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

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnove i regresija (6 pitanja)

1 (P) Optimizacija parametara modela temelji se na funkciji gubitka $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_0^+$, gdje je $L(y, h(\mathbf{x}))$ gubitak na primjeru $(\mathbf{x}, y)$. U većini primjena koristimo simetričan gubitak 0-1. Međutim, u nekim primjenama ima više smisla definirati asimetričan gubitak. Jedan takav primjer je zadatak detekcije neželjenih elektroničkih poruka (*spam filtering*). Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava da je poruka neželjena (spam), a $y = 0$ da ona to nije. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

- A $L(0,0) = L(1,1) < L(0,1) < L(1,0)$
- B $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$
- C $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) = 0$
- D $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$

2 (T) Hipoteza h je funkcija koja primjerima iz $\mathcal{X}$ pridjeljuje oznake iz $\mathcal{Y}$. Za h kažemo da je definirana “do na parametre $\boldsymbol{\theta}$”. **Što to znači?**

- A Funkcija h definirana je bez parametara, i njih treba odrediti naknadno postupkom odabira modela $\mathcal{H}$
- B Različite vrijednosti za $\boldsymbol{\theta}$ mogu dati različite funkcije h, a skup svih takvih različitih funkcija definira model $\mathcal{H}$
- C Funkcija h definira preslikavanja iz označenog skupa primjera $\mathcal{X} \times \mathcal{Y}$ u prostor parametara $\boldsymbol{\theta}$
- D Funkcija h jednoznačno određuje parametre $\boldsymbol{\theta}$ iz skupa svih mogućih parametara, koji nazivamo prostor parametara

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

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

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Odredite veličinu prostora inačica, $|VS_{\mathcal{H},\boldsymbol{\Phi}}|$, za ove kombinacije modela i funkcije preslikavanja. **Za koju kombinaciju modela i funkcije preslikavanja prostor inačica sadržava točno jednu hipotezu, $|VS_{\mathcal{H},\boldsymbol{\Phi}}| = 1$?**

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

4 (T) Optimizacijom regularizirane funkcije pogreške smanjuje se prenaučenost modela. **Kakve parametre modela nalazi optimizacija L_2-regularizirane pogreške?**

- A Parametre koji uz što veću magnitudu vrijednosti daju što manje očekivanje gubitka na skupu za ispitivanje
- B Parametre koji uz što veću magnitudu vrijednosti daju što veće očekivanje gubitka na skupu za učenje
- C Parametre koji uz što manju magnitudu vrijednosti daju što veće očekivanje gubitka na skupu za ispitivanje
- D Parametre koji uz što manju magnitudu vrijednosti daju što manje očekivanje gubitka na skupu za učenje

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

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

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

A 230.98 B 191.95 C 165.89 D 31.70

6 (P) Koristimo regresiju za predviđanje kreditne sposobnosti klijenata. Na raspolaganju imamo podatke o dobi x_1, ukupnom prinosu na račune x_2, ukupnoj ušteđevini u kunama x_3, ukupnoj deviznoj ušteđevini u eurima x_4, ukupnom iznosu odobrenih kredita x_5, ukupnom iznosu do sada otplaćenih kredita x_6 te preostalom ukupnom dugovanju po svim kreditima x_7. Ove podatke želimo iskoristiti kao značajke, ali pritom naravno moramo paziti da ne uvedemo multikolinearnost, pa možda nećemo upotrijebiti sve značajke. Za funkciju preslikavanja $\boldsymbol{\phi}$ koristimo preslikavanje s linearnim, kvadratnim i interakcijskim značajkama. Za interakcije razmatramo parove i trojke značajki, uključivo kombinacije kvadriranih i nekvadriranih značajki (npr., x_1x_2, $x_1^2x_2$ i $x_1x_2^2x_3^2$). **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 86 B 48 C 92 D 79

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

A 6.00 B 4.02 C 12.02 D 8.00

8 (T) Algoritam strojnog učenja idealno bi minimizirao gubitak 0-1. Međutim, funkciju gubitka 0-1 u praksi ne možemo koristiti za optimizaciju parametra modela. **Zašto gubitak 0-1 ne možemo koristiti za optimizaciju?**

A Gradijent gubitka 0-1 svugdje je nula osim za $h(\mathbf{x}) = 0$, pa funkcija pogreške ima zaravni po kojima se gradijentni spust ne može spuštati

B Funkcija gubitka 0-1 nije konveksna, pa ni funkcija pogreške nije konveksna već ima lokalne minimume te ne postoji minimizator u zatvorenoj formi

C Gubitak 0-1 pored neispravno klasificiranih primjera kažnjava i ispravno klasificirane primjere, i to tim više što su oni dalje od granice između klasa

D Funkcija gubitka 0-1 nije diferencijabilna, pa ne postoji rješenje u zatvorenoj formi i ne postoji gradijent

9 (P) Poopćeni linearni modeli mogu koristiti adaptivne bazne funkcije. Prednost toga je da ne moramo ručno definirati preslikavanje $\boldsymbol{\phi}$ u prostor značajki, već se to preslikavanje može naučiti na temelju podataka. Rasplažemo podatcima iz $K = 2$ klase u 100-dimenzijskome ulaznom prostoru. Za taj klasifikacijski problem koristimo binarnu logističku regresiju, ali s adaptivnim baznim funkcijama. Svaka adaptivna bazna funkcija ϕ_j je i sama jedan model logističke regresije, pri čemu $\phi_0(\mathbf{x}) = 1$. Naš je model onda definiran ovako:

$$h(\mathbf{x}) = \sigma\Big(\sum_{j=0}^{5} w_j \phi_j(\mathbf{x})\Big)$$

Koliko ukupno parametara ima ovaj model?

A 511 B 506 C 256 D 261

10 (P) Razmotrimo sljedeći skup označenih primjera:

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

Nad ovim skupom treniramo dva modela: neregulariziranu logističku regresiju (LR) i L_2-regulariziranu logističku regresiju (L2LR). Pored toga, razmatramo tri funkcije preslikavanja:

$$\begin{aligned}\boldsymbol{\phi}_{\mathbf{0}}(\mathbf{x}) &= (1, x_1, x_2)\\ \boldsymbol{\phi}_{\mathbf{1}}(\mathbf{x}) &= (1, x_1, x_2, x_1^2, x_2^2)\\ \boldsymbol{\phi}_{\mathbf{2}}(\mathbf{x}) &= (1, x_1, x_2, x_1 x_2)\end{aligned}$$

Ukupno, dakle, isprobavamo šest algoritama, odnosno šest kombinacija modela i funkcije preslikavanja. **Za koje algoritme (model+preslikavanje) optimizacijski postupak na skupu za učenje *ne* konvergira?**

A LR+$\boldsymbol{\phi}_{\mathbf{0}}$ B LR+$\boldsymbol{\phi}_{\mathbf{2}}$ C LR+$\boldsymbol{\phi}_{\mathbf{1}}$ D L2LR+$\boldsymbol{\phi}_{\mathbf{0}}$

11 (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 $\mathbf{w} = (0.5, 1.2, -1.1, 2.7)$. **Koliko u toj iteraciji iznosi promjena težine w_1 za primjer $(\mathbf{x}, y) = ((-1,2),1)$?**

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

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, koja je definirana na sljedeći način:

$$\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)$$

Početne težine perceptrona neka su sljedeće:

$$\mathbf{w} = (1, 0, 1, 2, -3, 0)$$

Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?

A 6   B 22   C 16   D 32

13 (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 1.19   B 7.11   C 2.54   D 4.03

14 (T) Poopćeni linearni modeli (linearna regresija, logistička regresija i multinomijalna regresija) probabilistički su algoritmi strojnog učenja. Njihova probabilistička priroda dolazi do izražaja kako kod modela tako i kod optimizacijskog postupka. **Koji je probabilistički princip ugrađen u optimizacijski postupak tih algoritama?**

A Maksimizirati $\sum_{i=1}^{N} \ln h(\mathbf{x}^{(i)}; \mathbf{w})$, gdje je $h(\mathbf{x}; \mathbf{w}) = f(\mathbf{w}^{\mathrm{T}}\mathbf{x})$

B Minimizirati $\sum_{i=1}^{N} \ln h(\mathbf{x}^{(i)}; \mathbf{w})$, gdje je $h(\mathbf{x}; \mathbf{w}) = f(\mathbf{w}^{\mathrm{T}}\mathbf{x})$

C Minimizirati $-\sum_{i=1}^{N} y^{(i)} p(y^{(i)}|\mathbf{x}^{(i)})$, gdje je $p(y^{(i)}|\mathbf{x}^{(i)}) = h(\mathbf{x}; \mathbf{w})$

D Maksimizirati $\sum_{i=1}^{N} \ln p(y^{(i)}|\mathbf{x}^{(i)})$, gdje je $\mathbb{E}[p(y^{(i)}|\mathbf{x}^{(i)})] = f(\mathbf{w}^{\mathrm{T}}\mathbf{x})$

Cjelina 3: Jezgrene i neparametarske metode (8 pitanja)

15 (P) Za binarnu klasifikaciju koristimo SVM s Gaussovom jezgrenom funkcijom. Razmatramo tri primjera u ulaznome prostoru: $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$ i $\mathbf{x}^{(3)}$. Euklidska udaljenosti između primjera $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(2)}$ manja je nego euklidska udaljenost između primjera $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(3)}$. Za izračun sličnosti između primjera koristimo Gaussovu jezgrenu funkciju $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ s preciznošću γ, gdje $\gamma = 1/2\sigma^2$. **Kakav bi efekt na sličnosti između primjera imalo smanjenje preciznosti γ Gaussove jezgrene funkcije?**

- A Sličnosti bi se povećale, ali $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se povećala više nego što bi se povećala $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$
- B Sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se smanjila, a sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$ bi se povećala, ali za veći iznos
- C Sličnosti bi se povećale, ali $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se povećala manje nego što bi se povećala $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$
- D Sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi postala manja od sličnosti $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$

16 (N) Rješavamo binarni klasifikacijski problem. Raspolažemo označenim skupom primjera. Odgovarajuća matrica dizajna je sljedeća:

$$\mathbf{X} = \begin{pmatrix} 1 & 3 & 16 & -8 & -11 \\ 1 & -5 & 4 & -8 & -7 \\ 1 & 7 & -4 & 11 & 9 \\ 1 & 15 & -20 & 25 & 25 \end{pmatrix}$$

Na ovom skupu treniramo model SVM-a s tvrdom marginom i linearnom jezgrenom funkcijom (tj. bez preslikavanja u prostor značajki). Model treniramo u primarnoj formulaciji. Za rješenje maksimalne margine dobili smo ovaj vektor težina (uključivo s težinom w_0):

$$\mathbf{w} = (-1.3214, +0.0714, +0.1071, 0, -0.0357)$$

Umjesto u primarnoj formulaciji, model smo mogli trenirati u dualnoj formulaciji, pa bismo umjesto vektora težina $\mathbf{w}$ dobili vektor dualnih parametara $\boldsymbol{\alpha}$, odnosno Lagrangeove multiplikatore. Prisjetite se da su vektori čiji su Lagrangeovi multiplikatori veći od nule potporni vektori. Premda to nije uvijek moguće, u ovom konkretnom slučaju dualni parametri modela mogu se izvesti iz rješenja primarnog modela. Izvedite vektor dualnih parametara $\boldsymbol{\alpha}$. **Koliko iznosi najveća vrijednost parametra u vektoru dualnih parametara $\boldsymbol{\alpha}$?** (Rezultate uspoređujte po prve tri decimale.)

A 0.0013 B 0.0089 C 0.0024 D 0.0045

17 (N) Pomoću SVM-a rješavamo problem binarne klasifikacije grafova. Budući da su primjeri $\mathbf{x}$ grafovi, koristimo SVM s jezgrenom funkcijom nad grafovima. Model treniramo na skupu od $N = 5$ označenih primjera, s vektorom oznaka jednakim $\mathbf{y} = (+1, +1, -1, -1, +1)$ i sa sljedećom jezgrenom matricom:

$$\mathbf{K} = \begin{pmatrix} 1.0 & 0.97 & -0.949 & -0.555 & -0.986 \\ 0.97 & 1.0 & -0.844 & -0.336 & -0.917 \\ -0.949 & -0.844 & 1.0 & 0.789 & 0.988 \\ -0.555 & -0.336 & 0.789 & 1.0 & 0.684 \\ -0.986 & -0.917 & 0.988 & 0.684 & 1.0 \end{pmatrix}$$

Treniranjem uz $C = 1$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (0, 0.754, 0.754, 1, 1)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za četvrti primjer, $L(y^{(4)}, h(\mathbf{x}^{(4)}))$?**

A 0.03 B 1.86 C 0.40 D 2.06

18 (N) Treniramo SVM s polinomijalnom jezgrom definiramo kao:

$$\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z})^2$$

Ova jezgra je Mercerova jezgra, što znači da postoji funkcija $\boldsymbol{\phi}$ takva da $\kappa(\mathbf{x}, \mathbf{z}) = \boldsymbol{\phi}(\mathbf{x})^T \boldsymbol{\phi}(\mathbf{z})$. Konkretno, u slučaju dvodimenzijskoga ulaznog prostora ($n = 2$), ova jezgra odgovara preslikavanju u šesterodimenzijski prostor. Međutim, postoji odstupanje u konkretnim koeficijentima polinoma. Razmotrite primjer $\mathbf{x} = (1, 0)$ te izračunajte $\boldsymbol{\phi}_\kappa(\mathbf{x})$, koji dobivamo preslikavanjem definiranim implicitno preko jezgre, te $\boldsymbol{\phi}_p(\mathbf{x})$, koji dobivamo preslikavanjem definiranim kao polinom drugog stupnja. **Koliko iznosi euklidska udaljenost između $\boldsymbol{\phi}_\kappa(\mathbf{x})$ i $\boldsymbol{\phi}_p(\mathbf{x})$?**

A $\sqrt{2}$ B $2\sqrt{2}$ C 0 D 4

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

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

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

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

20 (T) Problem prokletstva dimenzionalnosti (engl. *course of dimensionality*) pojavljuje se kod algoritama koji rade u visokodimenzijskome vektorskom prostoru i manifestira se na različite načine. **Kako se problem prokletstva dimenzionalnosti u visokodimenzijskim prostorima manifestira kod algoritma k-NN?**

A Broj susjeda k povećava se za svaki primjer te model postaje sve jednostavniji

B Udaljenost primjera do njegovih prvih k susjeda se smanjuje i svaki primjer postaje klasa za sebe

C Prvih k susjeda nekog primjera su vrlo udaljeni od dotičnog primjera

D Povećava se broj susjeda u okolini svakog primjera i model k-NN postaje sve jednostavniji

21 (P) Neka je $\mathcal{H}_{C,\gamma}$ model SVM-a s Gaussovom jezgrom. Hiperparametri tog modela su regularizacijski faktor C i preciznost jezgre γ. Odabir modela provodimo unakrsnom provjerom i to pretraživanjem po rešetci za sljedeće vrijednosti hiperparametara:

$$C = \{2^{-5}, 2^{-4}, \dots, 2^4, 2^5\}$$
$$\gamma = \{10^{-5}, 10^{-4}, \dots, 10^4, 10^5\}$$

Već ranije smo utvrdili da je model sa $C = 2^{-2}$ i $\gamma = 10^{-1}$ podnaučen, a da je model sa $C = 2^1$ i $\gamma = 10^1$ prenaučen. Za sve ostale modele ne znamo jesu li prenaučeni, podnaučeni ili optimalni. **Koliko modela još ima smisla ispitati jer su moguće optimalni?**

A 76 B 25 C 10 D 96

22 (T) Problem meke margine SVM-a s u primarnoj se formulaciji svodi na rješavanje sljedećeg optimizacijskog problema:

$$\underset{\mathbf{w},w_0,\boldsymbol{\xi}}{\operatorname{argmin}} \left\{\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i\right\}$$

uz određena linearna ograničenja. Ovaj optimizacijski problem odgovara optimizaciji regularizirane pogreške. Kod regularizirane pogreške u opreci su dva cilja: smanjenje vrijednosti funkcije gubitka i smanjenje složenosti modela. **Kako se ta opreka manifestira kod optimizacijskog problema meke margine SVM-a?**

A Što je manji hiperparametar C, to je manja važnost zbroja od ξ_i, pa je vrijednost $\|\mathbf{w}\|^2$ veća i model je veće složenosti

B Što je manja vrijednost $\|\mathbf{w}\|^2$, to je margina šira, no tim više primjera ulazi u marginu i to je veći zbroj od ξ_i

C Što je veći hiperparametar C, to se više kažnjava ulazak primjera u marginu, i to je manja vrijednost $\|\mathbf{w}\|^2$

D Što je veća vrijednost $\|\mathbf{w}\|^2$ to je model složeniji, no tim je veća nelinearnost granice i to je veći hiperparametar C

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

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnove i regresija (6 pitanja)

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

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

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Odredite veličinu prostora inačica, $|VS_{\mathcal{H},\mathbf{\Phi}}|$, za ove kombinacije modela i funkcije preslikavanja. **Za koju kombinaciju modela i funkcije preslikavanja prostor inačica sadržava točno jednu hipotezu, $|VS_{\mathcal{H},\mathbf{\Phi}}| = 1$?**

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

2 (T) Hipoteza h je funkcija koja primjerima iz $\mathcal{X}$ pridjeljuje oznake iz $\mathcal{Y}$. Za h kažemo da je definirana "do na parametre $\boldsymbol{\theta}$". **Što to znači?**

A Funkcija h jednoznačno određuje parametre $\boldsymbol{\theta}$ iz skupa svih mogućih parametara, koji nazivamo prostor parametara

B Funkcija h definirana je bez parametara, i njih treba odrediti naknadno postupkom odabira modela $\mathcal{H}$

C Funkcija h definira preslikavanja iz označenog skupa primjera $\mathcal{X} \times \mathcal{Y}$ u prostor parametara $\boldsymbol{\theta}$

D Svaka funkcija h ima svoje parametre $\boldsymbol{\theta}$ koji u potpunosti definiraju preslikavanje $\mathcal{X} \to \mathcal{Y}$

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

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

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

A 165.89 B 31.70 C 230.98 D 191.95

4 (P) Optimizacija parametara modela temelji se na funkciji gubitka $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_0^+$, gdje je $L(y, h(\mathbf{x}))$ gubitak na primjeru $(\mathbf{x}, y)$. U većini primjena koristimo simetričan gubitak 0-1. Međutim, u nekim primjenama ima više smisla definirati asimetričan gubitak. Jedan takav primjer je zadatak detekcije neželjenih elektroničkih poruka (*spam filtering*). Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava da je poruka neželjena (spam), a $y = 0$ da ona to nije. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

A $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) = 0$

B $L(0,0) = L(1,1) < L(0,1) < L(1,0)$

C $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$

D $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$

5 (T) Optimizacijom regularizirane funkcije pogreške smanjuje se prenaučenost modela. **Kakve parametre modela nalazi optimizacija L_2-regularizirane pogreške?**

- A Parametre koji uz što manju magnitudu vrijednosti daju što veće očekivanje gubitka na skupu za ispitivanje
- B Parametre koji uz što manju magnitudu vrijednosti daju što manje očekivanje gubitka na skupu za učenje
- C Parametre koji uz što veću magnitudu vrijednosti daju što manje očekivanje gubitka na skupu za ispitivanje
- D Parametre koji uz što veću magnitudu vrijednosti daju što veće očekivanje gubitka na skupu za učenje

6 (P) Koristimo regresiju za predviđanje kreditne sposobnosti klijenata. Na raspolaganju imamo podatke o dobi x_1, ukupnom prinosu na račune x_2, ukupnoj ušteđevini u kunama x_3, ukupnoj deviznoj ušteđevini u eurima x_4, ukupnom iznosu odobrenih kredita x_5, ukupnom iznosu do sada otplaćenih kredita x_6 te preostalom ukupnom dugovanju po svim kreditima x_7. Ove podatke želimo iskoristiti kao značajke, ali pritom naravno moramo paziti da ne uvedemo multikolinearnost, pa možda nećemo upotrijebiti sve značajke. Za funkciju preslikavanja $\boldsymbol{\phi}$ koristimo preslikavanje s linearnim, kvadratnim i interakcijskim značajkama. Za interakcije razmatramo parove i trojke značajki, uključivo kombinacije kvadriranih i nekvadriranih značajki (npr., x_1x_2, $x_1^2x_2$ i $x_1x_2^2x_3^2$). **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 48 B 92 C 86 D 79

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 $\mathbf{w} = (0.5, -2.2, -1.1, 2.7)$. **Koliko u toj iteraciji iznosi promjena težine w_1 za primjer $(\mathbf{x}, y) = ((-1, 2), 1)$?**

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

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 $\mathbf{y} = (0, 1, 0)$. **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

9 (P) Poopćeni linearni modeli mogu koristiti adaptivne bazne funkcije. Prednost toga je da ne moramo ručno definirati preslikavanje $\boldsymbol{\phi}$ u prostor značajki, već se to preslikavanje može naučiti na temelju podataka. Rasplažemo podatcima iz $K = 2$ klase u 50-dimenzijskome ulaznom prostoru. Za taj klasifikacijski problem koristimo binarnu logističku regresiju, ali s adaptivnim baznim funkcijama. Svaka adaptivna bazna funkcija ϕ_j je i sama jedan model logističke regresije, pri čemu $\phi_0(\mathbf{x}) = 1$. Naš je model onda definiran ovako:

$$h(\mathbf{x}) = \sigma\Big(\sum_{j=0}^{5} w_j \phi_j(\mathbf{x})\Big)$$

Koliko ukupno parametara ima ovaj model?

A 506 B 261 C 256 D 511

10 (P) Razmotrimo sljedeći skup označenih primjera:

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

Nad ovim skupom treniramo dva modela: neregulariziranu logističku regresiju (LR) i L_2-regulariziranu logističku regresiju (L2LR). Pored toga, razmatramo tri funkcije preslikavanja:

$$\boldsymbol{\phi}_\mathbf{0}(\mathbf{x}) = (1, x_1, x_2)$$
$$\boldsymbol{\phi}_\mathbf{1}(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_2^2)$$
$$\boldsymbol{\phi}_\mathbf{2}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$$

Ukupno, dakle, isprobavamo šest algoritama, odnosno šest kombinacija modela i funkcije preslikavanja. **Za koje algoritme (model+preslikavanje) optimizacijski postupak na skupu za učenje *ne* konvergira?**

A LR+$\boldsymbol{\phi}_\mathbf{2}$ B L2LR+$\boldsymbol{\phi}_\mathbf{2}$ C L2LR+$\boldsymbol{\phi}_\mathbf{0}$ D LR+$\boldsymbol{\phi}_\mathbf{1}$

11 (T) Algoritam strojnog učenja idealno bi minimizirao gubitak 0-1. Međutim, funkciju gubitka 0-1 u praksi ne možemo koristiti za optimizaciju parametra modela. **Zašto gubitak 0-1 ne možemo koristiti za optimizaciju?**

A Funkcija gubitka 0-1 nije diferencijabilna, pa ne postoji rješenje u zatvorenoj formi i ne postoji gradijent

B Funkcija gubitka 0-1 nije konveksna, pa ni funkcija pogreške nije konveksna već ima lokalne minimume te ne postoji minimizator u zatvorenoj formi

C Gubitak 0-1 pored neispravno klasificiranih primjera kažnjava i ispravno klasificirane primjere, i to tim više što su oni dalje od granice između klasa

D Gradijent gubitka 0-1 svugdje je nula osim za $h(\mathbf{x}) = 0$, pa funkcija pogreške ima zaravni po kojima se gradijentni spust ne može spuštati

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 1.19 B 4.03 C 2.54 D 7.11

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, koja je definirana na sljedeći način:

$$\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)$$

Početne težine perceptrona neka su sljedeće:

$$\mathbf{w} = (1, 0, 1, 2, -3, 0)$$

Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?

A 6 B 22 C 16 D 32

14 (T) Poopćeni linearni modeli (linearna regresija, logistička regresija i multinomijalna regresija) probabilistički su algoritmi strojnog učenja. Njihova probabilistička priroda dolazi do izražaja kako kod modela tako i kod optimizacijskog postupka. **Koji je probabilistički princip ugrađen u optimizacijski postupak tih algoritama?**

A Minimizirati $-\sum_{i=1}^N y^{(i)} p(y^{(i)}|\mathbf{x}^{(i)})$, gdje je $p(y^{(i)}|\mathbf{x}^{(i)}) = h(\mathbf{x};\mathbf{w})$

B Maksimizirati $\prod_{i=1}^N \ln p(y^{(i)}|\mathbf{x}^{(i)})$, gdje je $\mathbb{E}[p(y^{(i)}|\mathbf{x}^{(i)})] = \mathbf{w}^\mathrm{T}\mathbf{x}$

C Minimizirati $-\ln \prod_{i=1}^N p(y^{(i)}|\mathbf{x}^{(i)})$, gdje je $\mathbb{E}[p(y^{(i)}|\mathbf{x}^{(i)})] = f(\mathbf{w}^\mathrm{T}\mathbf{x})$

D Minimizirati $\sum_{i=1}^N \ln h(\mathbf{x}^{(i)};\mathbf{w})$, gdje je $h(\mathbf{x};\mathbf{w}) = f(\mathbf{w}^\mathrm{T}\mathbf{x})$

Cjelina 3: Jezgrene i neparametarske metode (8 pitanja)

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

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

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

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

16 (N) Pomoću SVM-a rješavamo problem binarne klasifikacije grafova. Budući da su primjeri $\mathbf{x}$ grafovi, koristimo SVM s jezgrenom funkcijom nad grafovima. Model treniramo na skupu od $N = 5$ označenih primjera, s vektorom oznaka jednakim $\mathbf{y} = (-1, -1, -1, +1, +1)$ i sa sljedećom jezgrenom matricom:

$$\mathbf{K} = \begin{pmatrix} 1.0 & 0.97 & -0.949 & -0.555 & -0.986 \\ 0.97 & 1.0 & -0.844 & -0.336 & -0.917 \\ -0.949 & -0.844 & 1.0 & 0.789 & 0.988 \\ -0.555 & -0.336 & 0.789 & 1.0 & 0.684 \\ -0.986 & -0.917 & 0.988 & 0.684 & 1.0 \end{pmatrix}$$

Treniranjem uz $C = 1$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (0, 0.733, 1, 1, 0.733)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za četvrti primjer, $L(y^{(4)}, h(\mathbf{x}^{(4)}))$?**

[A] 0.03 [B] 0.40 [C] 1.86 [D] 2.06

17 (T) Problem prokletstva dimenzionalnosti (engl. *course of dimensionality*) pojavljuje se kod algoritama koji rade u visokodimenzijskome vektorskom prostoru i manifestira se na različite načine. **Kako se problem prokletstva dimenzionalnosti u visokodimenzijskim prostorima manifestira kod algoritma k-NN?**

[A] Udaljenosti između primjera se smanjuju i model k-NN postaje sve složeniji

[B] Primjeri se grupiraju u središte vektorskog prostora i udaljenost postaje nediskriminativna

[C] Povećava se broj susjeda u okolini svakog primjera i model k-NN postaje sve jednostavniji

[D] Svi primjeri su međusobno vrlo udaljeni i gube se razlike u udaljenosti

18 (N) Treniramo SVM s polinomijalnom jezgrom definiramo kao:

$$\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z})^2$$

Ova jezgra je Mercerova jezgra, što znači da postoji funkcija $\boldsymbol{\phi}$ takva da $\kappa(\mathbf{x}, \mathbf{z}) = \boldsymbol{\phi}(\mathbf{x})^T\boldsymbol{\phi}(\mathbf{z})$. Konkretno, u slučaju dvodimenzijskoga ulaznog prostora ($n = 2$), ova jezgra odgovara preslikavanju u šesterodimenzijski prostor. Međutim, postoji odstupanje u konkretnim koeficijentima polinoma. Razmotrite primjer $\mathbf{x} = (1, 0)$ te izračunajte $\boldsymbol{\phi}_\kappa(\mathbf{x})$, koji dobivamo preslikavanjem definiranim implicitno preko jezgre, te $\boldsymbol{\phi}_p(\mathbf{x})$, koji dobivamo preslikavanjem definiranim kao polinom drugog stupnja. **Koliko iznosi euklidska udaljenost između $\boldsymbol{\phi}_\kappa(\mathbf{x})$ i $\boldsymbol{\phi}_p(\mathbf{x})$?**

[A] $2\sqrt{2}$ [B] $\sqrt{2}$ [C] 0 [D] 4

19 (P) Za binarnu klasifikaciju koristimo SVM s Gaussovom jezgrenom funkcijom. Razmatramo tri primjera u ulaznome prostoru: $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$ i $\mathbf{x}^{(3)}$. Euklidska udaljenosti između primjera $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(2)}$ manja je nego euklidska udaljenost između primjera $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(3)}$. Za izračun sličnosti između primjera koristimo Gaussovu jezgrenu funkciju $\kappa(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ s preciznošću γ, gdje $\gamma = 1/2\sigma^2$. **Kakav bi efekt na sličnosti između primjera imalo povećanje preciznosti γ Gaussove jezgrene funkcije?**

[A] Sličnosti bi se povećale, ali $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se povećala manje nego što bi se povećala $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$

[B] Sličnosti bi se smanjile, ali $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se smanjila više nego što bi se smanjila $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$

[C] Sličnosti bi se smanjile, ali $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se smanjila manje nego što bi se smanjila $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$

[D] Sličnosti bi se povećale, ali $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se povećala više nego što bi se povećala $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$

20 (P) Neka je $\mathcal{H}_{C,\gamma}$ model SVM-a s Gaussovom jezgrom. Hiperparametri tog modela su regularizacijski faktor C i preciznost jezgre γ. Odabir modela provodimo unakrsnom provjerom i to pretraživanjem po rešetci za sljedeće vrijednosti hiperparametara:

$$C = \{2^{-5}, 2^{-4}, \ldots, 2^4, 2^5\}$$
$$\gamma = \{10^{-5}, 10^{-4}, \ldots, 10^4, 10^5\}$$

Za model sa $C = 2^{-2}$ i $\gamma = 10^1$ utvrdili smo da je podnaučen, a za model sa $C = 2^1$ i $\gamma = 10^1$ utvrdili smo da je prenaučen. **Koliko modela još ima smisla ispitati jer su moguće optimalni?**

A 25 B 96 C 68 D 10

21 (N) Rješavamo binarni klasifikacijski problem. Raspolažemo označenim skupom primjera. Odgovarajuća matrica dizajna je sljedeća:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 22 & 0 & -5 \\ 1 & 6 & 28 & 8 & 2 \\ 1 & -6 & -15 & -9 & -6 \\ 1 & -11 & -34 & -16 & -9 \end{pmatrix}$$

Na ovom skupu treniramo model SVM-a s tvrdom marginom i linearnom jezgrenom funkcijom (tj. bez preslikavanja u prostor značajki). Model treniramo u primarnoj formulaciji. Za rješenje maksimalne margine dobili smo ovaj vektor težina (uključivo s težinom w_0):

$$\mathbf{w} = (+2.7838, +0.0225, +0.0856, +0.0315, +0.0135)$$

Umjesto u primarnoj formulaciji, model smo mogli trenirati u dualnoj formulaciji, pa bismo umjesto vektora težina $\mathbf{w}$ dobili vektor dualnih parametara $\boldsymbol{\alpha}$, odnosno Lagrangeove multiplikatore. Prisjetite se da su vektori čiji su Lagrangeovi multiplikatori veći od nule potporni vektori. Premda to nije uvijek moguće, u ovom konkretnom slučaju dualni parametri modela mogu se izvesti iz rješenja primarnog modela. Izvedite vektor dualnih parametara $\boldsymbol{\alpha}$. **Koliko iznosi najveća vrijednost parametra u vektoru dualnih parametara $\boldsymbol{\alpha}$?** (Rezultate uspoređujte po prve tri decimale.)

A 0.0089 B 0.0013 C 0.0024 D 0.0045

22 (T) Problem meke margine SVM-a s u primarnoj se formulaciji svodi na rješavanje sljedećeg optimizacijskog problema:

$$\underset{\mathbf{w},w_0,\boldsymbol{\xi}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N} \xi_i \right\}$$

uz određena linearna ograničenja. Ovaj optimizacijski problem odgovara optimizaciji regularizirane pogreške. Kod regularizirane pogreške u opreci su dva cilja: smanjenje vrijednosti funkcije gubitka i smanjenje složenosti modela. **Kako se ta opreka manifestira kod optimizacijskog problema meke margine SVM-a?**

A Što je manji zbroj od ξ_i, to više primjera može ući u marginu i tim je veća vrijednost $\|\mathbf{w}\|^2$ te je model manje složenosti

B Što je veći hiperparametar C, to se više kažnjava ulazak primjera u marginu, i to je manja vrijednost $\|\mathbf{w}\|^2$

C Što je veća vrijednost $\|\mathbf{w}\|^2$ to je model složeniji, no tim je veća nelinearnost granice i to je veći hiperparametar C

D Što je manja vrijednost $\|\mathbf{w}\|^2$, to je margina šira, no tim više primjera ulazi u marginu i to je veći zbroj od ξ_i

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

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnove i regresija (6 pitanja)

1 (T) Optimizacijom regularizirane funkcije pogreške smanjuje se prenaučenost modela. **Kakve parametre modela nalazi optimizacija L_2-regularizirane pogreške?**

A Parametre koji uz što veću magnitudu vrijednosti daju što veće očekivanje gubitka na skupu za učenje

B Parametre koji uz što veću magnitudu vrijednosti daju što manje očekivanje gubitka na skupu za ispitivanje

C Parametre koji uz što manju magnitudu vrijednosti daju što manje očekivanje gubitka na skupu za učenje

D Parametre koji uz što manju magnitudu vrijednosti daju što veće očekivanje gubitka na skupu za ispitivanje

2 (T) Hipoteza h je funkcija koja primjerima iz $\mathcal{X}$ pridjeljuje oznake iz $\mathcal{Y}$. Za h kažemo da je definirana "do na parametre $\boldsymbol{\theta}$". **Što to znači?**

A Funkcija h definirana je bez parametara, i njih treba odrediti naknadno postupkom odabira modela $\mathcal{H}$

B Svaka vrijednost parametara $\boldsymbol{\theta}$ daje jednu konkretnu funkciju h koja se razlikuje od svih drugih funkcija u modelu $\mathcal{H}$

C Funkcija h jednoznačno određuje parametre $\boldsymbol{\theta}$ iz skupa svih mogućih parametara, koji nazivamo prostor parametara

D Funkcija h izračunava oznaku $\mathcal{Y}$ na temelju parametra $\boldsymbol{\theta}$, koje treba podesiti optimizacijskim postupkom

3 (P) Koristimo regresiju za predviđanje kreditne sposobnosti klijenata. Na raspolaganju imamo podatke o dobi x_1, ukupnom prinosu na račune x_2, ukupnoj ušteđevini u kunama x_3, ukupnoj deviznoj ušteđevini u eurima x_4, ukupnom iznosu odobrenih kredita x_5, ukupnom iznosu do sada otplaćenih kredita x_6 te preostalom ukupnom dugovanju po svim kreditima x_7. Ove podatke želimo iskoristiti kao značajke, ali pritom naravno moramo paziti da ne uvedemo multikolinearnost, pa možda nećemo upotrijebiti sve značajke. Za funkciju preslikavanja $\boldsymbol{\phi}$ koristimo preslikavanje s linearnim, kvadratnim i interakcijskim značajkama. Za interakcije razmatramo parove i trojke značajki, uključivo kombinacije kvadriranih i nekvadriranih značajki (npr., x_1x_2, $x_1^2x_2$ i $x_1x_2^2x_3^2$). **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 48 B 86 C 92 D 79

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

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

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Odredite veličinu prostora inačica, $|VS_{\mathcal{H},\boldsymbol{\Phi}}|$, za ove kombinacije modela i funkcije preslikavanja. **Za koju kombinaciju modela i funkcije preslikavanja prostor inačica sadržava točno jednu hipotezu, $|VS_{\mathcal{H},\boldsymbol{\Phi}}| = 1$?**

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

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

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

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

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

6 (P) Optimizacija parametara modela temelji se na funkciji gubitka $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_0^+$, gdje je $L(y, h(\mathbf{x}))$ gubitak na primjeru $(\mathbf{x}, y)$. U većini primjena koristimo simetričan gubitak 0-1. Međutim, u nekim primjenama ima više smisla definirati asimetričan gubitak. Jedan takav primjer je zadatak detekcije neželjenih elektroničkih poruka (*spam filtering*). Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava da je poruka neželjena (spam), a $y = 0$ da ona to nije. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

[A] $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$

[B] $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$

[C] $L(0,0) = L(1,1) < L(0,1) < L(1,0)$

[D] $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) = 0$

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 $\mathbf{y} = (0, 0, 1)$. **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

8 (T) Algoritam strojnog učenja idealno bi minimizirao gubitak 0-1. Međutim, funkciju gubitka 0-1 u praksi ne možemo koristiti za optimizaciju parametra modela. **Zašto gubitak 0-1 ne možemo koristiti za optimizaciju?**

[A] Funkcija gubitka 0-1 nije diferencijabilna, pa ne postoji rješenje u zatvorenoj formi i ne postoji gradijent

[B] Gubitak 0-1 pored neispravno klasificiranih primjera kažnjava i ispravno klasificirane primjere, i to tim više što su oni dalje od granice između klasa

[C] Funkcija gubitka 0-1 nije konveksna, pa ni funkcija pogreške nije konveksna već ima lokalne minimume te ne postoji minimizator u zatvorenoj formi

[D] Gradijent gubitka 0-1 svugdje je nula osim za $h(\mathbf{x}) = 0$, pa funkcija pogreške ima zaravni po kojima se gradijentni spust ne može spuštati

9 (P) Poopćeni linearni modeli mogu koristiti adaptivne bazne funkcije. Prednost toga je da ne moramo ručno definirati preslikavanje $\boldsymbol{\phi}$ u prostor značajki, već se to preslikavanje može naučiti na temelju podataka. Rasplažemo podatcima iz $K = 2$ klase u 50-dimenzijskome ulaznom prostoru. Za taj klasifikacijski problem koristimo binarnu logističku regresiju, ali s adaptivnim baznim funkcijama. Svaka adaptivna bazna funkcija ϕ_j je i sama jedan model logističke regresije, pri čemu $\phi_0(\mathbf{x}) = 1$. Naš je model onda definiran ovako:

$$h(\mathbf{x}) = \sigma\Big(\sum_{j=0}^{5} w_j \phi_j(\mathbf{x})\Big)$$

Koliko ukupno parametara ima ovaj model?

[A] 506 [B] 511 [C] 261 [D] 256

10 (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 -12 B -5 C $+22$ D -2

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, koja je definirana na sljedeći način:

$$\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)$$

Početne težine perceptrona neka su sljedeće:

$$\mathbf{w} = (1, 0, -1, 2, -3, 0)$$

Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?

A 6 B 16 C 32 D 22

12 (T) Poopćeni linearni modeli (linearna regresija, logistička regresija i multinomijalna regresija) probabilistički su algoritmi strojnog učenja. Njihova probabilistička priroda dolazi do izražaja kako kod modela tako i kod optimizacijskog postupka. **Koji je probabilistički princip ugrađen u optimizacijski postupak tih algoritama?**

A Minimizirati $-\sum_{i=1}^{N} y^{(i)} p(y^{(i)}|\mathbf{x}^{(i)})$, gdje je $p(y^{(i)}|\mathbf{x}^{(i)}) = h(\mathbf{x}; \mathbf{w})$

B Minimizirati $\sum_{i=1}^{N} \ln h(\mathbf{x}^{(i)}; \mathbf{w})$, gdje je $h(\mathbf{x}; \mathbf{w}) = f(\mathbf{w}^{\mathrm{T}}\mathbf{x})$

C Minimizirati $-\sum_{i=1}^{N} \ln y^{(i)} h(\mathbf{x}^{(i)}; \mathbf{w})$, gdje je $h(\mathbf{x}; \mathbf{w}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$

D Minimizirati $-\ln \prod_{i=1}^{N} p(y^{(i)}|\mathbf{x}^{(i)})$, gdje je $\mathbb{E}[p(y^{(i)}|\mathbf{x}^{(i)})] = f(\mathbf{w}^{\mathrm{T}}\mathbf{x})$

13 (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 7.11 B 2.54 C 4.03 D 1.19

14 (P) Razmotrimo sljedeći skup označenih primjera:

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

Nad ovim skupom treniramo dva modela: neregulariziranu logističku regresiju (LR) i L_2-regulariziranu logističku regresiju (L2LR). Pored toga, razmatramo tri funkcije preslikavanja:

$$\begin{aligned}\boldsymbol{\phi_0}(\mathbf{x}) &= (1, x_1, x_2)\\ \boldsymbol{\phi_1}(\mathbf{x}) &= (1, x_1, x_2, x_1^2, x_2^2)\\ \boldsymbol{\phi_2}(\mathbf{x}) &= (1, x_1, x_2, x_1 x_2)\end{aligned}$$

Ukupno, dakle, isprobavamo šest algoritama, odnosno šest kombinacija modela i funkcije preslikavanja. **Za koje algoritme (model+preslikavanje) optimizacijski postupak na skupu za učenje *ne* konvergira?**

A L2LR+$\boldsymbol{\phi_2}$ B LR+$\boldsymbol{\phi_0}$ C LR+$\boldsymbol{\phi_1}$ D L2LR+$\boldsymbol{\phi_0}$

Cjelina 3: Jezgrene i neparametarske metode (8 pitanja)

15 (N) Treniramo SVM s polinomijalnom jezgrom definiramo kao:

$$\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z})^2$$

Ova jezgra je Mercerova jezgra, što znači da postoji funkcija $\boldsymbol{\phi}$ takva da $\kappa(\mathbf{x}, \mathbf{z}) = \boldsymbol{\phi}(\mathbf{x})^T \boldsymbol{\phi}(\mathbf{z})$. Konkretno, u slučaju dvodimenzijskoga ulaznog prostora ($n = 2$), ova jezgra odgovara preslikavanju u šesterodimenzijski prostor. Međutim, postoji odstupanje u konkretnim koeficijentima polinoma. Razmotrite primjer $\mathbf{x} = (1, 0)$ te izračunajte $\boldsymbol{\phi}_\kappa(\mathbf{x})$, koji dobivamo preslikavanjem definiranim implicitno preko jezgre, te $\boldsymbol{\phi}_p(\mathbf{x})$, koji dobivamo preslikavanjem definiranim kao polinom drugog stupnja. **Koliko iznosi euklidska udaljenost između $\boldsymbol{\phi}_\kappa(\mathbf{x})$ i $\boldsymbol{\phi}_p(\mathbf{x})$?**

A 4 B $2\sqrt{2}$ C 0 D $\sqrt{2}$

16 (N) Pomoću SVM-a rješavamo problem binarne klasifikacije grafova. Budući da su primjeri $\mathbf{x}$ grafovi, koristimo SVM s jezgrenom funkcijom nad grafovima. Model treniramo na skupu od $N = 5$ označenih primjera, s vektorom oznaka jednakim $\mathbf{y} = (-1, -1, -1, +1, +1)$ i sa sljedećom jezgrenom matricom:

$$\mathbf{K} = \begin{pmatrix} 1.0 & 0.97 & -0.949 & -0.555 & -0.986 \\ 0.97 & 1.0 & -0.844 & -0.336 & -0.917 \\ -0.949 & -0.844 & 1.0 & 0.789 & 0.988 \\ -0.555 & -0.336 & 0.789 & 1.0 & 0.684 \\ -0.986 & -0.917 & 0.988 & 0.684 & 1.0 \end{pmatrix}$$

Treniranjem uz $C = 1$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (0, 0.733, 1, 1, 0.733)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za četvrti primjer, $L(y^{(4)}, h(\mathbf{x}^{(4)}))$?**

A 0.03 B 2.06 C 1.86 D 0.40

17 (T) Problem meke margine SVM-a s u primarnoj se formulaciji svodi na rješavanje sljedećeg optimizacijskog problema:

$$\underset{\mathbf{w}, w_0, \boldsymbol{\xi}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{N} \xi_i \right\}$$

uz određena linearna ograničenja. Ovaj optimizacijski problem odgovara optimizaciji regularizirane pogreške. Kod regularizirane pogreške u opreci su dva cilja: smanjenje vrijednosti funkcije gubitka i smanjenje složenosti modela. **Kako se ta opreka manifestira kod optimizacijskog problema meke margine SVM-a?**

A Što je manji hiperparametar C, to je manja važnost zbroja od ξ_i, pa je vrijednost $\|\mathbf{w}\|^2$ veća i model je veće složenosti

B Što je veći zbroj od ξ_i, to više primjera ulazi u marginu i tim je manja vrijednost $\|\mathbf{w}\|^2$ te je model veće složenosti

C Što je manja vrijednost $\|\mathbf{w}\|^2$, to je margina šira, no tim više primjera ulazi u marginu i to je veći zbroj od ξ_i

D Što je veći hiperparametar C, to se više kažnjava ulazak primjera u marginu, i to je manja vrijednost $\|\mathbf{w}\|^2$

18 (T) Problem prokletstva dimenzionalnosti (engl. *course of dimensionality*) pojavljuje se kod algoritama koji rade u visokodimenzijskome vektorskom prostoru i manifestira se na različite načine. **Kako se problem prokletstva dimenzionalnosti u visokodimenzijskim prostorima manifestira kod algoritma k-NN?**

A Prvih k susjeda nekog primjera su vrlo udaljeni od dotičnog primjera

B Broj susjeda k nekog primjera se smanuje i gube se granice između klasa

C Primjeri se grupiraju u središte vektorskog prostora i udaljenost postaje nediskriminativna

D Udaljenosti između primjera se smanjuju i model k-NN postaje sve složeniji

19 (P) Neka je $\mathcal{H}_{C,\gamma}$ model SVM-a s Gaussovom jezgrom. Hiperparametri tog modela su regularizacijski faktor C i preciznost jezgre γ. Odabir modela provodimo unakrsnom provjerom i to pretraživanjem po rešetci za sljedeće vrijednosti hiperparametara:

$$C = \{2^{-5}, 2^{-4}, \dots, 2^4, 2^5\}$$
$$\gamma = \{10^{-5}, 10^{-4}, \dots, 10^4, 10^5\}$$

Za model sa $C = 2^{-2}$ i $\gamma = 10^1$ utvrdili smo da je podnaučen, a za model sa $C = 2^1$ i $\gamma = 10^1$ utvrdili smo da je prenaučen. **Koliko modela još ima smisla ispitati jer su moguće optimalni?**

A 68 B 96 C 25 D 10

20 (N) Rješavamo binarni klasifikacijski problem. Raspolažemo označenim skupom primjera. Odgovarajuća matrica dizajna je sljedeća:

$$\mathbf{X} = \begin{pmatrix} 1 & 3 & 16 & -8 & -11 \\ 1 & -5 & 4 & -8 & -7 \\ 1 & 7 & -4 & 11 & 9 \\ 1 & 15 & -20 & 25 & 25 \end{pmatrix}$$

Na ovom skupu treniramo model SVM-a s tvrdom marginom i linearnom jezgrenom funkcijom (tj. bez preslikavanja u prostor značajki). Model treniramo u primarnoj formulaciji. Za rješenje maksimalne margine dobili smo ovaj vektor težina (uključivo s težinom w_0):

$$\mathbf{w} = (+0.1370, -0.0290, +0.0194, -0.0461, -0.0388)$$

Umjesto u primarnoj formulaciji, model smo mogli trenirati u dualnoj formulaciji, pa bismo umjesto vektora težina $\mathbf{w}$ dobili vektor dualnih parametara $\boldsymbol{\alpha}$, odnosno Lagrangeove multiplikatore. Prisjetite se da su vektori čiji su Lagrangeovi multiplikatori veći od nule potporni vektori. Premda to nije uvijek moguće, u ovom konkretnom slučaju dualni parametri modela mogu se izvesti iz rješenja primarnog modela. Izvedite vektor dualnih parametara $\boldsymbol{\alpha}$. **Koliko iznosi najveća vrijednost parametra u vektoru dualnih parametara $\boldsymbol{\alpha}$?** (Rezultate uspoređujte po prve tri decimale.)

A 0.0045 B 0.0013 C 0.0024 D 0.0089

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

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

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

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

22 (P) Za binarnu klasifikaciju koristimo SVM s Gaussovom jezgrenom funkcijom. Razmatramo tri primjera u ulaznome prostoru: $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$ i $\mathbf{x}^{(3)}$. Euklidska udaljenosti između primjera $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(2)}$ manja je nego euklidska udaljenost između primjera $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(3)}$. Za izračun sličnosti između primjera koristimo Gaussovu jezgrenu funkciju $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ s preciznošću γ, gdje $\gamma = 1/2\sigma^2$. **Kakav bi efekt na sličnosti između primjera imalo smanjenje preciznosti γ Gaussove jezgrene funkcije?**

A Sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se povećala, a sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$ bi se smanjila, ali za veći iznos

B Sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se smanjila, a sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$ bi se povećala, ali za veći iznos

C Sličnosti bi se smanjile, ali $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se smanjila manje nego što bi se smanjila $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$

D Sličnosti bi se povećale, ali $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se povećala više nego što bi se povećala $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$

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

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnove i regresija (6 pitanja)

1 (T) Hipoteza h je funkcija koja primjerima iz $\mathcal{X}$ pridjeljuje oznake iz $\mathcal{Y}$. Za h kažemo da je definirana "do na parametre $\boldsymbol{\theta}$". **Što to znači?**

A Funkcija h definirana je bez parametara, i njih treba odrediti naknadno postupkom odabira modela $\mathcal{H}$

B Različite vrijednosti za $\boldsymbol{\theta}$ mogu dati različite funkcije h, a skup svih takvih različitih funkcija definira model $\mathcal{H}$

C Funkcija h jednoznačno određuje parametre $\boldsymbol{\theta}$ iz skupa svih mogućih parametara, koji nazivamo prostor parametara

D Svaka funkcija h definirana je uz pomoć parametara $\boldsymbol{\theta}$, pa različite vrijednosti za $\boldsymbol{\theta}$ daju različite funkcije h

2 (P) Optimizacija parametara modela temelji se na funkciji gubitka $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_0^+$, gdje je $L(y, h(\mathbf{x}))$ gubitak na primjeru $(\mathbf{x}, y)$. U većini primjena koristimo simetričan gubitak 0-1. Međutim, u nekim primjenama ima više smisla definirati asimetričan gubitak. Jedan takav primjer je zadatak detekcije neželjenih elektroničkih poruka (*spam filtering*). Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava da je poruka neželjena (spam), a $y = 0$ da ona to nije. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

A $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$

B $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$

C $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) = 0$

D $L(0,0) = L(1,1) < L(0,1) < L(1,0)$

3 (T) Optimizacijom regularizirane funkcije pogreške smanjuje se prenaučenost modela. **Kakve parametre modela nalazi optimizacija L_2-regularizirane pogreške?**

A Parametre koji uz što veću magnitudu vrijednosti daju što veće očekivanje gubitka na skupu za učenje

B Parametre koji uz što manju magnitudu vrijednosti daju što veće očekivanje gubitka na skupu za ispitivanje

C Parametre koji uz što manju magnitudu vrijednosti daju što manje očekivanje gubitka na skupu za učenje

D Parametre koji uz što veću magnitudu vrijednosti daju što manje očekivanje gubitka na skupu za ispitivanje

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

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

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

A 230.98 B 31.70 C 191.95 D 165.89

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

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

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Odredite veličinu prostora inačica, $|VS_{\mathcal{H},\mathbf{\Phi}}|$, za ove kombinacije modela i funkcije preslikavanja. **Za koju kombinaciju modela i funkcije preslikavanja prostor inačica sadržava točno jednu hipotezu, $|VS_{\mathcal{H},\mathbf{\Phi}}| = 1$?**

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

6 (P) Koristimo regresiju za predviđanje kreditne sposobnosti klijenata. Na raspolaganju imamo podatke o dobi x_1, ukupnom prinosu na račune x_2, ukupnoj ušteđevini u kunama x_3, ukupnoj deviznoj ušteđevini u eurima x_4, ukupnom iznosu odobrenih kredita x_5, ukupnom iznosu do sada otplaćenih kredita x_6 te preostalom ukupnom dugovanju po svim kreditima x_7. Ove podatke želimo iskoristiti kao značajke, ali pritom naravno moramo paziti da ne uvedemo multikolinearnost, pa možda nećemo upotrijebiti sve značajke. Za funkciju preslikavanja $\boldsymbol{\phi}$ koristimo preslikavanje s linearnim, kvadratnim i interakcijskim značajkama. Za interakcije razmatramo parove i trojke značajki, uključivo kombinacije kvadriranih i nekvadriranih značajki (npr., x_1x_2, $x_1^2x_2$ i $x_1x_2^2x_3^2$). **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

[A] 92 [B] 86 [C] 79 [D] 48

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

7 (T) Poopćeni linearni modeli (linearna regresija, logistička regresija i multinomijalna regresija) probabilistički su algoritmi strojnog učenja. Njihova probabilistička priroda dolazi do izražaja kako kod modela tako i kod optimizacijskog postupka. **Koji je probabilistički princip ugrađen u optimizacijski postupak tih algoritama?**

[A] Maksimizirati $\sum_{i=1}^{N} \ln h(\mathbf{x}^{(i)}; \mathbf{w})$, gdje je $h(\mathbf{x};\mathbf{w}) = f(\mathbf{w}^{\mathrm{T}}\mathbf{x})$

[B] Minimizirati $-\sum_{i=1}^{N} \ln y^{(i)} h(\mathbf{x}^{(i)}; \mathbf{w})$, gdje je $h(\mathbf{x};\mathbf{w}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$

[C] Maksimizirati $\sum_{i=1}^{N} \ln p(y^{(i)}|\mathbf{x}^{(i)})$, gdje je $\mathbb{E}[p(y^{(i)}|\mathbf{x}^{(i)})] = f(\mathbf{w}^{\mathrm{T}}\mathbf{x})$

[D] Minimizirati $-\sum_{i=1}^{N} y^{(i)} p(y^{(i)}|\mathbf{x}^{(i)})$, gdje je $p(y^{(i)}|\mathbf{x}^{(i)}) = h(\mathbf{x};\mathbf{w})$

8 (P) Razmotrimo sljedeći skup označenih primjera:

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

Nad ovim skupom treniramo dva modela: neregulariziranu logističku regresiju (LR) i L_2-regulariziranu logističku regresiju (L2LR). Pored toga, razmatramo tri funkcije preslikavanja:

$$\boldsymbol{\phi}_{\mathbf{0}}(\mathbf{x}) = (1, x_1, x_2)$$
$$\boldsymbol{\phi}_{\mathbf{1}}(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_2^2)$$
$$\boldsymbol{\phi}_{\mathbf{2}}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$$

Ukupno, dakle, isprobavamo šest algoritama, odnosno šest kombinacija modela i funkcije preslikavanja. **Za koje algoritme (model+preslikavanje) optimizacijski postupak na skupu za učenje *ne* konvergira?**

[A] L2LR+$\boldsymbol{\phi}_{\mathbf{1}}$ [B] L2LR+$\boldsymbol{\phi}_{\mathbf{2}}$ [C] LR+$\boldsymbol{\phi}_{\mathbf{2}}$ [D] LR+$\boldsymbol{\phi}_{\mathbf{1}}$

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] -5 [B] $+22$ [C] -12 [D] -2

10 (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 1.19 B 7.11 C 4.03 D 2.54

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

A 8.00 B 12.02 C 6.00 D 4.02

12 (T) Algoritam strojnog učenja idealno bi minimizirao gubitak 0-1. Međutim, funkciju gubitka 0-1 u praksi ne možemo koristiti za optimizaciju parametra modela. **Zašto gubitak 0-1 ne možemo koristiti za optimizaciju?**

A Funkcija gubitka 0-1 nije diferencijabilna, pa ne postoji rješenje u zatvorenoj formi i ne postoji gradijent

B Funkcija gubitka 0-1 nije konveksna, pa ni funkcija pogreške nije konveksna već ima lokalne minimume te ne postoji minimizator u zatvorenoj formi

C Gubitak 0-1 pored neispravno klasificiranih primjera kažnjava i ispravno klasificirane primjere, i to tim više što su oni dalje od granice između klasa

D Gradijent gubitka 0-1 svugdje je nula osim za $h(\mathbf{x}) = 0$, pa funkcija pogreške ima zaravni po kojima se gradijentni spust ne može spuštati

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, koja je definirana na sljedeći način:

$$\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)$$

Početne težine perceptrona neka su sljedeće:

$$\mathbf{w} = (1, 0, -1, 2, -3, 0)$$

Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?

A 6 B 22 C 16 D 32

14 (P) Poopćeni linearni modeli mogu koristiti adaptivne bazne funkcije. Prednost toga je da ne moramo ručno definirati preslikavanje $\boldsymbol{\phi}$ u prostor značajki, već se to preslikavanje može naučiti na temelju podataka. Rasplažemo podatcima iz $K = 2$ klase u 100-dimenzijskome ulaznom prostoru. Za taj klasifikacijski problem koristimo binarnu logističku regresiju, ali s adaptivnim baznim funkcijama. Svaka adaptivna bazna funkcija ϕ_j je i sama jedan model logističke regresije, pri čemu $\phi_0(\mathbf{x}) = 1$. Naš je model onda definiran ovako:

$$h(\mathbf{x}) = \sigma\Big(\sum_{j=0}^{5} w_j \phi_j(\mathbf{x})\Big)$$

Koliko ukupno parametara ima ovaj model?

A 511 B 506 C 256 D 261

Cjelina 3: Jezgrene i neparametarske metode (8 pitanja)

15 (N) Pomoću SVM-a rješavamo problem binarne klasifikacije grafova. Budući da su primjeri $\mathbf{x}$ grafovi, koristimo SVM s jezgrenom funkcijom nad grafovima. Model treniramo na skupu od $N = 5$ označenih primjera, s vektorom oznaka jednakim $\mathbf{y} = (+1, +1, -1, -1, +1)$ i sa sljedećom jezgrenom matricom:

$$\mathbf{K} = \begin{pmatrix} 1.0 & 0.97 & -0.949 & -0.555 & -0.986 \\ 0.97 & 1.0 & -0.844 & -0.336 & -0.917 \\ -0.949 & -0.844 & 1.0 & 0.789 & 0.988 \\ -0.555 & -0.336 & 0.789 & 1.0 & 0.684 \\ -0.986 & -0.917 & 0.988 & 0.684 & 1.0 \end{pmatrix}$$

Treniranjem uz $C = 1$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (0, 0.754, 0.754, 1, 1)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za četvrti primjer, $L(y^{(4)}, h(\mathbf{x}^{(4)}))$?**

A 2.06 B 1.86 C 0.03 D 0.40

16 (N) Treniramo SVM s polinomijalnom jezgrom definiramo kao:

$$\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z})^2$$

Ova jezgra je Mercerova jezgra, što znači da postoji funkcija $\boldsymbol{\phi}$ takva da $\kappa(\mathbf{x}, \mathbf{z}) = \boldsymbol{\phi}(\mathbf{x})^T\boldsymbol{\phi}(\mathbf{z})$. Konkretno, u slučaju dvodimenzijskoga ulaznog prostora ($n = 2$), ova jezgra odgovara preslikavanju u šesterodimenzijski prostor. Međutim, postoji odstupanje u konkretnim koeficijentima polinoma. Razmotrite primjer $\mathbf{x} = (1, 0)$ te izračunajte $\boldsymbol{\phi}_\kappa(\mathbf{x})$, koji dobivamo preslikavanjem definiranim implicitno preko jezgre, te $\boldsymbol{\phi}_p(\mathbf{x})$, koji dobivamo preslikavanjem definiranim kao polinom drugog stupnja. **Koliko iznosi euklidska udaljenost između $\boldsymbol{\phi}_\kappa(\mathbf{x})$ i $\boldsymbol{\phi}_p(\mathbf{x})$?**

A $\sqrt{2}$ B $2\sqrt{2}$ C 0 D 4

17 (T) Problem prokletstva dimenzionalnosti (engl. *course of dimensionality*) pojavljuje se kod algoritama koji rade u visokodimenzijskome vektorskom prostoru i manifestira se na različite načine. **Kako se problem prokletstva dimenzionalnosti u visokodimenzijskim prostorima manifestira kod algoritma k-NN?**

A Prvih k susjeda nekog primjera su vrlo udaljeni od dotičnog primjera

B Udaljenost primjera do njegovih prvih k susjeda se smanjuje i svaki primjer postaje klasa za sebe

C Udaljenosti između primjera se smanjuju i model k-NN postaje sve složeniji

D Povećava se broj susjeda u okolini svakog primjera i model k-NN postaje sve jednostavniji

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

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

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

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

19 (T) Problem meke margine SVM-a s u primarnoj se formulaciji svodi na rješavanje sljedećeg optimizacijskog problema:

$$\underset{\mathbf{w}, w_0, \boldsymbol{\xi}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N} \xi_i \right\}$$

uz određena linearna ograničenja. Ovaj optimizacijski problem odgovara optimizaciji regularizirane pogreške. Kod regularizirane pogreške u opreci su dva cilja: smanjenje vrijednosti funkcije gubitka i smanjenje složenosti modela.

Kako se ta opreka manifestira kod optimizacijskog problema meke margine SVM-a?

A Što je veća vrijednost $\|\mathbf{w}\|^2$ to je model složeniji, no tim je veća nelinearnost granice i to je veći hiperparametar C

B Što je veći hiperparametar C, to se više kažnjava ulazak primjera u marginu, i to je manja vrijednost $\|\mathbf{w}\|^2$

C Što je manji hiperparametar C, to je manja važnost zbroja od ξ_i, pa je vrijednost $\|\mathbf{w}\|^2$ veća i model je veće složenosti

D Što je veća vrijednost $\|\mathbf{w}\|^2$, to je margina uža i tim manje primjera ulazi u marginu, pa je tim manji zbroj od ξ_i

20 (P) Za binarnu klasifikaciju koristimo SVM s Gaussovom jezgrenom funkcijom. Razmatramo tri primjera u ulaznome prostoru: $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$ i $\mathbf{x}^{(3)}$. Euklidska udaljenosti između primjera $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(2)}$ manja je nego euklidska udaljenost između primjera $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(3)}$. Za izračun sličnosti između primjera koristimo Gaussovu jezgrenu funkciju $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ s preciznošću γ, gdje $\gamma = 1/2\sigma^2$. **Kakav bi efekt na sličnosti između primjera imalo smanjenje preciznosti γ Gaussove jezgrene funkcije?**

A Sličnosti bi se povećale, ali $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se povećala manje nego što bi se povećala $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$

B Sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se povećala, a sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$ bi se smanjila, ali za manji iznos

C Sličnosti bi se povećale, ali $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se povećala više nego što bi se povećala $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$

D Sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se smanjila, a sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$ bi se povećala, ali za manji iznos

21 (P) Neka je $\mathcal{H}_{C,\gamma}$ model SVM-a s Gaussovom jezgrom. Hiperparametri tog modela su regularizacijski faktor C i preciznost jezgre γ. Odabir modela provodimo unakrsnom provjerom i to pretraživanjem po rešetci za sljedeće vrijednosti hiperparametara:

$$C = \{2^{-5}, 2^{-4}, \dots, 2^4, 2^5\}$$
$$\gamma = \{10^{-5}, 10^{-4}, \dots, 10^4, 10^5\}$$

Za model sa $C = 2^{-2}$ i $\gamma = 10^1$ utvrdili smo da je podnaučen, a za model sa $C = 2^1$ i $\gamma = 10^1$ utvrdili smo da je prenaučen. **Koliko modela još ima smisla ispitati jer su moguće optimalni?**

A 25 B 10 C 96 D 68

22 (N) Rješavamo binarni klasifikacijski problem. Raspolažemo označenim skupom primjera. Odgovarajuća matrica dizajna je sljedeća:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 22 & 0 & -5 \\ 1 & 6 & 28 & 8 & 2 \\ 1 & -6 & -15 & -9 & -6 \\ 1 & -11 & -34 & -16 & -9 \end{pmatrix}$$

Na ovom skupu treniramo model SVM-a s tvrdom marginom i linearnom jezgrenom funkcijom (tj. bez preslikavanja u prostor značajki). Model treniramo u primarnoj formulaciji. Za rješenje maksimalne margine dobili smo ovaj vektor težina (uključivo s težinom w_0):

$$\mathbf{w} = (+2.7838, +0.0225, +0.0856, +0.0315, +0.0135)$$

Umjesto u primarnoj formulaciji, model smo mogli trenirati u dualnoj formulaciji, pa bismo umjesto vektora težina $\mathbf{w}$ dobili vektor dualnih parametara $\boldsymbol{\alpha}$, odnosno Lagrangeove multiplikatore. Prisjetite se da su vektori čiji su Lagrangeovi multiplikatori veći od nule potporni vektori. Premda to nije uvijek moguće, u ovom konkretnom slučaju dualni parametri modela mogu se izvesti iz rješenja primarnog modela. Izvedite vektor dualnih parametara $\boldsymbol{\alpha}$. **Koliko iznosi najveća vrijednost parametra u vektoru dualnih parametara $\boldsymbol{\alpha}$?** (Rezultate uspoređujte po prve tri decimale.)

A 0.0089 B 0.0024 C 0.0045 D 0.0013

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

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnove i regresija (6 pitanja)

1 (P) Koristimo regresiju za predviđanje kreditne sposobnosti klijenata. Na raspolaganju imamo podatke o dobi x_1, ukupnom prinosu na račune x_2, ukupnoj ušteđevini u kunama x_3, ukupnoj deviznoj ušteđevini u eurima x_4, ukupnom iznosu odobrenih kredita x_5, ukupnom iznosu do sada otplaćenih kredita x_6 te preostalom ukupnom dugovanju po svim kreditima x_7. Ove podatke želimo iskoristiti kao značajke, ali pritom naravno moramo paziti da ne uvedemo multikolinearnost, pa možda nećemo upotrijebiti sve značajke. Za funkciju preslikavanja $\boldsymbol{\phi}$ koristimo preslikavanje s linearnim, kvadratnim i interakcijskim značajkama. Za interakcije razmatramo parove i trojke značajki, uključivo kombinacije kvadriranih i nekvadriranih značajki (npr., x_1x_2, $x_1^2x_2$ i $x_1x_2^2x_3^2$). **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 48 B 79 C 86 D 92

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

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

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Odredite veličinu prostora inačica, $|VS_{\mathcal{H},\mathbf{\Phi}}|$, za ove kombinacije modela i funkcije preslikavanja. **Za koju kombinaciju modela i funkcije preslikavanja prostor inačica sadržava točno jednu hipotezu, $|VS_{\mathcal{H},\mathbf{\Phi}}| = 1$?**

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

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

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

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

A 31.70 B 191.95 C 165.89 D 230.98

4 (T) Optimizacijom regularizirane funkcije pogreške smanjuje se prenaučenost modela. **Kakve parametre modela nalazi optimizacija L_2-regularizirane pogreške?**

A Parametre koji uz što manju magnitudu vrijednosti daju što manje očekivanje gubitka na skupu za učenje

B Parametre koji uz što manju magnitudu vrijednosti daju što veće očekivanje gubitka na skupu za ispitivanje

C Parametre koji uz što veću magnitudu vrijednosti daju što veće očekivanje gubitka na skupu za učenje

D Parametre koji uz što veću magnitudu vrijednosti daju što manje očekivanje gubitka na skupu za ispitivanje

5 (T) Hipoteza h je funkcija koja primjerima iz $\mathcal{X}$ pridjeljuje oznake iz $\mathcal{Y}$. Za h kažemo da je definirana "do na parametre $\boldsymbol{\theta}$". **Što to znači?**

- A Funkcija h izračunava oznaku $\mathcal{Y}$ na temelju parametra $\boldsymbol{\theta}$, koje treba podesiti optimizacijskim postupkom
- B Funkcija h definirana je bez parametara, i njih treba odrediti naknadno postupkom odabira modela $\mathcal{H}$
- C Svaka vrijednost parametara $\boldsymbol{\theta}$ daje jednu konkretnu funkciju h koja se razlikuje od svih drugih funkcija u modelu $\mathcal{H}$
- D Funkcija h definira preslikavanja iz označenog skupa primjera $\mathcal{X} \times \mathcal{Y}$ u prostor parametara $\boldsymbol{\theta}$

6 (P) Optimizacija parametara modela temelji se na funkciji gubitka $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_0^+$, gdje je $L(y, h(\mathbf{x}))$ gubitak na primjeru $(\mathbf{x}, y)$. U većini primjena koristimo simetričan gubitak 0-1. Međutim, u nekim primjenama ima više smisla definirati asimetričan gubitak. Jedan takav primjer je zadatak detekcije neželjenih elektroničkih poruka (*spam filtering*). Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava da je poruka neželjena (spam), a $y = 0$ da ona to nije. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

- A $L(0,0) = L(1,1) < L(0,1) < L(1,0)$
- B $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) = 0$
- C $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$
- D $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

7 (T) Poopćeni linearni modeli (linearna regresija, logistička regresija i multinomijalna regresija) probabilistički su algoritmi strojnog učenja. Njihova probabilistička priroda dolazi do izražaja kako kod modela tako i kod optimizacijskog postupka. **Koji je probabilistički princip ugrađen u optimizacijski postupak tih algoritama?**

- A Maksimizirati $\sum_{i=1}^{N} \ln h(\mathbf{x}^{(i)}; \mathbf{w})$, gdje je $h(\mathbf{x}; \mathbf{w}) = f(\mathbf{w}^{\mathrm{T}}\mathbf{x})$
- B Maksimizirati $\sum_{i=1}^{N} \ln p(y^{(i)}|\mathbf{x}^{(i)})$, gdje je $\mathbb{E}[p(y^{(i)}|\mathbf{x}^{(i)})] = f(\mathbf{w}^{\mathrm{T}}\mathbf{x})$
- C Minimizirati $-\sum_{i=1}^{N} y^{(i)} p(y^{(i)}|\mathbf{x}^{(i)})$, gdje je $p(y^{(i)}|\mathbf{x}^{(i)}) = h(\mathbf{x}; \mathbf{w})$
- D Minimizirati $-\sum_{i=1}^{N} \ln y^{(i)} h(\mathbf{x}^{(i)}; \mathbf{w})$, gdje je $h(\mathbf{x}; \mathbf{w}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$

8 (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, koja je definirana na sljedeći način:

$$\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$$

Početne težine perceptrona neka su sljedeće:

$$\mathbf{w} = (1, 0, 1, 2, -3, 0)$$

Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?

A 16 B 32 C 6 D 22

9 (P) Poopćeni linearni modeli mogu koristiti adaptivne bazne funkcije. Prednost toga je da ne moramo ručno definirati preslikavanje $\boldsymbol{\phi}$ u prostor značajki, već se to preslikavanje može naučiti na temelju podataka. Rasplažemo podatcima iz $K = 2$ klase u 100-dimenzijskome ulaznom prostoru. Za taj klasifikacijski problem koristimo binarnu logističku regresiju, ali s adaptivnim baznim funkcijama. Svaka adaptivna bazna funkcija ϕ_j je i sama jedan model logističke regresije, pri čemu $\phi_0(\mathbf{x}) = 1$. Naš je model onda definiran ovako:

$$h(\mathbf{x}) = \sigma\Big(\sum_{j=0}^{5} w_j \phi_j(\mathbf{x})\Big)$$

Koliko ukupno parametara ima ovaj model?

A 506 B 256 C 261 D 511

10 (P) Razmotrimo sljedeći skup označenih primjera:

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

Nad ovim skupom treniramo dva modela: neregulariziranu logističku regresiju (LR) i L_2-regulariziranu logističku regresiju (L2LR). Pored toga, razmatramo tri funkcije preslikavanja:

$$\begin{aligned}\boldsymbol{\phi_0}(\mathbf{x}) &= (1, x_1, x_2)\\ \boldsymbol{\phi_1}(\mathbf{x}) &= (1, x_1, x_2, x_1^2, x_2^2)\\ \boldsymbol{\phi_2}(\mathbf{x}) &= (1, x_1, x_2, x_1x_2)\end{aligned}$$

Ukupno, dakle, isprobavamo šest algoritama, odnosno šest kombinacija modela i funkcije preslikavanja. **Za koje algoritme (model+preslikavanje) optimizacijski postupak na skupu za učenje *ne* konvergira?**

A LR+$\boldsymbol{\phi_1}$ B L2LR+$\boldsymbol{\phi_2}$ C LR+$\boldsymbol{\phi_2}$ D L2LR+$\boldsymbol{\phi_0}$

11 (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 $+22$ B -2 C -12 D -5

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 2.54 B 7.11 C 4.03 D 1.19

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 4.02 B 6.00 C 12.02 D 8.00

14 (T) Algoritam strojnog učenja idealno bi minimizirao gubitak 0-1. Međutim, funkciju gubitka 0-1 u praksi ne možemo koristiti za optimizaciju parametra modela. **Zašto gubitak 0-1 ne možemo koristiti za optimizaciju?**

A Gubitak 0-1 pored neispravno klasificiranih primjera kažnjava i ispravno klasificirane primjere, i to tim više što su oni dalje od granice između klasa

B Gradijent gubitka 0-1 svugdje je nula osim za $h(\mathbf{x}) = 0$, pa funkcija pogreške ima zaravni po kojima se gradijentni spust ne može spuštati

C Funkcija gubitka 0-1 nije konveksna, pa ni funkcija pogreške nije konveksna već ima lokalne minimume te ne postoji minimizator u zatvorenoj formi

D Funkcija gubitka 0-1 nije diferencijabilna, pa ne postoji rješenje u zatvorenoj formi i ne postoji gradijent

Cjelina 3: Jezgrene i neparametarske metode (8 pitanja)

15 (P) Neka je $\mathcal{H}_{C,\gamma}$ model SVM-a s Gaussovom jezgrom. Hiperparametri tog modela su regularizacijski faktor C i preciznost jezgre γ. Odabir modela provodimo unakrsnom provjerom i to pretraživanjem po rešetci za sljedeće vrijednosti hiperparametara:

$$C = \{2^{-5}, 2^{-4}, \ldots, 2^4, 2^5\}$$
$$\gamma = \{10^{-5}, 10^{-4}, \ldots, 10^4, 10^5\}$$

Već ranije smo utvrdili da je model sa $C = 2^{-2}$ i $\gamma = 10^{-1}$ podnaučen, a da je model sa $C = 2^1$ i $\gamma = 10^1$ prenaučen. Za sve ostale modele ne znamo jesu li prenaučeni, podnaučeni ili optimalni. **Koliko modela još ima smisla ispitati jer su mogući optimalni?**

A 10 B 96 C 25 D 76

16 (T) Problem prokletstva dimenzionalnosti (engl. *course of dimensionality*) pojavljuje se kod algoritama koji rade u visokodimenzijskome vektorskom prostoru i manifestira se na različite načine. **Kako se problem prokletstva dimenzionalnosti u visokodimenzijskim prostorima manifestira kod algoritma k-NN?**

A Primjeri se grupiraju u središte vektorskog prostora i udaljenost postaje nediskriminativna

B Svi primjeri su međusobno vrlo udaljeni i gube se razlike u udaljenosti

C Povećava se broj susjeda u okolini svakog primjera i model k-NN postaje sve jednostavniji

D Broj susjeda k nekog primjera se smanuje i gube se granice između klasa

17 (N) Pomoću SVM-a rješavamo problem binarne klasifikacije grafova. Budući da su primjeri $\mathbf{x}$ grafovi, koristimo SVM s jezgrenom funkcijom nad grafovima. Model treniramo na skupu od $N = 5$ označenih primjera, s vektorom oznaka jednakim $\mathbf{y} = (-1, -1, -1, +1, +1)$ i sa sljedećom jezgrenom matricom:

$$\mathbf{K} = \begin{pmatrix} 1.0 & 0.97 & -0.949 & -0.555 & -0.986 \\ 0.97 & 1.0 & -0.844 & -0.336 & -0.917 \\ -0.949 & -0.844 & 1.0 & 0.789 & 0.988 \\ -0.555 & -0.336 & 0.789 & 1.0 & 0.684 \\ -0.986 & -0.917 & 0.988 & 0.684 & 1.0 \end{pmatrix}$$

Treniranjem uz $C = 1$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (0, 0.733, 1, 1, 0.733)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za četvrti primjer, $L(y^{(4)}, h(\mathbf{x}^{(4)}))$?**

A 2.06 B 1.86 C 0.03 D 0.40

18 (T) Problem meke margine SVM-a s u primarnoj se formulaciji svodi na rješavanje sljedećeg optimizacijskog problema:

$$\underset{\mathbf{w}, w_0, \boldsymbol{\xi}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{N} \xi_i \right\}$$

uz određena linearna ograničenja. Ovaj optimizacijski problem odgovara optimizaciji regularizirane pogreške. Kod regularizirane pogreške u opreci su dva cilja: smanjenje vrijednosti funkcije gubitka i smanjenje složenosti modela. **Kako se ta opreka manifestira kod optimizacijskog problema meke margine SVM-a?**

A Što je veći zbroj od ξ_i, to više primjera ulazi u marginu i tim je manja vrijednost $\|\mathbf{w}\|^2$ te je model veće složenosti

B Što je veća vrijednost $\|\mathbf{w}\|^2$, to je margina uža i tim manje primjera ulazi u marginu, pa je tim manji zbroj od ξ_i

C Što je manji zbroj od ξ_i, to više primjera može ući u marginu i tim je veća vrijednost $\|\mathbf{w}\|^2$ te je model manje složenosti

D Što je veći hiperparametar C, to se više kažnjava ulazak primjera u marginu, i to je manja vrijednost $\|\mathbf{w}\|^2$

19 (P) Za binarnu klasifikaciju koristimo SVM s Gaussovom jezgrenom funkcijom. Razmatramo tri primjera u ulaznome prostoru: $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$ i $\mathbf{x}^{(3)}$. Euklidska udaljenosti između primjera $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(2)}$ manja je nego euklidska udaljenost između primjera $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(3)}$. Za izračun sličnosti između primjera koristimo Gaussovu jezgrenu funkciju $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ s preciznošću γ, gdje $\gamma = 1/2\sigma^2$. **Kakav bi efekt na sličnosti između primjera imalo smanjenje preciznosti γ Gaussove jezgrene funkcije?**

A Sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se povećala, a sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$ bi se smanjila, ali za manji iznos

B Sličnosti bi se povećale, ali $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se povećala više nego što bi se povećala $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$

C Sličnosti bi se smanjile, ali $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se smanjila više nego što bi se smanjila $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$

D Sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se smanjila, a sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$ bi se povećala, ali za manji iznos

20 (N) Rješavamo binarni klasifikacijski problem. Raspolažemo označenim skupom primjera. Odgovarajuća matrica dizajna je sljedeća:

$$\mathbf{X} = \begin{pmatrix} 1 & 3 & 16 & -8 & -11 \\ 1 & -5 & 4 & -8 & -7 \\ 1 & 7 & -4 & 11 & 9 \\ 1 & 15 & -20 & 25 & 25 \end{pmatrix}$$

Na ovom skupu treniramo model SVM-a s tvrdom marginom i linearnom jezgrenom funkcijom (tj. bez preslikavanja u prostor značajki). Model treniramo u primarnoj formulaciji. Za rješenje maksimalne margine dobili smo ovaj vektor težina (uključivo s težinom w_0):

$$\mathbf{w} = (+0.1370, -0.0290, +0.0194, -0.0461, -0.0388)$$

Umjesto u primarnoj formulaciji, model smo mogli trenirati u dualnoj formulaciji, pa bismo umjesto vektora težina $\mathbf{w}$ dobili vektor dualnih parametara $\boldsymbol{\alpha}$, odnosno Lagrangeove multiplikatore. Prisjetite se da su vektori čiji su Lagrangeovi multiplikatori veći od nule potporni vektori. Premda to nije uvijek moguće, u ovom konkretnom slučaju dualni parametri modela mogu se izvesti iz rješenja primarnog modela. Izvedite vektor dualnih parametara $\boldsymbol{\alpha}$. **Koliko iznosi najveća vrijednost parametra u vektoru dualnih parametara $\boldsymbol{\alpha}$?** (Rezultate uspoređujte po prve tri decimale.)

A 0.0024 B 0.0089 C 0.0045 D 0.0013

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

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

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

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

22 (N) Treniramo SVM s polinomijalnom jezgrom definiramo kao:

$$\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z})^2$$

Ova jezgra je Mercerova jezgra, što znači da postoji funkcija $\boldsymbol{\phi}$ takva da $\kappa(\mathbf{x}, \mathbf{z}) = \boldsymbol{\phi}(\mathbf{x})^T\boldsymbol{\phi}(\mathbf{z})$. Konkretno, u slučaju dvodimenzijskoga ulaznog prostora ($n = 2$), ova jezgra odgovara preslikavanju u šesterodimenzijski prostor. Međutim, postoji odstupanje u konkretnim koeficijentima polinoma. Razmotrite primjer $\mathbf{x} = (1, 0)$ te izračunajte $\boldsymbol{\phi}_\kappa(\mathbf{x})$, koji dobivamo preslikavanjem definiranim implicitno preko jezgre, te $\boldsymbol{\phi}_p(\mathbf{x})$, koji dobivamo preslikavanjem definiranim kao polinom drugog stupnja. **Koliko iznosi euklidska udaljenost između $\boldsymbol{\phi}_\kappa(\mathbf{x})$ i $\boldsymbol{\phi}_p(\mathbf{x})$?**

A 4 B $\sqrt{2}$ C 0 D $2\sqrt{2}$

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

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnove i regresija (6 pitanja)

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

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

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Odredite veličinu prostora inačica, $|VS_{\mathcal{H},\mathbf{\Phi}}|$, za ove kombinacije modela i funkcije preslikavanja. **Za koju kombinaciju modela i funkcije preslikavanja prostor inačica sadržava točno jednu hipotezu, $|VS_{\mathcal{H},\mathbf{\Phi}}| = 1$?**

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

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

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

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

A 191.95 B 165.89 C 230.98 D 31.70

3 (T) Optimizacijom regularizirane funkcije pogreške smanjuje se prenaučenost modela. **Kakve parametre modela nalazi optimizacija L_2-regularizirane pogreške?**

A Parametre koji uz što veću magnitudu vrijednosti daju što veće očekivanje gubitka na skupu za učenje

B Parametre koji uz što manju magnitudu vrijednosti daju što manje očekivanje gubitka na skupu za učenje

C Parametre koji uz što veću magnitudu vrijednosti daju što manje očekivanje gubitka na skupu za ispitivanje

D Parametre koji uz što manju magnitudu vrijednosti daju što veće očekivanje gubitka na skupu za ispitivanje

4 (P) Koristimo regresiju za predviđanje kreditne sposobnosti klijenata. Na raspolaganju imamo podatke o dobi x_1, ukupnom prinosu na račune x_2, ukupnoj ušteđevini u kunama x_3, ukupnoj deviznoj ušteđevini u eurima x_4, ukupnom iznosu odobrenih kredita x_5, ukupnom iznosu do sada otplaćenih kredita x_6 te preostalom ukupnom dugovanju po svim kreditima x_7. Ove podatke želimo iskoristiti kao značajke, ali pritom naravno moramo paziti da ne uvedemo multikolinearnost, pa možda nećemo upotrijebiti sve značajke. Za funkciju preslikavanja $\boldsymbol{\phi}$ koristimo preslikavanje s linearnim, kvadratnim i interakcijskim značajkama. Za interakcije razmatramo parove i trojke značajki, uključivo kombinacije kvadriranih i nekvadriranih značajki (npr., x_1x_2, $x_1^2x_2$ i $x_1x_2^2x_3^2$). **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 92 B 48 C 79 D 86

5 (T) Hipoteza h je funkcija koja primjerima iz $\mathcal{X}$ pridjeljuje oznake iz $\mathcal{Y}$. Za h kažemo da je definirana "do na parametre $\boldsymbol{\theta}$". **Što to znači?**

- A Svaka vrijednost parametara $\boldsymbol{\theta}$ daje jednu konkretnu funkciju h koja se razlikuje od svih drugih funkcija u modelu $\mathcal{H}$
- B Funkcija h definirana je bez parametara, i njih treba odrediti naknadno postupkom odabira modela $\mathcal{H}$
- C Funkcija h jednoznačno određuje parametre $\boldsymbol{\theta}$ iz skupa svih mogućih parametara, koji nazivamo prostor parametara
- D Različite vrijednosti za $\boldsymbol{\theta}$ mogu dati različite funkcije h, a skup svih takvih različitih funkcija definira model $\mathcal{H}$

6 (P) Optimizacija parametara modela temelji se na funkciji gubitka $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_0^+$, gdje je $L(y, h(\mathbf{x}))$ gubitak na primjeru $(\mathbf{x}, y)$. U većini primjena koristimo simetričan gubitak 0-1. Međutim, u nekim primjenama ima više smisla definirati asimetričan gubitak. Jedan takav primjer je zadatak detekcije neželjenih elektroničkih poruka (*spam filtering*). Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava da je poruka neželjena (spam), a $y = 0$ da ona to nije. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

- A $L(0,0) = L(1,1) < L(0,1) < L(1,0)$
- B $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$
- C $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) = 0$
- D $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$

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, koja je definirana na sljedeći način:

$$\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)$$

Početne težine perceptrona neka su sljedeće:

$$\mathbf{w} = (1, 0, 1, 2, -3, 0)$$

Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?

A 22 B 16 C 6 D 32

8 (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

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 $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$. 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 -12 B $+22$ C -2 D -5

10 (P) Razmotrimo sljedeći skup označenih primjera:

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

Nad ovim skupom treniramo dva modela: neregulariziranu logističku regresiju (LR) i L_2-regulariziranu logističku regresiju (L2LR). Pored toga, razmatramo tri funkcije preslikavanja:

$$\boldsymbol{\phi_0}(\mathbf{x}) = (1, x_1, x_2)$$
$$\boldsymbol{\phi_1}(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_2^2)$$
$$\boldsymbol{\phi_2}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$$

Ukupno, dakle, isprobavamo šest algoritama, odnosno šest kombinacija modela i funkcije preslikavanja. **Za koje algoritme (model+preslikavanje) optimizacijski postupak na skupu za učenje *ne* konvergira?**

A LR+$\boldsymbol{\phi_1}$ B L2LR+$\boldsymbol{\phi_1}$ C L2LR+$\boldsymbol{\phi_2}$ D LR+$\boldsymbol{\phi_0}$

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 12.02 D 6.00

12 (P) Poopćeni linearni modeli mogu koristiti adaptivne bazne funkcije. Prednost toga je da ne moramo ručno definirati preslikavanje $\boldsymbol{\phi}$ u prostor značajki, već se to preslikavanje može naučiti na temelju podataka. Rasplažemo podatcima iz $K = 2$ klase u 50-dimenzijskome ulaznom prostoru. Za taj klasifikacijski problem koristimo binarnu logističku regresiju, ali s adaptivnim baznim funkcijama. Svaka adaptivna bazna funkcija ϕ_j je i sama jedan model logističke regresije, pri čemu $\phi_0(\mathbf{x}) = 1$. Naš je model onda definiran ovako:

$$h(\mathbf{x}) = \sigma\Big(\sum_{j=0}^{5} w_j \phi_j(\mathbf{x})\Big)$$

Koliko ukupno parametara ima ovaj model?

A 261 B 506 C 256 D 511

13 (T) Algoritam strojnog učenja idealno bi minimizirao gubitak 0-1. Međutim, funkciju gubitka 0-1 u praksi ne možemo koristiti za optimizaciju parametra modela. **Zašto gubitak 0-1 ne možemo koristiti za optimizaciju?**

A Gradijent gubitka 0-1 svugdje je nula osim za $h(\mathbf{x}) = 0$, pa funkcija pogreške ima zaravni po kojima se gradijentni spust ne može spuštati

B Funkcija gubitka 0-1 nije diferencijabilna, pa ne postoji rješenje u zatvorenoj formi i ne postoji gradijent

C Funkcija gubitka 0-1 nije konveksna, pa ni funkcija pogreške nije konveksna već ima lokalne minimume te ne postoji minimizator u zatvorenoj formi

D Gubitak 0-1 pored neispravno klasificiranih primjera kažnjava i ispravno klasificirane primjere, i to tim više što su oni dalje od granice između klasa

14 (T) Poopćeni linearni modeli (linearna regresija, logistička regresija i multinomijalna regresija) probabilistički su algoritmi strojnog učenja. Njihova probabilistička priroda dolazi do izražaja kako kod modela tako i kod optimizacijskog postupka. **Koji je probabilistički princip ugrađen u optimizacijski postupak tih algoritama?**

A Minimizirati $-\sum_{i=1}^{N} \ln y^{(i)} h(\mathbf{x}^{(i)}; \mathbf{w})$, gdje je $h(\mathbf{x}; \mathbf{w}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$

B Maksimizirati $\sum_{i=1}^{N} \ln p(y^{(i)}|\mathbf{x}^{(i)})$, gdje je $\mathbb{E}[p(y^{(i)}|\mathbf{x}^{(i)})] = f(\mathbf{w}^{\mathrm{T}}\mathbf{x})$

C Maksimizirati $\prod_{i=1}^{N} \ln p(y^{(i)}|\mathbf{x}^{(i)})$, gdje je $\mathbb{E}[p(y^{(i)}|\mathbf{x}^{(i)})] = \mathbf{w}^{\mathrm{T}}\mathbf{x}$

D Maksimizirati $\sum_{i=1}^{N} \ln h(\mathbf{x}^{(i)}; \mathbf{w})$, gdje je $h(\mathbf{x}; \mathbf{w}) = f(\mathbf{w}^{\mathrm{T}}\mathbf{x})$

Cjelina 3: Jezgrene i neparametarske metode (8 pitanja)

15 (N) Treniramo SVM s polinomijalnom jezgrom definiramo kao:

$$\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z})^2$$

Ova jezgra je Mercerova jezgra, što znači da postoji funkcija $\boldsymbol{\phi}$ takva da $\kappa(\mathbf{x}, \mathbf{z}) = \boldsymbol{\phi}(\mathbf{x})^T\boldsymbol{\phi}(\mathbf{z})$. Konkretno, u slučaju dvodimenzijskoga ulaznog prostora ($n = 2$), ova jezgra odgovara preslikavanju u šesterodimenzijski prostor. Međutim, postoji odstupanje u konkretnim koeficijentima polinoma. Razmotrite primjer $\mathbf{x} = (1, 0)$ te izračunajte $\boldsymbol{\phi}_\kappa(\mathbf{x})$, koji dobivamo preslikavanjem definiranim implicitno preko jezgre, te $\boldsymbol{\phi}_p(\mathbf{x})$, koji dobivamo preslikavanjem definiranim kao polinom drugog stupnja. **Koliko iznosi euklidska udaljenost između $\boldsymbol{\phi}_\kappa(\mathbf{x})$ i $\boldsymbol{\phi}_p(\mathbf{x})$?**

A 0 B $\sqrt{2}$ C 4 D $2\sqrt{2}$

16 (P) Za binarnu klasifikaciju koristimo SVM s Gaussovom jezgrenom funkcijom. Razmatramo tri primjera u ulaznome prostoru: $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$ i $\mathbf{x}^{(3)}$. Euklidska udaljenosti između primjera $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(2)}$ veća je nego euklidska udaljenost između primjera $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(3)}$. Za izračun sličnosti između primjera koristimo Gaussovu jezgrenu funkciju $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ s preciznošću γ, gdje $\gamma = 1/2\sigma^2$. **Kakav bi efekt na sličnosti između primjera imalo povećanje preciznosti γ Gaussove jezgrene funkcije?**

A Sličnosti bi se smanjile, ali $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se smanjila manje nego što bi se smanjila $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$

B Sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se povećala, a sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$ bi se smanjila, ali za manji iznos

C Sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi postala manja od sličnosti $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$

D Sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ bi se povećala, a sličnost $\kappa(\mathbf{x}^{(1)}, \mathbf{x}^{(3)})$ bi se smanjila, ali za veći iznos

17 (T) Problem meke margine SVM-a s u primarnoj se formulaciji svodi na rješavanje sljedećeg optimizacijskog problema:

$$\underset{\mathbf{w}, w_0, \boldsymbol{\xi}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N} \xi_i \right\}$$

uz određena linearna ograničenja. Ovaj optimizacijski problem odgovara optimizaciji regularizirane pogreške. Kod regularizirane pogreške u opreci su dva cilja: smanjenje vrijednosti funkcije gubitka i smanjenje složenosti modela. **Kako se ta opreka manifestira kod optimizacijskog problema meke margine SVM-a?**

A Što je manji hiperparametar C, to je manja važnost zbroja od ξ_i, pa je vrijednost $\|\mathbf{w}\|^2$ veća i model je veće složenosti

B Što je veći zbroj od ξ_i, to više primjera ulazi u marginu i tim je manja vrijednost $\|\mathbf{w}\|^2$ te je model veće složenosti

C Što je manji zbroj od ξ_i, to više primjera može ući u marginu i tim je veća vrijednost $\|\mathbf{w}\|^2$ te je model manje složenosti

D Što je manja vrijednost $\|\mathbf{w}\|^2$, to je margina šira, no tim više primjera ulazi u marginu i to je veći zbroj od ξ_i

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

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

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

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

19 (N) Pomoću SVM-a rješavamo problem binarne klasifikacije grafova. Budući da su primjeri $\mathbf{x}$ grafovi, koristimo SVM s jezgrenom funkcijom nad grafovima. Model treniramo na skupu od $N = 5$ označenih primjera, s vektorom oznaka jednakim $\mathbf{y} = (+1, +1, -1, -1, +1)$ i sa sljedećom jezgrenom matricom:

$$\mathbf{K} = \begin{pmatrix} 1.0 & 0.97 & -0.949 & -0.555 & -0.986 \\ 0.97 & 1.0 & -0.844 & -0.336 & -0.917 \\ -0.949 & -0.844 & 1.0 & 0.789 & 0.988 \\ -0.555 & -0.336 & 0.789 & 1.0 & 0.684 \\ -0.986 & -0.917 & 0.988 & 0.684 & 1.0 \end{pmatrix}$$

Treniranjem uz $C = 5$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (0, 0.688, 5, 0.688, 5)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za peti primjer, $L(y^{(5)}, h(\mathbf{x}^{(5)}))$?**

A 0.40 B 2.06 C 1.86 D 0.03

20 (T) Problem prokletstva dimenzionalnosti (engl. *course of dimensionality*) pojavljuje se kod algoritama koji rade u visokodimenzijskome vektorskom prostoru i manifestira se na različite načine. **Kako se problem prokletstva dimenzionalnosti u visokodimenzijskim prostorima manifestira kod algoritma k-NN?**

A Broj susjeda k povećava se za svaki primjer te model postaje sve jednostavniji

B Svi primjeri su međusobno vrlo udaljeni i gube se razlike u udaljenosti

C Broj susjeda k nekog primjera se smanuje i gube se granice između klasa

D Udaljenosti između primjera se smanjuju i model k-NN postaje sve složeniji

21 (P) Neka je $\mathcal{H}_{C,\gamma}$ model SVM-a s Gaussovom jezgrom. Hiperparametri tog modela su regularizacijski faktor C i preciznost jezgre γ. Odabir modela provodimo unakrsnom provjerom i to pretraživanjem po rešetci za sljedeće vrijednosti hiperparametara:

$$C = \{2^{-5}, 2^{-4}, \dots, 2^4, 2^5\}$$
$$\gamma = \{10^{-5}, 10^{-4}, \dots, 10^4, 10^5\}$$

Već ranije smo utvrdili da je model sa $C = 2^{-2}$ i $\gamma = 10^{-1}$ podnaučen, a da je model sa $C = 2^1$ i $\gamma = 10^1$ prenaučen. Za sve ostale modele ne znamo jesu li prenaučeni, podnaučeni ili optimalni. **Koliko modela još ima smisla ispitati jer su moguće optimalni?**

A 25 B 10 C 76 D 96

22 (N) Rješavamo binarni klasifikacijski problem. Raspolažemo označenim skupom primjera. Odgovarajuća matrica dizajna je sljedeća:

$$\mathbf{X} = \begin{pmatrix} 1 & 3 & 16 & -8 & -11 \\ 1 & -5 & 4 & -8 & -7 \\ 1 & 7 & -4 & 11 & 9 \\ 1 & 15 & -20 & 25 & 25 \end{pmatrix}$$

Na ovom skupu treniramo model SVM-a s tvrdom marginom i linearnom jezgrenom funkcijom (tj. bez preslikavanja u prostor značajki). Model treniramo u primarnoj formulaciji. Za rješenje maksimalne margine dobili smo ovaj vektor težina (uključivo s težinom w_0):

$$\mathbf{w} = (-1.3214, +0.0714, +0.1071, 0, -0.0357)$$

Umjesto u primarnoj formulaciji, model smo mogli trenirati u dualnoj formulaciji, pa bismo umjesto vektora težina $\mathbf{w}$ dobili vektor dualnih parametara $\boldsymbol{\alpha}$, odnosno Lagrangeove multiplikatore. Prisjetite se da su vektori čiji su Lagrangeovi multiplikatori veći od nule potporni vektori. Premda to nije uvijek moguće, u ovom konkretnom slučaju dualni parametri modela mogu se izvesti iz rješenja primarnog modela. Izvedite vektor dualnih parametara $\boldsymbol{\alpha}$. **Koliko iznosi najveća vrijednost parametra u vektoru dualnih parametara $\boldsymbol{\alpha}$?** (Rezultate uspoređujte po prve tri decimale.)

A 0.0024 B 0.0013 C 0.0089 D 0.0045

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