Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

27. studenog 2025.

Strojno učenje 1 – Međuispit

*Ispit sadrži **20** pitanja i ukupno nosi najviše 18 bodova (za 30% bodova na predmetu). Točan odgovor donosi 1 bod, a 1/3 boda oduzima se za netočan odgovor. Ostvareni bodovi iznad 18 se zanemaruju. Trajanje ispita je **150 minuta**.* ***Uz primjerak ispita morate predati papire s jasno vidljivim postupkom rješavanja za sva pitanja na koja ste odgovorili, inače se ispit neće vrednovati.***

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

1 (P) Razmatramo tri modela L2-regularizirane linearne regresije s do na parametre definiranim hipotezama h_1, h_2 i h_3 i s nama nepoznatim regularizacijskim faktorima λ_1, λ_2 odnosno λ_3. Za sva tri modela mjerimo vrijednost (neregularizirane) funkcije kvadratne pogreške $E(h|\mathcal{D})$, i to na skupu za učenje $\mathcal{D}_u$ i na ispitnome skupu D_i. Na skupu za učenje opažamo $E(h_2|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$ i $E(h_3|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, a na skupu za ispitivanje $E(h_1|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$ i $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$. **Koji uređaj očekivano vrijedi između regularizacijskih faktora?**

A $\lambda_2 < \lambda_3 < \lambda_1$ B $\lambda_3 < \lambda_2 < \lambda_1$ C $\lambda_2 < \lambda_1 < \lambda_3$ D $\lambda_1 < \lambda_3 < \lambda_2$

2 (T) Optimizacija parametara hipoteze zasniva se na funkciji empirijske pogreške odnosno funkciji gubitka. **Koja je razlika između funkcije empirijske pogreške i funkcije gubitka?**

A Funkcija gubitka odnosi se na jedan primjer, a funkcija empirijske pogreške na skup primjera

B Funkcija empirijske pogreške uvijek je nenegativna, dok funkcija gubitka može biti negativna

C Funkcija gubitka definirana je kao očekivanje funkcije empirijske pogreške na skupu primjera

D Funkcija empirijske pogreške ovisi o konkretnoj hipotezi, dok funkcija gubitka ne ovisi

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

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

Za faktor regularizacije koristimo $\lambda = 10$. Parametre optimiramo u zatvorenoj formi, na temelju inverza modificirane Gramove matrice. Izračunajte parametre $\mathbf{w}^*$ te L2-regulariziranu pogrešku $E_R(\mathbf{w}|\mathcal{D})$. **Koliko iznosi regularizirana pogreška?**

A 6.695 B 8.258 C 12.525 D 10.252

4 (T) Rješenje za optimizacijski problem najmanjih kvadrata (OLS) izveli smo u zatvorenoj formi kao pseudoinverz $\mathbf{\Phi}^+$ matrice dizajna $\mathbf{\Phi}$. U nekim se slučajevima pseudoinverz $\mathbf{\Phi}^+$ može izračunati pomoću običnog inverza Gramove matrice. Neka je $m+1$ dimenzija prostora značajki, a N broj primjera u skupu za učenje. **Kada se pseudoinverz $\mathbf{\Phi}^+$ može izračunati pomoću inverza Gramove matrice?**

A Ako je $N = m+1$ i Gramova matrica je kvadratna

B Ako $N \geq m+1$ i stupci u $\mathbf{\Phi}$ nisu linearno zavisni

C Ako $N \leq m+1$ i nema multikolinearnosti značajki

D Ako $N < m+1$ i $\mathbf{\Phi}$ je punog ranga

5 (P) Rješavamo binaran klasifikacijski problem u prostoru $\mathbb{R}^3$. Skup označenih primjera je:

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

Razmatramo linearan model ($\mathcal{H}_1$) i model kugle ($\mathcal{H}_2$), definirane kao:

$$\mathcal{H}_1 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{x} \geq \theta_0\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}) \in \mathbb{R}^4$$

$$\mathcal{H}_2 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(\mathbf{x}-\tilde{\boldsymbol{\theta}})^{\mathrm{T}}(\mathbf{x}-\tilde{\boldsymbol{\theta}}) \geq \theta_0^2\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}), \theta_0 \in \big[0, \frac{1}{2}\big], \tilde{\boldsymbol{\theta}} \in \mathbb{R}^3$$

Optimizacijski postupak koristi funkciju pogreške $E(\boldsymbol{\theta}|\mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{h(\mathbf{x}^{(i)};\boldsymbol{\theta}) \neq y^{(i)}\}$. Odredite minimalne pogreške $E(\boldsymbol{\theta}_1^*|\mathcal{D})$ i $E(\boldsymbol{\theta}_2^*|\mathcal{D})$ modela $\mathcal{H}_1$ odnosno $\mathcal{H}_2$. **Koliko iznosi apsolutna razlika između vrijednosti tih pogrešaka?**

A 1/6 B 1/2 C 1/3 D 0

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

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

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

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

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

Cjelina 2: Linearni klasifikacijski modeli (7 pitanja)

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

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

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

A 32 B 16 C 22 D 6

8 (N) Raspolažemo označenim skupom primjera iz četiriju klasa ($K = 4$) u dvodimenzijskome ulaznom prostoru ($n = 2$). Na tom skupu treniramo model multinomijalne logističke regresije s funkcijom preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & -1 & -1 & 2 \\ 2 & -4 & -2 & 1 \\ 3 & -4 & 6 & 1 \\ -3 & 0 & 2 & 1 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (2, -3)$ s oznakom $\mathbf{y} = (0, 1, 0, 0)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi primjer x?**

A 64 B 21 C 51 D 13

9 (N) Raspolažemo sljedećim skupom označenih podataka u dvodimenzijskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((2,5),1),((-8,1),1),((3,9),0)\big\}$$

Na tom skupu treniramo model L2-regularizirane logističke regresije. Koristimo standardni (grupni) gradijentni spust uz faktor regularizacije $\lambda = 10$. Koristimo fiksnu stopu učenja $\eta = 0.01$ (ne koristimo linijsko pretraživanje). Vektor težina $\mathbf{w}$ početno je inicijaliziran na nul-vektor. Provedite jednu epohu gradijentnog spusta (jedan prolazak kroz sve primjere iz $\mathcal{D}$). **Koji se vektor težina w dobiva nakon prve epohe gradijentnog spusta?**

A $(-0.005, 0.045, 0.015)$ C $(0.005, -0.035, 0.025)$

B $(-0.005, 0.035, -0.025)$ D $(0.005, -0.045, -0.015)$

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

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

Na tom skupu trenirali smo dva modela L2-regularizirane logističke regresije, koristeći regularizacijske faktore $\lambda = 1$ odnosno $\lambda = 5$. Neka je $L_\lambda^{(i)}$ gubitak koji primjer $\mathbf{x}^{(i)}$ iz $\mathcal{D}$ nanosi modelu optimiranom uz regularizacijski faktor λ. **Što očekujemo da vrijedi?**

A $L_1^{(4)} > L_5^{(3)}, L_1^{(2)} > L_5^{(1)}$

B $L_1^{(5)} < L_5^{(6)}, L_1^{(4)} > L_5^{(3)}$

C $L_1^{(6)} > L_5^{(5)}, L_1^{(3)} < L_5^{(4)}$

D $L_1^{(3)} < L_5^{(4)}, L_1^{(1)} < L_5^{(2)}$

11 (P) Perceptronom rješavamo višeklasni klasifikacijski sa $K = 5$ klasa u ulaznome prostoru dimenzije $n = 6$. Koristimo shemu jedan-naspram-jedan (OVO). Kako bismo ostvarili linearnu razdvojivost klasa, koristimo funkciju preslikavanja $\boldsymbol{\phi}$ koja odgovara polinomu drugoga stupnja s interakcijskim parovima značajki. **Koliko iznosi ukupan broj parametara koji treba optimirati?**

A 140 B 315 C 126 D 280

12 (T) Jedan od nedostataka algoritma perceptrona jest taj što hipoteza ovisi o odabiru početnog vektora težina. **Zašto je tomu tako?**

A Gradijent funkcije pogreške jednak je nula ako i samo ako je vrijednost funkcije veća od nule

B Gradijent funkcije gubitka nije nula niti za jednu točku u prostoru parametara

C Funkcija pogreške nije konveksna funkcija parametara

D Funkcija pogreške ima više točaka u kojima ostvaruje minimum

13 (T) Za optimizaciju parametara logističke regresije umjesto gradijentnog spusta možemo primijeniti Newtonov postupak. **Konceptualno gledano, po čemu se Newtonov postupak razlikuje od gradijentnog spusta?**

A Umjesto gradijenta funkcije pogreške, izračunava gradijent njezine kvadratne aproksimacije

B Rekondicionira Gramovu matricu modifikacijom njezine glavne dijagonale

C Kod ažuriranja težina u obzir uzima informaciju o promjeni gradijenta

D Za izračun gradijenta koristi sve primjere iz skupa za učenje

Cjelina 3: SVM, jezgrene i neparametarske metode (7 pitanja)

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

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

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

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

15 (N) Raspolažemo sljedećim skupom označenih primjera iz trodimenzijskoga ulaznog prostora:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((9,1,3),+1),((-1,7,3),-1),((0,-8,1),+1)\big\}$$

Na ovome skupu trenirali smo SVM s inverznom kvadratnom jezgrom $\kappa(\mathbf{x},\mathbf{z}) = (1+\|\mathbf{x}-\mathbf{z}\|^2)^{-1}$ i hiperparametrom $C = 10$. Dobiveni Lagrangeovi koeficijenti su $\boldsymbol{\alpha} = (0.6728, 1.3400, 0.6678)$. Može se izračunati da vrijedi $w_0 = 0.333$. Zanima nas klasifikacija primjera $\mathbf{x} = (8,3,1)$. Primijenite jezgreni trik i izračunajte $h(\mathbf{x})$. **Koliko iznosi $h(\mathbf{x})$?**

A 0.231 B 0.550 C 0.391 D 0.118

16 (N) Raspolažemo sljedećim skupom označenih podataka:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} -15 & -30 & -15 & +1 \\ 6 & 12 & 6 & +1 \\ -5 & -8 & -3 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right)$$

Na ovom skupu treniramo SVM s linearnom jezgrom. Koristimo stohastički gradijentni spust. Vektor dualnih koeficijenata je $\boldsymbol{\alpha} = (0.048, 1.952, 0.0, 2.0)$. Jedan potporni vektor nalazi se unutar margine. Izračunajte vrijednost regularizacijskog izraza u funkciji pogreške SVM-a za parametre $\mathbf{w}$ koji odgovaraju dualnim koeficijentima $\boldsymbol{\alpha}$. **Koliko iznosi regularizacijski izraz?**

A 0.75 B 0.63 C 0.88 D 0.50

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

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

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

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

18 (T) Razina složenosti modela k-NN može se ugoditi hiperparametrom k. Neka je N broj primjera. **Koja je veza između k, N i razine složenosti modela k-NN?**

A Ako N pada, a k je fiksan, razina složenosti raste pa pada

B Ako N raste, s povećanjem k možemo zadržati istu razinu složenosti

C Ako je N fiksan, manji k daje model manje razine složenosti

D Neovisno o N, s porastom vrijednosti k razina složenosti raste

19 (T) Tradicionalan način optimizacije parametara stroja potpornih vektora jest primjena konveksne optimizacije uz ograničenja. Razmotrite problem meke margine sa N primjera u n-dimenzijskome ulaznom prostoru. Optimizacija se može provesti u primarnoj ili dualnoj formi. **Koliko se optimizacijskih varijabli te koliko ograničenja javlja u dualnoj formulaciji problema meke margine?**

A N varijabli, N ograničenja nejednakosti, 1 ograničenje jednakosti

B N varijabli, $2N$ ograničenja nejednakosti, 1 ograničenje jednakosti

C $n + 1$ varijabli, N ograničenja nejednakosti

D $n + N + 1$ varijabli, $2N$ ograničenja nejednakosti

20 (P) Rješavamo binaran klasifikacijski problem u ulaznome prostoru dimenzije $n = 100$. Raspolažemo sa $N = 1000$ primjera. Isprobavamo dva klasifikacijska algoritma: (1) L2-regulariziranu logističku regresiju s adaptivnim baznim funkcijama i (2) rijetki jezgreni stroj ostvaren kao L1-regularizirana logistička regresija. Kod prvog modela koristimo $m = 50$ adaptivnih baznih funkcija, pri čemu je svaka adaptivna bazna funkcija ϕ_j i sama jedan model logističke regresije. Kod drugog modela koristimo Gaussove jezgrene funkcije (svaka s istom preciznošću), a nakon treniranja naučen model sastoji se od $N' = 32$ prototipa. Razmotrite ukupan broj parametara i hiperparametara oba modela nakon što su ti modeli naučeni, tj. broj (hiper)parametara koje nužno trebamo pohraniti u datoteku kod serijalizacije modela, a da bismo modele mogli naknadno koristiti za predikciju. Jedan (hiper)parametar odgovara jednom realnom broju. **Koliko iznosi razlika u broju pohranjenih (hiper)parametara za ova dva modela?**

A 899 B 1867 C 1784 D 1749

Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

27. studenog 2025.

Strojno učenje 1 – Međuispit

*Ispit sadrži **20** pitanja i ukupno nosi najviše 18 bodova (za 30% bodova na predmetu). Točan odgovor donosi 1 bod, a 1/3 boda oduzima se za netočan odgovor. Ostvareni bodovi iznad 18 se zanemaruju. Trajanje ispita je **150 minuta**.* ***Uz primjerak ispita morate predati papire s jasno vidljivim postupkom rješavanja za sva pitanja na koja ste odgovorili, inače se ispit neće vrednovati.***

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

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

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

Za faktor regularizacije koristimo $\lambda = 10$. Parametre optimiramo u zatvorenoj formi, na temelju inverza modificirane Gramove matrice. Izračunajte parametre $\mathbf{w}^*$ te L2-regulariziranu pogrešku $E_R(\mathbf{w}|\mathcal{D})$. **Koliko iznosi regularizirana pogreška?**

A 10.252 B 6.695 C 8.258 D 12.525

2 (T) Optimizacija parametara hipoteze zasniva se na funkciji empirijske pogreške odnosno funkciji gubitka. **Koja je razlika između funkcije empirijske pogreške i funkcije gubitka?**

A Funkcija gubitka odnosi se na jedan primjer, a funkcija empirijske pogreške na skup primjera

B Funkcija gubitka definirana je kao očekivanje funkcije empirijske pogreške na skupu primjera

C Funkcija empirijske pogreške uvijek je nenegativna, dok funkcija gubitka može biti negativna

D Funkcija empirijske pogreške ovisi o konkretnoj hipotezi, dok funkcija gubitka ne ovisi

3 (P) Rješavamo binaran klasifikacijski problem u prostoru $\mathbb{R}^3$. Skup označenih primjera je:

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

Razmatramo linearan model ($\mathcal{H}_1$) i model kugle ($\mathcal{H}_2$), definirane kao:

$$\mathcal{H}_1 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{x} \geq \theta_0\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}) \in \mathbb{R}^4$$

$$\mathcal{H}_2 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(\mathbf{x}-\tilde{\boldsymbol{\theta}})^{\mathrm{T}}(\mathbf{x}-\tilde{\boldsymbol{\theta}}) \leq \theta_0^2\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}), \theta_0 \in \big[0, \frac{1}{2}\big], \tilde{\boldsymbol{\theta}} \in \mathbb{R}^3$$

Optimizacijski postupak koristi funkciju pogreške $E(\boldsymbol{\theta}|\mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{h(\mathbf{x}^{(i)};\boldsymbol{\theta}) \neq y^{(i)}\}$. Odredite minimalne pogreške $E(\boldsymbol{\theta}_1^*|\mathcal{D})$ i $E(\boldsymbol{\theta}_2^*|\mathcal{D})$ modela $\mathcal{H}_1$ odnosno $\mathcal{H}_2$. **Koliko iznosi apsolutna razlika između vrijednosti tih pogrešaka?**

A 1/2 B 1/3 C 0 D 1/6

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

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

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

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

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

5 (T) Rješenje za optimizacijski problem najmanjih kvadrata (OLS) izveli smo u zatvorenoj formi kao pseudoinverz $\mathbf{\Phi}^+$ matrice dizajna $\mathbf{\Phi}$. U nekim se slučajevima pseudoinverz $\mathbf{\Phi}^+$ može izračunati pomoću običnog inverza Gramove matrice. Neka je $m+1$ dimenzija prostora značajki, a N broj primjera u skupu za učenje. **Kada se pseudoinverz $\mathbf{\Phi}^+$ može izračunati pomoću inverza Gramove matrice?**

A Ako $N \leq m+1$ i nema multikolinearnosti značajki

B Ako $N \geq m+1$ i stupci u $\mathbf{\Phi}$ nisu linearno zavisni

C Ako je $N = m+1$ i Gramova matrica je kvadratna

D Ako $N < m+1$ i $\mathbf{\Phi}$ je punog ranga

6 (P) Razmatramo tri modela L2-regularizirane linearne regresije s do na parametre definiranim hipotezama h_1, h_2 i h_3 i s nama nepoznatim regularizacijskim faktorima λ_1, λ_2 odnosno λ_3. Za sva tri modela mjerimo vrijednost (neregularizirane) funkcije kvadratne pogreške $E(h|\mathcal{D})$, i to na skupu za učenje $\mathcal{D}_u$ i na ispitnome skupu D_i. Na skupu za učenje opažamo $E(h_2|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$ i $E(h_3|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, a na skupu za ispitivanje $E(h_1|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$ i $E(h_2|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$. **Koji uređaj očekivano vrijedi između regularizacijskih faktora?**

A $\lambda_2 < \lambda_3 < \lambda_1$ B $\lambda_3 < \lambda_2 < \lambda_1$ C $\lambda_1 < \lambda_2 < \lambda_3$ D $\lambda_3 < \lambda_1 < \lambda_2$

Cjelina 2: Linearni klasifikacijski modeli (7 pitanja)

7 (T) Za optimizaciju parametara logističke regresije umjesto gradijentnog spusta možemo primijeniti Newtonov postupak. **Konceptualno gledano, po čemu se Newtonov postupak razlikuje od gradijentnog spusta?**

A Umjesto gradijenta funkcije pogreške, izračunava gradijent njezine kvadratne aproksimacije

B Kod ažuriranja težina u obzir uzima informaciju o promjeni gradijenta

C Rekondicionira Gramovu matricu modifikacijom njezine glavne dijagonale

D Za izračun gradijenta koristi sve primjere iz skupa za učenje

8 (N) Raspolažemo sljedećim skupom označenih podataka u dvodimenzijskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((2,5),1),((-8,1),0),((3,9),0)\big\}$$

Na tom skupu treniramo model L2-regularizirane logističke regresije. Koristimo standardni (grupni) gradijentni spust uz faktor regularizacije $\lambda = 10$. Koristimo fiksnu stopu učenja $\eta = 0.01$ (ne koristimo linijsko pretraživanje). Vektor težina $\mathbf{w}$ početno je inicijaliziran na nul-vektor. Provedite jednu epohu gradijentnog spusta (jedan prolazak kroz sve primjere iz $\mathcal{D}$). **Koji se vektor težina $\mathbf{w}$ dobiva nakon prve epohe gradijentnog spusta?**

A $(-0.005, 0.045, 0.015)$ C $(0.005, -0.045, -0.015)$

B $(-0.005, 0.035, -0.025)$ D $(0.005, -0.035, 0.025)$

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

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

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

A 6 B 32 C 22 D 16

10 (T) Jedan od nedostataka algoritma perceptrona jest taj što hipoteza ovisi o odabiru početnog vektora težina. **Zašto je tomu tako?**

A Gradijent funkcije gubitka nije nula niti za jednu točku u prostoru parametara

B Gradijent funkcije pogreške jednak je nula ako i samo ako je vrijednost funkcije veća od nule

C Funkcija pogreške nije konveksna funkcija parametara

D Funkcija pogreške ima više točaka u kojima ostvaruje minimum

11 (P) Perceptronom rješavamo višeklasni klasifikacijski sa $K = 5$ klasa u ulaznome prostoru dimenzije $n = 6$. Koristimo shemu jedan-naspram-jedan (OVO). Kako bismo ostvarili linearnu razdvojivost klasa, koristimo funkciju preslikavanja $\boldsymbol{\phi}$ koja odgovara polinomu drugoga stupnja s interakcijskim parovima značajki. **Koliko iznosi ukupan broj parametara koji treba optimirati?**

A 140 B 315 C 280 D 126

12 (N) Raspolažemo označenim skupom primjera iz četiriju klasa ($K = 4$) u dvodimenzijskome ulaznom prostoru ($n = 2$). Na tom skupu treniramo model multinomijalne logističke regresije s funkcijom preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & -1 & -1 & 2 \\ 2 & -4 & -2 & 1 \\ 3 & -4 & 6 & 1 \\ -3 & 0 & 2 & 1 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (2, -3)$ s oznakom $\mathbf{y} = (0, 0, 0, 1)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi primjer x?**

A 64 B 51 C 13 D 21

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

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

Na tom skupu trenirali smo dva modela L2-regularizirane logističke regresije, koristeći regularizacijske faktore $\lambda = 1$ odnosno $\lambda = 5$. Neka je $L_\lambda^{(i)}$ gubitak koji primjer $\mathbf{x}^{(i)}$ iz $\mathcal{D}$ nanosi modelu optimiranom uz regularizacijski faktor λ. **Što očekujemo da vrijedi?**

A $L_1^{(3)} > L_5^{(6)}, L_1^{(5)} < L_5^{(2)}$ C $L_1^{(6)} < L_5^{(3)}, L_1^{(2)} > L_5^{(5)}$

B $L_1^{(5)} < L_5^{(2)}, L_1^{(4)} < L_5^{(1)}$ D $L_1^{(2)} > L_5^{(5)}, L_1^{(1)} > L_5^{(4)}$

Cjelina 3: SVM, jezgrene i neparametarske metode (7 pitanja)

14 (P) Rješavamo binaran klasifikacijski problem u ulaznome prostoru dimenzije $n = 100$. Raspolažemo sa $N = 1000$ primjera. Isprobavamo dva klasifikacijska algoritma: (1) L2-regulariziranu logističku regresiju s adaptivnim baznim funkcijama i (2) rijetki jezgreni stroj ostvaren kao L1-regularizirana logistička regresija. Kod prvog modela koristimo $m = 50$ adaptivnih baznih funkcija, pri čemu je svaka adaptivna bazna funkcija ϕ_j i sama jedan model logističke regresije. Kod drugog modela koristimo Gaussove jezgrene funkcije (svaka s istom preciznošću), a nakon treniranja naučen model sastoji se od $N' = 48$ prototipa. Razmotrite ukupan broj parametara i hiperparametara oba modela nakon što su ti modeli naučeni, tj. broj (hiper)parametara koje nužno trebamo pohraniti u datoteku kod serijalizacije modela, a da bismo modele mogli naknadno koristiti za predikciju. Jedan (hiper)parametar odgovara jednom realnom broju. **Koliko iznosi razlika u broju pohranjenih (hiper)parametara za ova dva modela?**

A 149 B 251 C 152 D 701

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

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

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

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

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

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

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

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

17 (T) Tradicionalan način optimizacije parametara stroja potpornih vektora jest primjena konveksne optimizacije uz ograničenja. Razmotrite problem meke margine sa N primjera u n-dimenzijskome ulaznom prostoru. Optimizacija se može provesti u primarnoj ili dualnoj formi. **Koliko se optimizacijskih varijabli te koliko ograničenja javlja u dualnoj formulaciji problema meke margine?**

A N varijabli, $2N$ ograničenja nejednakosti, 1 ograničenje jednakosti

B N varijabli, N ograničenja nejednakosti, 1 ograničenje jednakosti

C $n + 1$ varijabli, N ograničenja nejednakosti

D $n + N + 1$ varijabli, $2N$ ograničenja nejednakosti

18 (T) Razina složenosti modela k-NN može se ugoditi hiperparametrom k. Neka je N broj primjera. **Koja je veza između k, N i razine složenosti modela k-NN?**

A Ako N pada, a k je fiksan, razina složenosti raste pa pada

B Ako je N fiksan, manji k daje model manje razine složenosti

C Ako N raste, s povećanjem k možemo zadržati istu razinu složenosti

D Neovisno o N, s porastom vrijednosti k razina složenosti raste

19 (N) Raspolažemo sljedećim skupom označenih podataka:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} -15 & -30 & -15 & +1 \\ 6 & 12 & 6 & +1 \\ -5 & -8 & -3 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right)$$

Na ovom skupu treniramo SVM s linearnom jezgrom. Koristimo stohastički gradijentni spust. Vektor dualnih koeficijenata je $\boldsymbol{\alpha} = (0.048, 1.952, 0.0, 2.0)$. Jedan potporni vektor nalazi se unutar margine. Izračunajte vrijednost regularizacijskog izraza u funkciji pogreške SVM-a za parametre $\mathbf{w}$ koji odgovaraju dualnim koeficijentima $\boldsymbol{\alpha}$. **Koliko iznosi regularizacijski izraz?**

A 0.88 B 0.63 C 0.75 D 0.50

20 (N) Raspolažemo sljedećim skupom označenih primjera iz trodimenzijskoga ulaznog prostora:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((9, 1, 3), +1), ((-1, 7, 3), -1), ((0, -8, 1), +1)\big\}$$

Na ovome skupu trenirali smo SVM s inverznom kvadratnom jezgrom $\kappa(\mathbf{x}, \mathbf{z}) = (1 + \|\mathbf{x} - \mathbf{z}\|^2)^{-1}$ i hiperparametrom $C = 10$. Dobiveni Lagrangeovi koeficijenti su $\boldsymbol{\alpha} = (0.6728, 1.3400, 0.6678)$. Može se izračunati da vrijedi $w_0 = 0.333$. Zanima nas klasifikacija primjera $\mathbf{x} = (8, 3, 1)$. Primijenite jezgreni trik i izračunajte $h(\mathbf{x})$. **Koliko iznosi $h(\mathbf{x})$?**

A 0.118 B 0.391 C 0.550 D 0.231

Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

27. studenog 2025.

Strojno učenje 1 – Međuispit

*Ispit sadrži **20** pitanja i ukupno nosi najviše 18 bodova (za 30% bodova na predmetu). Točan odgovor donosi 1 bod, a 1/3 boda oduzima se za netočan odgovor. Ostvareni bodovi iznad 18 se zanemaruju. Trajanje ispita je **150 minuta**.* ***Uz primjerak ispita morate predati papire s jasno vidljivim postupkom rješavanja za sva pitanja na koja ste odgovorili, inače se ispit neće vrednovati.***

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

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

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

2 (P) Rješavamo binaran klasifikacijski problem u prostoru $\mathbb{R}^3$. Skup označenih primjera je:

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

Razmatramo linearan model ($\mathcal{H}_1$) i model kugle ($\mathcal{H}_2$), definirane kao:

$$\mathcal{H}_1 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{x} \geq \theta_0\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}) \in \mathbb{R}^4$$
$$\mathcal{H}_2 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(\mathbf{x}-\tilde{\boldsymbol{\theta}})^{\mathrm{T}}(\mathbf{x}-\tilde{\boldsymbol{\theta}}) \geq \theta_0^2\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}), \theta_0 \in \big[0, \frac{1}{2}\big], \tilde{\boldsymbol{\theta}} \in \mathbb{R}^3$$

Optimizacijski postupak koristi funkciju pogreške $E(\boldsymbol{\theta}|\mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{h(\mathbf{x}^{(i)};\boldsymbol{\theta}) \neq y^{(i)}\}$. Odredite minimalne pogreške $E(\boldsymbol{\theta}_1^*|\mathcal{D})$ i $E(\boldsymbol{\theta}_2^*|\mathcal{D})$ modela $\mathcal{H}_1$ odnosno $\mathcal{H}_2$. **Koliko iznosi apsolutna razlika između vrijednosti tih pogrešaka?**

A 0 B 1/3 C 1/2 D 1/6

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

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

Za faktor regularizacije koristimo $\lambda = 10$. Parametre optimiramo u zatvorenoj formi, na temelju inverza modificirane Gramove matrice. Izračunajte parametre $\mathbf{w}^*$ te L2-regulariziranu pogrešku $E_R(\mathbf{w}|\mathcal{D})$. **Koliko iznosi regularizirana pogreška?**

A 6.695 B 8.258 C 10.252 D 12.525

4 (T) Optimizacija parametara hipoteze zasniva se na funkciji empirijske pogreške odnosno funkciji gubitka. **Koja je razlika između funkcije empirijske pogreške i funkcije gubitka?**

- A Funkcija empirijske pogreške uvijek je nenegativna, dok funkcija gubitka može biti negativna
- B Funkcija empirijske pogreške ovisi o konkretnoj hipotezi, dok funkcija gubitka ne ovisi
- C Funkcija gubitka odnosi se na jedan primjer, a funkcija empirijske pogreške na skup primjera
- D Funkcija gubitka definirana je kao očekivanje funkcije empirijske pogreške na skupu primjera

5 (P) Razmatramo tri modela L2-regularizirane linearne regresije s do na parametre definiranim hipotezama h_1, h_2 i h_3 i s nama nepoznatim regularizacijskim faktorima λ_1, λ_2 odnosno λ_3. Za sva tri modela mjerimo vrijednost (neregularizirane) funkcije kvadratne pogreške $E(h|\mathcal{D})$, i to na skupu za učenje $\mathcal{D}_u$ i na ispitnome skupu D_i. Na skupu za učenje opažamo $E(h_2|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$ i $E(h_3|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, a na skupu za ispitivanje $E(h_1|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$ i $E(h_2|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$. **Koji uređaj očekivano vrijedi između regularizacijskih faktora?**

A $\lambda_3 < \lambda_1 < \lambda_2$ B $\lambda_2 < \lambda_3 < \lambda_1$ C $\lambda_3 < \lambda_2 < \lambda_1$ D $\lambda_1 < \lambda_2 < \lambda_3$

6 (T) Rješenje za optimizacijski problem najmanjih kvadrata (OLS) izveli smo u zatvorenoj formi kao pseudoinverz $\mathbf{\Phi}^+$ matrice dizajna $\mathbf{\Phi}$. U nekim se slučajevima pseudoinverz $\mathbf{\Phi}^+$ može izračunati pomoću običnog inverza Gramove matrice. Neka je $m+1$ dimenzija prostora značajki, a N broj primjera u skupu za učenje. **Kada se pseudoinverz $\mathbf{\Phi}^+$ može izračunati pomoću inverza Gramove matrice?**

A Ako je $N = m+1$ i Gramova matrica je kvadratna

B Ako $N \geq m+1$ i stupci u $\mathbf{\Phi}$ nisu linearno zavisni

C Ako $N \leq m+1$ i nema multikolinearnosti značajki

D Ako $N < m+1$ i $\mathbf{\Phi}$ je punog ranga

Cjelina 2: Linearni klasifikacijski modeli (7 pitanja)

7 (N) Raspolažemo sljedećim skupom označenih podataka u dvodimenzijskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((2,5),0),((-8,1),1),((3,9),1)\big\}$$

Na tom skupu treniramo model L2-regularizirane logističke regresije. Koristimo standardni (grupni) gradijentni spust uz faktor regularizacije $\lambda = 10$. Koristimo fiksnu stopu učenja $\eta = 0.01$ (ne koristimo linijsko pretraživanje). Vektor težina $\mathbf{w}$ početno je inicijaliziran na nul-vektor. Provedite jednu epohu gradijentnog spusta (jedan prolazak kroz sve primjere iz $\mathcal{D}$). **Koji se vektor težina $\mathbf{w}$ dobiva nakon prve epohe gradijentnog spusta?**

A $(-0.005, 0.045, 0.015)$ C $(0.005, -0.035, 0.025)$

B $(0.005, -0.045, -0.015)$ D $(-0.005, 0.035, -0.025)$

8 (N) Raspolažemo označenim skupom primjera iz četiriju klasa ($K = 4$) u dvodimenzijskome ulaznom prostoru ($n = 2$). Na tom skupu treniramo model multinomijalne logističke regresije s funkcijom preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & -1 & -1 & 2 \\ 2 & -4 & -2 & 1 \\ 3 & -4 & 6 & 1 \\ -3 & 0 & 2 & 1 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (2, -3)$ s oznakom $\mathbf{y} = (0, 1, 0, 0)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi primjer $\mathbf{x}$?**

A 51 B 64 C 21 D 13

9 (P) Perceptronom rješavamo višeklasni klasifikacijski sa $K = 5$ klasa u ulaznome prostoru dimenzije $n = 6$. Koristimo shemu jedan-naspram-jedan (OVO). Kako bismo ostvarili linearnu razdvojivost klasa, koristimo funkciju preslikavanja $\boldsymbol{\phi}$ koja odgovara polinomu drugoga stupnja s interakcijskim parovima značajki. **Koliko iznosi ukupan broj parametara koji treba optimirati?**

A 315 B 126 C 140 D 280

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

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

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

A 16 B 6 C 32 D 22

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

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

Na tom skupu trenirali smo dva modela L2-regularizirane logističke regresije, koristeći regularizacijske faktore $\lambda = 1$ odnosno $\lambda = 5$. Neka je $L_\lambda^{(i)}$ gubitak koji primjer $\mathbf{x}^{(i)}$ iz $\mathcal{D}$ nanosi modelu optimiranom uz regularizacijski faktor λ. **Što očekujemo da vrijedi?**

A $L_1^{(5)} < L_5^{(6)}, L_1^{(4)} > L_5^{(3)}$

B $L_1^{(4)} > L_5^{(3)}, L_1^{(2)} > L_5^{(1)}$

C $L_1^{(3)} < L_5^{(4)}, L_1^{(1)} < L_5^{(2)}$

D $L_1^{(6)} > L_5^{(5)}, L_1^{(3)} < L_5^{(4)}$

12 (T) Jedan od nedostataka algoritma perceptrona jest taj što hipoteza ovisi o odabiru početnog vektora težina. **Zašto je tomu tako?**

A Funkcija pogreške nije konveksna funkcija parametara

B Gradijent funkcije gubitka nije nula niti za jednu točku u prostoru parametara

C Funkcija pogreške ima više točaka u kojima ostvaruje minimum

D Gradijent funkcije pogreške jednak je nula ako i samo ako je vrijednost funkcije veća od nule

13 (T) Za optimizaciju parametara logističke regresije umjesto gradijentnog spusta možemo primijeniti Newtonov postupak. **Konceptualno gledano, po čemu se Newtonov postupak razlikuje od gradijentnog spusta?**

A Kod ažuriranja težina u obzir uzima informaciju o promjeni gradijenta

B Rekondicionira Gramovu matricu modifikacijom njezine glavne dijagonale

C Za izračun gradijenta koristi sve primjere iz skupa za učenje

D Umjesto gradijenta funkcije pogreške, izračunava gradijent njezine kvadratne aproksimacije

Cjelina 3: SVM, jezgrene i neparametarske metode (7 pitanja)

14 (T) Razina složenosti modela k-NN može se ugoditi hiperparametrom k. Neka je N broj primjera. **Koja je veza između k, N i razine složenosti modela k-NN?**

A Ako je N fiksan, manji k daje model manje razine složenosti

B Ako N pada, a k je fiksan, razina složenosti raste pa pada

C Ako N raste, s povećanjem k možemo zadržati istu razinu složenosti

D Neovisno o N, s porastom vrijednosti k razina složenosti raste

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

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

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

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

16 (N) Raspolažemo sljedećim skupom označenih podataka:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} -15 & -30 & -15 & +1 \\ 6 & 12 & 6 & +1 \\ -5 & -8 & -3 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right)$$

Na ovom skupu treniramo SVM s linearnom jezgrom. Koristimo stohastički gradijentni spust. Vektor dualnih koeficijenata je $\boldsymbol{\alpha} = (0.06, 2.44, 0.0, 2.5)$. Jedan potporni vektor nalazi se unutar margine. Izračunajte vrijednost regularizacijskog izraza u funkciji pogreške SVM-a za parametre $\mathbf{w}$ koji odgovaraju dualnim koeficijentima $\boldsymbol{\alpha}$. **Koliko iznosi regularizacijski izraz?**

A 0.50 B 0.88 C 0.63 D 0.75

17 (T) Tradicionalan način optimizacije parametara stroja potpornih vektora jest primjena konveksne optimizacije uz ograničenja. Razmotrite problem meke margine sa N primjera u n-dimenzijskome ulaznom prostoru. Optimizacija se može provesti u primarnoj ili dualnoj formi. **Koliko se optimizacijskih varijabli te koliko ograničenja javlja u dualnoj formulaciji problema meke margine?**

A $n+1$ varijabli, N ograničenja nejednakosti

B N varijabli, N ograničenja nejednakosti, 1 ograničenje jednakosti

C $n+N+1$ varijabli, $2N$ ograničenja nejednakosti

D N varijabli, $2N$ ograničenja nejednakosti, 1 ograničenje jednakosti

18 (N) Raspolažemo sljedećim skupom označenih primjera iz trodimenzijskoga ulaznog prostora:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((9,1,3),+1), ((-1,7,3),-1), ((0,-8,1),+1)\big\}$$

Na ovome skupu trenirali smo SVM s inverznom kvadratnom jezgrom $\kappa(\mathbf{x},\mathbf{z}) = (1+\|\mathbf{x}-\mathbf{z}\|^2)^{-1}$ i hiperparametrom $C = 10$. Dobiveni Lagrangeovi koeficijenti su $\boldsymbol{\alpha} = (0.6728, 1.3400, 0.6678)$. Može se izračunati da vrijedi $w_0 = 0.333$. Zanima nas klasifikacija primjera $\mathbf{x} = (8,3,1)$. Primijenite jezgreni trik i izračunajte $h(\mathbf{x})$. **Koliko iznosi $h(\mathbf{x})$?**

A 0.118 B 0.391 C 0.231 D 0.550

19 (P) Rješavamo binaran klasifikacijski problem u ulaznome prostoru dimenzije $n = 100$. Raspolažemo sa $N = 1000$ primjera. Isprobavamo dva klasifikacijska algoritma: (1) L2-regulariziranu logističku regresiju s adaptivnim baznim funkcijama i (2) rijetki jezgreni stroj ostvaren kao L1-regularizirana logistička regresija. Kod prvog modela koristimo $m = 50$ adaptivnih baznih funkcija, pri čemu je svaka adaptivna bazna funkcija ϕ_j i sama jedan model logističke regresije. Kod drugog modela koristimo Gaussove jezgrene funkcije (svaka s istom preciznošću), a nakon treniranja naučen model sastoji se od $N' = 48$ prototipa. Razmotrite ukupan broj parametara i hiperparametara oba modela nakon što su ti modeli naučeni, tj. broj (hiper)parametara koje nužno trebamo pohraniti u datoteku kod serijalizacije modela, a da bismo modele mogli naknadno koristiti za predikciju. Jedan (hiper)parametar odgovara jednom realnom broju. **Koliko iznosi razlika u broju pohranjenih (hiper)parametara za ova dva modela?**

A 152 B 251 C 149 D 701

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

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

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

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

Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

27. studenog 2025.

Strojno učenje 1 – Međuispit

*Ispit sadrži **20** pitanja i ukupno nosi najviše 18 bodova (za 30% bodova na predmetu). Točan odgovor donosi 1 bod, a 1/3 boda oduzima se za netočan odgovor. Ostvareni bodovi iznad 18 se zanemaruju. Trajanje ispita je **150 minuta**.* ***Uz primjerak ispita morate predati papire s jasno vidljivim postupkom rješavanja za sva pitanja na koja ste odgovorili, inače se ispit neće vrednovati.***

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

1 (P) Razmatramo tri modela L2-regularizirane linearne regresije s do na parametre definiranim hipotezama h_1, h_2 i h_3 i s nama nepoznatim regularizacijskim faktorima λ_1, λ_2 odnosno λ_3. Za sva tri modela mjerimo vrijednost (neregularizirane) funkcije kvadratne pogreške $E(h|\mathcal{D})$, i to na skupu za učenje $\mathcal{D}_u$ i na ispitnome skupu D_i. Na skupu za učenje opažamo $E(h_2|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$ i $E(h_3|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, a na skupu za ispitivanje $E(h_1|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$ i $E(h_2|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$. **Koji uređaj očekivano vrijedi između regularizacijskih faktora?**

A $\lambda_2 < \lambda_3 < \lambda_1$ B $\lambda_3 < \lambda_2 < \lambda_1$ C $\lambda_1 < \lambda_2 < \lambda_3$ D $\lambda_3 < \lambda_1 < \lambda_2$

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

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

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

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

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

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

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

Za faktor regularizacije koristimo $\lambda = 10$. Parametre optimiramo u zatvorenoj formi, na temelju inverza modificirane Gramove matrice. Izračunajte parametre $\mathbf{w}^*$ te L2-regulariziranu pogrešku $E_R(\mathbf{w}|\mathcal{D})$. **Koliko iznosi regularizirana pogreška?**

A 6.695 B 12.525 C 10.252 D 8.258

4 (T) Optimizacija parametara hipoteze zasniva se na funkciji empirijske pogreške odnosno funkciji gubitka. **Koja je razlika između funkcije empirijske pogreške i funkcije gubitka?**

A Funkcija empirijske pogreške ovisi o konkretnoj hipotezi, dok funkcija gubitka ne ovisi

B Funkcija gubitka definirana je kao očekivanje funkcije empirijske pogreške na skupu primjera

C Funkcija empirijske pogreške uvijek je nenegativna, dok funkcija gubitka može biti negativna

D Funkcija gubitka odnosi se na jedan primjer, a funkcija empirijske pogreške na skup primjera

5 (P) Rješavamo binaran klasifikacijski problem u prostoru $\mathbb{R}^3$. Skup označenih primjera je:

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

Razmatramo linearan model ($\mathcal{H}_1$) i model kugle ($\mathcal{H}_2$), definirane kao:

$$\mathcal{H}_1 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{x} \geq \theta_0\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}) \in \mathbb{R}^4$$

$$\mathcal{H}_2 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(\mathbf{x}-\tilde{\boldsymbol{\theta}})^{\mathrm{T}}(\mathbf{x}-\tilde{\boldsymbol{\theta}}) \leq \theta_0^2\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}), \theta_0 \in \left[0, \frac{1}{2}\right], \tilde{\boldsymbol{\theta}} \in \mathbb{R}^3$$

Optimizacijski postupak koristi funkciju pogreške $E(\boldsymbol{\theta}|\mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{h(\mathbf{x}^{(i)};\boldsymbol{\theta}) \neq y^{(i)}\}$. Odredite minimalne pogreške $E(\boldsymbol{\theta}_1^*|\mathcal{D})$ i $E(\boldsymbol{\theta}_2^*|\mathcal{D})$ modela $\mathcal{H}_1$ odnosno $\mathcal{H}_2$. **Koliko iznosi apsolutna razlika između vrijednosti tih pogrešaka?**

A 1/3 B 0 C 1/2 D 1/6

6 (T) Rješenje za optimizacijski problem najmanjih kvadrata (OLS) izveli smo u zatvorenoj formi kao pseudoinverz $\boldsymbol{\Phi}^+$ matrice dizajna $\boldsymbol{\Phi}$. U nekim se slučajevima pseudoinverz $\boldsymbol{\Phi}^+$ može izračunati pomoću običnog inverza Gramove matrice. Neka je $m+1$ dimenzija prostora značajki, a N broj primjera u skupu za učenje. **Kada se pseudoinverz $\boldsymbol{\Phi}^+$ može izračunati pomoću inverza Gramove matrice?**

A Ako $N < m+1$ i $\boldsymbol{\Phi}$ je punog ranga

B Ako $N \leq m+1$ i nema multikolinearnosti značajki

C Ako $N \geq m+1$ i stupci u $\boldsymbol{\Phi}$ nisu linearno zavisni

D Ako je $N = m+1$ i Gramova matrica je kvadratna

Cjelina 2: Linearni klasifikacijski modeli (7 pitanja)

7 (T) Za optimizaciju parametara logističke regresije umjesto gradijentnog spusta možemo primijeniti Newtonov postupak. **Konceptualno gledano, po čemu se Newtonov postupak razlikuje od gradijentnog spusta?**

A Kod ažuriranja težina u obzir uzima informaciju o promjeni gradijenta

B Za izračun gradijenta koristi sve primjere iz skupa za učenje

C Rekondicionira Gramovu matricu modifikacijom njezine glavne dijagonale

D Umjesto gradijenta funkcije pogreške, izračunava gradijent njezine kvadratne aproksimacije

8 (P) Perceptronom rješavamo višeklasni klasifikacijski sa $K = 5$ klasa u ulaznome prostoru dimenzije $n = 6$. Koristimo shemu jedan-naspram-jedan (OVO). Kako bismo ostvarili linearnu razdvojivost klasa, koristimo funkciju preslikavanja $\boldsymbol{\phi}$ koja odgovara polinomu drugoga stupnja s interakcijskim parovima značajki. **Koliko iznosi ukupan broj parametara koji treba optimirati?**

A 315 B 126 C 140 D 280

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

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

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

A 22 B 16 C 6 D 32

10 (N) Raspolažemo sljedećim skupom označenih podataka u dvodimenzijskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((2,5),0), ((-8,1),1), ((3,9),1)\big\}$$

Na tom skupu treniramo model L2-regularizirane logističke regresije. Koristimo standardni (grupni) gradijentni spust uz faktor regularizacije $\lambda = 10$. Koristimo fiksnu stopu učenja $\eta = 0.01$ (ne koristimo linijsko pretraživanje). Vektor težina $\mathbf{w}$ početno je inicijaliziran na nul-vektor. Provedite jednu epohu gradijentnog spusta (jedan prolazak kroz sve primjere iz $\mathcal{D}$). **Koji se vektor težina $\mathbf{w}$ dobiva nakon prve epohe gradijentnog spusta?**

A $(-0.005, 0.035, -0.025)$ C $(-0.005, 0.045, 0.015)$

B $(0.005, -0.035, 0.025)$ D $(0.005, -0.045, -0.015)$

11 (N) Raspolažemo označenim skupom primjera iz četiriju klasa ($K = 4$) u dvodimenzijskome ulaznom prostoru ($n = 2$). Na tom skupu treniramo model multinomijalne logističke regresije s funkcijom preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & -1 & -1 & 2 \\ 2 & -4 & -2 & 1 \\ 3 & -4 & 6 & 1 \\ -3 & 0 & 2 & 1 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (2, -3)$ s oznakom $\mathbf{y} = (0, 0, 1, 0)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi primjer x?**

A 21 B 13 C 64 D 51

12 (T) Jedan od nedostataka algoritma perceptrona jest taj što hipoteza ovisi o odabiru početnog vektora težina. **Zašto je tomu tako?**

A Gradijent funkcije pogreške jednak je nula ako i samo ako je vrijednost funkcije veća od nule

B Funkcija pogreške nije konveksna funkcija parametara

C Funkcija pogreške ima više točaka u kojima ostvaruje minimum

D Gradijent funkcije gubitka nije nula niti za jednu točku u prostoru parametara

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

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

Na tom skupu trenirali smo dva modela L2-regularizirane logističke regresije, koristeći regularizacijske faktore $\lambda = 1$ odnosno $\lambda = 5$. Neka je $L_\lambda^{(i)}$ gubitak koji primjer $\mathbf{x}^{(i)}$ iz $\mathcal{D}$ nanosi modelu optimiranom uz regularizacijski faktor λ. **Što očekujemo da vrijedi?**

A $L_1^{(3)} > L_5^{(6)}, L_1^{(5)} < L_5^{(2)}$ C $L_1^{(6)} < L_5^{(3)}, L_1^{(2)} > L_5^{(5)}$

B $L_1^{(5)} < L_5^{(2)}, L_1^{(4)} < L_5^{(1)}$ D $L_1^{(2)} > L_5^{(5)}, L_1^{(1)} > L_5^{(4)}$

Cjelina 3: SVM, jezgrene i neparametarske metode (7 pitanja)

14 (T) Tradicionalan način optimizacije parametara stroja potpornih vektora jest primjena konveksne optimizacije uz ograničenja. Razmotrite problem meke margine sa N primjera u n-dimenzijskome ulaznom prostoru. Optimizacija se može provesti u primarnoj ili dualnoj formi. **Koliko se optimizacijskih varijabli te koliko ograničenja javlja u dualnoj formulaciji problema meke margine?**

A $n + N + 1$ varijabli, $2N$ ograničenja nejednakosti

B N varijabli, N ograničenja nejednakosti, 1 ograničenje jednakosti

C N varijabli, $2N$ ograničenja nejednakosti, 1 ograničenje jednakosti

D $n + 1$ varijabli, N ograničenja nejednakosti

15 (P) Rješavamo binaran klasifikacijski problem u ulaznome prostoru dimenzije $n = 100$. Raspolažemo sa $N = 1000$ primjera. Isprobavamo dva klasifikacijska algoritma: (1) L2-regulariziranu logističku regresiju s adaptivnim baznim funkcijama i (2) rijetki jezgreni stroj ostvaren kao L1-regularizirana logistička regresija. Kod prvog modela koristimo $m = 50$ adaptivnih baznih funkcija, pri čemu je svaka adaptivna bazna funkcija ϕ_j i sama jedan model logističke regresije. Kod drugog modela koristimo Gaussove jezgrene funkcije (svaka s istom preciznošću), a nakon treniranja naučen model sastoji se od $N' = 48$ prototipa. Razmotrite ukupan broj parametara i hiperparametara oba modela nakon što su ti modeli naučeni, tj. broj (hiper)parametara koje nužno trebamo pohraniti u datoteku kod serijalizacije modela, a da bismo modele mogli naknadno koristiti za predikciju. Jedan (hiper)parametar odgovara jednom realnom broju. **Koliko iznosi razlika u broju pohranjenih (hiper)parametara za ova dva modela?**

A 149 B 152 C 701 D 251

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

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

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

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

17 (T) Razina složenosti modela k-NN može se ugoditi hiperparametrom k. Neka je N broj primjera. **Koja je veza između k, N i razine složenosti modela k-NN?**

A Ako je N fiksan, manji k daje model manje razine složenosti

B Ako N raste, s povećanjem k možemo zadržati istu razinu složenosti

C Ako N pada, a k je fiksan, razina složenosti raste pa pada

D Neovisno o N, s porastom vrijednosti k razina složenosti raste

18 (N) Raspolažemo sljedećim skupom označenih primjera iz trodimenzijskoga ulaznog prostora:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((9,1,3),+1),((-1,7,3),-1),((0,-8,1),+1)\big\}$$

Na ovome skupu trenirali smo SVM s inverznom kvadratnom jezgrom $\kappa(\mathbf{x},\mathbf{z}) = (1+\|\mathbf{x}-\mathbf{z}\|^2)^{-1}$ i hiperparametrom $C = 10$. Dobiveni Lagrangeovi koeficijenti su $\boldsymbol{\alpha} = (0.6728, 1.3400, 0.6678)$. Može se izračunati da vrijedi $w_0 = 0.333$. Zanima nas klasifikacija primjera $\mathbf{x} = (8,3,1)$. Primijenite jezgreni trik i izračunajte $h(\mathbf{x})$. **Koliko iznosi $h(\mathbf{x})$?**

A 0.550 B 0.391 C 0.231 D 0.118

19 (N) Raspolažemo sljedećim skupom označenih podataka:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} -15 & -30 & -15 & +1 \\ 6 & 12 & 6 & +1 \\ -5 & -8 & -3 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right)$$

Na ovom skupu treniramo SVM s linearnom jezgrom. Koristimo stohastički gradijentni spust. Vektor dualnih koeficijenata je $\boldsymbol{\alpha} = (0.083, 3.417, 0.0, 3.5)$. Jedan potporni vektor je unutar margine. Izračunajte vrijednost regularizacijskog izraza u funkciji pogreške SVM-a za parametre $\mathbf{w}$ koji odgovaraju dualnim koeficijentima $\boldsymbol{\alpha}$. **Koliko iznosi regularizacijski izraz?**

A 0.75 B 0.63 C 0.88 D 0.50

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

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

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

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

Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

27. studenog 2025.

Strojno učenje 1 – Međuispit

*Ispit sadrži **20** pitanja i ukupno nosi najviše 18 bodova (za 30% bodova na predmetu). Točan odgovor donosi 1 bod, a 1/3 boda oduzima se za netočan odgovor. Ostvareni bodovi iznad 18 se zanemaruju. Trajanje ispita je **150 minuta**.* ***Uz primjerak ispita morate predati papire s jasno vidljivim postupkom rješavanja za sva pitanja na koja ste odgovorili, inače se ispit neće vrednovati.***

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

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

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

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

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

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

2 (P) Razmatramo tri modela L2-regularizirane linearne regresije s do na parametre definiranim hipotezama h_1, h_2 i h_3 i s nama nepoznatim regularizacijskim faktorima λ_1, λ_2 odnosno λ_3. Za sva tri modela mjerimo vrijednost (neregularizirane) funkcije kvadratne pogreške $E(h|\mathcal{D})$, i to na skupu za učenje $\mathcal{D}_u$ i na ispitnome skupu D_i. Na skupu za učenje opažamo $E(h_2|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$ i $E(h_3|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, a na skupu za ispitivanje $E(h_1|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$ i $E(h_2|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$. **Koji uređaj očekivano vrijedi između regularizacijskih faktora?**

A $\lambda_3 < \lambda_2 < \lambda_1$ B $\lambda_3 < \lambda_1 < \lambda_2$ C $\lambda_1 < \lambda_2 < \lambda_3$ D $\lambda_2 < \lambda_3 < \lambda_1$

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

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

Za faktor regularizacije koristimo $\lambda = 10$. Parametre optimiramo u zatvorenoj formi, na temelju inverza modificirane Gramove matrice. Izračunajte parametre $\mathbf{w}^*$ te L2-regulariziranu pogrešku $E_R(\mathbf{w}|\mathcal{D})$. **Koliko iznosi regularizirana pogreška?**

A 8.258 B 10.252 C 12.525 D 6.695

4 (T) Rješenje za optimizacijski problem najmanjih kvadrata (OLS) izveli smo u zatvorenoj formi kao pseudoinverz $\mathbf{\Phi}^+$ matrice dizajna $\mathbf{\Phi}$. U nekim se slučajevima pseudoinverz $\mathbf{\Phi}^+$ može izračunati pomoću običnog inverza Gramove matrice. Neka je $m+1$ dimenzija prostora značajki, a N broj primjera u skupu za učenje. **Kada se pseudoinverz $\mathbf{\Phi}^+$ može izračunati pomoću inverza Gramove matrice?**

A Ako je $N = m+1$ i Gramova matrica je kvadratna

B Ako $N \geq m+1$ i stupci u $\mathbf{\Phi}$ nisu linearno zavisni

C Ako $N < m+1$ i $\mathbf{\Phi}$ je punog ranga

D Ako $N \leq m+1$ i nema multikolinearnosti značajki

5 (P) Rješavamo binaran klasifikacijski problem u prostoru $\mathbb{R}^3$. Skup označenih primjera je:

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

Razmatramo linearan model ($\mathcal{H}_1$) i model kugle ($\mathcal{H}_2$), definirane kao:

$$\mathcal{H}_1 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{x} \geq \theta_0\big\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}) \in \mathbb{R}^4$$

$$\mathcal{H}_2 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{(\mathbf{x}-\tilde{\boldsymbol{\theta}})^{\mathrm{T}}(\mathbf{x}-\tilde{\boldsymbol{\theta}}) \leq \theta_0^2\big\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}), \theta_0 \in \big[0, \frac{1}{2}\big], \tilde{\boldsymbol{\theta}} \in \mathbb{R}^3$$

Optimizacijski postupak koristi funkciju pogreške $E(\boldsymbol{\theta}|\mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{h(\mathbf{x}^{(i)};\boldsymbol{\theta}) \neq y^{(i)}\}$. Odredite minimalne pogreške $E(\boldsymbol{\theta}_1^*|\mathcal{D})$ i $E(\boldsymbol{\theta}_2^*|\mathcal{D})$ modela $\mathcal{H}_1$ odnosno $\mathcal{H}_2$. **Koliko iznosi apsolutna razlika između vrijednosti tih pogrešaka?**

A 1/3 B 1/6 C 0 D 1/2

6 (T) Optimizacija parametara hipoteze zasniva se na funkciji empirijske pogreške odnosno funkciji gubitka. **Koja je razlika između funkcije empirijske pogreške i funkcije gubitka?**

A Funkcija gubitka odnosi se na jedan primjer, a funkcija empirijske pogreške na skup primjera

B Funkcija empirijske pogreške uvijek je nenegativna, dok funkcija gubitka može biti negativna

C Funkcija empirijske pogreške ovisi o konkretnoj hipotezi, dok funkcija gubitka ne ovisi

D Funkcija gubitka definirana je kao očekivanje funkcije empirijske pogreške na skupu primjera

Cjelina 2: Linearni klasifikacijski modeli (7 pitanja)

7 (T) Za optimizaciju parametara logističke regresije umjesto gradijentnog spusta možemo primijeniti Newtonov postupak. **Konceptualno gledano, po čemu se Newtonov postupak razlikuje od gradijentnog spusta?**

A Kod ažuriranja težina u obzir uzima informaciju o promjeni gradijenta

B Umjesto gradijenta funkcije pogreške, izračunava gradijent njezine kvadratne aproksimacije

C Rekondicionira Gramovu matricu modifikacijom njezine glavne dijagonale

D Za izračun gradijenta koristi sve primjere iz skupa za učenje

8 (N) Raspolažemo označenim skupom primjera iz četiriju klasa ($K = 4$) u dvodimenzijskome ulaznom prostoru ($n = 2$). Na tom skupu treniramo model multinomijalne logističke regresije s funkcijom preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & -1 & -1 & 2 \\ 2 & -4 & -2 & 1 \\ 3 & -4 & 6 & 1 \\ -3 & 0 & 2 & 1 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (2, -4)$ s oznakom $\mathbf{y} = (0, 0, 1, 0)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi primjer x?**

A 13 B 21 C 64 D 51

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

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

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

A 6 B 16 C 32 D 22

10 (N) Raspolažemo sljedećim skupom označenih podataka u dvodimenzijskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((2,5),1),((-8,1),1),((3,9),0)\big\}$$

Na tom skupu treniramo model L2-regularizirane logističke regresije. Koristimo standardni (grupni) gradijentni spust uz faktor regularizacije $\lambda = 10$. Koristimo fiksnu stopu učenja $\eta = 0.01$ (ne koristimo linijsko pretraživanje). Vektor težina $\mathbf{w}$ početno je inicijaliziran na nul-vektor. Provedite jednu epohu gradijentnog spusta (jedan prolazak kroz sve primjere iz $\mathcal{D}$). **Koji se vektor težina w dobiva nakon prve epohe gradijentnog spusta?**

A $(-0.005, 0.035, -0.025)$ C $(-0.005, 0.045, 0.015)$
B $(0.005, -0.045, -0.015)$ D $(0.005, -0.035, 0.025)$

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

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

Na tom skupu trenirali smo dva modela L2-regularizirane logističke regresije, koristeći regularizacijske faktore $\lambda = 1$ odnosno $\lambda = 5$. Neka je $L_\lambda^{(i)}$ gubitak koji primjer $\mathbf{x}^{(i)}$ iz $\mathcal{D}$ nanosi modelu optimiranom uz regularizacijski faktor λ. **Što očekujemo da vrijedi?**

A $L_1^{(6)} < L_5^{(3)}, L_1^{(2)} > L_5^{(5)}$ C $L_1^{(2)} > L_5^{(5)}, L_1^{(1)} > L_5^{(4)}$
B $L_1^{(3)} > L_5^{(6)}, L_1^{(5)} < L_5^{(2)}$ D $L_1^{(5)} < L_5^{(2)}, L_1^{(4)} < L_5^{(1)}$

12 (P) Perceptronom rješavamo višeklasni klasifikacijski sa $K = 6$ klasa u ulaznome prostoru dimenzije $n = 5$. Koristimo shemu jedan-naspram-jedan (OVO). Kako bismo ostvarili linearnu razdvojivost klasa, koristimo funkciju preslikavanja $\boldsymbol{\phi}$ koja odgovara polinomu drugoga stupnja s interakcijskim parovima značajki. **Koliko iznosi ukupan broj parametara koji treba optimirati?**

A 140 B 315 C 280 D 126

13 (T) Jedan od nedostataka algoritma perceptrona jest taj što hipoteza ovisi o odabiru početnog vektora težina. **Zašto je tomu tako?**

A Funkcija pogreške nije konveksna funkcija parametara

B Funkcija pogreške ima više točaka u kojima ostvaruje minimum

C Gradijent funkcije pogreške jednak je nula ako i samo ako je vrijednost funkcije veća od nule

D Gradijent funkcije gubitka nije nula niti za jednu točku u prostoru parametara

Cjelina 3: SVM, jezgrene i neparametarske metode (7 pitanja)

14 (P) Rješavamo binaran klasifikacijski problem u ulaznome prostoru dimenzije $n = 100$. Raspolažemo sa $N = 1000$ primjera. Isprobavamo dva klasifikacijska algoritma: (1) L2-regulariziranu logističku regresiju s adaptivnim baznim funkcijama i (2) rijetki jezgreni stroj ostvaren kao L1-regularizirana logistička regresija. Kod prvog modela koristimo $m = 50$ adaptivnih baznih funkcija, pri čemu je svaka adaptivna bazna funkcija ϕ_j i sama jedan model logističke regresije. Kod drugog modela koristimo Gaussove jezgrene funkcije (svaka s istom preciznošću), a nakon treniranja naučen model sastoji se od $N' = 48$ prototipa. Razmotrite ukupan broj parametara i hiperparametara oba modela nakon što su ti modeli naučeni, tj. broj (hiper)parametara koje nužno trebamo pohraniti u datoteku kod serijalizacije modela, a da bismo modele mogli naknadno koristiti za predikciju. Jedan (hiper)parametar odgovara jednom realnom broju. **Koliko iznosi razlika u broju pohranjenih (hiper)parametara za ova dva modela?**

A 701 B 149 C 152 D 251

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

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

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

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

16 (T) Razina složenosti modela k-NN može se ugoditi hiperparametrom k. Neka je N broj primjera. **Koja je veza između k, N i razine složenosti modela k-NN?**

A Ako N raste, s povećanjem k možemo zadržati istu razinu složenosti

B Neovisno o N, s porastom vrijednosti k razina složenosti raste

C Ako N pada, a k je fiksan, razina složenosti raste pa pada

D Ako je N fiksan, manji k daje model manje razine složenosti

17 (N) Raspolažemo sljedećim skupom označenih podataka:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} -15 & -30 & -15 & +1 \\ 6 & 12 & 6 & +1 \\ -5 & -8 & -3 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right)$$

Na ovom skupu treniramo SVM s linearnom jezgrom. Koristimo stohastički gradijentni spust. Vektor dualnih koeficijenata je $\boldsymbol{\alpha} = (0.083, 3.417, 0.0, 3.5)$. Jedan potporni vektor je unutar margine. Izračunajte vrijednost regularizacijskog izraza u funkciji pogreške SVM-a za parametre $\mathbf{w}$ koji odgovaraju dualnim koeficijentima $\boldsymbol{\alpha}$. **Koliko iznosi regularizacijski izraz?**

A 0.88 B 0.50 C 0.75 D 0.63

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

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

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

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

19 (N) Raspolažemo sljedećim skupom označenih primjera iz trodimenzijskoga ulaznog prostora:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \left\{((9,1,3),+1), ((-1,7,3),-1), ((0,-8,1),+1)\right\}$$

Na ovome skupu trenirali smo SVM s inverznom kvadratnom jezgrom $\kappa(\mathbf{x}, \mathbf{z}) = (1+\|\mathbf{x}-\mathbf{z}\|^2)^{-1}$ i hiperparametrom $C = 10$. Dobiveni Lagrangeovi koeficijenti su $\boldsymbol{\alpha} = (0.6728, 1.3400, 0.6678)$. Može se izračunati da vrijedi $w_0 = 0.333$. Zanima nas klasifikacija primjera $\mathbf{x} = (-1, 6, 5)$. Primijenite jezgreni trik i izračunajte $h(\mathbf{x})$. **Koliko iznosi $h(\mathbf{x})$?**

A 0.118 B 0.550 C 0.231 D 0.391

20 (T) Tradicionalan način optimizacije parametara stroja potpornih vektora jest primjena konveksne optimizacije uz ograničenja. Razmotrite problem meke margine sa N primjera u n-dimenzijskome ulaznom prostoru. Optimizacija se može provesti u primarnoj ili dualnoj formi. **Koliko se optimizacijskih varijabli te koliko ograničenja javlja u dualnoj formulaciji problema meke margine?**

A N varijabli, N ograničenja nejednakosti, 1 ograničenje jednakosti

B N varijabli, $2N$ ograničenja nejednakosti, 1 ograničenje jednakosti

C $n + 1$ varijabli, N ograničenja nejednakosti

D $n + N + 1$ varijabli, $2N$ ograničenja nejednakosti

Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

27. studenog 2025.

Strojno učenje 1 – Međuispit

*Ispit sadrži **20** pitanja i ukupno nosi najviše 18 bodova (za 30% bodova na predmetu). Točan odgovor donosi 1 bod, a 1/3 boda oduzima se za netočan odgovor. Ostvareni bodovi iznad 18 se zanemaruju. Trajanje ispita je **150 minuta**.* ***Uz primjerak ispita morate predati papire s jasno vidljivim postupkom rješavanja za sva pitanja na koja ste odgovorili, inače se ispit neće vrednovati.***

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

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

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

2 (T) Optimizacija parametara hipoteze zasniva se na funkciji empirijske pogreške odnosno funkciji gubitka. **Koja je razlika između funkcije empirijske pogreške i funkcije gubitka?**

- A Funkcija empirijske pogreške ovisi o konkretnoj hipotezi, dok funkcija gubitka ne ovisi
- B Funkcija gubitka odnosi se na jedan primjer, a funkcija empirijske pogreške na skup primjera
- C Funkcija empirijske pogreške uvijek je nenegativna, dok funkcija gubitka može biti negativna
- D Funkcija gubitka definirana je kao očekivanje funkcije empirijske pogreške na skupu primjera

3 (T) Rješenje za optimizacijski problem najmanjih kvadrata (OLS) izveli smo u zatvorenoj formi kao pseudoinverz $\mathbf{\Phi}^+$ matrice dizajna $\mathbf{\Phi}$. U nekim se slučajevima pseudoinverz $\mathbf{\Phi}^+$ može izračunati pomoću običnog inverza Gramove matrice. Neka je $m+1$ dimenzija prostora značajki, a N broj primjera u skupu za učenje. **Kada se pseudoinverz $\mathbf{\Phi}^+$ može izračunati pomoću inverza Gramove matrice?**

- A Ako $N \leq m+1$ i nema multikolinearnosti značajki
- B Ako $N < m+1$ i $\mathbf{\Phi}$ je punog ranga
- C Ako $N \geq m+1$ i stupci u $\mathbf{\Phi}$ nisu linearno zavisni
- D Ako je $N = m+1$ i Gramova matrica je kvadratna

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

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

Za faktor regularizacije koristimo $\lambda = 10$. Parametre optimiramo u zatvorenoj formi, na temelju inverza modificirane Gramove matrice. Izračunajte parametre $\mathbf{w}^*$ te L2-regulariziranu pogrešku $E_R(\mathbf{w}|\mathcal{D})$. **Koliko iznosi regularizirana pogreška?**

A 8.258 B 6.695 C 10.252 D 12.525

5 (P) Rješavamo binaran klasifikacijski problem u prostoru $\mathbb{R}^3$. Skup označenih primjera je:

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

Razmatramo linearan model ($\mathcal{H}_1$) i model kugle ($\mathcal{H}_2$), definirane kao:

$$\mathcal{H}_1 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{x} \geq \theta_0\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}) \in \mathbb{R}^4$$

$$\mathcal{H}_2 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{(\mathbf{x}-\tilde{\boldsymbol{\theta}})^{\mathrm{T}}(\mathbf{x}-\tilde{\boldsymbol{\theta}}) \geq \theta_0^2\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}), \theta_0 \in \big[0, \frac{1}{2}\big], \tilde{\boldsymbol{\theta}} \in \mathbb{R}^3$$

Optimizacijski postupak koristi funkciju pogreške $E(\boldsymbol{\theta}|\mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{h(\mathbf{x}^{(i)};\boldsymbol{\theta}) \neq y^{(i)}\}$. Odredite minimalne pogreške $E(\boldsymbol{\theta}_1^*|\mathcal{D})$ i $E(\boldsymbol{\theta}_2^*|\mathcal{D})$ modela $\mathcal{H}_1$ odnosno $\mathcal{H}_2$. **Koliko iznosi apsolutna razlika između vrijednosti tih pogrešaka?**

A 1/6 B 0 C 1/2 D 1/3

6 (P) Razmatramo tri modela L2-regularizirane linearne regresije s do na parametre definiranim hipotezama h_1, h_2 i h_3 i s nama nepoznatim regularizacijskim faktorima λ_1, λ_2 odnosno λ_3. Za sva tri modela mjerimo vrijednost (neregularizirane) funkcije kvadratne pogreške $E(h|\mathcal{D})$, i to na skupu za učenje $\mathcal{D}_u$ i na ispitnome skupu D_i. Na skupu za učenje opažamo $E(h_2|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$ i $E(h_3|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, a na skupu za ispitivanje $E(h_1|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$ i $E(h_2|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$. **Koji uređaj očekivano vrijedi između regularizacijskih faktora?**

A $\lambda_3 < \lambda_1 < \lambda_2$ B $\lambda_1 < \lambda_2 < \lambda_3$ C $\lambda_2 < \lambda_3 < \lambda_1$ D $\lambda_3 < \lambda_2 < \lambda_1$

Cjelina 2: Linearni klasifikacijski modeli (7 pitanja)

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

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

Na tom skupu trenirali smo dva modela L2-regularizirane logističke regresije, koristeći regularizacijske faktore $\lambda = 1$ odnosno $\lambda = 5$. Neka je $L_\lambda^{(i)}$ gubitak koji primjer $\mathbf{x}^{(i)}$ iz $\mathcal{D}$ nanosi modelu optimiranom uz regularizacijski faktor λ. **Što očekujemo da vrijedi?**

A $L_1^{(6)} < L_5^{(3)}, L_1^{(2)} > L_5^{(5)}$ C $L_1^{(3)} > L_5^{(6)}, L_1^{(5)} < L_5^{(2)}$

B $L_1^{(2)} > L_5^{(5)}, L_1^{(1)} > L_5^{(4)}$ D $L_1^{(5)} < L_5^{(2)}, L_1^{(4)} < L_5^{(1)}$

8 (T) Jedan od nedostataka algoritma perceptrona jest taj što hipoteza ovisi o odabiru početnog vektora težina. **Zašto je tomu tako?**

A Gradijent funkcije pogreške jednak je nula ako i samo ako je vrijednost funkcije veća od nule

B Funkcija pogreške nije konveksna funkcija parametara

C Gradijent funkcije gubitka nije nula niti za jednu točku u prostoru parametara

D Funkcija pogreške ima više točaka u kojima ostvaruje minimum

9 (T) Za optimizaciju parametara logističke regresije umjesto gradijentnog spusta možemo primijeniti Newtonov postupak. **Konceptualno gledano, po čemu se Newtonov postupak razlikuje od gradijentnog spusta?**

A Za izračun gradijenta koristi sve primjere iz skupa za učenje

B Rekondicionira Gramovu matricu modifikacijom njezine glavne dijagonale

C Kod ažuriranja težina u obzir uzima informaciju o promjeni gradijenta

D Umjesto gradijenta funkcije pogreške, izračunava gradijent njezine kvadratne aproksimacije

10 (N) Raspolažemo označenim skupom primjera iz četiriju klasa ($K = 4$) u dvodimenzijskome ulaznom prostoru ($n = 2$). Na tom skupu treniramo model multinomijalne logističke regresije s funkcijom preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & -1 & -1 & 2 \\ 2 & -4 & -2 & 1 \\ 3 & -4 & 6 & 1 \\ -3 & 0 & 2 & 1 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (2, -3)$ s oznakom $\mathbf{y} = (0, 0, 0, 1)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi primjer x?**

A 51 B 21 C 13 D 64

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

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

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

A 32 B 6 C 22 D 16

12 (P) Perceptronom rješavamo višeklasni klasifikacijski sa $K = 6$ klasa u ulaznome prostoru dimenzije $n = 5$. Koristimo shemu jedan-naspram-jedan (OVO). Kako bismo ostvarili linearnu razdvojivost klasa, koristimo funkciju preslikavanja $\boldsymbol{\phi}$ koja odgovara polinomu drugoga stupnja s interakcijskim parovima značajki. **Koliko iznosi ukupan broj parametara koji treba optimirati?**

A 126 B 140 C 315 D 280

13 (N) Raspolažemo sljedećim skupom označenih podataka u dvodimenzijskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((2, 5), 1), ((-8, 1), 0), ((3, 9), 0)\big\}$$

Na tom skupu treniramo model L2-regularizirane logističke regresije. Koristimo standardni (grupni) gradijentni spust uz faktor regularizacije $\lambda = 10$. Koristimo fiksnu stopu učenja $\eta = 0.01$ (ne koristimo linijsko pretraživanje). Vektor težina $\mathbf{w}$ početno je inicijaliziran na nul-vektor. Provedite jednu epohu gradijentnog spusta (jedan prolazak kroz sve primjere iz $\mathcal{D}$). **Koji se vektor težina w dobiva nakon prve epohe gradijentnog spusta?**

A $(0.005, -0.045, -0.015)$ C $(-0.005, 0.035, -0.025)$
B $(0.005, -0.035, 0.025)$ D $(-0.005, 0.045, 0.015)$

Cjelina 3: SVM, jezgrene i neparametarske metode (7 pitanja)

14 (N) Raspolažemo sljedećim skupom označenih primjera iz trodimenzijskoga ulaznog prostora:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((9, 1, 3), +1), ((-1, 7, 3), -1), ((0, -8, 1), +1)\big\}$$

Na ovome skupu trenirali smo SVM s inverznom kvadratnom jezgrom $\kappa(\mathbf{x}, \mathbf{z}) = (1+\|\mathbf{x}-\mathbf{z}\|^2)^{-1}$ i hiperparametrom $C = 10$. Dobiveni Lagrangeovi koeficijenti su $\boldsymbol{\alpha} = (0.6728, 1.3400, 0.6678)$. Može se izračunati da vrijedi $w_0 = 0.333$. Zanima nas klasifikacija primjera $\mathbf{x} = (9, 2, 4)$. Primijenite jezgreni trik i izračunajte $h(\mathbf{x})$. **Koliko iznosi $h(\mathbf{x})$?**

A 0.231 B 0.550 C 0.391 D 0.118

15 (T) Razina složenosti modela k-NN može se ugoditi hiperparametrom k. Neka je N broj primjera. **Koja je veza između k, N i razine složenosti modela k-NN?**

A Ako N pada, a k je fiksan, razina složenosti raste pa pada

B Neovisno o N, s porastom vrijednosti k razina složenosti raste

C Ako je N fiksan, manji k daje model manje razine složenosti

D Ako N raste, s povećanjem k možemo zadržati istu razinu složenosti

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

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

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

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

17 (T) Tradicionalan način optimizacije parametara stroja potpornih vektora jest primjena konveksne optimizacije uz ograničenja. Razmotrite problem meke margine sa N primjera u n-dimenzijskome ulaznom prostoru. Optimizacija se može provesti u primarnoj ili dualnoj formi. **Koliko se optimizacijskih varijabli te koliko ograničenja javlja u dualnoj formulaciji problema meke margine?**

A $n + 1$ varijabli, N ograničenja nejednakosti

B N varijabli, $2N$ ograničenja nejednakosti, 1 ograničenje jednakosti

C N varijabli, N ograničenja nejednakosti, 1 ograničenje jednakosti

D $n + N + 1$ varijabli, $2N$ ograničenja nejednakosti

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

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

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

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

19 (P) Rješavamo binaran klasifikacijski problem u ulaznome prostoru dimenzije $n = 100$. Raspolažemo sa $N = 1000$ primjera. Isprobavamo dva klasifikacijska algoritma: (1) L2-regulariziranu logističku regresiju s adaptivnim baznim funkcijama i (2) rijetki jezgreni stroj ostvaren kao L1-regularizirana logistička regresija. Kod prvog modela koristimo $m = 50$ adaptivnih baznih funkcija, pri čemu je svaka adaptivna bazna funkcija ϕ_j i sama jedan model logističke regresije. Kod drugog modela koristimo Gaussove jezgrene funkcije (svaka s istom preciznošću), a nakon treniranja naučen model sastoji se od $N' = 48$ prototipa. Razmotrite ukupan broj parametara i hiperparametara oba modela nakon što su ti modeli naučeni, tj. broj (hiper)parametara koje nužno trebamo pohraniti u datoteku kod serijalizacije modela, a da bismo modele mogli naknadno koristiti za predikciju. Jedan (hiper)parametar odgovara jednom realnom broju. **Koliko iznosi razlika u broju pohranjenih (hiper)parametara za ova dva modela?**

A 251 B 149 C 701 D 152

20 (N) Raspolažemo sljedećim skupom označenih podataka:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} -15 & -30 & -15 & +1 \\ 6 & 12 & 6 & +1 \\ -5 & -8 & -3 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right)$$

Na ovom skupu treniramo SVM s linearnom jezgrom. Koristimo stohastički gradijentni spust. Vektor dualnih koeficijenata je $\boldsymbol{\alpha} = (0.083, 3.417, 0.0, 3.5)$. Jedan potporni vektor je unutar margine. Izračunajte vrijednost regularizacijskog izraza u funkciji pogreške SVM-a za parametre $\mathbf{w}$ koji odgovaraju dualnim koeficijentima $\boldsymbol{\alpha}$. **Koliko iznosi regularizacijski izraz?**

A 0.50 B 0.63 C 0.75 D 0.88

Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

27. studenog 2025.

Strojno učenje 1 – Međuispit

*Ispit sadrži **20** pitanja i ukupno nosi najviše 18 bodova (za 30% bodova na predmetu). Točan odgovor donosi 1 bod, a 1/3 boda oduzima se za netočan odgovor. Ostvareni bodovi iznad 18 se zanemaruju. Trajanje ispita je **150 minuta**.* ***Uz primjerak ispita morate predati papire s jasno vidljivim postupkom rješavanja za sva pitanja na koja ste odgovorili, inače se ispit neće vrednovati.***

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

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

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

Za faktor regularizacije koristimo $\lambda = 10$. Parametre optimiramo u zatvorenoj formi, na temelju inverza modificirane Gramove matrice. Izračunajte parametre $\mathbf{w}^*$ te L2-regulariziranu pogrešku $E_R(\mathbf{w}|\mathcal{D})$. **Koliko iznosi regularizirana pogreška?**

A 6.695 B 8.258 C 12.525 D 10.252

2 (T) Rješenje za optimizacijski problem najmanjih kvadrata (OLS) izveli smo u zatvorenoj formi kao pseudoinverz $\mathbf{\Phi}^+$ matrice dizajna $\mathbf{\Phi}$. U nekim se slučajevima pseudoinverz $\mathbf{\Phi}^+$ može izračunati pomoću običnog inverza Gramove matrice. Neka je $m+1$ dimenzija prostora značajki, a N broj primjera u skupu za učenje. **Kada se pseudoinverz $\mathbf{\Phi}^+$ može izračunati pomoću inverza Gramove matrice?**

A Ako $N \leq m+1$ i nema multikolinearnosti značajki

B Ako je $N = m+1$ i Gramova matrica je kvadratna

C Ako $N \geq m+1$ i stupci u $\mathbf{\Phi}$ nisu linearno zavisni

D Ako $N < m+1$ i $\mathbf{\Phi}$ je punog ranga

3 (T) Optimizacija parametara hipoteze zasniva se na funkciji empirijske pogreške odnosno funkciji gubitka. **Koja je razlika između funkcije empirijske pogreške i funkcije gubitka?**

A Funkcija gubitka definirana je kao očekivanje funkcije empirijske pogreške na skupu primjera

B Funkcija empirijske pogreške uvijek je nenegativna, dok funkcija gubitka može biti negativna

C Funkcija empirijske pogreške ovisi o konkretnoj hipotezi, dok funkcija gubitka ne ovisi

D Funkcija gubitka odnosi se na jedan primjer, a funkcija empirijske pogreške na skup primjera

4 (P) Razmatramo tri modela L2-regularizirane linearne regresije s do na parametre definiranim hipotezama h_1, h_2 i h_3 i s nama nepoznatim regularizacijskim faktorima λ_1, λ_2 odnosno λ_3. Za sva tri modela mjerimo vrijednost (neregularizirane) funkcije kvadratne pogreške $E(h|\mathcal{D})$, i to na skupu za učenje $\mathcal{D}_u$ i na ispitnome skupu D_i. Na skupu za učenje opažamo $E(h_2|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$ i $E(h_3|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, a na skupu za ispitivanje $E(h_1|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$ i $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$. **Koji uređaj očekivano vrijedi između regularizacijskih faktora?**

A $\lambda_1 < \lambda_3 < \lambda_2$ B $\lambda_3 < \lambda_2 < \lambda_1$ C $\lambda_2 < \lambda_1 < \lambda_3$ D $\lambda_2 < \lambda_3 < \lambda_1$

5 (P) Rješavamo binaran klasifikacijski problem u prostoru $\mathbb{R}^3$. Skup označenih primjera je:

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

Razmatramo linearan model ($\mathcal{H}_1$) i model kugle ($\mathcal{H}_2$), definirane kao:

$$\mathcal{H}_1 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{x} \geq \theta_0\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}) \in \mathbb{R}^4$$

$$\mathcal{H}_2 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(\mathbf{x}-\tilde{\boldsymbol{\theta}})^{\mathrm{T}}(\mathbf{x}-\tilde{\boldsymbol{\theta}}) \geq \theta_0^2\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}), \theta_0 \in \big[0, \frac{1}{2}\big], \tilde{\boldsymbol{\theta}} \in \mathbb{R}^3$$

Optimizacijski postupak koristi funkciju pogreške $E(\boldsymbol{\theta}|\mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{h(\mathbf{x}^{(i)};\boldsymbol{\theta}) \neq y^{(i)}\}$. Odredite minimalne pogreške $E(\boldsymbol{\theta}_1^*|\mathcal{D})$ i $E(\boldsymbol{\theta}_2^*|\mathcal{D})$ modela $\mathcal{H}_1$ odnosno $\mathcal{H}_2$. **Koliko iznosi apsolutna razlika između vrijednosti tih pogrešaka?**

A 1/6 B 0 C 1/3 D 1/2

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

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

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

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

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

Cjelina 2: Linearni klasifikacijski modeli (7 pitanja)

7 (N) Raspolažemo označenim skupom primjera iz četiriju klasa ($K = 4$) u dvodimenzijskome ulaznom prostoru ($n = 2$). Na tom skupu treniramo model multinomijalne logističke regresije s funkcijom preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & -1 & -1 & 2 \\ 2 & -4 & -2 & 1 \\ 3 & -4 & 6 & 1 \\ -3 & 0 & 2 & 1 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (2, -3)$ s oznakom $\mathbf{y} = (0, 0, 0, 1)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi primjer x?**

A 51 B 13 C 21 D 64

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

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

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

A 32 B 22 C 6 D 16

9 (T) Za optimizaciju parametara logističke regresije umjesto gradijentnog spusta možemo primijeniti Newtonov postupak. **Konceptualno gledano, po čemu se Newtonov postupak razlikuje od gradijentnog spusta?**

A Rekondicionira Gramovu matricu modifikacijom njezine glavne dijagonale

B Umjesto gradijenta funkcije pogreške, izračunava gradijent njezine kvadratne aproksimacije

C Kod ažuriranja težina u obzir uzima informaciju o promjeni gradijenta

D Za izračun gradijenta koristi sve primjere iz skupa za učenje

10 (N) Raspolažemo sljedećim skupom označenih podataka u dvodimenzijskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((2,5),1), ((-8,1),1), ((3,9),0)\big\}$$

Na tom skupu treniramo model L2-regularizirane logističke regresije. Koristimo standardni (grupni) gradijentni spust uz faktor regularizacije $\lambda = 10$. Koristimo fiksnu stopu učenja $\eta = 0.01$ (ne koristimo linijsko pretraživanje). Vektor težina $\mathbf{w}$ početno je inicijaliziran na nul-vektor. Provedite jednu epohu gradijentnog spusta (jedan prolazak kroz sve primjere iz $\mathcal{D}$). **Koji se vektor težina w dobiva nakon prve epohe gradijentnog spusta?**

A $(-0.005, 0.035, -0.025)$ C $(0.005, -0.035, 0.025)$
B $(-0.005, 0.045, 0.015)$ D $(0.005, -0.045, -0.015)$

11 (P) Perceptronom rješavamo višeklasni klasifikacijski sa $K = 5$ klasa u ulaznome prostoru dimenzije $n = 6$. Koristimo shemu jedan-naspram-jedan (OVO). Kako bismo ostvarili linearnu razdvojivost klasa, koristimo funkciju preslikavanja $\boldsymbol{\phi}$ koja odgovara polinomu drugoga stupnja s interakcijskim parovima značajki. **Koliko iznosi ukupan broj parametara koji treba optimirati?**

A 315 B 280 C 140 D 126

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

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

Na tom skupu trenirali smo dva modela L2-regularizirane logističke regresije, koristeći regularizacijske faktore $\lambda = 1$ odnosno $\lambda = 5$. Neka je $L_\lambda^{(i)}$ gubitak koji primjer $\mathbf{x}^{(i)}$ iz $\mathcal{D}$ nanosi modelu optimiranom uz regularizacijski faktor λ. **Što očekujemo da vrijedi?**

A $L_1^{(6)} < L_5^{(3)}, L_1^{(2)} > L_5^{(5)}$ C $L_1^{(2)} > L_5^{(5)}, L_1^{(1)} > L_5^{(4)}$
B $L_1^{(3)} > L_5^{(6)}, L_1^{(5)} < L_5^{(2)}$ D $L_1^{(5)} < L_5^{(2)}, L_1^{(4)} < L_5^{(1)}$

13 (T) Jedan od nedostataka algoritma perceptrona jest taj što hipoteza ovisi o odabiru početnog vektora težina. **Zašto je tomu tako?**

A Funkcija pogreške ima više točaka u kojima ostvaruje minimum

B Funkcija pogreške nije konveksna funkcija parametara

C Gradijent funkcije pogreške jednak je nula ako i samo ako je vrijednost funkcije veća od nule

D Gradijent funkcije gubitka nije nula niti za jednu točku u prostoru parametara

Cjelina 3: SVM, jezgrene i neparametarske metode (7 pitanja)

14 (P) Rješavamo binaran klasifikacijski problem u ulaznome prostoru dimenzije $n = 100$. Raspolažemo sa $N = 1000$ primjera. Isprobavamo dva klasifikacijska algoritma: (1) L2-regulariziranu logističku regresiju s adaptivnim baznim funkcijama i (2) rijetki jezgreni stroj ostvaren kao L1-regularizirana logistička regresija. Kod prvog modela koristimo $m = 50$ adaptivnih baznih funkcija, pri čemu je svaka adaptivna bazna funkcija ϕ_j i sama jedan model logističke regresije. Kod drugog modela koristimo Gaussove jezgrene funkcije (svaka s istom preciznošću), a nakon treniranja naučen model sastoji se od $N' = 32$ prototipa. Razmotrite ukupan broj parametara i hiperparametara oba modela nakon što su ti modeli naučeni, tj. broj (hiper)parametara koje nužno trebamo pohraniti u datoteku kod serijalizacije modela, a da bismo modele mogli naknadno koristiti za predikciju. Jedan (hiper)parametar odgovara jednom realnom broju. **Koliko iznosi razlika u broju pohranjenih (hiper)parametara za ova dva modela?**

A 899 B 1749 C 1867 D 1784

15 (N) Raspolažemo sljedećim skupom označenih podataka:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} -15 & -30 & -15 & +1 \\ 6 & 12 & 6 & +1 \\ -5 & -8 & -3 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right)$$

Na ovom skupu treniramo SVM s linearnom jezgrom. Koristimo stohastički gradijentni spust. Vektor dualnih koeficijenata je $\boldsymbol{\alpha} = (0.071, 2.929, 0.0, 3.0)$. Jedan potporni vektor nalazi se unutar margine. Izračunajte vrijednost regularizacijskog izraza u funkciji pogreške SVM-a za parametre $\mathbf{w}$ koji odgovaraju dualnim koeficijentima $\boldsymbol{\alpha}$. **Koliko iznosi regularizacijski izraz?**

A 0.88 B 0.75 C 0.50 D 0.63

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

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

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

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

17 (N) Raspolažemo sljedećim skupom označenih primjera iz trodimenzijskoga ulaznog prostora:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((9,1,3),+1), ((-1,7,3),-1), ((0,-8,1),+1)\big\}$$

Na ovome skupu trenirali smo SVM s inverznom kvadratnom jezgrom $\kappa(\mathbf{x}, \mathbf{z}) = (1+\|\mathbf{x}-\mathbf{z}\|^2)^{-1}$ i hiperparametrom $C = 10$. Dobiveni Lagrangeovi koeficijenti su $\boldsymbol{\alpha} = (0.6728, 1.3400, 0.6678)$. Može se izračunati da vrijedi $w_0 = 0.333$. Zanima nas klasifikacija primjera $\mathbf{x} = (-2, 4, 4)$. Primijenite jezgreni trik i izračunajte $h(\mathbf{x})$. **Koliko iznosi $h(\mathbf{x})$?**

A 0.550 B 0.118 C 0.231 D 0.391

18 (T) Tradicionalan način optimizacije parametara stroja potpornih vektora jest primjena konveksne optimizacije uz ograničenja. Razmotrite problem meke margine sa N primjera u n-dimenzijskome ulaznom prostoru. Optimizacija se može provesti u primarnoj ili dualnoj formi. **Koliko se optimizacijskih varijabli te koliko ograničenja javlja u dualnoj formulaciji problema meke margine?**

A N varijabli, N ograničenja nejednakosti, 1 ograničenje jednakosti

B $n + 1$ varijabli, N ograničenja nejednakosti

C N varijabli, $2N$ ograničenja nejednakosti, 1 ograničenje jednakosti

D $n + N + 1$ varijabli, $2N$ ograničenja nejednakosti

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

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

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

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

20 (T) Razina složenosti modela k-NN može se ugoditi hiperparametrom k. Neka je N broj primjera. **Koja je veza između k, N i razine složenosti modela k-NN?**

A Ako N pada, a k je fiksan, razina složenosti raste pa pada

B Ako je N fiksan, manji k daje model manje razine složenosti

C Ako N raste, s povećanjem k možemo zadržati istu razinu složenosti

D Neovisno o N, s porastom vrijednosti k razina složenosti raste

Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

27. studenog 2025.

Strojno učenje 1 – Međuispit

*Ispit sadrži **20** pitanja i ukupno nosi najviše 18 bodova (za 30% bodova na predmetu). Točan odgovor donosi 1 bod, a 1/3 boda oduzima se za netočan odgovor. Ostvareni bodovi iznad 18 se zanemaruju. Trajanje ispita je **150 minuta**.* ***Uz primjerak ispita morate predati papire s jasno vidljivim postupkom rješavanja za sva pitanja na koja ste odgovorili, inače se ispit neće vrednovati.***

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

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

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

Za faktor regularizacije koristimo $\lambda = 10$. Parametre optimiramo u zatvorenoj formi, na temelju inverza modificirane Gramove matrice. Izračunajte parametre $\mathbf{w}^*$ te L2-regulariziranu pogrešku $E_R(\mathbf{w}|\mathcal{D})$. **Koliko iznosi regularizirana pogreška?**

A 8.258 B 10.252 C 6.695 D 12.525

2 (P) Rješavamo binaran klasifikacijski problem u prostoru $\mathbb{R}^3$. Skup označenih primjera je:

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

Razmatramo linearan model ($\mathcal{H}_1$) i model kugle ($\mathcal{H}_2$), definirane kao:

$$\mathcal{H}_1 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{x} \geq \theta_0\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}) \in \mathbb{R}^4$$

$$\mathcal{H}_2 : h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(\mathbf{x}-\tilde{\boldsymbol{\theta}})^{\mathrm{T}}(\mathbf{x}-\tilde{\boldsymbol{\theta}}) \leq \theta_0^2\},\ \boldsymbol{\theta} = (\theta_0, \tilde{\boldsymbol{\theta}}), \theta_0 \in \big[0, \frac{1}{2}\big], \tilde{\boldsymbol{\theta}} \in \mathbb{R}^3$$

Optimizacijski postupak koristi funkciju pogreške $E(\boldsymbol{\theta}|\mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{h(\mathbf{x}^{(i)};\boldsymbol{\theta}) \neq y^{(i)}\}$. Odredite minimalne pogreške $E(\boldsymbol{\theta}_1^*|\mathcal{D})$ i $E(\boldsymbol{\theta}_2^*|\mathcal{D})$ modela $\mathcal{H}_1$ odnosno $\mathcal{H}_2$. **Koliko iznosi apsolutna razlika između vrijednosti tih pogrešaka?**

A 1/3 B 1/2 C 0 D 1/6

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

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

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

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

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

4 (P) Razmatramo tri modela L2-regularizirane linearne regresije s do na parametre definiranim hipotezama h_1, h_2 i h_3 i s nama nepoznatim regularizacijskim faktorima λ_1, λ_2 odnosno λ_3. Za sva tri modela mjerimo vrijednost (neregularizirane) funkcije kvadratne pogreške $E(h|\mathcal{D})$, i to na skupu za učenje $\mathcal{D}_u$ i na ispitnome skupu D_i. Na skupu za učenje opažamo $E(h_2|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$ i $E(h_3|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, a na skupu za ispitivanje $E(h_1|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$ i $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$. **Koji uređaj očekivano vrijedi između regularizacijskih faktora?**

A $\lambda_2 < \lambda_1 < \lambda_3$ B $\lambda_1 < \lambda_3 < \lambda_2$ C $\lambda_3 < \lambda_2 < \lambda_1$ D $\lambda_2 < \lambda_3 < \lambda_1$

5 (T) Optimizacija parametara hipoteze zasniva se na funkciji empirijske pogreške odnosno funkciji gubitka. **Koja je razlika između funkcije empirijske pogreške i funkcije gubitka?**

A Funkcija gubitka odnosi se na jedan primjer, a funkcija empirijske pogreške na skup primjera

B Funkcija empirijske pogreške uvijek je nenegativna, dok funkcija gubitka može biti negativna

C Funkcija empirijske pogreške ovisi o konkretnoj hipotezi, dok funkcija gubitka ne ovisi

D Funkcija gubitka definirana je kao očekivanje funkcije empirijske pogreške na skupu primjera

6 (T) Rješenje za optimizacijski problem najmanjih kvadrata (OLS) izveli smo u zatvorenoj formi kao pseudoinverz $\mathbf{\Phi}^+$ matrice dizajna $\mathbf{\Phi}$. U nekim se slučajevima pseudoinverz $\mathbf{\Phi}^+$ može izračunati pomoću običnog inverza Gramove matrice. Neka je $m+1$ dimenzija prostora značajki, a N broj primjera u skupu za učenje. **Kada se pseudoinverz $\mathbf{\Phi}^+$ može izračunati pomoću inverza Gramove matrice?**

A Ako $N \leq m+1$ i nema multikolinearnosti značajki

B Ako $N \geq m+1$ i stupci u $\mathbf{\Phi}$ nisu linearno zavisni

C Ako je $N = m+1$ i Gramova matrica je kvadratna

D Ako $N < m+1$ i $\mathbf{\Phi}$ je punog ranga

Cjelina 2: Linearni klasifikacijski modeli (7 pitanja)

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

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

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

A 16 B 32 C 6 D 22

8 (T) Jedan od nedostataka algoritma perceptrona jest taj što hipoteza ovisi o odabiru početnog vektora težina. **Zašto je tomu tako?**

A Gradijent funkcije gubitka nije nula niti za jednu točku u prostoru parametara

B Funkcija pogreške nije konveksna funkcija parametara

C Gradijent funkcije pogreške jednak je nula ako i samo ako je vrijednost funkcije veća od nule

D Funkcija pogreške ima više točaka u kojima ostvaruje minimum

9 (P) Perceptronom rješavamo višeklasni klasifikacijski sa $K = 6$ klasa u ulaznome prostoru dimenzije $n = 5$. Koristimo shemu jedan-naspram-jedan (OVO). Kako bismo ostvarili linearnu razdvojivost klasa, koristimo funkciju preslikavanja $\boldsymbol{\phi}$ koja odgovara polinomu drugoga stupnja s interakcijskim parovima značajki. **Koliko iznosi ukupan broj parametara koji treba optimirati?**

A 126 B 315 C 140 D 280

10 (N) Raspolažemo označenim skupom primjera iz četiriju klasa ($K = 4$) u dvodimenzijskome ulaznom prostoru ($n = 2$). Na tom skupu treniramo model multinomijalne logističke regresije s funkcijom preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & -1 & -1 & 2 \\ 2 & -4 & -2 & 1 \\ 3 & -4 & 6 & 1 \\ -3 & 0 & 2 & 1 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (2, -3)$ s oznakom $\mathbf{y} = (0, 0, 0, 1)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi primjer x?**

A 64 B 51 C 13 D 21

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

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

Na tom skupu trenirali smo dva modela L2-regularizirane logističke regresije, koristeći regularizacijske faktore $\lambda = 1$ odnosno $\lambda = 5$. Neka je $L_\lambda^{(i)}$ gubitak koji primjer $\mathbf{x}^{(i)}$ iz $\mathcal{D}$ nanosi modelu optimiranom uz regularizacijski faktor λ. **Što očekujemo da vrijedi?**

A $L_1^{(5)} < L_5^{(2)}, L_1^{(4)} < L_5^{(1)}$

B $L_1^{(2)} > L_5^{(5)}, L_1^{(1)} > L_5^{(4)}$

C $L_1^{(6)} < L_5^{(3)}, L_1^{(2)} > L_5^{(5)}$

D $L_1^{(3)} > L_5^{(6)}, L_1^{(5)} < L_5^{(2)}$

12 (T) Za optimizaciju parametara logističke regresije umjesto gradijentnog spusta možemo primijeniti Newtonov postupak. **Konceptualno gledano, po čemu se Newtonov postupak razlikuje od gradijentnog spusta?**

A Kod ažuriranja težina u obzir uzima informaciju o promjeni gradijenta

B Rekondicionira Gramovu matricu modifikacijom njezine glavne dijagonale

C Za izračun gradijenta koristi sve primjere iz skupa za učenje

D Umjesto gradijenta funkcije pogreške, izračunava gradijent njezine kvadratne aproksimacije

13 (N) Raspolažemo sljedećim skupom označenih podataka u dvodimenzijskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((2,5),0),((-8,1),0),((3,9),1)\big\}$$

Na tom skupu treniramo model L2-regularizirane logističke regresije. Koristimo standardni (grupni) gradijentni spust uz faktor regularizacije $\lambda = 10$. Koristimo fiksnu stopu učenja $\eta = 0.01$ (ne koristimo linijsko pretraživanje). Vektor težina $\mathbf{w}$ početno je inicijaliziran na nul-vektor. Provedite jednu epohu gradijentnog spusta (jedan prolazak kroz sve primjere iz $\mathcal{D}$). **Koji se vektor težina $\mathbf{w}$ dobiva nakon prve epohe gradijentnog spusta?**

A $(0.005, -0.045, -0.015)$

B $(-0.005, 0.045, 0.015)$

C $(-0.005, 0.035, -0.025)$

D $(0.005, -0.035, 0.025)$

Cjelina 3: SVM, jezgrene i neparametarske metode (7 pitanja)

14 (T) Razina složenosti modela k-NN može se ugoditi hiperparametrom k. Neka je N broj primjera. **Koja je veza između k, N i razine složenosti modela k-NN?**

A Neovisno o N, s porastom vrijednosti k razina složenosti raste

B Ako N raste, s povećanjem k možemo zadržati istu razinu složenosti

C Ako N pada, a k je fiksan, razina složenosti raste pa pada

D Ako je N fiksan, manji k daje model manje razine složenosti

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

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

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

A $\frac{1}{2}\sqrt{10}$ puta

B $\frac{5}{4}\sqrt{2}$ puta

C $\frac{1}{2}\sqrt{13}$ puta

D $\frac{1}{2}\sqrt{17}$ puta

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

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

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

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

17 (P) Rješavamo binaran klasifikacijski problem u ulaznome prostoru dimenzije $n = 100$. Raspolažemo sa $N = 1000$ primjera. Isprobavamo dva klasifikacijska algoritma: (1) L2-regulariziranu logističku regresiju s adaptivnim baznim funkcijama i (2) rijetki jezgreni stroj ostvaren kao L1-regularizirana logistička regresija. Kod prvog modela koristimo $m = 50$ adaptivnih baznih funkcija, pri čemu je svaka adaptivna bazna funkcija ϕ_j i sama jedan model logističke regresije. Kod drugog modela koristimo Gaussove jezgrene funkcije (svaka s istom preciznošću), a nakon treniranja naučen model sastoji se od $N' = 32$ prototipa. Razmotrite ukupan broj parametara i hiperparametara oba modela nakon što su ti modeli naučeni, tj. broj (hiper)parametara koje nužno trebamo pohraniti u datoteku kod serijalizacije modela, a da bismo modele mogli naknadno koristiti za predikciju. Jedan (hiper)parametar odgovara jednom realnom broju. **Koliko iznosi razlika u broju pohranjenih (hiper)parametara za ova dva modela?**

A 899 B 1867 C 1749 D 1784

18 (N) Raspolažemo sljedećim skupom označenih podataka:

$$\mathbf{X}|\mathbf{y} = \left(\begin{array}{rrr|r} -15 & -30 & -15 & +1 \\ 6 & 12 & 6 & +1 \\ -5 & -8 & -3 & -1 \\ 5 & 11 & 6 & -1 \end{array}\right)$$

Na ovom skupu treniramo SVM s linearnom jezgrom. Koristimo stohastički gradijentni spust. Vektor dualnih koeficijenata je $\boldsymbol{\alpha} = (0.071, 2.929, 0.0, 3.0)$. Jedan potporni vektor nalazi se unutar margine. Izračunajte vrijednost regularizacijskog izraza u funkciji pogreške SVM-a za parametre $\mathbf{w}$ koji odgovaraju dualnim koeficijentima $\boldsymbol{\alpha}$. **Koliko iznosi regularizacijski izraz?**

A 0.88 B 0.63 C 0.50 D 0.75

19 (N) Raspolažemo sljedećim skupom označenih primjera iz trodimenzijskoga ulaznog prostora:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((9,1,3),+1), ((-1,7,3),-1), ((0,-8,1),+1)\big\}$$

Na ovome skupu trenirali smo SVM s inverznom kvadratnom jezgrom $\kappa(\mathbf{x}, \mathbf{z}) = (1+\|\mathbf{x}-\mathbf{z}\|^2)^{-1}$ i hiperparametrom $C = 10$. Dobiveni Lagrangeovi koeficijenti su $\boldsymbol{\alpha} = (0.6728, 1.3400, 0.6678)$. Može se izračunati da vrijedi $w_0 = 0.333$. Zanima nas klasifikacija primjera $\mathbf{x} = (-2, 4, 4)$. Primijenite jezgreni trik i izračunajte $h(\mathbf{x})$. **Koliko iznosi $h(\mathbf{x})$?**

A 0.231 B 0.550 C 0.391 D 0.118

20 (T) Tradicionalan način optimizacije parametara stroja potpornih vektora jest primjena konveksne optimizacije uz ograničenja. Razmotrite problem meke margine sa N primjera u n-dimenzijskome ulaznom prostoru. Optimizacija se može provesti u primarnoj ili dualnoj formi. **Koliko se optimizacijskih varijabli te koliko ograničenja javlja u dualnoj formulaciji problema meke margine?**

A N varijabli, N ograničenja nejednakosti, 1 ograničenje jednakosti

B $n + N + 1$ varijabli, $2N$ ograničenja nejednakosti

C $n + 1$ varijabli, N ograničenja nejednakosti

D N varijabli, $2N$ ograničenja nejednakosti, 1 ograničenje jednakosti

Strojno učenje 1 -- Međuispit
27. studenog 2025.

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

	A	B	C	D	E	F	G	H
1	A	D	D	B	A	C	B	B
2	A	A	A	C	A	B	C	D
3	B	D	B	D	A	C	D	C
4	B	D	C	D	B	D	D	D
5	D	B	C	D	B	B	B	A
6	A	B	B	C	A	D	D	B

Cjelina 2: Linearni klasifikacijski modeli (7 pitanja)

	A	B	C	D	E	F	G	H
7	C	B	C	A	A	A	C	C
8	D	B	D	D	C	D	B	D
9	D	B	D	B	A	C	C	B
10	B	D	A	B	B	B	D	D
11	D	C	A	D	A	B	B	C
12	D	D	C	C	B	C	A	A
13	C	C	A	C	B	C	A	B

Cjelina 3: SVM, jezgrene i neparametarske metode (7 pitanja)

	A	B	C	D	E	F	G	H
14	D	B	C	C	D	B	C	B
15	C	C	B	D	A	D	B	A
16	D	C	C	C	A	B	C	A
17	D	A	D	B	A	B	C	B
18	B	C	B	B	D	B	C	D
19	B	D	B	C	A	A	A	A
20	B	B	B	B	B	D	C	D