 0.795 B 0.456 C 0.552 D 0.359

18 (T) Razmotrite problem maksimalne margine kao standardni problem optimizacije kriterijske funkcije $f(\mathbf{x})$ uz ograničenja nejednakosti $g_i(\mathbf{x}) \leq 0$ i jednakosti $h_i(\mathbf{x}) = 0$. Razmotrite taj problem za slučaj tvrde margine i za slučaj meke margine, oba u dualu. **Po čemu se dualni optimizacijski problem meke margine razlikuje od dualnog optimizacijskog problema tvrde margine?**

A Ima dodatnih N optimizacijskih varijabli i dodatnih $2N$ ograničenja nejednakosti

B Ima dodatnih N ograničenja nejednakosti

C Ima dodatnih n optimizacijskih varijabli i dodatnih N ograničenja jednakosti

D Ima dodatnih n ograničenja jednakosti

19 (P) Razmatramo troklasni ($K = 3$) klasifikacijski problem sa $N = 100$ primjera u ulaznom prostoru dimenzije $n = 5$ (ne računajući značajku x_0). Razmatramo dva modela. Model $\mathcal{H}_1$ je rijetki jezgreni stroj koji za bazne funkcije koristi Gaussove jezgre. Model $\mathcal{H}_2$ je multinomijalna regresija s adaptivnim baznim funkcijama. Adaptivnih baznih funkcija ima ukupno deset, uključujući baznu funkciju $\phi_0(\mathbf{x}) = 1$. Za oba modela izračunajte njihov ukupan broj parametara prije učenja. **Koliko iznosi razlika u broju parametara između ova dva modela?**

A 219 B 725 C 819 D 719

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

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

Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(2, 0)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina veća od stare?**

A $\frac{1}{2}\sqrt{13}$ puta B $\frac{5}{4}\sqrt{2}$ puta C $\frac{1}{2}\sqrt{10}$ puta D $\frac{1}{2}\sqrt{17}$ puta

21 (N) Za klasifikacijski problem koristimo stroj potpornih vektora s linearnom jezgrom i $C = 2$. Koristimo optimizaciju u dualu. Označen skup podataka i vektor dualnih koeficijenata jesu sljedeći:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} 15 & 30 & 15 & +1 \\ 6 & 12 & 6 & +1 \\ 11 & 32 & 21 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right) \qquad \boldsymbol{\alpha} = \begin{pmatrix} 0 \\ 2 \\ 0.077 \\ 1.923 \end{pmatrix}$$

Izračunajte parametre primarnog modela. **Koliko je primjer $\mathbf{x}^{(1)}$ udaljen od margine?**

A 4.09 B 3.23 C 5.25 D 0.44

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

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

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 0 D 1/4

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija (6 pitanja)

1 (P) Razmatramo binarni klasifikacijski problem sa $\mathcal{Y} = \{0,1\}$ u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((0,0),1),((2,2),1),((0,1),0),((0,-1),0),((1,1),0)\big\}$. Razmatramo sljedeće modele, parametrizirane sa $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$:

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geq 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = h_1(\mathbf{x};\boldsymbol{\theta}_1)\cdot h_2(\mathbf{x};\boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \geq \theta_0^2\} \qquad \mathcal{H}_4 : h_4(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \min\big(1, h_1(\mathbf{x};\boldsymbol{\theta}_1) + h_2(\mathbf{x};\boldsymbol{\theta}_2)\big)$$

Neka je E_α minimalna empirijska pogreška modela $\mathcal{H}_\alpha$ na skupu $\mathcal{D}$, tj. $E_\alpha = \operatorname{argmin}_{h\in\mathcal{H}_\alpha} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

[A] $E_1 = E_3 > E_2 = E_4$ [B] $E_1 > E_4 > E_2 = E_3$ [C] $E_1 = E_2 = E_4 > E_3$ [D] $E_2 > E_1 = E_3 > E_4$

2 (N) U dvodimenzijskome ulaznom prostoru postupkom najmanjih kvadrata (OLS) treniramo L2-regularizirani model linearne regresije. Koristimo funkciju preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_2^2)$. Označen skup podataka je:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rr|r} 1 & 5 & 1 \\ 2 & -2 & -5 \\ 3 & 1 & 2 \\ -5 & 6 & -2 \end{array}\right) \qquad \mathbf{G}^+ = \begin{pmatrix} 0.722 & -0.035 & 0.005 & -0.018 & -0.018 \\ -0.035 & 0.01 & 0 & 0.001 & 0.001 \\ 0.005 & 0 & 0.009 & 0 & -0.002 \\ -0.018 & 0.001 & 0 & 0.003 & -0.001 \\ -0.018 & 0.001 & -0.002 & -0.001 & 0.002 \end{pmatrix}$$

gdje je $\mathbf{G}^+$ pseudoinverz odgovarajuće Gramove matrice u prostoru značajki modificirane za regularizaciju uz $\lambda = 100$. **Koliko iznosi L_2-regularizirana pogreška naučenog modela?** (Napomena: matrica $\mathbf{G}^+$ dana je sa zaokruženim vrijednostima i koristite tu matricu.)

[A] 19.37 [B] 4.38 [C] 2.92 [D] 13.74

3 (P) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka su $\mathcal{H}_1$ i $\mathcal{H}_2$ dva modela različite složenosti te neka su $h_1^* \in \mathcal{H}_1$ i $h_2^* \in \mathcal{H}_2$ dvije hipoteze koje minimiziraju pogrešku $E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h)$ razlika između pogreške na ispitnom skupu i pogreške na skupu za učenje za hipotezu h, tj. $\Delta(h) = E(h|\mathcal{D}_{\text{test}}) - E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h_2^*) > \Delta(h_1^*)$. **Što možemo zaključiti o modelima $\mathcal{H}_1$ i $\mathcal{H}_2$?**

[A] Ako $E(h_1^*|\mathcal{D}_{\text{train}}) < E(h_2^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_1$ podnaučen
[B] Ako $E(h_2^*|\mathcal{D}_{\text{test}}) > E(h_1^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_2$ prenaučen
[C] Ako $E(h_2^*|\mathcal{D}_{\text{train}}) < E(h_1^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_2$ podnaučen
[D] Ako $E(h_1^*|\mathcal{D}_{\text{test}}) > E(h_2^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_1$ prenaučen

4 (T) Premda je linearna regresija linearan model, ona može modelirati nelinearne hipoteze. Nelinearnost možemo ostvariti uporabom prikladne funkcije preslikavanja $\boldsymbol{\phi}$ koja primjere iz ulaznog prostora preslikava u prostor značajki. **Koja pretpostavka stoji iza takvog postupka?**

[A] Primjeri koji su linearno odvojivi u ulaznome prostoru postaju linearno neodvojivi u prostoru značajki dovoljno visoke dimenzije
[B] Hiperravnina koja minimizira pogrešku u ulaznome prostoru odgovara nelinearnoj hiperpovršini u prostoru značajki
[C] Primjeri koji u ulaznome prostoru leže na nelinearnoj hiperpovršini u prostoru značajki leže na hiperravnini
[D] Primjeri koji se u prostoru značajki nalaze blizu nelinearne hiperpovršine u ulaznome prostoru leže na hiperavnini

5 (T) Algoritam linearne regresije koristi kvadratni gubitak. **Koje je probabilističko opravdanje za korištenje kvadratnog gubitka?**

- A Minimizacija kvadratnog gubitka istovjetna je maksimizaciji negativnog logaritma vjerojatnosti parametara
- B Vjerojatnost parametara za fiksni označeni skup podataka raste kvadratno s očekivanjem gubitka
- C Primjeri za koje je predviđanje hipoteze pogrešno pokoravaju se normalnoj razdiobi sa središtem u nuli
- D Vjerojatnost skupa primjera najveća je za parametre koji minimiziraju očekivanje kvadratnog gubitka

6 (P) Koristimo regresiju za predviđanje uspjeha studija na temelju prosjeka ocjena u četiri razreda srednje škole te uspjeha iz matematike, fizike i informatike na državnoj maturi (ukupno 7 značajki). Za preslikavanje u prostor značajki koristimo polinom drugog stupnja s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki. Pretpostavite da nema multikolinearnosti. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 56 B 22 C 36 D 28

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

7 (N) Na skupu označenih primjera $\mathcal{D}$ trenirali smo model logističke regresije. Dobili smo neki vektor težina $\mathbf{w}$ i pomak $w_0 = 3.15$. Tako naučenom modelu neki primjer $\mathbf{x}$, čija je oznaka u skupu primjera $y = 0$, nanosi gubitak unakrsne entropije od $L(0, h(\mathbf{x})) = 0.5$. **Koliki gubitak unakrsne entropije bi nanosio primjer x kada bismo njegove značajke pomnožili sa dva i promijenili mu oznaku**?

A 2.54 B 1.19 C 4.03 D 7.11

8 (N) Model regularizirane logističke regresije treniramo stohastičkim gradijentnim spustom. Koristimo faktor regularizacije $\lambda = 1000$ i stopu učenja $\eta = 0.01$. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.5, -2.2, -1.1, 2.7)$. **Koliko u toj iteraciji iznosi promjena težine w_1 za primjer $(\mathbf{x}, y) = ((-1, 2), 1)$?**

A -5 B $+22$ C -2 D -12

9 (T) Multinomijalna logistička regresija primjere klasificira u jednu od K klasa. Neka je $\mathbf{y}^{(i)} = (y_1^i, \ldots, y_k^i, \ldots, y_K^i)$ vektor indikatorskih varijabli koji odgovara oznaci primjera $\mathbf{x}^{(i)}$. Neka je $h_k(\mathbf{x}; \mathbf{W})$ vrijednost funkcije *softmax* za klasu k. Optimizacija parametara $\mathbf{W}$ odgovara maksimizaciji logaritma vjerojatnost oznaka. **Čemu je jednaka ta vjerojatnost?**

A $\sum_{i=1}^N \sum_{k=1}^K \ln y_k^i h_k(\mathbf{x}^{(i)}; \mathbf{W})$ C $\ln \prod_{i=1}^N \prod_{k=1}^K y_k^i h_k(\mathbf{x}^{(i)}; \mathbf{W})$

B $\sum_{i=1}^N \ln \prod_{k=1}^K h_k(\mathbf{x}^{(i)}; \mathbf{W})^{y_k^i}$ D $\prod_{i=1}^N \sum_{k=1}^K \ln h_k(\mathbf{x}^{(i)}; \mathbf{W})^{y_k^i}$

10 (P) Algoritmom perceptrona rješavamo klasifikacijski problem sa $K = 10$ klasa. Koristimo pet značajki, $n = 5$. Za preslikavanje $\boldsymbol{\phi}$ koristimo polinom drugog stupnja sa svim interakcijskim parovima značajki. Koristimo shemu jedan-naspram-jedan (OVO). **Koliko iznosi ukupan broj parametara takvog višeklasnog klasifikatora?**

A 900 B 945 C 495 D 210

11 (N) Raspolažemo sljedećim skupom za učenje u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijski prostor značajki, definiranu kao $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$. Početne težine perceptrona su $\mathbf{w} = (1, 0, 1, 2, -3, 0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 32 B 22 C 16 D 6

12 (T) Jedan od nedostataka algoritma perceptrona jest taj što se za linearno neodovojive probleme algoritam ne zaustavlja. **Zašto je tomu tako?**

A Svaki minimum funkcije pogreške u prostoru parametara veći je od nule

B Gradijent funkcije gubitka uvijek je različit od nul-vektora za barem jedan označeni primjer

C Površina funkcije pogreške u prostoru parametara po dijelovima je linearna i stoga nederivabilna

D Sve točke minimuma funkcije pogreške u prostoru parametara leže na istoj ravnini

13 (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 6.00 C 4.02 D 8.00

14 (P) Algoritmom perceptrona rješavamo klasifikacijski problem. Oznake su iz skupa $\{-1, +1\}$. Za stopu učenja koristimo $\eta = \frac{1}{2}$. U jednoj iteraciji algoritma, kod ažuriranja težina na temelju primjera $\mathbf{x}$, vektor težina $\mathbf{w}$ mijenja se tako da mu se duljina prepolavlja, ali se zadržavaju smjer i orijentacija. Razmotite da, umjesto perceptrona, za klasifikaciju koristimo linearnu regresiju treniranu na oznakama $\{-1, +1\}$. **Koliko bi iznosio gubitak za isti primjer x kada bismo koristili linearnu regresiju s težinama w?**

A $(y - \|\mathbf{x}\|^2)^2$ B $(\|\mathbf{w}\|^2 + 1)^2$ C $1 - \|\mathbf{x}\|^2$ D $1 - y\|\mathbf{w}\|^2$

Cjelina 3: SVM, jezgrene i neparametarske metode (8 pitanja)

15 (T) Razmotrite problem maksimalne margine kao standardni problem optimizacije kriterijske funkcije $f(\mathbf{x})$ uz ograničenja nejednakosti $g_i(\mathbf{x}) \leq 0$ i jednakosti $h_i(\mathbf{x}) = 0$. Razmotrite taj problem za slučaj tvrde margine i za slučaj meke margine, oba u dualu. **Po čemu se dualni optimizacijski problem meke margine razlikuje od dualnog optimizacijskog problema tvrde margine?**

A Ima dodatnih N optimizacijskih varijabli i dodatnih $2N$ ograničenja nejednakosti

B Ima dodatnih n optimizacijskih varijabli i dodatnih N ograničenja jednakosti

C Ima dodatnih n ograničenja jednakosti

D Ima dodatnih N ograničenja nejednakosti

16 (N) Za klasifikacijski problem koristimo stroj potpornih vektora s linearnom jezgrom i $C = 2$. Koristimo optimizaciju u dualu. Označen skup podataka i vektor dualnih koeficijenata jesu sljedeći:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} 25 & 70 & 45 & +1 \\ 6 & 12 & 6 & +1 \\ 11 & 32 & 21 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right) \qquad \boldsymbol{\alpha} = \begin{pmatrix} 0 \\ 2 \\ 0.077 \\ 1.923 \end{pmatrix}$$

Izračunajte parametre primarnog modela. **Koliko je primjer $\mathbf{x}^{(1)}$ udaljen od margine?**

A 3.23 B 5.25 C 0.44 D 4.09

17 (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 0 C 1/4 D 3/4

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

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

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

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

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

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

Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(2, 0)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina veća od stare?**

A $\frac{1}{2}\sqrt{17}$ puta B $\frac{1}{2}\sqrt{13}$ puta C $\frac{5}{4}\sqrt{2}$ puta D $\frac{1}{2}\sqrt{10}$ puta

20 (N) Rješavamo problem određivanja podrijetla pojedinih riječi u jeziku: za svaku riječ trebamo odrediti je li engleskog ($y = 1$) ili francuskog ($y = 0$) podrijetla. Problem rješavamo logističkom regresijom izvedenom kao rijetki jezgreni stroj, gdje za bazne funkcije koristimo jezgru κ nad znakovnim nizovima. Funkcija κ definirana je kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Na primjer, $\kappa(\text{water}, \text{eau}) = 2/6 = 0.33$. Skup za učenje je sljedeći:

$$\mathcal{D} = \{(\mathbf{x}, y)\}_i = \big\{(\text{water}, 1), (\text{eau}, 0), (\text{dog}, 1), (\text{chien}, 0), (\text{paperclip}, 1), (\text{trombone}), 0), (\text{chance}, 1), (\text{hasard}, 0)\big\}$$

Treniranjem rijetkoga jezgrenog stroja dobili smo vektor težina $\mathbf{w} = (-0.5, 0, 0, 0, -3.5, 0, 1, 0, 1)$. Razmotrite primjer $(\mathbf{x}, y) = (\text{nounours}, 0)$. **Koliko iznosi gubitak modela na primjeru $(\mathbf{x}, y)$?**

A 0.552 B 0.359 C 0.456 D 0.795

21 (T) Koristimo algoritam SVM s jezgrenom funkcijom κ. Optimizaciju provodimo u dualu. Ako je κ Mercerova jezgra, onda takva jezgra implicitno definira preslikavanje $\boldsymbol{\phi}$. Međutim, u nekim slučajevima trebamo eksplicitno izraziti funkciju $\boldsymbol{\phi}$ na temelju korištene jezgre κ. **Kada je potrebno eksplicitno izraziti $\boldsymbol{\phi}$?**

A Za izračun funkcije gubitka
B Za izračun širine margine
C Za izračun težine w_0
D Za izračun dualnih koeficijenata

22 (P) Razmatramo troklasni ($K = 3$) klasifikacijski problem sa $N = 100$ primjera u ulaznom prostoru dimenzije $n = 5$ (ne računajući značajku x_0). Razmatramo dva modela. Model $\mathcal{H}_1$ je rijetki jezgreni stroj koji za bazne funkcije koristi Gaussove jezgre. Model $\mathcal{H}_2$ je multinomijalna regresija s adaptivnim baznim funkcijama. Adaptivnih baznih funkcija ima ukupno deset, uključujući baznu funkciju $\phi_0(\mathbf{x}) = 1$. Za oba modela izračunajte njihov ukupan broj parametara prije učenja. **Koliko iznosi razlika u broju parametara između ova dva modela?**

A 219 B 725 C 819 D 719

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija (6 pitanja)

1 (T) Algoritam linearne regresije koristi kvadratni gubitak. **Koje je probabilističko opravdanje za korištenje kvadratnog gubitka?**

- A Vjerojatnost parametara za fiksni označeni skup podataka raste kvadratno s očekivanjem gubitka
- B Vjerojatnost skupa primjera najveća je za parametre koji minimiziraju očekivanje kvadratnog gubitka
- C Minimizacija kvadratnog gubitka istovjetna je maksimizaciji negativnog logaritma vjerojatnosti parametara
- D Primjeri za koje je predviđanje hipoteze pogrešno pokoravaju se normalnoj razdiobi sa središtem u nuli

2 (N) U dvodimenzijskome ulaznom prostoru postupkom najmanjih kvadrata (OLS) treniramo L2-regularizirani model linearne regresije. Koristimo funkciju preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_2^2)$. Označen skup podataka je:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rr|r} 1 & 5 & -2 \\ 2 & -2 & 5 \\ 3 & 1 & 3 \\ -5 & 6 & -3 \end{array}\right) \qquad \mathbf{G}^+ = \begin{pmatrix} 0.722 & -0.035 & 0.005 & -0.018 & -0.018 \\ -0.035 & 0.01 & 0 & 0.001 & 0.001 \\ 0.005 & 0 & 0.009 & 0 & -0.002 \\ -0.018 & 0.001 & 0 & 0.003 & -0.001 \\ -0.018 & 0.001 & -0.002 & -0.001 & 0.002 \end{pmatrix}$$

gdje je $\mathbf{G}^+$ pseudoinverz odgovarajuće Gramove matrice u prostoru značajki modificirane za regularizaciju uz $\lambda = 100$. **Koliko iznosi L_2-regularizirana pogreška naučenog modela?** (Napomena: matrica $\mathbf{G}^+$ dana je sa zaokruženim vrijednostima i koristite tu matricu.)

A 19.37 B 13.74 C 4.38 D 2.92

3 (T) Premda je linearna regresija linearan model, ona može modelirati nelinearne hipoteze. Nelinearnost možemo ostvariti uporabom prikladne funkcije preslikavanja $\boldsymbol{\phi}$ koja primjere iz ulaznog prostora preslikava u prostor značajki. **Koja pretpostavka stoji iza takvog postupka?**

- A Primjeri koji u ulaznome prostoru leže na nelinearnoj hiperpovršini u prostoru značajki leže na hiperravnini
- B Primjeri koji se u prostoru značajki nalaze blizu nelinearne hiperpovršine u ulaznome prostoru leže na hiperavnini
- C Primjeri koji su linearno odvojivi u ulaznome prostoru postaju linearno neodvojivi u prostoru značajki dovoljno visoke dimenzije
- D Hiperravnina koja minimizira pogrešku u ulaznome prostoru odgovara nelinearnoj hiperpovršini u prostoru značajki

4 (P) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka su $\mathcal{H}_1$ i $\mathcal{H}_2$ dva modela različite složenosti te neka su $h_1^* \in \mathcal{H}_1$ i $h_2^* \in \mathcal{H}_2$ dvije hipoteze koje minimiziraju pogrešku $E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h)$ razlika između pogreške na ispitnom skupu i pogreške na skupu za učenje za hipotezu h, tj. $\Delta(h) = E(h|\mathcal{D}_{\text{test}}) - E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h_2^*) > \Delta(h_1^*)$. **Što možemo zaključiti o modelima $\mathcal{H}_1$ i $\mathcal{H}_2$?**

- A Ako $E(h_2^*|\mathcal{D}_{\text{test}}) > E(h_1^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_2$ prenaučen
- B Ako $E(h_1^*|\mathcal{D}_{\text{train}}) < E(h_2^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_1$ podnaučen
- C Ako $E(h_1^*|\mathcal{D}_{\text{test}}) > E(h_2^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_1$ prenaučen
- D Ako $E(h_2^*|\mathcal{D}_{\text{train}}) < E(h_1^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_2$ podnaučen

5 (P) Razmatramo binarni klasifikacijski problem sa $\mathcal{Y} = \{0, 1\}$ u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((0,0),1),((2,2),1),((0,1),0),((0,-1),0),((1,1),0)\big\}$. Razmatramo sljedeće modele, parametrizirane sa $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$:

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geq 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = h_1(\mathbf{x};\boldsymbol{\theta}_1)\cdot h_2(\mathbf{x};\boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \geq \theta_0^2\big\} \qquad \mathcal{H}_4 : h_4(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \min\big(1, h_1(\mathbf{x};\boldsymbol{\theta}_1) + h_2(\mathbf{x};\boldsymbol{\theta}_2)\big)$$

Neka je E_α minimalna empirijska pogreška modela $\mathcal{H}_\alpha$ na skupu $\mathcal{D}$, tj. $E_\alpha = \operatorname{argmin}_{h\in\mathcal{H}_\alpha} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

A $E_2 > E_1 = E_3 > E_4$ B $E_1 = E_2 = E_4 > E_3$ C $E_1 > E_4 > E_2 = E_3$ D $E_1 = E_3 > E_2 = E_4$

6 (P) Koristimo regresiju za predviđanje uspjeha studija na temelju prosjeka ocjena u četiri razreda srednje škole te uspjeha iz matematike, fizike i informatike na državnoj maturi (ukupno 7 značajki). Za preslikavanje u prostor značajki koristimo polinom drugog stupnja s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki. Pretpostavite da nema multikolinearnosti. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 28 B 56 C 36 D 22

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

7 (T) Jedan od nedostataka algoritma perceptrona jest taj što se za linearno neodovojive probleme algoritam ne zaustavlja. **Zašto je tomu tako?**

A Gradijent funkcije gubitka uvijek je različit od nul-vektora za barem jedan označeni primjer

B Površina funkcije pogreške u prostoru parametara po dijelovima je linearna i stoga nederivabilna

C Sve točke minimuma funkcije pogreške u prostoru parametara leže na istoj ravnini

D Svaki minimum funkcije pogreške u prostoru parametara veći je od nule

8 (N) Model regularizirane logističke regresije treniramo stohastičkim gradijentnim spustom. Koristimo faktor regularizacije $\lambda = 1000$ i stopu učenja $\eta = 0.01$. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.5, -2.2, -1.1, 2.7)$. **Koliko u toj iteraciji iznosi promjena težine w_1 za primjer $(\mathbf{x}, y) = ((-1,2),1)$?**

A -5 B $+22$ C -2 D -12

9 (P) Algoritmom perceptrona rješavamo klasifikacijski problem sa $K = 10$ klasa. Koristimo pet značajki, $n = 5$. Za preslikavanje $\boldsymbol{\phi}$ koristimo polinom drugog stupnja sa svim interakcijskim parovima značajki. Koristimo shemu jedan-naspram-jedan (OVO). **Koliko iznosi ukupan broj parametara takvog višeklasnog klasifikatora?**

A 210 B 900 C 495 D 945

10 (T) Multinomijalna logistička regresija primjere klasificira u jednu od K klasa. Neka je $\mathbf{y}^{(i)} = (y_1^i, \ldots, y_k^i, \ldots, y_K^i)$ vektor indikatorskih varijabli koji odgovara oznaci primjera $\mathbf{x}^{(i)}$. Neka je $h_k(\mathbf{x};\mathbf{W})$ vrijednost funkcije *softmax* za klasu k. Optimizacija parametara $\mathbf{W}$ odgovara maksimizaciji logaritma vjerojatnost oznaka. **Čemu je jednaka ta vjerojatnost?**

A $\prod_{i=1}^N \sum_{k=1}^K \ln h_k(\mathbf{x}^{(i)};\mathbf{W})^{y_k^i}$ C $\sum_{i=1}^N \ln \prod_{k=1}^K h_k(\mathbf{x}^{(i)};\mathbf{W})^{y_k^i}$

B $\ln \prod_{i=1}^N \prod_{k=1}^K y_k^i h_k(\mathbf{x}^{(i)};\mathbf{W})$ D $\sum_{i=1}^N \sum_{k=1}^K \ln y_k^i h_k(\mathbf{x}^{(i)};\mathbf{W})$

11 (N) Na skupu označenih primjera $\mathcal{D}$ trenirali smo model logističke regresije. Dobili smo neki vektor težina $\mathbf{w}$ i pomak $w_0 = 3.15$. Tako naučenom modelu neki primjer $\mathbf{x}$, čija je oznaka u skupu primjera $y = 0$, nanosi gubitak unakrsne entropije od $L(0, h(\mathbf{x})) = 0.5$. **Koliki gubitak unakrsne entropije bi nanosio primjer x kada bismo njegove značajke pomnožili sa dva i promijenili mu oznaku**?

A 7.11 B 2.54 C 4.03 D 1.19

12 (N) Raspolažemo sljedećim skupom za učenje u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijski prostor značajki, definiranu kao $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$. Početne težine perceptrona su $\mathbf{w} = (1,0,-1,2,3,0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 22 B 6 C 16 D 32

13 (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 6.00 B 8.00 C 4.02 D 12.02

14 (P) Algoritmom perceptrona rješavamo klasifikacijski problem. Oznake su iz skupa $\{-1, +1\}$. Za stopu učenja koristimo $\eta = \frac{1}{2}$. U jednoj iteraciji algoritma, kod ažuriranja težina na temelju primjera $\mathbf{x}$, vektor težina $\mathbf{w}$ mijenja se tako da mu se duljina prepolavlja, ali se zadržavaju smjer i orijentacija. Razmotite da, umjesto perceptrona, za klasifikaciju koristimo linearnu regresiju treniranu na oznakama $\{-1, +1\}$. **Koliko bi iznosio gubitak za isti primjer $\mathbf{x}$ kada bismo koristili linearnu regresiju s težinama $\mathbf{w}$?**

A $(y - \|\mathbf{x}\|^2)^2$ B $1 - \|\mathbf{x}\|^2$ C $1 - y\|\mathbf{w}\|^2$ D $(\|\mathbf{w}\|^2 + 1)^2$

Cjelina 3: SVM, jezgrene i neparametarske metode (8 pitanja)

15 (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 0 B 1/4 C 3/4 D 1/2

16 (T) Razmotrite problem maksimalne margine kao standardni problem optimizacije kriterijske funkcije $f(\mathbf{x})$ uz ograničenja nejednakosti $g_i(\mathbf{x}) \leq 0$ i jednakosti $h_i(\mathbf{x}) = 0$. Razmotrite taj problem za slučaj tvrde margine i za slučaj meke margine, oba u dualu. **Po čemu se dualni optimizacijski problem meke margine razlikuje od dualnog optimizacijskog problema tvrde margine?**

A Ima dodatnih n ograničenja jednakosti

B Ima dodatnih n optimizacijskih varijabli i dodatnih N ograničenja jednakosti

C Ima dodatnih N ograničenja nejednakosti

D Ima dodatnih N optimizacijskih varijabli i dodatnih $2N$ ograničenja nejednakosti

17 (N) Rješavamo problem određivanja podrijetla pojedinih riječi u jeziku: za svaku riječ trebamo odrediti je li engleskog ($y = 1$) ili francuskog ($y = 0$) podrijetla. Problem rješavamo logističkom regresijom izvedenom kao rijetki jezgreni stroj, gdje za bazne funkcije koristimo jezgru κ nad znakovnim nizovima. Funkcija κ definirana je kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Na primjer, $\kappa(\text{water}, \text{eau}) = 2/6 = 0.33$. Skup za učenje je sljedeći:

$$\mathcal{D} = \{(\mathbf{x}, y)\}_i = \big\{(\text{water}, 1), (\text{eau}, 0), (\text{dog}, 1), (\text{chien}, 0), (\text{paperclip}, 1), (\text{trombone}), 0), (\text{chance}, 1), (\text{hasard}, 0)\big\}$$

Treniranjem rijetkoga jezgrenog stroja dobili smo vektor težina $\mathbf{w} = (-0.5, 0, 0, -3.5, 0, -1, 0, 0, 1)$. Razmotrite primjer $(\mathbf{x}, y) = (\text{nounours}, 0)$. **Koliko iznosi gubitak modela na primjeru $(\mathbf{x}, y)$?**

A 0.552 B 0.456 C 0.795 D 0.359

18 (T) Koristimo algoritam SVM s jezgrenom funkcijom κ. Optimizaciju provodimo u dualu. Ako je κ Mercerova jezgra, onda takva jezgra implicitno definira preslikavanje $\boldsymbol{\phi}$. Međutim, u nekim slučajevima trebamo eksplicitno izraziti funkciju $\boldsymbol{\phi}$ na temelju korištene jezgre κ. **Kada je potrebno eksplicitno izraziti $\boldsymbol{\phi}$?**

A Za izračun funkcije gubitka

B Za izračun širine margine

C Za izračun dualnih koeficijenata

D Za izračun težine w_0

19 (P) Razmatramo troklasni ($K = 3$) klasifikacijski problem sa $N = 100$ primjera u ulaznom prostoru dimenzije $n = 5$ (ne računajući značajku x_0). Razmatramo dva modela. Model $\mathcal{H}_1$ je rijetki jezgreni stroj koji za bazne funkcije koristi Gaussove jezgre. Model $\mathcal{H}_2$ je multinomijalna regresija s adaptivnim baznim funkcijama. Adaptivnih baznih funkcija ima ukupno deset, uključujući baznu funkciju $\phi_0(\mathbf{x}) = 1$. Za oba modela izračunajte njihov ukupan broj parametara prije učenja. **Koliko iznosi razlika u broju parametara između ova dva modela?**

A 219 B 719 C 725 D 819

20 (N) Za klasifikacijski problem koristimo stroj potpornih vektora s linearnom jezgrom i $C = 2$. Koristimo optimizaciju u dualu. Označen skup podataka i vektor dualnih koeficijenata jesu sljedeći:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{ccc|c} 25 & 70 & 45 & +1 \\ 6 & 12 & 6 & +1 \\ 11 & 32 & 21 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right) \qquad \boldsymbol{\alpha} = \begin{pmatrix} 0 \\ 2 \\ 0.077 \\ 1.923 \end{pmatrix}$$

Izračunajte parametre primarnog modela. **Koliko je primjer $\mathbf{x}^{(1)}$ udaljen od margine?**

A 0.44 B 4.09 C 5.25 D 3.23

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

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

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

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

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

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

Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(2, 0)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina veća od stare?**

A $\frac{1}{2}\sqrt{17}$ puta B $\frac{5}{4}\sqrt{2}$ puta C $\frac{1}{2}\sqrt{13}$ puta D $\frac{1}{2}\sqrt{10}$ puta

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija (6 pitanja)

1 (P) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka su $\mathcal{H}_1$ i $\mathcal{H}_2$ dva modela različite složenosti te neka su $h_1^* \in \mathcal{H}_1$ i $h_2^* \in \mathcal{H}_2$ dvije hipoteze koje minimiziraju pogrešku $E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h)$ razlika između pogreške na ispitnom skupu i pogreške na skupu za učenje za hipotezu h, tj. $\Delta(h) = E(h|\mathcal{D}_{\text{test}}) - E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h_2^*) < \Delta(h_1^*)$. **Što možemo zaključiti o modelima $\mathcal{H}_1$ i $\mathcal{H}_2$?**

- A Ako $E(h_1^*|\mathcal{D}_{\text{train}}) < E(h_2^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_1$ podnaučen
- B Ako $E(h_2^*|\mathcal{D}_{\text{test}}) > E(h_1^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_2$ prenaučen
- C Ako $E(h_2^*|\mathcal{D}_{\text{train}}) < E(h_1^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_2$ podnaučen
- D Ako $E(h_1^*|\mathcal{D}_{\text{test}}) > E(h_2^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_1$ prenaučen

2 (P) Razmatramo binarni klasifikacijski problem sa $\mathcal{Y} = \{0, 1\}$ u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((0,0),1),((2,2),1),((0,1),0),((0,-1),0),((1,1),0)\big\}$. Razmatramo sljedeće modele, parametrizirane sa $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$:

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geq 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = h_1(\mathbf{x};\boldsymbol{\theta}_1) \cdot h_2(\mathbf{x};\boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \geq \theta_0^2\} \qquad \mathcal{H}_4 : h_4(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \min\big(1, h_1(\mathbf{x};\boldsymbol{\theta}_1) + h_2(\mathbf{x};\boldsymbol{\theta}_2)\big)$$

Neka je E_α minimalna empirijska pogreška modela $\mathcal{H}_\alpha$ na skupu $\mathcal{D}$, tj. $E_\alpha = \operatorname{argmin}_{h\in\mathcal{H}_\alpha} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

A $E_2 > E_1 = E_3 > E_4$ B $E_1 > E_4 > E_2 = E_3$ C $E_1 = E_3 > E_2 = E_4$ D $E_1 = E_2 = E_4 > E_3$

3 (T) Algoritam linearne regresije koristi kvadratni gubitak. **Koje je probabilističko opravdanje za korištenje kvadratnog gubitka?**

- A Vjerojatnost skupa primjera najveća je za parametre koji minimiziraju očekivanje kvadratnog gubitka
- B Primjeri za koje je predviđanje hipoteze pogrešno pokoravaju se normalnoj razdiobi sa središtem u nuli
- C Minimizacija kvadratnog gubitka istovjetna je maksimizaciji negativnog logaritma vjerojatnosti parametara
- D Vjerojatnost parametara za fiksni označeni skup podataka raste kvadratno s očekivanjem gubitka

4 (T) Premda je linearna regresija linearan model, ona može modelirati nelinearne hipoteze. Nelinearnost možemo ostvariti uporabom prikladne funkcije preslikavanja $\boldsymbol{\phi}$ koja primjere iz ulaznog prostora preslikava u prostor značajki. **Koja pretpostavka stoji iza takvog postupka?**

- A Primjeri koji se u prostoru značajki nalaze blizu nelinearne hiperpovršine u ulaznome prostoru leže na hiperravnini
- B Primjeri koji u ulaznome prostoru leže na nelinearnoj hiperpovršini u prostoru značajki leže na hiperravnini
- C Primjeri koji su linearno odvojivi u ulaznome prostoru postaju linearno neodvojivi u prostoru značajki dovoljno visoke dimenzije
- D Hiperravnina koja minimizira pogrešku u ulaznome prostoru odgovara nelinearnoj hiperpovršini u prostoru značajki

5 (P) Koristimo regresiju za predviđanje uspjeha studija na temelju prosjeka ocjena u četiri razreda srednje škole te uspjeha iz matematike, fizike i informatike na državnoj maturi (ukupno 7 značajki). Za preslikavanje u prostor značajki koristimo polinom drugog stupnja s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki. Pretpostavite da nema multikolinearnosti. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 22 B 56 C 28 D 36

6 (N) U dvodimenzijskome ulaznom prostoru postupkom najmanjih kvadrata (OLS) treniramo L2-regularizirani model linearne regresije. Koristimo funkciju preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_2^2)$. Označen skup podataka je:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rr|r} 1 & 5 & 1 \\ 2 & -2 & -5 \\ 3 & 1 & 2 \\ -5 & 6 & -2 \end{array}\right) \qquad \mathbf{G}^+ = \begin{pmatrix} 0.722 & -0.035 & 0.005 & -0.018 & -0.018 \\ -0.035 & 0.01 & 0 & 0.001 & 0.001 \\ 0.005 & 0 & 0.009 & 0 & -0.002 \\ -0.018 & 0.001 & 0 & 0.003 & -0.001 \\ -0.018 & 0.001 & -0.002 & -0.001 & 0.002 \end{pmatrix}$$

gdje je $\mathbf{G}^+$ pseudoinverz odgovarajuće Gramove matrice u prostoru značajki modificirane za regularizaciju uz $\lambda = 100$. **Koliko iznosi L_2-regularizirana pogreška naučenog modela?** (Napomena: matrica $\mathbf{G}^+$ dana je sa zaokruženim vrijednostima i koristite tu matricu.)

A 2.92 B 13.74 C 4.38 D 19.37

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

7 (N) Model regularizirane logističke regresije treniramo stohastičkim gradijentnim spustom. Koristimo faktor regularizacije $\lambda = 1000$ i stopu učenja $\eta = 0.01$. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.5, 0.2, -1.1, 2.7)$. **Koliko u toj iteraciji iznosi promjena težine w_1 za primjer $(\mathbf{x}, y) = ((-1, 2), 1)$?**

A $+22$ B -5 C -2 D -12

8 (T) Jedan od nedostataka algoritma perceptrona jest taj što se za linearno neodovojive probleme algoritam ne zaustavlja. **Zašto je tomu tako?**

A Gradijent funkcije gubitka uvijek je različit od nul-vektora za barem jedan označeni primjer

B Površina funkcije pogreške u prostoru parametara po dijelovima je linearna i stoga nederivabilna

C Svaki minimum funkcije pogreške u prostoru parametara veći je od nule

D Sve točke minimuma funkcije pogreške u prostoru parametara leže na istoj ravnini

9 (P) Algoritmom perceptrona rješavamo klasifikacijski problem. Oznake su iz skupa $\{-1, +1\}$. Za stopu učenja koristimo $\eta = \frac{1}{2}$. U jednoj iteraciji algoritma, kod ažuriranja težina na temelju primjera $\mathbf{x}$, vektor težina $\mathbf{w}$ mijenja se tako da mu se duljina prepolavlja, ali se zadržavaju smjer i orijentacija. Razmotite da, umjesto perceptrona, za klasifikaciju koristimo linearnu regresiju treniranu na oznakama $\{-1, +1\}$. **Koliko bi iznosio gubitak za isti primjer x kada bismo koristili linearnu regresiju s težinama w?**

A $(\|\mathbf{w}\|^2 + 1)^2$ B $1 - y\|\mathbf{w}\|^2$ C $1 - \|\mathbf{x}\|^2$ D $(y - \|\mathbf{x}\|^2)^2$

10 (N) Raspolažemo sljedećim skupom za učenje u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijski prostor značajki, definiranu kao $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$. Početne težine perceptrona su $\mathbf{w} = (1, 0, -1, 2, -3, 0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 22 B 32 C 16 D 6

11 (P) Algoritmom perceptrona rješavamo klasifikacijski problem sa $K = 10$ klasa. Koristimo pet značajki, $n = 5$. Za preslikavanje $\boldsymbol{\phi}$ koristimo polinom drugog stupnja sa svim interakcijskim parovima značajki. Koristimo shemu jedan-naspram-jedan (OVO). **Koliko iznosi ukupan broj parametara takvog višeklasnog klasifikatora?**

A 900 B 210 C 495 D 945

12 (N) Na skupu označenih primjera $\mathcal{D}$ trenirali smo model logističke regresije. Dobili smo neki vektor težina $\mathbf{w}$ i pomak $w_0 = 0.15$. Tako naučenom modelu neki primjer $\mathbf{x}$, čija je oznaka u skupu primjera $y = 0$, nanosi gubitak unakrsne entropije od $L(0, h(\mathbf{x})) = 0.274$. **Koliki gubitak unakrsne entropije bi nanosio primjer x kada bismo njegove značajke pomnožili sa dva i promijenili mu oznaku**?

A 4.03 B 1.19 C 2.54 D 7.11

13 (T) Multinomijalna logistička regresija primjere klasificira u jednu od K klasa. Neka je $\mathbf{y}^{(i)} = (y_1^i, \ldots, y_k^i, \ldots, y_K^i)$ vektor indikatorskih varijabli koji odgovara oznaci primjera $\mathbf{x}^{(i)}$. Neka je $h_k(\mathbf{x}; \mathbf{W})$ vrijednost funkcije *softmax* za klasu k. Optimizacija parametara $\mathbf{W}$ odgovara maksimizaciji logaritma vjerojatnost oznaka. **Čemu je jednaka ta vjerojatnost?**

A $\sum_{i=1}^{N} \sum_{k=1}^{K} \ln y_k^i h_k(\mathbf{x}^{(i)}; \mathbf{W})$ C $\sum_{i=1}^{N} \ln \prod_{k=1}^{K} h_k(\mathbf{x}^{(i)}; \mathbf{W})^{y_k^i}$
B $\prod_{i=1}^{N} \sum_{k=1}^{K} \ln h_k(\mathbf{x}^{(i)}; \mathbf{W})^{y_k^i}$ D $\ln \prod_{i=1}^{N} \prod_{k=1}^{K} y_k^i h_k(\mathbf{x}^{(i)}; \mathbf{W})$

14 (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 4.02 B 6.00 C 8.00 D 12.02

Cjelina 3: SVM, jezgrene i neparametarske metode (8 pitanja)

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

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

Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(2, 0)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina veća od stare?**

A $\frac{5}{4}\sqrt{2}$ puta B $\frac{1}{2}\sqrt{13}$ puta C $\frac{1}{2}\sqrt{10}$ puta D $\frac{1}{2}\sqrt{17}$ puta

16 (P) Razmatramo troklasni ($K = 3$) klasifikacijski problem sa $N = 100$ primjera u ulaznom prostoru dimenzije $n = 5$ (ne računajući značajku x_0). Razmatramo dva modela. Model $\mathcal{H}_1$ je rijetki jezgreni stroj koji za bazne funkcije koristi Gaussove jezgre. Model $\mathcal{H}_2$ je multinomijalna regresija s adaptivnim baznim funkcijama. Adaptivnih baznih funkcija ima ukupno deset, uključujući baznu funkciju $\phi_0(\mathbf{x}) = 1$. Za oba modela izračunajte njihov ukupan broj parametara prije učenja. **Koliko iznosi razlika u broju parametara između ova dva modela?**

A 819 B 719 C 725 D 219

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

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

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

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

18 (N) Rješavamo problem određivanja podrijetla pojedinih riječi u jeziku: za svaku riječ trebamo odrediti je li engleskog ($y = 1$) ili francuskog ($y = 0$) podrijetla. Problem rješavamo logističkom regresijom izvedenom kao rijetki jezgreni stroj, gdje za bazne funkcije koristimo jezgru κ nad znakovnim nizovima. Funkcija κ definirana je kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Na primjer, $\kappa(\text{water}, \text{eau}) = 2/6 = 0.33$. Skup za učenje je sljedeći:

$$\mathcal{D} = \{(\mathbf{x}, y)\}_i = \big\{(\text{water}, 1), (\text{eau}, 0), (\text{dog}, 1), (\text{chien}, 0), (\text{paperclip}, 1), (\text{trombone}), 0), (\text{chance}, 1), (\text{hasard}, 0)\big\}$$

Treniranjem rijetkoga jezgrenog stroja dobili smo vektor težina $\mathbf{w} = (0.5, 0, 0, 0, -3.5, 0, 1, 0, -1)$. Razmotrite primjer $(\mathbf{x}, y) = (\text{nounours}, 0)$. **Koliko iznosi gubitak modela na primjeru $(\mathbf{x}, y)$?**

A 0.456 B 0.795 C 0.359 D 0.552

19 (N) Za klasifikacijski problem koristimo stroj potpornih vektora s linearnom jezgrom i $C = 2$. Koristimo optimizaciju u dualu. Označen skup podataka i vektor dualnih koeficijenata jesu sljedeći:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} 15 & 30 & 15 & +1 \\ 6 & 12 & 6 & +1 \\ 11 & 32 & 21 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right) \qquad \boldsymbol{\alpha} = \begin{pmatrix} 0 \\ 2 \\ 0.077 \\ 1.923 \end{pmatrix}$$

Izračunajte parametre primarnog modela. **Koliko je primjer $\mathbf{x}^{(1)}$ udaljen od margine?**

A 5.25 B 3.23 C 0.44 D 4.09

20 (T) Koristimo algoritam SVM s jezgrenom funkcijom κ. Optimizaciju provodimo u dualu. Ako je κ Mercerova jezgra, onda takva jezgra implicitno definira preslikavanje $\boldsymbol{\phi}$. Međutim, u nekim slučajevima trebamo eksplicitno izraziti funkciju $\boldsymbol{\phi}$ na temelju korištene jezgre κ. **Kada je potrebno eksplicitno izraziti $\boldsymbol{\phi}$?**

A Za izračun težine w_0
C Za izračun širine margine
B Za izračun funkcije gubitka
D Za izračun dualnih koeficijenata

21 (T) Razmotrite problem maksimalne margine kao standardni problem optimizacije kriterijske funkcije $f(\mathbf{x})$ uz ograničenja nejednakosti $g_i(\mathbf{x}) \leq 0$ i jednakosti $h_i(\mathbf{x}) = 0$. Razmotrite taj problem za slučaj tvrde margine i za slučaj meke margine, oba u dualu. **Po čemu se dualni optimizacijski problem meke margine razlikuje od dualnog optimizacijskog problema tvrde margine?**

A Ima dodatnih n optimizacijskih varijabli i dodatnih N ograničenja jednakosti

B Ima dodatnih n ograničenja jednakosti

C Ima dodatnih N ograničenja nejednakosti

D Ima dodatnih N optimizacijskih varijabli i dodatnih $2N$ ograničenja nejednakosti

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

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

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

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija (6 pitanja)

1 (N) U dvodimenzijskome ulaznom prostoru postupkom najmanjih kvadrata (OLS) treniramo L2-regularizirani model linearne regresije. Koristimo funkciju preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_2^2)$. Označen skup podataka je:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rr|r} 1 & 5 & 1 \\ 2 & -2 & -5 \\ 3 & 1 & 2 \\ -5 & 6 & -2 \end{array}\right) \qquad \mathbf{G}^+ = \begin{pmatrix} 0.722 & -0.035 & 0.005 & -0.018 & -0.018 \\ -0.035 & 0.01 & 0 & 0.001 & 0.001 \\ 0.005 & 0 & 0.009 & 0 & -0.002 \\ -0.018 & 0.001 & 0 & 0.003 & -0.001 \\ -0.018 & 0.001 & -0.002 & -0.001 & 0.002 \end{pmatrix}$$

gdje je $\mathbf{G}^+$ pseudoinverz odgovarajuće Gramove matrice u prostoru značajki modificirane za regularizaciju uz $\lambda = 100$. **Koliko iznosi L_2-regularizirana pogreška naučenog modela?** (Napomena: matrica $\mathbf{G}^+$ dana je sa zaokruženim vrijednostima i koristite tu matricu.)

[A] 2.92 [B] 13.74 [C] 19.37 [D] 4.38

2 (P) Razmatramo binarni klasifikacijski problem sa $\mathcal{Y} = \{0, 1\}$ u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((0,0),1), ((2,2),1), ((0,1),0), ((0,-1),0), ((1,1),0)\big\}$. Razmatramo sljedeće modele, parametrizirane sa $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$:

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geq 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = h_1(\mathbf{x};\boldsymbol{\theta}_1) \cdot h_2(\mathbf{x};\boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \geq \theta_0^2\big\} \qquad \mathcal{H}_4 : h_4(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \min\big(1, h_1(\mathbf{x};\boldsymbol{\theta}_1) + h_2(\mathbf{x};\boldsymbol{\theta}_2)\big)$$

Neka je E_α minimalna empirijska pogreška modela $\mathcal{H}_\alpha$ na skupu $\mathcal{D}$, tj. $E_\alpha = \mathrm{argmin}_{h\in\mathcal{H}_\alpha} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

[A] $E_1 = E_3 > E_2 = E_4$ [B] $E_1 > E_4 > E_2 = E_3$ [C] $E_1 = E_2 = E_4 > E_3$ [D] $E_2 > E_1 = E_3 > E_4$

3 (T) Algoritam linearne regresije koristi kvadratni gubitak. **Koje je probabilističko opravdanje za korištenje kvadratnog gubitka?**

[A] Primjeri za koje je predviđanje hipoteze pogrešno pokoravaju se normalnoj razdiobi sa središtem u nuli

[B] Vjerojatnost skupa primjera najveća je za parametre koji minimiziraju očekivanje kvadratnog gubitka

[C] Minimizacija kvadratnog gubitka istovjetna je maksimizaciji negativnog logaritma vjerojatnosti parametara

[D] Vjerojatnost parametara za fiksni označeni skup podataka raste kvadratno s očekivanjem gubitka

4 (P) Koristimo regresiju za predviđanje uspjeha studija na temelju prosjeka ocjena u četiri razreda srednje škole te uspjeha iz matematike, fizike i informatike na državnoj maturi (ukupno 7 značajki). Za preslikavanje u prostor značajki koristimo polinom drugog stupnja s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki. Pretpostavite da nema multikolinearnosti. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

[A] 22 [B] 28 [C] 36 [D] 56

5 (P) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka su $\mathcal{H}_1$ i $\mathcal{H}_2$ dva modela različite složenosti te neka su $h_1^* \in \mathcal{H}_1$ i $h_2^* \in \mathcal{H}_2$ dvije hipoteze koje minimiziraju pogrešku $E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h)$ razlika između pogreške na ispitnom skupu i pogreške na skupu za učenje za hipotezu h, tj. $\Delta(h) = E(h|\mathcal{D}_{\text{test}}) - E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h_2^*) > \Delta(h_1^*)$. **Što možemo zaključiti o modelima $\mathcal{H}_1$ i $\mathcal{H}_2$?**

A Ako $E(h_1^*|\mathcal{D}_{\text{train}}) < E(h_2^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_1$ podnaučen

B Ako $E(h_1^*|\mathcal{D}_{\text{test}}) > E(h_2^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_1$ prenaučen

C Ako $E(h_2^*|\mathcal{D}_{\text{train}}) < E(h_1^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_2$ podnaučen

D Ako $E(h_2^*|\mathcal{D}_{\text{test}}) > E(h_1^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_2$ prenaučen

6 (T) Premda je linearna regresija linearan model, ona može modelirati nelinearne hipoteze. Nelinearnost možemo ostvariti uporabom prikladne funkcije preslikavanja $\boldsymbol{\phi}$ koja primjere iz ulaznog prostora preslikava u prostor značajki. **Koja pretpostavka stoji iza takvog postupka?**

A Primjeri koji se u prostoru značajki nalaze blizu nelinearne hiperpovršine u ulaznome prostoru leže na hiperavnini

B Hiperravnina koja minimizira pogrešku u ulaznome prostoru odgovara nelinearnoj hiperpovršini u prostoru značajki

C Primjeri koji su linearno odvojivi u ulaznome prostoru postaju linearno neodvojivi u prostoru značajki dovoljno visoke dimenzije

D Primjeri koji u ulaznome prostoru leže na nelinearnoj hiperpovršini u prostoru značajki leže na hiperravnini

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

7 (P) Algoritmom perceptrona rješavamo klasifikacijski problem sa $K = 10$ klasa. Koristimo pet značajki, $n = 5$. Za preslikavanje $\boldsymbol{\phi}$ koristimo polinom drugog stupnja sa svim interakcijskim parovima značajki. Koristimo shemu jedan-naspram-jedan (OVO). **Koliko iznosi ukupan broj parametara takvog višeklasnog klasifikatora?**

A 900 B 495 C 210 D 945

8 (T) Multinomijalna logistička regresija primjere klasificira u jednu od K klasa. Neka je $\mathbf{y}^{(i)} = (y_1^i, \dots, y_k^i, \dots, y_K^i)$ vektor indikatorskih varijabli koji odgovara oznaci primjera $\mathbf{x}^{(i)}$. Neka je $h_k(\mathbf{x}; \mathbf{W})$ vrijednost funkcije *softmax* za klasu k. Optimizacija parametara $\mathbf{W}$ odgovara maksimizaciji logaritma vjerojatnost oznaka. **Čemu je jednaka ta vjerojatnost?**

A $\prod_{i=1}^N \sum_{k=1}^K \ln h_k(\mathbf{x}^{(i)}; \mathbf{W})^{y_k^i}$ C $\sum_{i=1}^N \ln \prod_{k=1}^K h_k(\mathbf{x}^{(i)}; \mathbf{W})^{y_k^i}$

B $\ln \prod_{i=1}^N \prod_{k=1}^K y_k^i h_k(\mathbf{x}^{(i)}; \mathbf{W})$ D $\sum_{i=1}^N \sum_{k=1}^K \ln y_k^i h_k(\mathbf{x}^{(i)}; \mathbf{W})$

9 (N) Raspolažemo sljedećim skupom za učenje u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijski prostor značajki, definiranu kao $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$. Početne težine perceptrona su $\mathbf{w} = (1, 0, -1, 2, 3, 0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 16 B 22 C 32 D 6

10 (P) Algoritmom perceptrona rješavamo klasifikacijski problem. Oznake su iz skupa $\{-1, +1\}$. Za stopu učenja koristimo $\eta = \frac{1}{2}$. U jednoj iteraciji algoritma, kod ažuriranja težina na temelju primjera $\mathbf{x}$, vektor težina $\mathbf{w}$ mijenja se tako da mu se duljina prepolavlja, ali se zadržavaju smjer i orijentacija. Razmotite da, umjesto perceptrona, za klasifikaciju koristimo linearnu regresiju treniranu na oznakama $\{-1, +1\}$. **Koliko bi iznosio gubitak za isti primjer x kada bismo koristili linearnu regresiju s težinama w?**

A $1 - \|\mathbf{x}\|^2$ B $(\|\mathbf{w}\|^2 + 1)^2$ C $1 - y\|\mathbf{w}\|^2$ D $(y - \|\mathbf{x}\|^2)^2$

11 (N) Na skupu označenih primjera $\mathcal{D}$ trenirali smo model logističke regresije. Dobili smo neki vektor težina $\mathbf{w}$ i pomak $w_0 = 3.15$. Tako naučenom modelu neki primjer $\mathbf{x}$, čija je oznaka u skupu primjera $y = 0$, nanosi gubitak unakrsne entropije od $L(0, h(\mathbf{x})) = 0.5$. **Koliki gubitak unakrsne entropije bi nanosio primjer x kada bismo njegove značajke pomnožili sa dva i promijenili mu oznaku**?

A 2.54 B 1.19 C 7.11 D 4.03

12 (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 8.00 B 12.02 C 4.02 D 6.00

13 (T) Jedan od nedostataka algoritma perceptrona jest taj što se za linearno neodovojive probleme algoritam ne zaustavlja. **Zašto je tomu tako?**

A Površina funkcije pogreške u prostoru parametara po dijelovima je linearna i stoga nederivabilna

B Sve točke minimuma funkcije pogreške u prostoru parametara leže na istoj ravnini

C Svaki minimum funkcije pogreške u prostoru parametara veći je od nule

D Gradijent funkcije gubitka uvijek je različit od nul-vektora za barem jedan označeni primjer

14 (N) Model regularizirane logističke regresije treniramo stohastičkim gradijentnim spustom. Koristimo faktor regularizacije $\lambda = 1000$ i stopu učenja $\eta = 0.01$. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.2, 0.5, -1.1, 2.7)$. **Koliko u toj iteraciji iznosi promjena težine w_1 za primjer $(\mathbf{x}, y) = ((-1, 2), 1)$?**

A -5 B -2 C $+22$ D -12

Cjelina 3: SVM, jezgrene i neparametarske metode (8 pitanja)

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

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

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

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

16 (T) Koristimo algoritam SVM s jezgrenom funkcijom κ. Optimizaciju provodimo u dualu. Ako je κ Mercerova jezgra, onda takva jezgra implicitno definira preslikavanje $\boldsymbol{\phi}$. Međutim, u nekim slučajevima trebamo eksplicitno izraziti funkciju $\boldsymbol{\phi}$ na temelju korištene jezgre κ. **Kada je potrebno eksplicitno izraziti $\boldsymbol{\phi}$?**

A Za izračun širine margine

B Za izračun dualnih koeficijenata

C Za izračun težine w_0

D Za izračun funkcije gubitka

17 (T) Razmotrite problem maksimalne margine kao standardni problem optimizacije kriterijske funkcije $f(\mathbf{x})$ uz ograničenja nejednakosti $g_i(\mathbf{x}) \leq 0$ i jednakosti $h_i(\mathbf{x}) = 0$. Razmotrite taj problem za slučaj tvrde margine i za slučaj meke margine, oba u dualu. **Po čemu se dualni optimizacijski problem meke margine razlikuje od dualnog optimizacijskog problema tvrde margine?**

- A Ima dodatnih N ograničenja nejednakosti
- B Ima dodatnih N optimizacijskih varijabli i dodatnih $2N$ ograničenja nejednakosti
- C Ima dodatnih n optimizacijskih varijabli i dodatnih N ograničenja jednakosti
- D Ima dodatnih n ograničenja jednakosti

18 (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 $1/2$ B $1/4$ C 0 D $3/4$

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

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

Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(2,0)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina veća od stare?**

A $\frac{1}{2}\sqrt{10}$ puta B $\frac{5}{4}\sqrt{2}$ puta C $\frac{1}{2}\sqrt{17}$ puta D $\frac{1}{2}\sqrt{13}$ puta

20 (P) Razmatramo troklasni ($K = 3$) klasifikacijski problem sa $N = 100$ primjera u ulaznom prostoru dimenzije $n = 5$ (ne računajući značajku x_0). Razmatramo dva modela. Model $\mathcal{H}_1$ je rijetki jezgreni stroj koji za bazne funkcije koristi Gaussove jezgre. Model $\mathcal{H}_2$ je multinomijalna regresija s adaptivnim baznim funkcijama. Adaptivnih baznih funkcija ima ukupno deset, uključujući baznu funkciju $\phi_0(\mathbf{x}) = 1$. Za oba modela izračunajte njihov ukupan broj parametara prije učenja. **Koliko iznosi razlika u broju parametara između ova dva modela?**

A 725 B 219 C 719 D 819

21 (N) Rješavamo problem određivanja podrijetla pojedinih riječi u jeziku: za svaku riječ trebamo odrediti je li engleskog ($y = 1$) ili francuskog ($y = 0$) podrijetla. Problem rješavamo logističkom regresijom izvedenom kao rijetki jezgreni stroj, gdje za bazne funkcije koristimo jezgru κ nad znakovnim nizovima. Funkcija κ definirana je kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2| / |\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Na primjer, $\kappa(\text{water}, \text{eau}) = 2/6 = 0.33$. Skup za učenje je sljedeći:

$$\mathcal{D} = \{(\mathbf{x}, y)\}_i = \{(\text{water}, 1), (\text{eau}, 0), (\text{dog}, 1), (\text{chien}, 0), (\text{paperclip}, 1), (\text{trombone}), 0), (\text{chance}, 1), (\text{hasard}, 0)\}$$

Treniranjem rijetkoga jezgrenog stroja dobili smo vektor težina $\mathbf{w} = (-0.5, 0, 0, 0, -3.5, 1, 0, 0, 1)$. Razmotrite primjer $(\mathbf{x}, y) = (\text{nounours}, 0)$. **Koliko iznosi gubitak modela na primjeru $(\mathbf{x}, y)$?**

A 0.456 B 0.359 C 0.552 D 0.795

22 (N) Za klasifikacijski problem koristimo stroj potpornih vektora s linearnom jezgrom i $C = 2$. Koristimo optimizaciju u dualu. Označen skup podataka i vektor dualnih koeficijenata jesu sljedeći:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} 15 & 30 & 15 & +1 \\ 6 & 12 & 6 & +1 \\ 11 & 32 & 21 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right) \qquad \boldsymbol{\alpha} = \begin{pmatrix} 0 \\ 2 \\ 0.077 \\ 1.923 \end{pmatrix}$$

Izračunajte parametre primarnog modela. **Koliko je primjer $\mathbf{x}^{(1)}$ udaljen od margine?**

A 0.44 B 4.09 C 5.25 D 3.23

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija (6 pitanja)

1 (P) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka su $\mathcal{H}_1$ i $\mathcal{H}_2$ dva modela različite složenosti te neka su $h_1^* \in \mathcal{H}_1$ i $h_2^* \in \mathcal{H}_2$ dvije hipoteze koje minimiziraju pogrešku $E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h)$ razlika između pogreške na ispitnom skupu i pogreške na skupu za učenje za hipotezu h, tj. $\Delta(h) = E(h|\mathcal{D}_{\text{test}}) - E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h_2^*) < \Delta(h_1^*)$. **Što možemo zaključiti o modelima $\mathcal{H}_1$ i $\mathcal{H}_2$?**

- A Ako $E(h_1^*|\mathcal{D}_{\text{test}}) > E(h_2^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_1$ prenaučen
- B Ako $E(h_1^*|\mathcal{D}_{\text{train}}) < E(h_2^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_1$ podnaučen
- C Ako $E(h_2^*|\mathcal{D}_{\text{test}}) > E(h_1^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_2$ prenaučen
- D Ako $E(h_2^*|\mathcal{D}_{\text{train}}) < E(h_1^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_2$ podnaučen

2 (T) Premda je linearna regresija linearan model, ona može modelirati nelinearne hipoteze. Nelinearnost možemo ostvariti uporabom prikladne funkcije preslikavanja $\boldsymbol{\phi}$ koja primjere iz ulaznog prostora preslikava u prostor značajki. **Koja pretpostavka stoji iza takvog postupka?**

- A Hiperravnina koja minimizira pogrešku u ulaznome prostoru odgovara nelinearnoj hiperpovršini u prostoru značajki
- B Primjeri koji su linearno odvojivi u ulaznome prostoru postaju linearno neodvojivi u prostoru značajki dovoljno visoke dimenzije
- C Primjeri koji se u prostoru značajki nalaze blizu nelinearne hiperpovršine u ulaznome prostoru leže na hiperravnini
- D Primjeri koji u ulaznome prostoru leže na nelinearnoj hiperpovršini u prostoru značajki leže na hiperravnini

3 (P) Koristimo regresiju za predviđanje uspjeha studija na temelju prosjeka ocjena u četiri razreda srednje škole te uspjeha iz matematike, fizike i informatike na državnoj maturi (ukupno 7 značajki). Za preslikavanje u prostor značajki koristimo polinom drugog stupnja s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki. Pretpostavite da nema multikolinearnosti. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 28 B 56 C 36 D 22

4 (N) U dvodimenzijskome ulaznom prostoru postupkom najmanjih kvadrata (OLS) treniramo L2-regularizirani model linearne regresije. Koristimo funkciju preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_2^2)$. Označen skup podataka je:

$$
\mathbf{X}|\mathbf{y} = \left(\begin{array}{rr|r} 1 & 5 & 2 \\ 2 & -2 & -5 \\ 3 & 1 & 3 \\ -5 & 6 & -3 \end{array}\right) \qquad \mathbf{G}^+ = \begin{pmatrix} 0.722 & -0.035 & 0.005 & -0.018 & -0.018 \\ -0.035 & 0.01 & 0 & 0.001 & 0.001 \\ 0.005 & 0 & 0.009 & 0 & -0.002 \\ -0.018 & 0.001 & 0 & 0.003 & -0.001 \\ -0.018 & 0.001 & -0.002 & -0.001 & 0.002 \end{pmatrix}
$$

gdje je $\mathbf{G}^+$ pseudoinverz odgovarajuće Gramove matrice u prostoru značajki modificirane za regularizaciju uz $\lambda = 100$. **Koliko iznosi L_2-regularizirana pogreška naučenog modela?** (Napomena: matrica $\mathbf{G}^+$ dana je sa zaokruženim vrijednostima i koristite tu matricu.)

A 13.74 B 4.38 C 2.92 D 19.37

5 (T) Algoritam linearne regresije koristi kvadratni gubitak. **Koje je probabilističko opravdanje za korištenje kvadratnog gubitka?**

A Vjerojatnost parametara za fiksni označeni skup podataka raste kvadratno s očekivanjem gubitka

B Primjeri za koje je predviđanje hipoteze pogrešno pokoravaju se normalnoj razdiobi sa središtem u nuli

C Minimizacija kvadratnog gubitka istovjetna je maksimizaciji negativnog logaritma vjerojatnosti parametara

D Vjerojatnost skupa primjera najveća je za parametre koji minimiziraju očekivanje kvadratnog gubitka

6 (P) Razmatramo binarni klasifikacijski problem sa $\mathcal{Y} = \{0, 1\}$ u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((0,0),1), ((2,2),1), ((0,1),0), ((0,-1),0), ((1,1),0)\big\}$. Razmatramo sljedeće modele, parametrizirane sa $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$:

$$\mathcal{H}_1 : h_1(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geq 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = h_1(\mathbf{x}; \boldsymbol{\theta}_1) \cdot h_2(\mathbf{x}; \boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{1}\big\{(x_1 - \theta_1)^2 + (x_2 - \theta_2)^2 \geq \theta_0^2\big\} \qquad \mathcal{H}_4 : h_4(\mathbf{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \min\big(1, h_1(\mathbf{x}; \boldsymbol{\theta}_1) + h_2(\mathbf{x}; \boldsymbol{\theta}_2)\big)$$

Neka je E_α minimalna empirijska pogreška modela $\mathcal{H}_\alpha$ na skupu $\mathcal{D}$, tj. $E_\alpha = \mathrm{argmin}_{h \in \mathcal{H}_\alpha} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

A $E_1 = E_2 = E_4 > E_3$ B $E_1 = E_3 > E_2 = E_4$ C $E_1 > E_4 > E_2 = E_3$ D $E_2 > E_1 = E_3 > E_4$

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

7 (N) Raspolažemo sljedećim skupom za učenje u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijski prostor značajki, definiranu kao $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$. Početne težine perceptrona su $\mathbf{w} = (1, 0, -1, 2, -3, 0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 32 B 22 C 16 D 6

8 (N) Model regularizirane logističke regresije treniramo stohastičkim gradijentnim spustom. Koristimo faktor regularizacije $\lambda = 1000$ i stopu učenja $\eta = 0.01$. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.5, 1.2, -1.1, 2.7)$. **Koliko u toj iteraciji iznosi promjena težine w_1 za primjer $(\mathbf{x}, y) = ((-1, 2), 1)$?**

A $+22$ B -5 C -12 D -2

9 (T) Multinomijalna logistička regresija primjere klasificira u jednu od K klasa. Neka je $\mathbf{y}^{(i)} = (y_1^i, \ldots, y_k^i, \ldots, y_K^i)$ vektor indikatorskih varijabli koji odgovara oznaci primjera $\mathbf{x}^{(i)}$. Neka je $h_k(\mathbf{x}; \mathbf{W})$ vrijednost funkcije *softmax* za klasu k. Optimizacija parametara $\mathbf{W}$ odgovara maksimizaciji logaritma vjerojatnost oznaka. **Čemu je jednaka ta vjerojatnost?**

A $\prod_{i=1}^{N} \sum_{k=1}^{K} \ln h_k(\mathbf{x}^{(i)}; \mathbf{W})^{y_k^i}$ C $\ln \prod_{i=1}^{N} \prod_{k=1}^{K} y_k^i h_k(\mathbf{x}^{(i)}; \mathbf{W})$

B $\sum_{i=1}^{N} \sum_{k=1}^{K} \ln y_k^i h_k(\mathbf{x}^{(i)}; \mathbf{W})$ D $\sum_{i=1}^{N} \ln \prod_{k=1}^{K} h_k(\mathbf{x}^{(i)}; \mathbf{W})^{y_k^i}$

10 (T) Jedan od nedostataka algoritma perceptrona jest taj što se za linearno neodovojive probleme algoritam ne zaustavlja. **Zašto je tomu tako?**

A Sve točke minimuma funkcije pogreške u prostoru parametara leže na istoj ravnini

B Gradijent funkcije gubitka uvijek je različit od nul-vektora za barem jedan označeni primjer

C Svaki minimum funkcije pogreške u prostoru parametara veći je od nule

D Površina funkcije pogreške u prostoru parametara po dijelovima je linearna i stoga nederivabilna

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

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

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

A 4.02   B 8.00   C 6.00   D 12.02

12 (P) Algoritmom perceptrona rješavamo klasifikacijski problem. Oznake su iz skupa $\{-1, +1\}$. Za stopu učenja koristimo $\eta = \frac{1}{2}$. U jednoj iteraciji algoritma, kod ažuriranja težina na temelju primjera $\mathbf{x}$, vektor težina $\mathbf{w}$ mijenja se tako da mu se duljina prepolavlja, ali se zadržavaju smjer i orijentacija. Razmotite da, umjesto perceptrona, za klasifikaciju koristimo linearnu regresiju treniranu na oznakama $\{-1, +1\}$. **Koliko bi iznosio gubitak za isti primjer x kada bismo koristili linearnu regresiju s težinama w?**

A $1 - \|\mathbf{x}\|^2$ B $(y - \|\mathbf{x}\|^2)^2$ C $(\|\mathbf{w}\|^2 + 1)^2$ D $1 - y\|\mathbf{w}\|^2$

13 (P) Algoritmom perceptrona rješavamo klasifikacijski problem sa $K = 10$ klasa. Koristimo pet značajki, $n = 5$. Za preslikavanje $\boldsymbol{\phi}$ koristimo polinom drugog stupnja sa svim interakcijskim parovima značajki. Koristimo shemu jedan-naspram-jedan (OVO). **Koliko iznosi ukupan broj parametara takvog višeklasnog klasifikatora?**

A 900   B 945   C 495   D 210

14 (N) Na skupu označenih primjera $\mathcal{D}$ trenirali smo model logističke regresije. Dobili smo neki vektor težina $\mathbf{w}$ i pomak $w_0 = 3.15$. Tako naučenom modelu neki primjer $\mathbf{x}$, čija je oznaka u skupu primjera $y = 0$, nanosi gubitak unakrsne entropije od $L(0, h(\mathbf{x})) = 0.5$. **Koliki gubitak unakrsne entropije bi nanosio primjer x kada bismo njegove značajke pomnožili sa dva i promijenili mu oznaku**?

A 4.03   B 7.11   C 1.19   D 2.54

Cjelina 3: SVM, jezgrene i neparametarske metode (8 pitanja)

15 (T) Koristimo algoritam SVM s jezgrenom funkcijom κ. Optimizaciju provodimo u dualu. Ako je κ Mercerova jezgra, onda takva jezgra implicitno definira preslikavanje $\boldsymbol{\phi}$. Međutim, u nekim slučajevima trebamo eksplicitno izraziti funkciju $\boldsymbol{\phi}$ na temelju korištene jezgre κ. **Kada je potrebno eksplicitno izraziti $\boldsymbol{\phi}$?**

A Za izračun širine margine
B Za izračun funkcije gubitka
C Za izračun težine w_0
D Za izračun dualnih koeficijenata

16 (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 0   B 1/2   C 3/4   D 1/4

17 (T) Razmotrite problem maksimalne margine kao standardni problem optimizacije kriterijske funkcije $f(\mathbf{x})$ uz ograničenja nejednakosti $g_i(\mathbf{x}) \leq 0$ i jednakosti $h_i(\mathbf{x}) = 0$. Razmotrite taj problem za slučaj tvrde margine i za slučaj meke margine, oba u dualu. **Po čemu se dualni optimizacijski problem meke margine razlikuje od dualnog optimizacijskog problema tvrde margine?**

A Ima dodatnih N optimizacijskih varijabli i dodatnih $2N$ ograničenja nejednakosti

B Ima dodatnih n ograničenja jednakosti

C Ima dodatnih N ograničenja nejednakosti

D Ima dodatnih n optimizacijskih varijabli i dodatnih N ograničenja jednakosti

18 (P) Razmatramo troklasni ($K = 3$) klasifikacijski problem sa $N = 100$ primjera u ulaznom prostoru dimenzije $n = 5$ (ne računajući značajku x_0). Razmatramo dva modela. Model $\mathcal{H}_1$ je rijetki jezgreni stroj koji za bazne funkcije koristi Gaussove jezgre. Model $\mathcal{H}_2$ je multinomijalna regresija s adaptivnim baznim funkcijama. Adaptivnih baznih funkcija ima ukupno deset, uključujući baznu funkciju $\phi_0(\mathbf{x}) = 1$. Za oba modela izračunajte njihov ukupan broj parametara prije učenja. **Koliko iznosi razlika u broju parametara između ova dva modela?**

A 719 B 819 C 725 D 219

19 (N) Rješavamo problem određivanja podrijetla pojedinih riječi u jeziku: za svaku riječ trebamo odrediti je li engleskog ($y = 1$) ili francuskog ($y = 0$) podrijetla. Problem rješavamo logističkom regresijom izvedenom kao rijetki jezgreni stroj, gdje za bazne funkcije koristimo jezgru κ nad znakovnim nizovima. Funkcija κ definirana je kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Na primjer, $\kappa(\text{water}, \text{eau}) = 2/6 = 0.33$. Skup za učenje je sljedeći:

$$\mathcal{D} = \{(\mathbf{x}, y)\}_i = \big\{(\text{water}, 1), (\text{eau}, 0), (\text{dog}, 1), (\text{chien}, 0), (\text{paperclip}, 1), (\text{trombone}), 0), (\text{chance}, 1), (\text{hasard}, 0)\big\}$$

Treniranjem rijetkoga jezgrenog stroja dobili smo vektor težina $\mathbf{w} = (-0.5, 0, 0, -3.5, 0, -1, 0, 0, 1)$. Razmotrite primjer $(\mathbf{x}, y) = (\text{nounours}, 0)$. **Koliko iznosi gubitak modela na primjeru $(\mathbf{x}, y)$?**

A 0.552 B 0.456 C 0.795 D 0.359

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

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

Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(2, 0)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina veća od stare?**

A $\frac{1}{2}\sqrt{13}$ puta B $\frac{1}{2}\sqrt{17}$ puta C $\frac{5}{4}\sqrt{2}$ puta D $\frac{1}{2}\sqrt{10}$ puta

21 (N) Za klasifikacijski problem koristimo stroj potpornih vektora s linearnom jezgrom i $C = 2$. Koristimo optimizaciju u dualu. Označen skup podataka i vektor dualnih koeficijenata jesu sljedeći:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} 15 & 30 & 15 & +1 \\ 6 & 12 & 6 & +1 \\ 11 & 32 & 21 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right) \qquad \boldsymbol{\alpha} = \begin{pmatrix} 0 \\ 2 \\ 0.077 \\ 1.923 \end{pmatrix}$$

Izračunajte parametre primarnog modela. **Koliko je primjer $\mathbf{x}^{(1)}$ udaljen od margine?**

A 3.23 B 4.09 C 0.44 D 5.25

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

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

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

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

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 30% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija (6 pitanja)

1 (P) Razmatramo binarni klasifikacijski problem sa $\mathcal{Y} = \{0, 1\}$ u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((0,0),1),((2,2),1),((0,1),0),((0,-1),0),((1,1),0)\big\}$. Razmatramo sljedeće modele, parametrizirane sa $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$:

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geq 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = h_1(\mathbf{x};\boldsymbol{\theta}_1)\cdot h_2(\mathbf{x};\boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \geq \theta_0^2\big\} \qquad \mathcal{H}_4 : h_4(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \min\big(1, h_1(\mathbf{x};\boldsymbol{\theta}_1) + h_2(\mathbf{x};\boldsymbol{\theta}_2)\big)$$

Neka je E_α minimalna empirijska pogreška modela $\mathcal{H}_\alpha$ na skupu $\mathcal{D}$, tj. $E_\alpha = \mathrm{argmin}_{h\in\mathcal{H}_\alpha} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

[A] $E_2 > E_1 = E_3 > E_4$ [B] $E_1 = E_3 > E_2 = E_4$ [C] $E_1 > E_4 > E_2 = E_3$ [D] $E_1 = E_2 = E_4 > E_3$

2 (P) Koristimo regresiju za predviđanje uspjeha studija na temelju prosjeka ocjena u četiri razreda srednje škole te uspjeha iz matematike, fizike i informatike na državnoj maturi (ukupno 7 značajki). Za preslikavanje u prostor značajki koristimo polinom drugog stupnja s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki. Pretpostavite da nema multikolinearnosti. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

[A] 22 [B] 36 [C] 56 [D] 28

3 (P) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka su $\mathcal{H}_1$ i $\mathcal{H}_2$ dva modela različite složenosti te neka su $h_1^* \in \mathcal{H}_1$ i $h_2^* \in \mathcal{H}_2$ dvije hipoteze koje minimiziraju pogrešku $E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h)$ razlika između pogreške na ispitnom skupu i pogreške na skupu za učenje za hipotezu h, tj. $\Delta(h) = E(h|\mathcal{D}_{\text{test}}) - E(h|\mathcal{D}_{\text{train}})$. Neka je $\Delta(h_2^*) < \Delta(h_1^*)$. **Što možemo zaključiti o modelima $\mathcal{H}_1$ i $\mathcal{H}_2$?**

[A] Ako $E(h_2^*|\mathcal{D}_{\text{train}}) < E(h_1^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_2$ podnaučen

[B] Ako $E(h_1^*|\mathcal{D}_{\text{train}}) < E(h_2^*|\mathcal{D}_{\text{train}})$, onda je $\mathcal{H}_1$ podnaučen

[C] Ako $E(h_1^*|\mathcal{D}_{\text{test}}) > E(h_2^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_1$ prenaučen

[D] Ako $E(h_2^*|\mathcal{D}_{\text{test}}) > E(h_1^*|\mathcal{D}_{\text{test}})$, onda je $\mathcal{H}_2$ prenaučen

4 (N) U dvodimenzijskome ulaznom prostoru postupkom najmanjih kvadrata (OLS) treniramo L2-regularizirani model linearne regresije. Koristimo funkciju preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_2^2)$. Označen skup podataka je:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rr|r} 1 & 5 & 2 \\ 2 & -2 & -5 \\ 3 & 1 & 3 \\ -5 & 6 & -3 \end{array}\right) \qquad \mathbf{G}^{+} = \begin{pmatrix} 0.722 & -0.035 & 0.005 & -0.018 & -0.018 \\ -0.035 & 0.01 & 0 & 0.001 & 0.001 \\ 0.005 & 0 & 0.009 & 0 & -0.002 \\ -0.018 & 0.001 & 0 & 0.003 & -0.001 \\ -0.018 & 0.001 & -0.002 & -0.001 & 0.002 \end{pmatrix}$$

gdje je $\mathbf{G}^{+}$ pseudoinverz odgovarajuće Gramove matrice u prostoru značajki modificirane za regularizaciju uz $\lambda = 100$. **Koliko iznosi L_2-regularizirana pogreška naučenog modela?** (Napomena: matrica $\mathbf{G}^{+}$ dana je sa zaokruženim vrijednostima i koristite tu matricu.)

[A] 4.38 [B] 2.92 [C] 13.74 [D] 19.37

5 (T) Premda je linearna regresija linearan model, ona može modelirati nelinearne hipoteze. Nelinearnost možemo ostvariti uporabom prikladne funkcije preslikavanja $\boldsymbol{\phi}$ koja primjere iz ulaznog prostora preslikava u prostor značajki. **Koja pretpostavka stoji iza takvog postupka?**

- A Hiperravnina koja minimizira pogrešku u ulaznome prostoru odgovara nelinearnoj hiperpovršini u prostoru značajki
- B Primjeri koji su linearno odvojivi u ulaznome prostoru postaju linearno neodvojivi u prostoru značajki dovoljno visoke dimenzije
- C Primjeri koji u ulaznome prostoru leže na nelinearnoj hiperpovršini u prostoru značajki leže na hiperravnini
- D Primjeri koji se u prostoru značajki nalaze blizu nelinearne hiperpovršine u ulaznome prostoru leže na hiperravnini

6 (T) Algoritam linearne regresije koristi kvadratni gubitak. **Koje je probabilističko opravdanje za korištenje kvadratnog gubitka?**

- A Vjerojatnost skupa primjera najveća je za parametre koji minimiziraju očekivanje kvadratnog gubitka
- B Primjeri za koje je predviđanje hipoteze pogrešno pokoravaju se normalnoj razdiobi sa središtem u nuli
- C Minimizacija kvadratnog gubitka istovjetna je maksimizaciji negativnog logaritma vjerojatnosti parametara
- D Vjerojatnost parametara za fiksni označeni skup podataka raste kvadratno s očekivanjem gubitka

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

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

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

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

A 8.00 B 4.02 C 6.00 D 12.02

8 (N) Model regularizirane logističke regresije treniramo stohastičkim gradijentnim spustom. Koristimo faktor regularizacije $\lambda = 1000$ i stopu učenja $\eta = 0.01$. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.5, 0.2, -1.1, 2.7)$. **Koliko u toj iteraciji iznosi promjena težine w_1 za primjer $(\mathbf{x}, y) = ((-1, 2), 1)$?**

A -12 B -5 C -2 D $+22$

9 (N) Raspolažemo sljedećim skupom za učenje u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijski prostor značajki, definiranu kao $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)$. Početne težine perceptrona su $\mathbf{w} = (1, 0, -1, 2, -3, 0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 16 B 6 C 22 D 32

10 (T) Jedan od nedostataka algoritma perceptrona jest taj što se za linearno neodovojive probleme algoritam ne zaustavlja. **Zašto je tomu tako?**

- A Svaki minimum funkcije pogreške u prostoru parametara veći je od nule
- B Gradijent funkcije gubitka uvijek je različit od nul-vektora za barem jedan označeni primjer
- C Sve točke minimuma funkcije pogreške u prostoru parametara leže na istoj ravnini
- D Površina funkcije pogreške u prostoru parametara po dijelovima je linearna i stoga nederivabilna

11 (T) Multinomijalna logistička regresija primjere klasificira u jednu od K klasa. Neka je $\mathbf{y}^{(i)} = (y_1^i, \dots, y_k^i, \dots, y_K^i)$ vektor indikatorskih varijabli koji odgovara oznaci primjera $\mathbf{x}^{(i)}$. Neka je $h_k(\mathbf{x}; \mathbf{W})$ vrijednost funkcije *softmax* za klasu k. Optimizacija parametara $\mathbf{W}$ odgovara maksimizaciji logaritma vjerojatnost oznaka. **Čemu je jednaka ta vjerojatnost?**

A $\prod_{i=1}^N \sum_{k=1}^K \ln h_k(\mathbf{x}^{(i)}; \mathbf{W})^{y_k^i}$ C $\sum_{i=1}^N \sum_{k=1}^K \ln y_k^i h_k(\mathbf{x}^{(i)}; \mathbf{W})$
B $\sum_{i=1}^N \ln \prod_{k=1}^K h_k(\mathbf{x}^{(i)}; \mathbf{W})^{y_k^i}$ D $\ln \prod_{i=1}^N \prod_{k=1}^K y_k^i h_k(\mathbf{x}^{(i)}; \mathbf{W})$

12 (P) Algoritmom perceptrona rješavamo klasifikacijski problem. Oznake su iz skupa $\{-1, +1\}$. Za stopu učenja koristimo $\eta = \frac{1}{2}$. U jednoj iteraciji algoritma, kod ažuriranja težina na temelju primjera $\mathbf{x}$, vektor težina $\mathbf{w}$ mijenja se tako da mu se duljina prepolavlja, ali se zadržavaju smjer i orijentacija. Razmotite da, umjesto perceptrona, za klasifikaciju koristimo linearnu regresiju treniranu na oznakama $\{-1, +1\}$. **Koliko bi iznosio gubitak za isti primjer $\mathbf{x}$ kada bismo koristili linearnu regresiju s težinama $\mathbf{w}$?**

A $1 - \|\mathbf{x}\|^2$ B $(\|\mathbf{w}\|^2 + 1)^2$ C $(y - \|\mathbf{x}\|^2)^2$ D $1 - y\|\mathbf{w}\|^2$

13 (N) Na skupu označenih primjera $\mathcal{D}$ trenirali smo model logističke regresije. Dobili smo neki vektor težina $\mathbf{w}$ i pomak $w_0 = 3.15$. Tako naučenom modelu neki primjer $\mathbf{x}$, čija je oznaka u skupu primjera $y = 0$, nanosi gubitak unakrsne entropije od $L(0, h(\mathbf{x})) = 0.5$. **Koliki gubitak unakrsne entropije bi nanosio primjer $\mathbf{x}$ kada bismo njegove značajke pomnožili sa dva i promijenili mu oznaku**?

A 2.54 B 4.03 C 7.11 D 1.19

14 (P) Algoritmom perceptrona rješavamo klasifikacijski problem sa $K = 10$ klasa. Koristimo pet značajki, $n = 5$. Za preslikavanje $\boldsymbol{\phi}$ koristimo polinom drugog stupnja sa svim interakcijskim parovima značajki. Koristimo shemu jedan-naspram-jedan (OVO). **Koliko iznosi ukupan broj parametara takvog višeklasnog klasifikatora?**

A 945 B 495 C 210 D 900

Cjelina 3: SVM, jezgrene i neparametarske metode (8 pitanja)

15 (T) Razmotrite problem maksimalne margine kao standardni problem optimizacije kriterijske funkcije $f(\mathbf{x})$ uz ograničenja nejednakosti $g_i(\mathbf{x}) \leq 0$ i jednakosti $h_i(\mathbf{x}) = 0$. Razmotrite taj problem za slučaj tvrde margine i za slučaj meke margine, oba u dualu. **Po čemu se dualni optimizacijski problem meke margine razlikuje od dualnog optimizacijskog problema tvrde margine?**

A Ima dodatnih n optimizacijskih varijabli i dodatnih N ograničenja jednakosti

B Ima dodatnih N ograničenja nejednakosti

C Ima dodatnih n ograničenja jednakosti

D Ima dodatnih N optimizacijskih varijabli i dodatnih $2N$ ograničenja nejednakosti

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

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

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

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

17 (P) Razmatramo troklasni ($K = 3$) klasifikacijski problem sa $N = 100$ primjera u ulaznom prostoru dimenzije $n = 5$ (ne računajući značajku x_0). Razmatramo dva modela. Model $\mathcal{H}_1$ je rijetki jezgreni stroj koji za bazne funkcije koristi Gaussove jezgre. Model $\mathcal{H}_2$ je multinomijalna regresija s adaptivnim baznim funkcijama. Adaptivnih baznih funkcija ima ukupno deset, uključujući baznu funkciju $\phi_0(\mathbf{x}) = 1$. Za oba modela izračunajte njihov ukupan broj parametara prije učenja. **Koliko iznosi razlika u broju parametara između ova dva modela?**

A 725 B 719 C 819 D 219

18 (T) Koristimo algoritam SVM s jezgrenom funkcijom κ. Optimizaciju provodimo u dualu. Ako je κ Mercerova jezgra, onda takva jezgra implicitno definira preslikavanje $\boldsymbol{\phi}$. Međutim, u nekim slučajevima trebamo eksplicitno izraziti funkciju $\boldsymbol{\phi}$ na temelju korištene jezgre κ. **Kada je potrebno eksplicitno izraziti $\boldsymbol{\phi}$?**

A Za izračun težine w_0
B Za izračun dualnih koeficijenata
C Za izračun širine margine
D Za izračun funkcije gubitka

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

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

Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(2, 0)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina veća od stare?**

A $\frac{5}{4}\sqrt{2}$ puta B $\frac{1}{2}\sqrt{13}$ puta C $\frac{1}{2}\sqrt{10}$ puta D $\frac{1}{2}\sqrt{17}$ puta

20 (N) Za klasifikacijski problem koristimo stroj potpornih vektora s linearnom jezgrom i $C = 2$. Koristimo optimizaciju u dualu. Označen skup podataka i vektor dualnih koeficijenata jesu sljedeći:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} 25 & 70 & 45 & +1 \\ 6 & 12 & 6 & +1 \\ 11 & 32 & 21 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right) \qquad \boldsymbol{\alpha} = \begin{pmatrix} 0 \\ 2 \\ 0.077 \\ 1.923 \end{pmatrix}$$

Izračunajte parametre primarnog modela. **Koliko je primjer $\mathbf{x}^{(1)}$ udaljen od margine?**

A 4.09 B 3.23 C 0.44 D 5.25

21 (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 3/4 B 1/4 C 0 D 1/2

22 (N) Rješavamo problem određivanja podrijetla pojedinih riječi u jeziku: za svaku riječ trebamo odrediti je li engleskog ($y = 1$) ili francuskog ($y = 0$) podrijetla. Problem rješavamo logističkom regresijom izvedenom kao rijetki jezgreni stroj, gdje za bazne funkcije koristimo jezgru κ nad znakovnim nizovima. Funkcija κ definirana je kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Na primjer, $\kappa(\text{water}, \text{eau}) = 2/6 = 0.33$. Skup za učenje je sljedeći:

$$\mathcal{D} = \{(\mathbf{x}, y)\}_i = \{(\text{water}, 1), (\text{eau}, 0), (\text{dog}, 1), (\text{chien}, 0), (\text{paperclip}, 1), (\text{trombone}), 0), (\text{chance}, 1), (\text{hasard}, 0)\}$$

Treniranjem rijetkoga jezgrenog stroja dobili smo vektor težina $\mathbf{w} = (-0.5, 0, 0, 0, -3.5, 0, 1, 0, 1)$. Razmotrite primjer $(\mathbf{x}, y) = (\text{nounours}, 0)$. **Koliko iznosi gubitak modela na primjeru $(\mathbf{x}, y)$?**

A 0.359 B 0.456 C 0.795 D 0.552

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