Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

2. prosinca 2022.

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

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

1 (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable y o zavisnoj varijabli x. Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka $\mathcal{D} = \{(x^{(i)}, y^{(i)})\} = \{(0,0),(2,0),(3,2),(5,2)\}$. Neka je h hipoteza koju dobivamo treniranjem modela te neka je L^i gubitak hipoteze h na primjeru $x^{(i)}$, tj. $L^i = L(y^{(i)}, h(x^{(i)}))$. **Što vrijedi za gubitke hipoteze na pojedinim primjerima?**

A $L^1 = L^3 = 0,\ L^2 = L^4 < 1$

C $L^1 = L^2 = 1 < L^3 < L^4$

B $L^1 = L^4 < L^2 = L^3$

D $L^1 = L^4 = 1,\ L^2 < L^3$

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

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geq 0\}$$
$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \geq \theta_0^2\big\}$$

Pored ova dva modela, razmatramo i njihove kombinacije, modele $\mathcal{H}_3$ i $\mathcal{H}_4$. Neka je $\mathcal{H}_3 = \mathcal{H}_1 \cup \mathcal{H}_2$ te neka je $\mathcal{H}_4$ skup funkcija definiranih kao $h_4(\mathbf{x};\boldsymbol{\theta}) = h_1(\mathbf{x}) \cdot h_2(\mathbf{x})$. Neka je E_k minimalna empirijska pogreška koja se modelom $\mathcal{H}_k$ može ostvariti na skupu $\mathcal{D}$, tj. $E_k = \operatorname{argmin}_{h\in\mathcal{H}_k} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

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

3 (T) Modeli strojnog učenja općenito su različite složenosti. S porastom složenosti modela raste vjerojatnost da model bude prenaučen. Ta vjerojatnost raste s količinom šuma u podatcima. **Zašto šum u podatcima za učenje može dovesti do prenaučenosti klasifikacijskog modela?**

A Povećanjem količine šuma granica između klasa postaje sve nelinearnija, pa raste i složenost modela te dobivena hipoteza očekivano neće odgovarati granici između klasa na ispitnom skupu

B Zbog šuma granica između klasa izgleda nelinearnijom nego što ona to zapravo jest, pa primjeri blizu granice znatno više doprinose pogrešci učenja nego primjeri koji su udaljeni od granice

C Zbog šuma su oznake nekih primjera u skupu za učenje pogrešne, pa sve hipoteze iz modela imaju na tom skupu pogrešku koja je veća od nula, a još veća na ispitnom skupu

D Efekt šuma ja slučajan, pa će hipoteza koja se previše prilagodi šumu na skupu za učenje očekivano imati veliku pogrešku na ispitnom skupu gdje je šum drugačiji ili ga nema

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

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

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

A 2.69 B 1.58 C 0.29 D 7.10

5 (T) Multikolinearnost značajki jedan je od problema koji može nastupiti kod primjene modela regresije na stvarnim podatcima. Efekt multikolinearnosti i savršene multikolinearnosti dobro je uočljiv kod optimizacijskoga postupka običnih najmanjih kvadrata (OLS) kada se on provodi izračunom pseudoinverza matrice dizajna. Neka je m broj značajki, Φ je matrica dizajna i $\mathbf{G} = \Phi^{\mathrm{T}}\Phi$ je Gramova matrica. **Koji je efekt savršene multikolinearnosti kod postupka OLS?**

- A Φ ima puni rang, $\mathrm{rang}(\mathbf{G}) > m$ i $\mathbf{G}$ ima inverz, ali s visokim kondicijskim brojem
- B $\mathrm{rang}(\Phi) < m+1$, $\mathbf{G}$ nema puni rang i nema inverz, no ima pseudoinverz koji nije numerički stabilan
- C $\mathrm{rang}(\Phi) = N$, no $\mathrm{rang}(\mathbf{G}) < N$, pa $\mathbf{G}$ ima pseudoinverz, ali nema numerički stabilan inverz
- D Φ nema puni rang, $\mathrm{rang}(\mathbf{G}) < m+1$ i $\mathbf{G}$ nema pseudoinverz

6 (P) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi:

$$E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$$

Što iz toga možemo zaključiti?

- A Model $\mathcal{H}_{\alpha_1}$ je podnaučen
- B Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$
- C Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$
- D Model $\mathcal{H}_{\alpha_2}$ je prenaučen

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

7 (P) Skup za učenje čine sljedeći označeni primjeri:

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

Na skupu $\mathcal{D}$ treniramo logističku regresiju (LR) i stroj potpornih vektora s tvrdom marginom (SVM). Dodatno, treniramo model linearne regresije (LINR), gdje izlaz tog modela koristimo za klasifikaciju, tj. $h(\mathbf{x}) = \mathbf{1}\{\mathbf{w}^{\mathrm{T}}\mathbf{x} \geq 0\}$. Za modele SVM i LINR umjesto oznake $y = 0$ koristimo oznaku $y = -1$. Za treniranje modela LR koristimo dovoljan broj iteracija tako da možemo pretpostaviti da je dobivena pogreška unakrsne entropije praktički jednaka nuli. Razmotrite primjer $\mathbf{x}^{(7)} = (2,0)$. Neka je $d(m)$ udaljenost primjera $\mathbf{x}^{(7)}$ od granice između klasa dobivene modelom m. **Što od navedenog vrijedi za tu udaljenost?**

- A $d(\mathrm{SVM}) < d(\mathrm{LINR}) < d(\mathrm{LR})$
- C $d(\mathrm{SVM}) < d(\mathrm{LR}) < d(\mathrm{LINR})$
- B $d(\mathrm{LR}) < d(\mathrm{SVM}) < d(\mathrm{LINR})$
- D $d(\mathrm{LINR}) < d(\mathrm{LR}) < d(\mathrm{SVM})$

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

A 2.54 B 7.11 C 4.03 D 1.19

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

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

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

A 12.02 B 4.02 C 8.00 D 6.00

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu na sljedeći način:

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

Početne težine perceptrona neka su $\mathbf{w} = (1, 0, -1, 2, 3, 2)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 18 B 100 C 96 D 40

11 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.1, 0.5, -2, -0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1,-1), 1)$?**

A 0.50 B 0.09 C 1.42 D 0.22

12 (T) Višeklasni problem može se riješiti binarnim klasifikatorom uz primjenu sheme OVO ili sheme OVR. Obje sheme imaju svoje prednosti i nedostatke. Pretpostavite da raspolažemo sa K klasa i da svaka klasa ima N/K primjera, gdje je N ukupan broj primjera u skupu za učenje. **Što su prednosti odnosno nedostatci OVO i OVR sheme u takvom slučaju?**

A OVO iziskuje $(K-1)/2$ puta više parametara nego OVR, ali svaki OVR klasifikator ima $K-1$ puta manje pozitivnih primjera nego negativnih

B OVR iziskuje $K-1$ puta više klasifikatora od sheme OVO, ali kod OVO pozitivne klase imaju $K-1$ puta manje primjera nego kod OVR

C OVO svaki klasifikator trenira s $K/2$ puta manje primjera nego OVR, ali pozitivne klase kod OVR imaju K puta manje primjera nego kod OVO

D OVR treba K puta više klasifikatora nego OVO, ali su kod OVO pozitivne klase $K/2$ puta manje zastupljene nego OVR

13 (P) Na primjerima iz dvodimenzijskoga ulaznog prostora treniramo L_2-regulariziranu logističku regresiju. Neka su $\mathbf{w}_0 = (1,-4,4)$, $\mathbf{w}_1 = (1,-4,6)$, $\mathbf{w}_2 = (1,-1,7)$, $\mathbf{w}_3 = (1,-7,1)$ i $\mathbf{w}_4 = (1,-7,-3)$ vektori u prostoru parametara. Neka je $E(\mathbf{w}|\mathcal{D})$ neregularizirana pogreška unakrsne entropije na skupu za učenje $\mathcal{D}$. Pritom je $\mathbf{w}_0$ minimizator funkcije $E(\mathbf{w}|\mathcal{D})$ te vrijedi $E(\mathbf{w}_1|\mathcal{D}) = E(\mathbf{w}_2|\mathcal{D}) = E(\mathbf{w}_3|\mathcal{D})$. Napravite skicu izokontura funkcije pogreške u potprostoru $w_1 \times w_2$. Za treniranje modela koristimo gradijentni spust s linijskim pretraživanjem uz regularizacijski faktor $\lambda = 100$. Za tako naučen model vrijednost regularizacijskog izraza $\frac{\lambda}{2}\|\mathbf{w}\|^2$ jednaka je 400. Međutim, broj koraka gradijentnog spusta (broj poziva linijskog pretraživanja) ovisi o tome koliko će spust krivudati, a to ovisi o odabiru inicijalnih parametara. Kao moguće inicijalne parametre razmotrite vektore $\mathbf{w}_1$–$\mathbf{w}_4$. **S kojim inicijalnim parametarima će algoritam gradijentnog spusta konvergirati u najmanjem broju koraka?**

A $\mathbf{w}_1$ B $\mathbf{w}_3$ C $\mathbf{w}_4$ D $\mathbf{w}_2$

14 (T) Za optimizaciju parametara poopćenih linearnih modela može se koristiti stohastički gradijentni spust, odnosno pravilo LMS. Neka je $(\mathbf{x}, y)$ označeni primjer za koji radimo ažuriranje težina pomoću pravila LMS. **Što možemo reći o razlici između novih (ažuriranih) i starih težina (težina prije ažuriranja)?**

A Razlika je to manja što je vektor $\boldsymbol{\phi}(\mathbf{x})$ bliži ishodištu

B Razlika je to veća što je stopa učenja η bliža jedinici

C Razlika je to manja što je oznaka y bliže jedinici

D Razlika je to veća što je izlaz modela $h(\mathbf{x})$ bliži nuli

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

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

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

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

A 0.359   B 0.552   C 0.456   D 0.795

16 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 5$ označenih primjera. Vektor oznaka je $\mathbf{y} = (+1, +1, -1, -1, +1)$. Euklidske udaljenosti između primjera dane su sljedećom matricom udaljenosti:

$$\mathbf{D} = \begin{pmatrix} 0.0 & 7.48 & 6.16 & 13.42 & 12.21 \\ 7.48 & 0.0 & 12.73 & 20.1 & 14.18 \\ 6.16 & 12.73 & 0.0 & 10.49 & 9.95 \\ 13.42 & 20.1 & 10.49 & 0.0 & 20.02 \\ 12.21 & 14.18 & 9.95 & 20.02 & 0.0 \end{pmatrix}$$

Treniranjem uz $C = 10$ i $\gamma = 0.0001$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (10, 1.052, 10, 10, 8.948)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za treći primjer, $L(y^{(3)}, h(\mathbf{x}^{(3)}))$?**

A 1.64   B 1.18   C 0.03   D 0.24

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

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

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

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

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

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

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

A 10   B 25   C 96   D 76

19 (P) Na 900 primjera sa 50 značajki treniramo rijetki jezgreni stroj s Gaussovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 52 prototipa. **Koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

A Optimiramo 901 parametar, a naučeni model ima 3501 parametar

B Optimiramo 51 parametar, a naučeni model ima 2653 parametara

C Optimiramo 51 parametar, a naučeni model ima 2600 parametara

D Optimiramo 901 parametar, a naučeni model ima 2653 parametara

20 (T) Optimizacijski problem algoritma SVM može se postaviti u formulaciji meke ili tvrde margine te u primarnoj ili dualnoj formulaciji. Ovisno o formulaciji, kvadratni program sadrži različit broj varijabli po kojima optimiramo (optimizacijske varijable). **Ako matrica dizajna ima više redaka nego stupaca, koja formulacija ima najmanje optimizacijskih varijabli?**

A Dualni problem tvrde margine
C Primarni problem tvrde margine
B Primarni problem meke margine
D Dualni problem meke margine

21 (T) Algoritmi strojnog učenja mogu biti parametarski ili neparametarski. **Što je karakteristika neparametarskih algoritama strojnog učenja?**

A Broj parametara ovisi o broju primjera

B Složenost modela raste s normom vektora parametara

C Pretpostavljaju teorijsku distribuciju podataka

D Hiperparametri nemaju utjecaja na složenost modela

22 (N) U ulaznome prostoru dimenzije $n = 3$ trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri $((2, -5, 15), -1)$, $((1, 8, -305), -1)$ i $((1, -6, 225), +1)$, a njima odgovarajući dualni koeficijenti su $\alpha_1 = 0.5$, $\alpha_2 = 0.8$ i $\alpha_3 = 0.9$. Treniranje smo proveli na skaliranim značajkama: svaku smo značajku x_j standardizirali primjenom transformacije $\frac{x_j - \mu_j}{\sigma_j}$, gdje su μ_j i σ_j srednja vrijednost odnosno varijanca značajke x_j u skupu označenih podataka $\mathcal{D}$. Parametri skaliranja su $\boldsymbol{\mu} = (15, -2, 100)$ i $\boldsymbol{\sigma} = (4, 1, 12)$. Model SVM-a koristimo za predikciju klase primjera $\mathbf{x} = (1, 2, -30)$. **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera x?**

A $+907.43$ B $+541.53$ C -739.13 D -373.22

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

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

1 (T) Modeli strojnog učenja općenito su različite složenosti. S porastom složenosti modela raste vjerojatnost da model bude prenaučen. Ta vjerojatnost raste s količinom šuma u podatcima. **Zašto šum u podatcima za učenje može dovesti do prenaučenosti klasifikacijskog modela?**

- A Zbog šuma granica između klasa izgleda nelinearnijom nego što ona to zapravo jest, pa primjeri blizu granice znatno više doprinose pogrešci učenja nego primjeri koji su udaljeni od granice
- B Povećanjem količine šuma granica između klasa postaje sve nelinearnija, pa raste i složenost modela te dobivena hipoteza očekivano neće odgovarati granici između klasa na ispitnom skupu
- C Efekt šuma ja slučajan, pa će hipoteza koja se previše prilagodi šumu na skupu za učenje očekivano imati veliku pogrešku na ispitnom skupu gdje je šum drugačiji ili ga nema
- D Zbog šuma su oznake nekih primjera u skupu za učenje pogrešne, pa sve hipoteze iz modela imaju na tom skupu pogrešku koja je veća od nula, a još veća na ispitnom skupu

2 (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable y o zavisnoj varijabli x. Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka $\mathcal{D} = \{(x^{(i)}, y^{(i)})\} = \{(0,0),(2,0),(3,2),(5,2)\}$. Neka je h hipoteza koju dobivamo treniranjem modela te neka je L^i gubitak hipoteze h na primjeru $x^{(i)}$, tj. $L^i = L(y^{(i)}, h(x^{(i)}))$. **Što vrijedi za gubitke hipoteze na pojedinim primjerima?**

- A $L^1 = L^4 = 1,\ L^2 < L^3$
- B $L^1 = L^4 < L^2 = L^3$
- C $L^1 = L^2 = 1 < L^3 < L^4$
- D $L^1 = L^3 = 0,\ L^2 = L^4 < 1$

3 (T) Multikolinearnost značajki jedan je od problema koji može nastupiti kod primjene modela regresije na stvarnim podatcima. Efekt multikolinearnosti i savršene multikolinearnosti dobro je uočljiv kod optimizacijskoga postupka običnih najmanjih kvadrata (OLS) kada se on provodi izračunom pseudoinverza matrice dizajna. Neka je m broj značajki, Φ je matrica dizajna i $\mathbf{G} = \Phi^{\mathrm{T}}\Phi$ je Gramova matrica. **Koji je efekt savršene multikolinearnosti kod postupka OLS?**

- A Φ ima puni rang, $\mathrm{rang}(\mathbf{G}) > m$ i $\mathbf{G}$ ima inverz, ali s visokim kondicijskim brojem
- B Φ nema puni rang, $\mathrm{rang}(\mathbf{G}) < m+1$ i $\mathbf{G}$ nema pseudoinverz
- C $\mathrm{rang}(\Phi) = N$, no $\mathrm{rang}(\mathbf{G}) < N$, pa $\mathbf{G}$ ima pseudoinverz, ali nema numerički stabilan inverz
- D $\mathrm{rang}(\Phi) < m+1$, $\mathbf{G}$ nema puni rang i nema inverz, no ima pseudoinverz koji nije numerički stabilan

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

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geq 0\big\}$$
$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \geq \theta_0^2\big\}$$

Pored ova dva modela, razmatramo i njihove kombinacije, modele $\mathcal{H}_3$ i $\mathcal{H}_4$. Neka je $\mathcal{H}_3 = \mathcal{H}_1 \cup \mathcal{H}_2$ te neka je $\mathcal{H}_4$ skup funkcija definiranih kao $h_4(\mathbf{x};\boldsymbol{\theta}) = h_1(\mathbf{x}) \cdot h_2(\mathbf{x})$. Neka je E_k minimalna empirijska pogreška koja se modelom $\mathcal{H}_k$ može ostvariti na skupu $\mathcal{D}$, tj. $E_k = \mathrm{argmin}_{h\in\mathcal{H}_k} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

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

5 (P) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi:

$$E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$$

Što iz toga možemo zaključiti?

A Model $\mathcal{H}_{\alpha_1}$ je podnaučen

B Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$

C Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$

D Model $\mathcal{H}_{\alpha_2}$ je prenaučen

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

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

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

A 0.29 B 7.10 C 1.58 D 2.69

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

7 (P) Na primjerima iz dvodimenzijskoga ulaznog prostora treniramo L_2-regulariziranu logističku regresiju. Neka su $\mathbf{w}_0 = (1,-4,4)$, $\mathbf{w}_1 = (1,-4,6)$, $\mathbf{w}_2 = (1,-1,7)$, $\mathbf{w}_3 = (1,-7,1)$ i $\mathbf{w}_4 = (1,-7,-3)$ vektori u prostoru parametara. Neka je $E(\mathbf{w}|\mathcal{D})$ neregularizirana pogreška unakrsne entropije na skupu za učenje $\mathcal{D}$. Pritom je $\mathbf{w}_0$ minimizator funkcije $E(\mathbf{w}|\mathcal{D})$ te vrijedi $E(\mathbf{w}_1|\mathcal{D}) = E(\mathbf{w}_2|\mathcal{D}) = E(\mathbf{w}_3|\mathcal{D})$. Napravite skicu izokontura funkcije pogreške u potprostoru $w_1 \times w_2$. Za treniranje modela koristimo gradijentni spust s linijskim pretraživanjem uz regularizacijski faktor $\lambda = 100$. Za tako naučen model vrijednost regularizacijskog izraza $\frac{\lambda}{2}\|\mathbf{w}\|^2$ jednaka je 400. Međutim, broj koraka gradijentnog spusta (broj poziva linijskog pretraživanja) ovisi o tome koliko će spust krivudati, a to ovisi o odabiru inicijalnih parametara. Kao moguće inicijalne parametre razmotrite vektore $\mathbf{w}_1$–$\mathbf{w}_4$. **S kojim inicijalnim parametarima će algoritam gradijentnog spusta konvergirati u najmanjem broju koraka?**

A $\mathbf{w}_2$ B $\mathbf{w}_3$ C $\mathbf{w}_1$ D $\mathbf{w}_4$

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

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

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

A 4.02 B 6.00 C 8.00 D 12.02

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

A 1.19 B 2.54 C 7.11 D 4.03

10 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.1, 0.5, -2, 0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1, -1), 1)$?**

A 0.09 B 0.22 C 1.42 D 0.50

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu na sljedeći način:

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

Početne težine perceptrona neka su $\mathbf{w} = (1, 0, -1, 2, 3, 0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 100 B 18 C 96 D 40

12 (P) Skup za učenje čine sljedeći označeni primjeri:

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

Na skupu $\mathcal{D}$ treniramo logističku regresiju (LR) i stroj potpornih vektora s tvrdom marginom (SVM). Dodatno, treniramo model linearne regresije (LINR), gdje izlaz tog modela koristimo za klasifikaciju, tj. $h(\mathbf{x}) = \mathbf{1}\{\mathbf{w}^{\mathrm{T}}\mathbf{x} \geq 0\}$. Za modele SVM i LINR umjesto oznake $y = 0$ koristimo oznaku $y = -1$. Za treniranje modela LR koristimo dovoljan broj iteracija tako da možemo pretpostaviti da je dobivena pogreška unakrsne entropije praktički jednaka nuli. Razmotrite primjer $\mathbf{x}^{(7)} = (2, 0)$. Neka je $d(m)$ udaljenost primjera $\mathbf{x}^{(7)}$ od granice između klasa dobivene modelom m. **Što od navedenog vrijedi za tu udaljenost?**

A $d(\text{LR}) < d(\text{SVM}) < d(\text{LINR})$
C $d(\text{LINR}) < d(\text{LR}) < d(\text{SVM})$
B $d(\text{SVM}) < d(\text{LR}) < d(\text{LINR})$
D $d(\text{SVM}) < d(\text{LINR}) < d(\text{LR})$

13 (T) Za optimizaciju parametara poopćenih linearnih modela može se koristiti stohastički gradijentni spust, odnosno pravilo LMS. Neka je $(\mathbf{x}, y)$ označeni primjer za koji radimo ažuriranje težina pomoću pravila LMS. **Što možemo reći o razlici između novih (ažuriranih) i starih težina (težina prije ažuriranja)?**

A Razlika je to manja što je vektor $\boldsymbol{\phi}(\mathbf{x})$ bliži ishodištu

B Razlika je to veća što je izlaz modela $h(\mathbf{x})$ bliži nuli

C Razlika je to veća što je stopa učenja η bliža jedinici

D Razlika je to manja što je oznaka y bliže jedinici

14 (T) Višeklasni problem može se riješiti binarnim klasifikatorom uz primjenu sheme OVO ili sheme OVR. Obje sheme imaju svoje prednosti i nedostatke. Pretpostavite da raspolažemo sa K klasa i da svaka klasa ima N/K primjera, gdje je N ukupan broj primjera u skupu za učenje. **Što su prednosti odnosno nedostatci OVO i OVR sheme u takvom slučaju?**

A OVR treba K puta više klasifikatora nego OVO, ali su kod OVO pozitivne klase $K/2$ puta manje zastupljene nego OVR

B OVO iziskuje $(K-1)/2$ puta više parametara nego OVR, ali svaki OVR klasifikator ima $K-1$ puta manje pozitivnih primjera nego negativnih

C OVO svaki klasifikator trenira s $K/2$ puta manje primjera nego OVR, ali pozitivne klase kod OVR imaju K puta manje primjera nego kod OVO

D OVR iziskuje $K-1$ puta više klasifikatora od sheme OVO, ali kod OVO pozitivne klase imaju $K-1$ puta manje primjera nego kod OVR

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

15 (P) Na 500 primjera sa 80 značajki treniramo rijetki jezgreni stroj s Gaussovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 38 prototipa. **Koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

- A Optimiramo 501 parametara, a naučeni model ima 3541 parametar
- B Optimiramo 81 parametar, a naučeni model ima 3040 parametara
- C Optimiramo 81 parametar, a naučeni model ima 3079 parametara
- D Optimiramo 501 parametara, a naučeni model ima 3079 parametara

16 (N) U ulaznome prostoru dimenzije $n = 3$ trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri $((2,-5,15),-1)$, $((1,8,-305),-1)$ i $((1,-6,225),+1)$, a njima odgovarajući dualni koeficijenti su $\alpha_1 = 0.5$, $\alpha_2 = 0.8$ i $\alpha_3 = 0.9$. Treniranje smo proveli na skaliranim značajkama: svaku smo značajku x_j standardizirali primjenom transformacije $\frac{x_j-\mu_j}{\sigma_j}$, gdje su μ_j i σ_j srednja vrijednost odnosno varijanca značajke x_j u skupu označenih podataka $\mathcal{D}$. Parametri skaliranja su $\boldsymbol{\mu} = (15,-2,100)$ i $\boldsymbol{\sigma} = (4,1,12)$. Model SVM-a koristimo za predikciju klase primjera $\mathbf{x} = (1,-2,5)$. **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera x?**

A -373.22 B $+907.43$ C $+541.53$ D -739.13

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

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

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

A 76 B 25 C 96 D 10

18 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 5$ označenih primjera. Vektor oznaka je $\mathbf{y} = (+1,+1,-1,-1,+1)$. Euklidske udaljenosti između primjera dane su sljedećom matricom udaljenosti:

$$\mathbf{D} = \begin{pmatrix} 0.0 & 7.48 & 6.16 & 13.42 & 12.21 \\ 7.48 & 0.0 & 12.73 & 20.1 & 14.18 \\ 6.16 & 12.73 & 0.0 & 10.49 & 9.95 \\ 13.42 & 20.1 & 10.49 & 0.0 & 20.02 \\ 12.21 & 14.18 & 9.95 & 20.02 & 0.0 \end{pmatrix}$$

Treniranjem uz $C = 10$ i $\gamma = 0.0001$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (10, 1.052, 10, 10, 8.948)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za prvi primjer, $L(y^{(1)}, h(\mathbf{x}^{(1)}))$?**

A 1.64 B 0.03 C 0.24 D 1.18

19 (T) Algoritmi strojnog učenja mogu biti parametarski ili neparametarski. **Što je karakteristika neparametarskih algoritama strojnog učenja?**

- A Broj parametara ovisi o broju primjera
- B Broj značajki ne ovisi o dimenziji ulaznog prostora
- C Svaki primjer ima globalan utjecaj na izgled hipoteze
- D Hiperparametri nemaju utjecaja na složenost modela

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

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

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

A 0.552 B 0.795 C 0.456 D 0.359

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

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

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

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

22 (T) Optimizacijski problem algoritma SVM može se postaviti u formulaciji meke ili tvrde margine te u primarnoj ili dualnoj formulaciji. Ovisno o formulaciji, kvadratni program sadrži različit broj varijabli po kojima optimiramo (optimizacijske varijable). **Ako matrica dizajna ima više redaka nego stupaca, koja formulacija ima najmanje optimizacijskih varijabli?**

A Primarni problem tvrde margine C Dualni problem meke margine

B Primarni problem meke margine D Dualni problem tvrde margine

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

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

1 (P) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi:

$$E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$$

Što iz toga možemo zaključiti?

A Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$

B Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$

C Model $\mathcal{H}_{\alpha_1}$ je podnaučen

D Model $\mathcal{H}_{\alpha_2}$ je prenaučen

2 (T) Modeli strojnog učenja općenito su različite složenosti. S porastom složenosti modela raste vjerojatnost da model bude prenaučen. Ta vjerojatnost raste s količinom šuma u podatcima. **Zašto šum u podatcima za učenje može dovesti do prenaučenosti klasifikacijskog modela?**

A Zbog šuma granica između klasa izgleda nelinearnijom nego što ona to zapravo jest, pa primjeri blizu granice znatno više doprinose pogrešci učenja nego primjeri koji su udaljeni od granice

B Povećanjem količine šuma granica između klasa postaje sve nelinearnija, pa raste i složenost modela te dobivena hipoteza očekivano neće odgovarati granici između klasa na ispitnom skupu

C Efekt šuma ja slučajan, pa će hipoteza koja se previše prilagodi šumu na skupu za učenje očekivano imati veliku pogrešku na ispitnom skupu gdje je šum drugačiji ili ga nema

D Zbog šuma su oznake nekih primjera u skupu za učenje pogrešne, pa sve hipoteze iz modela imaju na tom skupu pogrešku koja je veća od nula, a još veća na ispitnom skupu

3 (T) Multikolinearnost značajki jedan je od problema koji može nastupiti kod primjene modela regresije na stvarnim podatcima. Efekt multikolinearnosti i savršene multikolinearnosti dobro je uočljiv kod optimizacijskoga postupka običnih najmanjih kvadrata (OLS) kada se on provodi izračunom pseudoinverza matrice dizajna. Neka je m broj značajki, Φ je matrica dizajna i $\mathbf{G} = \Phi^\mathrm{T}\Phi$ je Gramova matrica. **Koji je efekt savršene multikolinearnosti kod postupka OLS?**

A Φ ima puni rang, $\mathrm{rang}(\mathbf{G}) > m$ i $\mathbf{G}$ ima inverz, ali s visokim kondicijskim brojem

B $\mathrm{rang}(\Phi) = N$, no $\mathrm{rang}(\mathbf{G}) < N$, pa $\mathbf{G}$ ima pseudoinverz, ali nema numerički stabilan inverz

C $\mathrm{rang}(\Phi) < m + 1$, $\mathbf{G}$ nema puni rang i nema inverz, no ima pseudoinverz koji nije numerički stabilan

D Φ nema puni rang, $\mathrm{rang}(\mathbf{G}) < m + 1$ i $\mathbf{G}$ nema pseudoinverz

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

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{\boldsymbol{\theta}^\mathrm{T}\mathbf{x} \geq 0\big\}$$

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

Pored ova dva modela, razmatramo i njihove kombinacije, modele $\mathcal{H}_3$ i $\mathcal{H}_4$. Neka je $\mathcal{H}_3 = \mathcal{H}_1 \cup \mathcal{H}_2$ te neka je $\mathcal{H}_4$ skup funkcija definiranih kao $h_4(\mathbf{x};\boldsymbol{\theta}) = h_1(\mathbf{x}) \cdot h_2(\mathbf{x})$. Neka je E_k minimalna empirijska pogreška koja se modelom $\mathcal{H}_k$ može ostvariti na skupu $\mathcal{D}$, tj. $E_k = \text{argmin}_{h\in\mathcal{H}_k} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

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

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

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

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

A 0.29 B 2.69 C 1.58 D 7.10

6 (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable y o zavisnoj varijabli x. Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka $\mathcal{D} = \{(x^{(i)}, y^{(i)})\} = \{(0,0),(2,0),(3,2),(5,2)\}$. Neka je h hipoteza koju dobivamo treniranjem modela te neka je L^i gubitak hipoteze h na primjeru $x^{(i)}$, tj. $L^i = L(y^{(i)}, h(x^{(i)}))$. **Što vrijedi za gubitke hipoteze na pojedinim primjerima?**

A $L^1 = L^3 = 0$, $L^2 = L^4 < 1$ C $L^1 = L^2 = 1 < L^3 < L^4$

B $L^1 = L^4 < L^2 = L^3$ D $L^1 = L^4 = 1$, $L^2 < L^3$

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

7 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.1, 0.5, -2, -0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1,-1),1)$?**

A 1.42 B 0.22 C 0.09 D 0.50

8 (T) Za optimizaciju parametara poopćenih linearnih modela može se koristiti stohastički gradijentni spust, odnosno pravilo LMS. Neka je $(\mathbf{x}, y)$ označeni primjer za koji radimo ažuriranje težina pomoću pravila LMS. **Što možemo reći o razlici između novih (ažuriranih) i starih težina (težina prije ažuriranja)?**

A Razlika je to veća što je izlaz modela $h(\mathbf{x})$ bliži nuli

B Razlika je to manja što je oznaka y bliže jedinici

C Razlika je to veća što je stopa učenja η bliža jedinici

D Razlika je to manja što je vektor $\boldsymbol{\phi}(\mathbf{x})$ bliži ishodištu

9 (T) Višeklasni problem može se riješiti binarnim klasifikatorom uz primjenu sheme OVO ili sheme OVR. Obje sheme imaju svoje prednosti i nedostatke. Pretpostavite da raspolažemo sa K klasa i da svaka klasa ima N/K primjera, gdje je N ukupan broj primjera u skupu za učenje. **Što su prednosti odnosno nedostatci OVO i OVR sheme u takvom slučaju?**

A OVR treba K puta više klasifikatora nego OVO, ali su kod OVO pozitivne klase $K/2$ puta manje zastupljene nego OVR

B OVO iziskuje $(K-1)/2$ puta više parametara nego OVR, ali svaki OVR klasifikator ima $K-1$ puta manje pozitivnih primjera nego negativnih

C OVO svaki klasifikator trenira s $K/2$ puta manje primjera nego OVR, ali pozitivne klase kod OVR imaju K puta manje primjera nego kod OVO

D OVR iziskuje $K-1$ puta više klasifikatora od sheme OVO, ali kod OVO pozitivne klase imaju $K-1$ puta manje primjera nego kod OVR

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu na sljedeći način:

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

Početne težine perceptrona neka su $\mathbf{w} = (1, 0, -1, 2, 3, 0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 40 B 96 C 100 D 18

11 (P) Na primjerima iz dvodimenzijskoga ulaznog prostora treniramo L_2-regulariziranu logističku regresiju. Neka su $\mathbf{w}_0 = (1,-4,4)$, $\mathbf{w}_1 = (1,-4,6)$, $\mathbf{w}_2 = (1,-1,7)$, $\mathbf{w}_3 = (1,-7,1)$ i $\mathbf{w}_4 = (1,-7,-3)$ vektori u prostoru parametara. Neka je $E(\mathbf{w}|\mathcal{D})$ neregularizirana pogreška unakrsne entropije na skupu za učenje $\mathcal{D}$. Pritom je $\mathbf{w}_0$ minimizator funkcije $E(\mathbf{w}|\mathcal{D})$ te vrijedi $E(\mathbf{w}_1|\mathcal{D}) = E(\mathbf{w}_2|\mathcal{D}) = E(\mathbf{w}_3|\mathcal{D})$. Napravite skicu izokontura funkcije pogreške u potprostoru $w_1 \times w_2$. Za treniranje modela koristimo gradijentni spust s linijskim pretraživanjem uz regularizacijski faktor $\lambda = 100$. Za tako naučen model vrijednost regularizacijskog izraza $\frac{\lambda}{2}\|\mathbf{w}\|^2$ jednaka je 400. Međutim, broj koraka gradijentnog spusta (broj poziva linijskog pretraživanja) ovisi o tome koliko će spust krivudati, a to ovisi o odabiru inicijalnih parametara. Kao moguće inicijalne parametre razmotrite vektore $\mathbf{w}_1$–$\mathbf{w}_4$. **S kojim inicijalnim parametarima će algoritam gradijentnog spusta konvergirati u najmanjem broju koraka?**

A $\mathbf{w}_3$ B $\mathbf{w}_1$ C $\mathbf{w}_2$ D $\mathbf{w}_4$

12 (P) Skup za učenje čine sljedeći označeni primjeri:

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

Na skupu $\mathcal{D}$ treniramo logističku regresiju (LR) i stroj potpornih vektora s tvrdom marginom (SVM). Dodatno, treniramo model linearne regresije (LINR), gdje izlaz tog modela koristimo za klasifikaciju, tj. $h(\mathbf{x}) = \mathbf{1}\{\mathbf{w}^{\mathrm{T}}\mathbf{x} \geq 0\}$. Za modele SVM i LINR umjesto oznake $y = 0$ koristimo oznaku $y = -1$. Za treniranje modela LR koristimo dovoljan broj iteracija tako da možemo pretpostaviti da je dobivena pogreška unakrsne entropije praktički jednaka nuli. Razmotrite primjer $\mathbf{x}^{(7)} = (2, 0)$. Neka je $d(m)$ udaljenost primjera $\mathbf{x}^{(7)}$ od granice između klasa dobivene modelom m. **Što od navedenog vrijedi za tu udaljenost?**

A $d(\text{LINR}) < d(\text{LR}) < d(\text{SVM})$ C $d(\text{LR}) < d(\text{SVM}) < d(\text{LINR})$

B $d(\text{SVM}) < d(\text{LINR}) < d(\text{LR})$ D $d(\text{SVM}) < d(\text{LR}) < d(\text{LINR})$

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

A 1.19 B 7.11 C 2.54 D 4.03

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

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

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

A 6.00 B 8.00 C 4.02 D 12.02

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

15 (T) Algoritmi strojnog učenja mogu biti parametarski ili neparametarski. **Što je karakteristika neparametarskih algoritama strojnog učenja?**

- A Složenost modela raste s normom vektora parametara
- B Pretpostavljaju teorijsku distribuciju podataka
- C Eksplicitno modeliraju granicu između primjera
- D Broj parametara ovisi o broju primjera

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

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

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

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

17 (T) Optimizacijski problem algoritma SVM može se postaviti u formulaciji meke ili tvrde margine te u primarnoj ili dualnoj formulaciji. Ovisno o formulaciji, kvadratni program sadrži različit broj varijabli po kojima optimiramo (optimizacijske varijable). **Ako matrica dizajna ima više redaka nego stupaca, koja formulacija ima najmanje optimizacijskih varijabli?**

A Dualni problem tvrde margine   C Dualni problem meke margine
B Primarni problem meke margine   D Primarni problem tvrde margine

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

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

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

A 96   B 76   C 25   D 10

19 (P) Na 500 primjera sa 80 značajki treniramo rijetki jezgreni stroj s Gaussovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 38 prototipa. **Koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

- A Optimiramo 81 parametar, a naučeni model ima 3079 parametara
- B Optimiramo 501 parametara, a naučeni model ima 3040 parametara
- C Optimiramo 501 parametara, a naučeni model ima 3079 parametara
- D Optimiramo 501 parametara, a naučeni model ima 3541 parametar

20 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 5$ označenih primjera. Vektor oznaka je $\mathbf{y} = (+1, +1, -1, -1, +1)$. Euklidske udaljenosti između primjera dane su sljedećom matricom udaljenosti:

$$\mathbf{D} = \begin{pmatrix} 0.0 & 7.48 & 6.16 & 13.42 & 12.21 \\ 7.48 & 0.0 & 12.73 & 20.1 & 14.18 \\ 6.16 & 12.73 & 0.0 & 10.49 & 9.95 \\ 13.42 & 20.1 & 10.49 & 0.0 & 20.02 \\ 12.21 & 14.18 & 9.95 & 20.02 & 0.0 \end{pmatrix}$$

Treniranjem uz $C = 10$ i $\gamma = 0.0001$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (10, 1.052, 10, 10, 8.948)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za prvi primjer, $L(y^{(1)}, h(\mathbf{x}^{(1)}))$?**

A 0.24 B 1.64 C 0.03 D 1.18

21 (N) U ulaznome prostoru dimenzije $n = 3$ trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri $((2, -5, 15), -1)$, $((1, 8, -305), -1)$ i $((1, -6, 225), +1)$, a njima odgovarajući dualni koeficijenti su $\alpha_1 = 0.5$, $\alpha_2 = 0.8$ i $\alpha_3 = 0.9$. Treniranje smo proveli na skaliranim značajkama: svaku smo značajku x_j standardizirali primjenom transformacije $\frac{x_j - \mu_j}{\sigma_j}$, gdje su μ_j i σ_j srednja vrijednost odnosno varijanca značajke x_j u skupu označenih podataka $\mathcal{D}$. Parametri skaliranja su $\boldsymbol{\mu} = (15, -2, 100)$ i $\boldsymbol{\sigma} = (4, 1, 12)$. Model SVM-a koristimo za predikciju klase primjera $\mathbf{x} = (1, 2, -30)$. **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera $\mathbf{x}$?**

A $+541.53$ B -739.13 C $+907.43$ D -373.22

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

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

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

A 0.552 B 0.359 C 0.795 D 0.456

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

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

1 (T) Modeli strojnog učenja općenito su različite složenosti. S porastom složenosti modela raste vjerojatnost da model bude prenaučen. Ta vjerojatnost raste s količinom šuma u podatcima. **Zašto šum u podatcima za učenje može dovesti do prenaučenosti klasifikacijskog modela?**

- A Efekt šuma ja slučajan, pa će hipoteza koja se previše prilagodi šumu na skupu za učenje očekivano imati veliku pogrešku na ispitnom skupu gdje je šum drugačiji ili ga nema
- B Zbog šuma granica između klasa izgleda nelinearnijom nego što ona to zapravo jest, pa primjeri blizu granice znatno više doprinose pogrešci učenja nego primjeri koji su udaljeni od granice
- C Povećanjem količine šuma granica između klasa postaje sve nelinearnija, pa raste i složenost modela te dobivena hipoteza očekivano neće odgovarati granici između klasa na ispitnom skupu
- D Zbog šuma su oznake nekih primjera u skupu za učenje pogrešne, pa sve hipoteze iz modela imaju na tom skupu pogrešku koja je veća od nula, a još veća na ispitnom skupu

2 (T) Multikolinearnost značajki jedan je od problema koji može nastupiti kod primjene modela regresije na stvarnim podatcima. Efekt multikolinearnosti i savršene multikolinearnosti dobro je uočljiv kod optimizacijskoga postupka običnih najmanjih kvadrata (OLS) kada se on provodi izračunom pseudoinverza matrice dizajna. Neka je m broj značajki, Φ je matrica dizajna i $\mathbf{G} = \Phi^\mathrm{T}\Phi$ je Gramova matrica. **Koji je efekt savršene multikolinearnosti kod postupka OLS?**

- A $\mathrm{rang}(\Phi) < m + 1$, $\mathbf{G}$ nema puni rang i nema inverz, no ima pseudoinverz koji nije numerički stabilan
- B $\mathrm{rang}(\Phi) = N$, no $\mathrm{rang}(\mathbf{G}) < N$, pa $\mathbf{G}$ ima pseudoinverz, ali nema numerički stabilan inverz
- C Φ ima puni rang, $\mathrm{rang}(\mathbf{G}) > m$ i $\mathbf{G}$ ima inverz, ali s visokim kondicijskim brojem
- D Φ nema puni rang, $\mathrm{rang}(\mathbf{G}) < m + 1$ i $\mathbf{G}$ nema pseudoinverz

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

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

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

A 7.10 B 1.58 C 0.29 D 2.69

4 (P) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi:

$$E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$$

Što iz toga možemo zaključiti?

- A Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$
- B Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$
- C Model $\mathcal{H}_{\alpha_1}$ je podnaučen
- D Model $\mathcal{H}_{\alpha_2}$ je prenaučen

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

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geq 0\}$$
$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \geq \theta_0^2\}$$

Pored ova dva modela, razmatramo i njihove kombinacije, modele $\mathcal{H}_3$ i $\mathcal{H}_4$. Neka je $\mathcal{H}_3 = \mathcal{H}_1 \cup \mathcal{H}_2$ te neka je $\mathcal{H}_4$ skup funkcija definiranih kao $h_4(\mathbf{x};\boldsymbol{\theta}) = h_1(\mathbf{x}) \cdot h_2(\mathbf{x})$. Neka je E_k minimalna empirijska pogreška koja se modelom $\mathcal{H}_k$ može ostvariti na skupu $\mathcal{D}$, tj. $E_k = \mathrm{argmin}_{h\in\mathcal{H}_k} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

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

6 (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable y o zavisnoj varijabli x. Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka $\mathcal{D} = \{(x^{(i)}, y^{(i)})\} = \{(0,0),(2,0),(3,2),(5,2)\}$. Neka je h hipoteza koju dobivamo treniranjem modela te neka je L^i gubitak hipoteze h na primjeru $x^{(i)}$, tj. $L^i = L(y^{(i)}, h(x^{(i)}))$. **Što vrijedi za gubitke hipoteze na pojedinim primjerima?**

A $L^1 = L^4 < L^2 = L^3$ C $L^1 = L^2 = 1 < L^3 < L^4$
B $L^1 = L^3 = 0,\ L^2 = L^4 < 1$ D $L^1 = L^4 = 1,\ L^2 < L^3$

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

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

A 2.54 B 1.19 C 4.03 D 7.11

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu na sljedeći način:

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

Početne težine perceptrona neka su $\mathbf{w} = (-1,0,-1,2,3,2)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 40 B 100 C 18 D 96

9 (P) Na primjerima iz dvodimenzijskoga ulaznog prostora treniramo L_2-regulariziranu logističku regresiju. Neka su $\mathbf{w}_0 = (1,-4,4)$, $\mathbf{w}_1 = (1,-4,6)$, $\mathbf{w}_2 = (1,-1,7)$, $\mathbf{w}_3 = (1,-7,1)$ i $\mathbf{w}_4 = (1,-7,-3)$ vektori u prostoru parametara. Neka je $E(\mathbf{w}|\mathcal{D})$ neregularizirana pogreška unakrsne entropije na skupu za učenje $\mathcal{D}$. Pritom je $\mathbf{w}_0$ minimizator funkcije $E(\mathbf{w}|\mathcal{D})$ te vrijedi $E(\mathbf{w}_1|\mathcal{D}) = E(\mathbf{w}_2|\mathcal{D}) = E(\mathbf{w}_3|\mathcal{D})$. Napravite skicu izokontura funkcije pogreške u potprostoru $w_1 \times w_2$. Za treniranje modela koristimo gradijentni spust s linijskim pretraživanjem uz regularizacijski faktor $\lambda = 100$. Za tako naučen model vrijednost regularizacijskog izraza $\frac{\lambda}{2}\|\mathbf{w}\|^2$ jednaka je 400. Međutim, broj koraka gradijentnog spusta (broj poziva linijskog pretraživanja) ovisi o tome koliko će spust krivudati, a to ovisi o odabiru inicijalnih parametara. Kao moguće inicijalne parametre razmotrite vektore $\mathbf{w}_1$–$\mathbf{w}_4$. **S kojim inicijalnim parametarima će algoritam gradijentnog spusta konvergirati u najmanjem broju koraka?**

A $\mathbf{w}_3$ B $\mathbf{w}_1$ C $\mathbf{w}_2$ D $\mathbf{w}_4$

10 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.1, 0.5, -2, 0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1,-1),1)$?**

A 1.42 B 0.50 C 0.09 D 0.22

11 (T) Višeklasni problem može se riješiti binarnim klasifikatorom uz primjenu sheme OVO ili sheme OVR. Obje sheme imaju svoje prednosti i nedostatke. Pretpostavite da raspolažemo sa K klasa i da svaka klasa ima N/K primjera, gdje je N ukupan broj primjera u skupu za učenje. **Što su prednosti odnosno nedostatci OVO i OVR sheme u takvom slučaju?**

- A OVO svaki klasifikator trenira s $K/2$ puta manje primjera nego OVR, ali pozitivne klase kod OVR imaju K puta manje primjera nego kod OVO
- B OVR treba K puta više klasifikatora nego OVO, ali su kod OVO pozitivne klase $K/2$ puta manje zastupljene nego OVR
- C OVR iziskuje $K-1$ puta više klasifikatora od sheme OVO, ali kod OVO pozitivne klase imaju $K-1$ puta manje primjera nego kod OVR
- D OVO iziskuje $(K-1)/2$ puta više parametara nego OVR, ali svaki OVR klasifikator ima $K-1$ puta manje pozitivnih primjera nego negativnih

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

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

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

A 6.00 B 12.02 C 8.00 D 4.02

13 (P) Skup za učenje čine sljedeći označeni primjeri:

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

Na skupu $\mathcal{D}$ treniramo logističku regresiju (LR) i stroj potpornih vektora s tvrdom marginom (SVM). Dodatno, treniramo model linearne regresije (LINR), gdje izlaz tog modela koristimo za klasifikaciju, tj. $h(\mathbf{x}) = \mathbf{1}\{\mathbf{w}^{\mathrm{T}}\mathbf{x} \geq 0\}$. Za modele SVM i LINR umjesto oznake $y=0$ koristimo oznaku $y=-1$. Za treniranje modela LR koristimo dovoljan broj iteracija tako da možemo pretpostaviti da je dobivena pogreška unakrsne entropije praktički jednaka nuli. Razmotrite primjer $\mathbf{x}^{(7)} = (2,0)$. Neka je $d(m)$ udaljenost primjera $\mathbf{x}^{(7)}$ od granice između klasa dobivene modelom m. **Što od navedenog vrijedi za tu udaljenost?**

A $d(\mathrm{SVM}) < d(\mathrm{LINR}) < d(\mathrm{LR})$ C $d(\mathrm{SVM}) < d(\mathrm{LR}) < d(\mathrm{LINR})$

B $d(\mathrm{LINR}) < d(\mathrm{LR}) < d(\mathrm{SVM})$ D $d(\mathrm{LR}) < d(\mathrm{SVM}) < d(\mathrm{LINR})$

14 (T) Za optimizaciju parametara poopćenih linearnih modela može se koristiti stohastički gradijentni spust, odnosno pravilo LMS. Neka je $(\mathbf{x}, y)$ označeni primjer za koji radimo ažuriranje težina pomoću pravila LMS. **Što možemo reći o razlici između novih (ažuriranih) i starih težina (težina prije ažuriranja)?**

- A Razlika je to manja što je oznaka y bliže jedinici
- B Razlika je to veća što je stopa učenja η bliža jedinici
- C Razlika je to veća što je izlaz modela $h(\mathbf{x})$ bliži nuli
- D Razlika je to manja što je vektor $\boldsymbol{\phi}(\mathbf{x})$ bliži ishodištu

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

15 (T) Optimizacijski problem algoritma SVM može se postaviti u formulaciji meke ili tvrde margine te u primarnoj ili dualnoj formulaciji. Ovisno o formulaciji, kvadratni program sadrži različit broj varijabli po kojima optimiramo (optimizacijske varijable). **Ako matrica dizajna ima više redaka nego stupaca, koja formulacija ima najmanje optimizacijskih varijabli?**

A Dualni problem tvrde margine C Primarni problem tvrde margine

B Dualni problem meke margine D Primarni problem meke margine

16 (P) Na 800 primjera sa 50 značajki treniramo rijetki jezgreni stroj s Gaussovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 28 prototipa. **Koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

- A Optimiramo 51 parametar, a naučeni model ima 1429 parametara
- B Optimiramo 51 parametar, a naučeni model ima 1400 parametara
- C Optimiramo 801 parametar, a naučeni model ima 1429 parametara
- D Optimiramo 801 parametar, a naučeni model ima 1400 parametara

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

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

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

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

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

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

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

A 0.359 B 0.456 C 0.552 D 0.795

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

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

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

A 10 B 25 C 96 D 68

20 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 5$ označenih primjera. Vektor oznaka je $\mathbf{y} = (+1, +1, -1, -1, +1)$. Euklidske udaljenosti između primjera dane su sljedećom matricom udaljenosti:

$$\mathbf{D} = \begin{pmatrix} 0.0 & 7.48 & 6.16 & 13.42 & 12.21 \\ 7.48 & 0.0 & 12.73 & 20.1 & 14.18 \\ 6.16 & 12.73 & 0.0 & 10.49 & 9.95 \\ 13.42 & 20.1 & 10.49 & 0.0 & 20.02 \\ 12.21 & 14.18 & 9.95 & 20.02 & 0.0 \end{pmatrix}$$

Treniranjem uz $C = 10$ i $\gamma = 0.0001$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (10, 1.052, 10, 10, 8.948)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za treći primjer, $L(y^{(3)}, h(\mathbf{x}^{(3)}))$?**

A 1.64 B 0.03 C 1.18 D 0.24

21 (T) Algoritmi strojnog učenja mogu biti parametarski ili neparametarski. **Što je karakteristika neparametarskih algoritama strojnog učenja?**

- A Hiperparametri nemaju utjecaja na složenost modela
- B Eksplicitno modeliraju granicu između primjera
- C Broj parametara ovisi o broju primjera
- D Pretpostavljaju teorijsku distribuciju podataka

22 (N) U ulaznome prostoru dimenzije $n = 3$ trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri $((2, -5, 15), -1)$, $((1, 8, -305), -1)$ i $((1, -6, 225), +1)$, a njima odgovarajući dualni koeficijenti su $\alpha_1 = 0.5$, $\alpha_2 = 0.8$ i $\alpha_3 = 0.9$. Treniranje smo proveli na skaliranim značajkama: svaku smo značajku x_j standardizirali primjenom transformacije $\frac{x_j - \mu_j}{\sigma_j}$, gdje su μ_j i σ_j srednja vrijednost odnosno varijanca značajke x_j u skupu označenih podataka $\mathcal{D}$. Parametri skaliranja su $\boldsymbol{\mu} = (15, -2, 100)$ i $\boldsymbol{\sigma} = (4, 1, 12)$. Model SVM-a koristimo za predikciju klase primjera $\mathbf{x} = (1, -2, 5)$. **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera x?**

A -739.13 B $+907.43$ C -373.22 D $+541.53$

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

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

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

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

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

A 2.69 B 1.58 C 0.29 D 7.10

2 (T) Modeli strojnog učenja općenito su različite složenosti. S porastom složenosti modela raste vjerojatnost da model bude prenaučen. Ta vjerojatnost raste s količinom šuma u podatcima. **Zašto šum u podatcima za učenje može dovesti do prenaučenosti klasifikacijskog modela?**

A Efekt šuma ja slučajan, pa će hipoteza koja se previše prilagodi šumu na skupu za učenje očekivano imati veliku pogrešku na ispitnom skupu gdje je šum drugačiji ili ga nema

B Zbog šuma su oznake nekih primjera u skupu za učenje pogrešne, pa sve hipoteze iz modela imaju na tom skupu pogrešku koja je veća od nula, a još veća na ispitnom skupu

C Povećanjem količine šuma granica između klasa postaje sve nelinearnija, pa raste i složenost modela te dobivena hipoteza očekivano neće odgovarati granici između klasa na ispitnom skupu

D Zbog šuma granica između klasa izgleda nelinearnijom nego što ona to zapravo jest, pa primjeri blizu granice znatno više doprinose pogrešci učenja nego primjeri koji su udaljeni od granice

3 (P) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi:

$$E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$$

Što iz toga možemo zaključiti?

A Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$

B Model $\mathcal{H}_{\alpha_1}$ je podnaučen

C Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$

D Model $\mathcal{H}_{\alpha_2}$ je prenaučen

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

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geq 0\}$$
$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \geq \theta_0^2\}$$

Pored ova dva modela, razmatramo i njihove kombinacije, modele $\mathcal{H}_3$ i $\mathcal{H}_4$. Neka je $\mathcal{H}_3 = \mathcal{H}_1 \cup \mathcal{H}_2$ te neka je $\mathcal{H}_4$ skup funkcija definiranih kao $h_4(\mathbf{x};\boldsymbol{\theta}) = h_1(\mathbf{x}) \cdot h_2(\mathbf{x})$. Neka je E_k minimalna empirijska pogreška koja se modelom $\mathcal{H}_k$ može ostvariti na skupu $\mathcal{D}$, tj. $E_k = \mathrm{argmin}_{h\in\mathcal{H}_k} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

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

5 (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable y o zavisnoj varijabli x. Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka $\mathcal{D} = \{(x^{(i)}, y^{(i)})\} = \{(0,0), (2,0), (3,2), (5,2)\}$. Neka je h hipoteza koju dobivamo treniranjem modela te neka je L^i gubitak hipoteze h na primjeru $x^{(i)}$, tj. $L^i = L(y^{(i)}, h(x^{(i)}))$. **Što vrijedi za gubitke hipoteze na pojedinim primjerima?**

A $L^1 = L^3 = 0$, $L^2 = L^4 < 1$ C $L^1 = L^4 = 1$, $L^2 < L^3$
B $L^1 = L^4 < L^2 = L^3$ D $L^1 = L^2 = 1 < L^3 < L^4$

6 (T) Multikolinearnost značajki jedan je od problema koji može nastupiti kod primjene modela regresije na stvarnim podatcima. Efekt multikolinearnosti i savršene multikolinearnosti dobro je uočljiv kod optimizacijskoga postupka običnih najmanjih kvadrata (OLS) kada se on provodi izračunom pseudoinverza matrice dizajna. Neka je m broj značajki, Φ je matrica dizajna i $\mathbf{G} = \Phi^{\mathrm{T}}\Phi$ je Gramova matrica. **Koji je efekt savršene multikolinearnosti kod postupka OLS?**

A $\mathrm{rang}(\Phi) = N$, no $\mathrm{rang}(\mathbf{G}) < N$, pa $\mathbf{G}$ ima pseudoinverz, ali nema numerički stabilan inverz

B $\mathrm{rang}(\Phi) < m + 1$, $\mathbf{G}$ nema puni rang i nema inverz, no ima pseudoinverz koji nije numerički stabilan

C Φ ima puni rang, $\mathrm{rang}(\mathbf{G}) > m$ i $\mathbf{G}$ ima inverz, ali s visokim kondicijskim brojem

D Φ nema puni rang, $\mathrm{rang}(\mathbf{G}) < m + 1$ i $\mathbf{G}$ nema pseudoinverz

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

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

A 2.54 B 7.11 C 4.03 D 1.19

8 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.1, 0.5, -2, -0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1, -1), 1)$?**

A 0.09 B 0.22 C 1.42 D 0.50

9 (P) Na primjerima iz dvodimenzijskoga ulaznog prostora treniramo L_2-regulariziranu logističku regresiju. Neka su $\mathbf{w}_0 = (1, -4, 4)$, $\mathbf{w}_1 = (1, -4, 6)$, $\mathbf{w}_2 = (1, -1, 7)$, $\mathbf{w}_3 = (1, -7, 1)$ i $\mathbf{w}_4 = (1, -7, -3)$ vektori u prostoru parametara. Neka je $E(\mathbf{w}|\mathcal{D})$ neregularizirana pogreška unakrsne entropije na skupu za učenje $\mathcal{D}$. Pritom je $\mathbf{w}_0$ minimizator funkcije $E(\mathbf{w}|\mathcal{D})$ te vrijedi $E(\mathbf{w}_1|\mathcal{D}) = E(\mathbf{w}_2|\mathcal{D}) = E(\mathbf{w}_3|\mathcal{D})$. Napravite skicu izokontura funkcije pogreške u potprostoru $w_1 \times w_2$. Za treniranje modela koristimo gradijentni spust s linijskim pretraživanjem uz regularizacijski faktor $\lambda = 100$. Za tako naučen model vrijednost regularizacijskog izraza $\frac{\lambda}{2}\|\mathbf{w}\|^2$ jednaka je 400. Međutim, broj koraka gradijentnog spusta (broj poziva linijskog pretraživanja) ovisi o tome koliko će spust krivudati, a to ovisi o odabiru inicijalnih parametara. Kao moguće inicijalne parametre razmotrite vektore $\mathbf{w}_1$–$\mathbf{w}_4$. **S kojim inicijalnim parametarima će algoritam gradijentnog spusta konvergirati u najmanjem broju koraka?**

A $\mathbf{w}_2$ B $\mathbf{w}_3$ C $\mathbf{w}_1$ D $\mathbf{w}_4$

10 (T) Za optimizaciju parametara poopćenih linearnih modela može se koristiti stohastički gradijentni spust, odnosno pravilo LMS. Neka je $(\mathbf{x}, y)$ označeni primjer za koji radimo ažuriranje težina pomoću pravila LMS. **Što možemo reći o razlici između novih (ažuriranih) i starih težina (težina prije ažuriranja)?**

A Razlika je to manja što je oznaka y bliže jedinici

B Razlika je to manja što je vektor $\boldsymbol{\phi}(\mathbf{x})$ bliži ishodištu

C Razlika je to veća što je stopa učenja η bliža jedinici

D Razlika je to veća što je izlaz modela $h(\mathbf{x})$ bliži nuli

11 (T) Višeklasni problem može se riješiti binarnim klasifikatorom uz primjenu sheme OVO ili sheme OVR. Obje sheme imaju svoje prednosti i nedostatke. Pretpostavite da raspolažemo sa K klasa i da svaka klasa ima N/K primjera, gdje je N ukupan broj primjera u skupu za učenje. **Što su prednosti odnosno nedostatci OVO i OVR sheme u takvom slučaju?**

- A OVR treba K puta više klasifikatora nego OVO, ali su kod OVO pozitivne klase $K/2$ puta manje zastupljene nego OVR
- B OVO iziskuje $(K-1)/2$ puta više parametara nego OVR, ali svaki OVR klasifikator ima $K-1$ puta manje pozitivnih primjera nego negativnih
- C OVR iziskuje $K-1$ puta više klasifikatora od sheme OVO, ali kod OVO pozitivne klase imaju $K-1$ puta manje primjera nego kod OVR
- D OVO svaki klasifikator trenira s $K/2$ puta manje primjera nego OVR, ali pozitivne klase kod OVR imaju K puta manje primjera nego kod OVO

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

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

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

A 6.00 B 4.02 C 12.02 D 8.00

13 (P) Skup za učenje čine sljedeći označeni primjeri:

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

Na skupu $\mathcal{D}$ treniramo logističku regresiju (LR) i stroj potpornih vektora s tvrdom marginom (SVM). Dodatno, treniramo model linearne regresije (LINR), gdje izlaz tog modela koristimo za klasifikaciju, tj. $h(\mathbf{x}) = \mathbf{1}\{\mathbf{w}^{\mathrm{T}}\mathbf{x} \geq 0\}$. Za modele SVM i LINR umjesto oznake $y = 0$ koristimo oznaku $y = -1$. Za treniranje modela LR koristimo dovoljan broj iteracija tako da možemo pretpostaviti da je dobivena pogreška unakrsne entropije praktički jednaka nuli. Razmotrite primjer $\mathbf{x}^{(7)} = (2, 0)$. Neka je $d(m)$ udaljenost primjera $\mathbf{x}^{(7)}$ od granice između klasa dobivene modelom m. **Što od navedenog vrijedi za tu udaljenost?**

A $d(\text{SVM}) < d(\text{LR}) < d(\text{LINR})$ C $d(\text{LINR}) < d(\text{LR}) < d(\text{SVM})$
B $d(\text{SVM}) < d(\text{LINR}) < d(\text{LR})$ D $d(\text{LR}) < d(\text{SVM}) < d(\text{LINR})$

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu na sljedeći način:

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

Početne težine perceptrona neka su $\mathbf{w} = (1, 0, -1, 2, -3, 0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 96 B 100 C 18 D 40

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

15 (T) Optimizacijski problem algoritma SVM može se postaviti u formulaciji meke ili tvrde margine te u primarnoj ili dualnoj formulaciji. Ovisno o formulaciji, kvadratni program sadrži različit broj varijabli po kojima optimiramo (optimizacijske varijable). **Ako matrica dizajna ima više redaka nego stupaca, koja formulacija ima najmanje optimizacijskih varijabli?**

[A] Dualni problem meke margine [C] Primarni problem tvrde margine

[B] Dualni problem tvrde margine [D] Primarni problem meke margine

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

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

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

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

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

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

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

[A] 0.359 [B] 0.552 [C] 0.795 [D] 0.456

18 (T) Algoritmi strojnog učenja mogu biti parametarski ili neparametarski. **Što je karakteristika neparametarskih algoritama strojnog učenja?**

[A] Složenost modela raste s normom vektora parametara

[B] Broj značajki ne ovisi o dimenziji ulaznog prostora

[C] Svaki primjer ima globalan utjecaj na izgled hipoteze

[D] Broj parametara ovisi o broju primjera

19 (N) U ulaznome prostoru dimenzije $n = 3$ trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri $((2,-5,15),-1)$, $((1,8,-305),-1)$ i $((1,-6,225),+1)$, a njima odgovarajući dualni koeficijenti su $\alpha_1 = 0.5$, $\alpha_2 = 0.8$ i $\alpha_3 = 0.9$. Treniranje smo proveli na skaliranim značajkama: svaku smo značajku x_j standardizirali primjenom transformacije $\frac{x_j - \mu_j}{\sigma_j}$, gdje su μ_j i σ_j srednja vrijednost odnosno varijanca značajke x_j u skupu označenih podataka $\mathcal{D}$. Parametri skaliranja su $\boldsymbol{\mu} = (15, -2, 100)$ i $\boldsymbol{\sigma} = (4, 1, 12)$. Model SVM-a koristimo za predikciju klase primjera $\mathbf{x} = (1, 2, -30)$. **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera x?**

[A] -373.22 [B] -739.13 [C] $+907.43$ [D] $+541.53$

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

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

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

[A] 10 [B] 96 [C] 68 [D] 25

21 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 5$ označenih primjera. Vektor oznaka je $\mathbf{y} = (+1, +1, -1, -1, +1)$. Euklidske udaljenosti između primjera dane su sljedećom matricom udaljenosti:

$$\mathbf{D} = \begin{pmatrix} 0.0 & 7.48 & 6.16 & 13.42 & 12.21 \\ 7.48 & 0.0 & 12.73 & 20.1 & 14.18 \\ 6.16 & 12.73 & 0.0 & 10.49 & 9.95 \\ 13.42 & 20.1 & 10.49 & 0.0 & 20.02 \\ 12.21 & 14.18 & 9.95 & 20.02 & 0.0 \end{pmatrix}$$

Treniranjem uz $C = 10$ i $\gamma = 0.001$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (10, 0, 10, 2.779, 2.779)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za prvi primjer, $L(y^{(1)}, h(\mathbf{x}^{(1)}))$?**

A 0.24 B 1.64 C 1.18 D 0.03

22 (P) Na 500 primjera sa 80 značajki treniramo rijetki jezgreni stroj s Gaussovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 38 prototipa. **Koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

A Optimiramo 81 parametar, a naučeni model ima 3040 parametara

B Optimiramo 81 parametar, a naučeni model ima 3079 parametara

C Optimiramo 501 parametara, a naučeni model ima 3079 parametara

D Optimiramo 501 parametara, a naučeni model ima 3040 parametara

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

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

1 (T) Modeli strojnog učenja općenito su različite složenosti. S porastom složenosti modela raste vjerojatnost da model bude prenaučen. Ta vjerojatnost raste s količinom šuma u podatcima. **Zašto šum u podatcima za učenje može dovesti do prenaučenosti klasifikacijskog modela?**

A Zbog šuma granica između klasa izgleda nelinearnijom nego što ona to zapravo jest, pa primjeri blizu granice znatno više doprinose pogrešci učenja nego primjeri koji su udaljeni od granice

B Zbog šuma su oznake nekih primjera u skupu za učenje pogrešne, pa sve hipoteze iz modela imaju na tom skupu pogrešku koja je veća od nula, a još veća na ispitnom skupu

C Povećanjem količine šuma granica između klasa postaje sve nelinearnija, pa raste i složenost modela te dobivena hipoteza očekivano neće odgovarati granici između klasa na ispitnom skupu

D Efekt šuma ja slučajan, pa će hipoteza koja se previše prilagodi šumu na skupu za učenje očekivano imati veliku pogrešku na ispitnom skupu gdje je šum drugačiji ili ga nema

2 (T) Multikolinearnost značajki jedan je od problema koji može nastupiti kod primjene modela regresije na stvarnim podatcima. Efekt multikolinearnosti i savršene multikolinearnosti dobro je uočljiv kod optimizacijskoga postupka običnih najmanjih kvadrata (OLS) kada se on provodi izračunom pseudoinverza matrice dizajna. Neka je m broj značajki, Φ je matrica dizajna i $\mathbf{G} = \Phi^{\mathrm{T}}\Phi$ je Gramova matrica. **Koji je efekt savršene multikolinearnosti kod postupka OLS?**

A Φ ima puni rang, $\mathrm{rang}(\mathbf{G}) > m$ i $\mathbf{G}$ ima inverz, ali s visokim kondicijskim brojem

B $\mathrm{rang}(\Phi) = N$, no $\mathrm{rang}(\mathbf{G}) < N$, pa $\mathbf{G}$ ima pseudoinverz, ali nema numerički stabilan inverz

C Φ nema puni rang, $\mathrm{rang}(\mathbf{G}) < m+1$ i $\mathbf{G}$ nema pseudoinverz

D $\mathrm{rang}(\Phi) < m+1$, $\mathbf{G}$ nema puni rang i nema inverz, no ima pseudoinverz koji nije numerički stabilan

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

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geq 0\}$$
$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \geq \theta_0^2\}$$

Pored ova dva modela, razmatramo i njihove kombinacije, modele $\mathcal{H}_3$ i $\mathcal{H}_4$. Neka je $\mathcal{H}_3 = \mathcal{H}_1 \cup \mathcal{H}_2$ te neka je $\mathcal{H}_4$ skup funkcija definiranih kao $h_4(\mathbf{x};\boldsymbol{\theta}) = h_1(\mathbf{x}) \cdot h_2(\mathbf{x})$. Neka je E_k minimalna empirijska pogreška koja se modelom $\mathcal{H}_k$ može ostvariti na skupu $\mathcal{D}$, tj. $E_k = \mathrm{argmin}_{h\in\mathcal{H}_k} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

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

4 (P) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi:

$$E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$$

Što iz toga možemo zaključiti?

A Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$

B Model $\mathcal{H}_{\alpha_1}$ je podnaučen

C Model $\mathcal{H}_{\alpha_2}$ je prenaučen

D Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$

5 (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable y o zavisnoj varijabli x. Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka $\mathcal{D} = \{(x^{(i)}, y^{(i)})\} = \{(0,0), (2,0), (3,2), (5,2)\}$. Neka je h hipoteza koju dobivamo treniranjem modela te neka je L^i gubitak hipoteze h na primjeru $x^{(i)}$, tj. $L^i = L(y^{(i)}, h(x^{(i)}))$. **Što vrijedi za gubitke hipoteze na pojedinim primjerima?**

A $L^1 = L^2 = 1 < L^3 < L^4$ C $L^1 = L^4 < L^2 = L^3$

B $L^1 = L^4 = 1,\ L^2 < L^3$ D $L^1 = L^3 = 0,\ L^2 = L^4 < 1$

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

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

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

A 7.10 B 1.58 C 0.29 D 2.69

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

7 (P) Skup za učenje čine sljedeći označeni primjeri:

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

Na skupu $\mathcal{D}$ treniramo logističku regresiju (LR) i stroj potpornih vektora s tvrdom marginom (SVM). Dodatno, treniramo model linearne regresije (LINR), gdje izlaz tog modela koristimo za klasifikaciju, tj. $h(\mathbf{x}) = \mathbf{1}\{\mathbf{w}^{\mathrm{T}}\mathbf{x} \geq 0\}$. Za modele SVM i LINR umjesto oznake $y = 0$ koristimo oznaku $y = -1$. Za treniranje modela LR koristimo dovoljan broj iteracija tako da možemo pretpostaviti da je dobivena pogreška unakrsne entropije praktički jednaka nuli. Razmotrite primjer $\mathbf{x}^{(7)} = (2,0)$. Neka je $d(m)$ udaljenost primjera $\mathbf{x}^{(7)}$ od granice između klasa dobivene modelom m. **Što od navedenog vrijedi za tu udaljenost?**

A $d(\text{LR}) < d(\text{SVM}) < d(\text{LINR})$ C $d(\text{SVM}) < d(\text{LR}) < d(\text{LINR})$

B $d(\text{SVM}) < d(\text{LINR}) < d(\text{LR})$ D $d(\text{LINR}) < d(\text{LR}) < d(\text{SVM})$

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

A 4.03 B 2.54 C 7.11 D 1.19

9 (T) Za optimizaciju parametara poopćenih linearnih modela može se koristiti stohastički gradijentni spust, odnosno pravilo LMS. Neka je $(\mathbf{x}, y)$ označeni primjer za koji radimo ažuriranje težina pomoću pravila LMS. **Što možemo reći o razlici između novih (ažuriranih) i starih težina (težina prije ažuriranja)?**

A Razlika je to manja što je oznaka y bliže jedinici

B Razlika je to veća što je izlaz modela $h(\mathbf{x})$ bliži nuli

C Razlika je to veća što je stopa učenja η bliža jedinici

D Razlika je to manja što je vektor $\boldsymbol{\phi}(\mathbf{x})$ bliži ishodištu

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

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

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

A 12.02 B 6.00 C 4.02 D 8.00

11 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.1, 0.5, -2, -0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1, -1), 1)$?**

A 0.09 B 0.22 C 0.50 D 1.42

12 (P) Na primjerima iz dvodimenzijskoga ulaznog prostora treniramo L_2-regulariziranu logističku regresiju. Neka su $\mathbf{w}_0 = (1, -4, 4)$, $\mathbf{w}_1 = (1, -4, 6)$, $\mathbf{w}_2 = (1, -1, 7)$, $\mathbf{w}_3 = (1, -7, 1)$ i $\mathbf{w}_4 = (1, -7, -3)$ vektori u prostoru parametara. Neka je $E(\mathbf{w}|\mathcal{D})$ neregularizirana pogreška unakrsne entropije na skupu za učenje $\mathcal{D}$. Pritom je $\mathbf{w}_0$ minimizator funkcije $E(\mathbf{w}|\mathcal{D})$ te vrijedi $E(\mathbf{w}_1|\mathcal{D}) = E(\mathbf{w}_2|\mathcal{D}) = E(\mathbf{w}_3|\mathcal{D})$. Napravite skicu izokontura funkcije pogreške u potprostoru $w_1 \times w_2$. Za treniranje modela koristimo gradijentni spust s linijskim pretraživanjem uz regularizacijski faktor $\lambda = 100$. Za tako naučen model vrijednost regularizacijskog izraza $\frac{\lambda}{2}\|\mathbf{w}\|^2$ jednaka je 400. Međutim, broj koraka gradijentnog spusta (broj poziva linijskog pretraživanja) ovisi o tome koliko će spust krivudati, a to ovisi o odabiru inicijalnih parametara. Kao moguće inicijalne parametre razmotrite vektore $\mathbf{w}_1$–$\mathbf{w}_4$. **S kojim inicijalnim parametarima će algoritam gradijentnog spusta konvergirati u najmanjem broju koraka?**

A $\mathbf{w}_1$ B $\mathbf{w}_2$ C $\mathbf{w}_4$ D $\mathbf{w}_3$

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu na sljedeći način:

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

Početne težine perceptrona neka su $\mathbf{w} = (1, 0, -1, 2, 3, 2)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 96 B 40 C 18 D 100

14 (T) Višeklasni problem može se riješiti binarnim klasifikatorom uz primjenu sheme OVO ili sheme OVR. Obje sheme imaju svoje prednosti i nedostatke. Pretpostavite da raspolažemo sa K klasa i da svaka klasa ima N/K primjera, gdje je N ukupan broj primjera u skupu za učenje. **Što su prednosti odnosno nedostatci OVO i OVR sheme u takvom slučaju?**

A OVR iziskuje $K - 1$ puta više klasifikatora od sheme OVO, ali kod OVO pozitivne klase imaju $K - 1$ puta manje primjera nego kod OVR

B OVO svaki klasifikator trenira s $K/2$ puta manje primjera nego OVR, ali pozitivne klase kod OVR imaju K puta manje primjera nego kod OVO

C OVO iziskuje $(K - 1)/2$ puta više parametara nego OVR, ali svaki OVR klasifikator ima $K - 1$ puta manje pozitivnih primjera nego negativnih

D OVR treba K puta više klasifikatora nego OVO, ali su kod OVO pozitivne klase $K/2$ puta manje zastupljene nego OVR

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

15 (T) Optimizacijski problem algoritma SVM može se postaviti u formulaciji meke ili tvrde margine te u primarnoj ili dualnoj formulaciji. Ovisno o formulaciji, kvadratni program sadrži različit broj varijabli po kojima optimiramo (optimizacijske varijable). **Ako matrica dizajna ima više redaka nego stupaca, koja formulacija ima najmanje optimizacijskih varijabli?**

A Dualni problem meke margine
C Primarni problem meke margine
B Primarni problem tvrde margine
D Dualni problem tvrde margine

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

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

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

A 10 B 25 C 76 D 96

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

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

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

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

18 (P) Na 900 primjera sa 50 značajki treniramo rijetki jezgreni stroj s Gaussovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 52 prototipa. **Koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

A Optimiramo 51 parametar, a naučeni model ima 2653 parametara

B Optimiramo 901 parametar, a naučeni model ima 2600 parametara

C Optimiramo 51 parametar, a naučeni model ima 2600 parametara

D Optimiramo 901 parametar, a naučeni model ima 2653 parametara

19 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 5$ označenih primjera. Vektor oznaka je $\mathbf{y} = (+1, +1, -1, -1, +1)$. Euklidske udaljenosti između primjera dane su sljedećom matricom udaljenosti:

$$\mathbf{D} = \begin{pmatrix} 0.0 & 7.48 & 6.16 & 13.42 & 12.21 \\ 7.48 & 0.0 & 12.73 & 20.1 & 14.18 \\ 6.16 & 12.73 & 0.0 & 10.49 & 9.95 \\ 13.42 & 20.1 & 10.49 & 0.0 & 20.02 \\ 12.21 & 14.18 & 9.95 & 20.02 & 0.0 \end{pmatrix}$$

Treniranjem uz $C = 10$ i $\gamma = 0.001$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (10, 0, 10, 2.779, 2.779)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za prvi primjer, $L(y^{(1)}, h(\mathbf{x}^{(1)}))$?**

A 0.03 B 0.24 C 1.64 D 1.18

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

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

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

A 0.552 B 0.456 C 0.359 D 0.795

21 (N) U ulaznome prostoru dimenzije $n = 3$ trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri $((2, -5, 15), -1)$, $(((1, 8, -305), -1)$ i $((1, -6, 225), +1)$, a njima odgovarajući dualni koeficijenti su $\alpha_1 = 0.5$, $\alpha_2 = 0.8$ i $\alpha_3 = 0.9$. Treniranje smo proveli na skaliranim značajkama: svaku smo značajku x_j standardizirali primjenom transformacije $\frac{x_j - \mu_j}{\sigma_j}$, gdje su μ_j i σ_j srednja vrijednost odnosno varijanca značajke x_j u skupu označenih podataka $\mathcal{D}$. Parametri skaliranja su $\boldsymbol{\mu} = (15, -2, 100)$ i $\boldsymbol{\sigma} = (4, 1, 12)$. Model SVM-a koristimo za predikciju klase primjera $\mathbf{x} = (1, -2, 5)$. **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera x?**

A -373.22 B $+541.53$ C $+907.43$ D -739.13

22 (T) Algoritmi strojnog učenja mogu biti parametarski ili neparametarski. **Što je karakteristika neparametarskih algoritama strojnog učenja?**

A Hiperparametri nemaju utjecaja na složenost modela

B Broj parametara ovisi o broju primjera

C Broj značajki ne ovisi o dimenziji ulaznog prostora

D Složenost modela raste s normom vektora parametara

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

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

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

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^\mathrm{T}\mathbf{x} \geq 0\}$$
$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \geq \theta_0^2\}$$

Pored ova dva modela, razmatramo i njihove kombinacije, modele $\mathcal{H}_3$ i $\mathcal{H}_4$. Neka je $\mathcal{H}_3 = \mathcal{H}_1 \cup \mathcal{H}_2$ te neka je $\mathcal{H}_4$ skup funkcija definiranih kao $h_4(\mathbf{x};\boldsymbol{\theta}) = h_1(\mathbf{x}) \cdot h_2(\mathbf{x})$. Neka je E_k minimalna empirijska pogreška koja se modelom $\mathcal{H}_k$ može ostvariti na skupu $\mathcal{D}$, tj. $E_k = \mathrm{argmin}_{h\in\mathcal{H}_k} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

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

2 (T) Multikolinearnost značajki jedan je od problema koji može nastupiti kod primjene modela regresije na stvarnim podatcima. Efekt multikolinearnosti i savršene multikolinearnosti dobro je uočljiv kod optimizacijskoga postupka običnih najmanjih kvadrata (OLS) kada se on provodi izračunom pseudoinverza matrice dizajna. Neka je m broj značajki, Φ je matrica dizajna i $\mathbf{G} = \Phi^\mathrm{T}\Phi$ je Gramova matrica. **Koji je efekt savršene multikolinearnosti kod postupka OLS?**

A $\mathrm{rang}(\Phi) = N$, no $\mathrm{rang}(\mathbf{G}) < N$, pa $\mathbf{G}$ ima pseudoinverz, ali nema numerički stabilan inverz

B Φ nema puni rang, $\mathrm{rang}(\mathbf{G}) < m+1$ i $\mathbf{G}$ nema pseudoinverz

C Φ ima puni rang, $\mathrm{rang}(\mathbf{G}) > m$ i $\mathbf{G}$ ima inverz, ali s visokim kondicijskim brojem

D $\mathrm{rang}(\Phi) < m+1$, $\mathbf{G}$ nema puni rang i nema inverz, no ima pseudoinverz koji nije numerički stabilan

3 (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable y o zavisnoj varijabli x. Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka $\mathcal{D} = \{(x^{(i)}, y^{(i)})\} = \{(0,0),(2,0),(3,2),(5,2)\}$. Neka je h hipoteza koju dobivamo treniranjem modela te neka je L^i gubitak hipoteze h na primjeru $x^{(i)}$, tj. $L^i = L(y^{(i)}, h(x^{(i)}))$. **Što vrijedi za gubitke hipoteze na pojedinim primjerima?**

A $L^1 = L^3 = 0$, $L^2 = L^4 < 1$ C $L^1 = L^4 < L^2 = L^3$

B $L^1 = L^4 = 1$, $L^2 < L^3$ D $L^1 = L^2 = 1 < L^3 < L^4$

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

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

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

A 0.29 B 2.69 C 7.10 D 1.58

5 (P) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi:

$$E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$$

Što iz toga možemo zaključiti?

- A Model $\mathcal{H}_{\alpha_2}$ je prenaučen
- B Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$
- C Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$
- D Model $\mathcal{H}_{\alpha_1}$ je podnaučen

6 (T) Modeli strojnog učenja općenito su različite složenosti. S porastom složenosti modela raste vjerojatnost da model bude prenaučen. Ta vjerojatnost raste s količinom šuma u podatcima. **Zašto šum u podatcima za učenje može dovesti do prenaučenosti klasifikacijskog modela?**

- A Efekt šuma ja slučajan, pa će hipoteza koja se previše prilagodi šumu na skupu za učenje očekivano imati veliku pogrešku na ispitnom skupu gdje je šum drugačiji ili ga nema
- B Zbog šuma granica između klasa izgleda nelinearnijom nego što ona to zapravo jest, pa primjeri blizu granice znatno više doprinose pogrešci učenja nego primjeri koji su udaljeni od granice
- C Zbog šuma su oznake nekih primjera u skupu za učenje pogrešne, pa sve hipoteze iz modela imaju na tom skupu pogrešku koja je veća od nula, a još veća na ispitnom skupu
- D Povećanjem količine šuma granica između klasa postaje sve nelinearnija, pa raste i složenost modela te dobivena hipoteza očekivano neće odgovarati granici između klasa na ispitnom skupu

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

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

A 4.03 B 1.19 C 7.11 D 2.54

8 (T) Višeklasni problem može se riješiti binarnim klasifikatorom uz primjenu sheme OVO ili sheme OVR. Obje sheme imaju svoje prednosti i nedostatke. Pretpostavite da raspolažemo sa K klasa i da svaka klasa ima N/K primjera, gdje je N ukupan broj primjera u skupu za učenje. **Što su prednosti odnosno nedostatci OVO i OVR sheme u takvom slučaju?**

- A OVR treba K puta više klasifikatora nego OVO, ali su kod OVO pozitivne klase $K/2$ puta manje zastupljene nego OVR
- B OVR iziskuje $K-1$ puta više klasifikatora od sheme OVO, ali kod OVO pozitivne klase imaju $K-1$ puta manje primjera nego kod OVR
- C OVO iziskuje $(K-1)/2$ puta više parametara nego OVR, ali svaki OVR klasifikator ima $K-1$ puta manje pozitivnih primjera nego negativnih
- D OVO svaki klasifikator trenira s $K/2$ puta manje primjera nego OVR, ali pozitivne klase kod OVR imaju K puta manje primjera nego kod OVO

9 (T) Za optimizaciju parametara poopćenih linearnih modela može se koristiti stohastički gradijentni spust, odnosno pravilo LMS. Neka je $(\mathbf{x}, y)$ označeni primjer za koji radimo ažuriranje težina pomoću pravila LMS. **Što možemo reći o razlici između novih (ažuriranih) i starih težina (težina prije ažuriranja)?**

- A Razlika je to manja što je oznaka y bliže jedinici
- B Razlika je to veća što je stopa učenja η bliža jedinici
- C Razlika je to manja što je vektor $\boldsymbol{\phi}(\mathbf{x})$ bliži ishodištu
- D Razlika je to veća što je izlaz modela $h(\mathbf{x})$ bliži nuli

10 (P) Skup za učenje čine sljedeći označeni primjeri:

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

Na skupu $\mathcal{D}$ treniramo logističku regresiju (LR) i stroj potpornih vektora s tvrdom marginom (SVM). Dodatno, treniramo model linearne regresije (LINR), gdje izlaz tog modela koristimo za klasifikaciju, tj. $h(\mathbf{x}) = \mathbf{1}\{\mathbf{w}^{\mathrm{T}}\mathbf{x} \geq 0\}$. Za modele SVM i LINR umjesto oznake $y = 0$ koristimo oznaku $y = -1$. Za treniranje modela LR koristimo dovoljan broj iteracija tako da možemo pretpostaviti da je dobivena pogreška unakrsne entropije praktički jednaka nuli. Razmotrite primjer $\mathbf{x}^{(7)} = (2,0)$. Neka je $d(m)$ udaljenost primjera $\mathbf{x}^{(7)}$ od granice između klasa dobivene modelom m. **Što od navedenog vrijedi za tu udaljenost?**

A $d(\mathrm{LR}) < d(\mathrm{SVM}) < d(\mathrm{LINR})$ C $d(\mathrm{SVM}) < d(\mathrm{LINR}) < d(\mathrm{LR})$

B $d(\mathrm{LINR}) < d(\mathrm{LR}) < d(\mathrm{SVM})$ D $d(\mathrm{SVM}) < d(\mathrm{LR}) < d(\mathrm{LINR})$

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu na sljedeći način:

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

Početne težine perceptrona neka su $\mathbf{w} = (1,0,-1,2,3,2)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 100 B 18 C 40 D 96

12 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.1, -0.5, -2, 0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1,-1),1)$?**

A 0.50 B 0.22 C 1.42 D 0.09

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

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

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

A 6.00 B 8.00 C 12.02 D 4.02

14 (P) Na primjerima iz dvodimenzijskoga ulaznog prostora treniramo L_2-regulariziranu logističku regresiju. Neka su $\mathbf{w}_0 = (1,-4,4)$, $\mathbf{w}_1 = (1,-4,6)$, $\mathbf{w}_2 = (1,-1,7)$, $\mathbf{w}_3 = (1,-7,1)$ i $\mathbf{w}_4 = (1,-7,-3)$ vektori u prostoru parametara. Neka je $E(\mathbf{w}|\mathcal{D})$ neregularizirana pogreška unakrsne entropije na skupu za učenje $\mathcal{D}$. Pritom je $\mathbf{w}_0$ minimizator funkcije $E(\mathbf{w}|\mathcal{D})$ te vrijedi $E(\mathbf{w}_1|\mathcal{D}) = E(\mathbf{w}_2|\mathcal{D}) = E(\mathbf{w}_3|\mathcal{D})$. Napravite skicu izokontura funkcije pogreške u potprostoru $w_1 \times w_2$. Za treniranje modela koristimo gradijentni spust s linijskim pretraživanjem uz regularizacijski faktor $\lambda = 100$. Za tako naučen model vrijednost regularizacijskog izraza $\frac{\lambda}{2}\|\mathbf{w}\|^2$ jednaka je 400. Međutim, broj koraka gradijentnog spusta (broj poziva linijskog pretraživanja) ovisi o tome koliko će spust krivudati, a to ovisi o odabiru inicijalnih parametara. Kao moguće inicijalne parametre razmotrite vektore $\mathbf{w}_1$–$\mathbf{w}_4$. **S kojim inicijalnim parametarima će algoritam gradijentnog spusta konvergirati u najmanjem broju koraka?**

A $\mathbf{w}_3$ B $\mathbf{w}_1$ C $\mathbf{w}_2$ D $\mathbf{w}_4$

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

15 (T) Algoritmi strojnog učenja mogu biti parametarski ili neparametarski. **Što je karakteristika neparametarskih algoritama strojnog učenja?**

- A Broj značajki ne ovisi o dimenziji ulaznog prostora
- B Eksplicitno modeliraju granicu između primjera
- C Svaki primjer ima globalan utjecaj na izgled hipoteze
- D Broj parametara ovisi o broju primjera

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

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

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

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

17 (N) U ulaznome prostoru dimenzije $n = 3$ trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri $((2,-5,15),-1)$, $((1,8,-305),-1)$ i $((1,-6,225),+1)$, a njima odgovarajući dualni koeficijenti su $\alpha_1 = 0.5$, $\alpha_2 = 0.8$ i $\alpha_3 = 0.9$. Treniranje smo proveli na skaliranim značajkama: svaku smo značajku x_j standardizirali primjenom transformacije $\frac{x_j - \mu_j}{\sigma_j}$, gdje su μ_j i σ_j srednja vrijednost odnosno varijanca značajke x_j u skupu označenih podataka $\mathcal{D}$. Parametri skaliranja su $\boldsymbol{\mu} = (15,-2,100)$ i $\boldsymbol{\sigma} = (4,1,12)$. Model SVM-a koristimo za predikciju klase primjera $\mathbf{x} = (1,2,-20)$. **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera x?**

A -373.22 B -739.13 C $+907.43$ D $+541.53$

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

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

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

A 10 B 76 C 96 D 25

19 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 5$ označenih primjera. Vektor oznaka je $\mathbf{y} = (+1,+1,-1,-1,+1)$. Euklidske udaljenosti između primjera dane su sljedećom matricom udaljenosti:

$$\mathbf{D} = \begin{pmatrix} 0.0 & 7.48 & 6.16 & 13.42 & 12.21 \\ 7.48 & 0.0 & 12.73 & 20.1 & 14.18 \\ 6.16 & 12.73 & 0.0 & 10.49 & 9.95 \\ 13.42 & 20.1 & 10.49 & 0.0 & 20.02 \\ 12.21 & 14.18 & 9.95 & 20.02 & 0.0 \end{pmatrix}$$

Treniranjem uz $C = 10$ i $\gamma = 0.0001$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (10, 1.052, 10, 10, 8.948)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za prvi primjer, $L(y^{(1)}, h(\mathbf{x}^{(1)}))$?**

A 0.03 B 0.24 C 1.64 D 1.18

20 (P) Na 700 primjera sa 100 značajki treniramo rijetki jezgreni stroj s Gaussovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 42 prototipa. **Koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

- A Optimiramo 701 parametar, a naučeni model ima 4200 parametara
- B Optimiramo 701 parametar, a naučeni model ima 4243 parametara
- C Optimiramo 101 parametar, a naučeni model ima 4243 parametara
- D Optimiramo 101 parametar, a naučeni model ima 4200 parametara

21 (T) Optimizacijski problem algoritma SVM može se postaviti u formulaciji meke ili tvrde margine te u primarnoj ili dualnoj formulaciji. Ovisno o formulaciji, kvadratni program sadrži različit broj varijabli po kojima optimiramo (optimizacijske varijable). **Ako matrica dizajna ima više redaka nego stupaca, koja formulacija ima najmanje optimizacijskih varijabli?**

- A Dualni problem tvrde margine
- B Primarni problem tvrde margine
- C Dualni problem meke margine
- D Primarni problem meke margine

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

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

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

A 0.552 B 0.359 C 0.795 D 0.456

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

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

1 (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable y o zavisnoj varijabli x. Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka $\mathcal{D} = \{(x^{(i)}, y^{(i)})\} = \{(0,0),(2,0),(3,2),(5,2)\}$. Neka je h hipoteza koju dobivamo treniranjem modela te neka je L^i gubitak hipoteze h na primjeru $x^{(i)}$, tj. $L^i = L(y^{(i)}, h(x^{(i)}))$. **Što vrijedi za gubitke hipoteze na pojedinim primjerima?**

A $L^1 = L^4 < L^2 = L^3$
C $L^1 = L^2 = 1 < L^3 < L^4$
B $L^1 = L^4 = 1,\ L^2 < L^3$
D $L^1 = L^3 = 0,\ L^2 = L^4 < 1$

2 (T) Modeli strojnog učenja općenito su različite složenosti. S porastom složenosti modela raste vjerojatnost da model bude prenaučen. Ta vjerojatnost raste s količinom šuma u podatcima. **Zašto šum u podatcima za učenje može dovesti do prenaučenosti klasifikacijskog modela?**

A Zbog šuma granica između klasa izgleda nelinearnijom nego što ona to zapravo jest, pa primjeri blizu granice znatno više doprinose pogrešci učenja nego primjeri koji su udaljeni od granice

B Povećanjem količine šuma granica između klasa postaje sve nelinearnija, pa raste i složenost modela te dobivena hipoteza očekivano neće odgovarati granici između klasa na ispitnom skupu

C Zbog šuma su oznake nekih primjera u skupu za učenje pogrešne, pa sve hipoteze iz modela imaju na tom skupu pogrešku koja je veća od nula, a još veća na ispitnom skupu

D Efekt šuma ja slučajan, pa će hipoteza koja se previše prilagodi šumu na skupu za učenje očekivano imati veliku pogrešku na ispitnom skupu gdje je šum drugačiji ili ga nema

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

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

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

A 2.69 B 0.29 C 1.58 D 7.10

4 (T) Multikolinearnost značajki jedan je od problema koji može nastupiti kod primjene modela regresije na stvarnim podatcima. Efekt multikolinearnosti i savršene multikolinearnosti dobro je uočljiv kod optimizacijskoga postupka običnih najmanjih kvadrata (OLS) kada se on provodi izračunom pseudoinverza matrice dizajna. Neka je m broj značajki, Φ je matrica dizajna i $\mathbf{G} = \Phi^{\mathrm{T}}\Phi$ je Gramova matrica. **Koji je efekt savršene multikolinearnosti kod postupka OLS?**

A $\mathrm{rang}(\Phi) = N$, no $\mathrm{rang}(\mathbf{G}) < N$, pa $\mathbf{G}$ ima pseudoinverz, ali nema numerički stabilan inverz

B Φ ima puni rang, $\mathrm{rang}(\mathbf{G}) > m$ i $\mathbf{G}$ ima inverz, ali s visokim kondicijskim brojem

C Φ nema puni rang, $\mathrm{rang}(\mathbf{G}) < m+1$ i $\mathbf{G}$ nema pseudoinverz

D $\mathrm{rang}(\Phi) < m+1$, $\mathbf{G}$ nema puni rang i nema inverz, no ima pseudoinverz koji nije numerički stabilan

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

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geq 0\}$$
$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \geq \theta_0^2\}$$

Pored ova dva modela, razmatramo i njihove kombinacije, modele $\mathcal{H}_3$ i $\mathcal{H}_4$. Neka je $\mathcal{H}_3 = \mathcal{H}_1 \cup \mathcal{H}_2$ te neka je $\mathcal{H}_4$ skup funkcija definiranih kao $h_4(\mathbf{x};\boldsymbol{\theta}) = h_1(\mathbf{x}) \cdot h_2(\mathbf{x})$. Neka je E_k minimalna empirijska pogreška koja se modelom $\mathcal{H}_k$ može ostvariti na skupu $\mathcal{D}$, tj. $E_k = \mathrm{argmin}_{h \in \mathcal{H}_k} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

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

6 (P) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi:

$$E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$$

Što iz toga možemo zaključiti?

A Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$

B Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$

C Model $\mathcal{H}_{\alpha_2}$ je prenaučen

D Model $\mathcal{H}_{\alpha_1}$ je podnaučen

Cjelina 2: Linearni klasifikacijski modeli (8 pitanja)

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

A 7.11 B 2.54 C 4.03 D 1.19

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu na sljedeći način:

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

Početne težine perceptrona neka su $\mathbf{w} = (1, 0, -1, 2, 3, 0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 18 B 100 C 96 D 40

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

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

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

A 8.00 B 4.02 C 6.00 D 12.02

10 (T) Za optimizaciju parametara poopćenih linearnih modela može se koristiti stohastički gradijentni spust, odnosno pravilo LMS. Neka je $(\mathbf{x}, y)$ označeni primjer za koji radimo ažuriranje težina pomoću pravila LMS. **Što možemo reći o razlici između novih (ažuriranih) i starih težina (težina prije ažuriranja)?**

- A Razlika je to manja što je oznaka y bliže jedinici
- B Razlika je to manja što je vektor $\boldsymbol{\phi}(\mathbf{x})$ bliži ishodištu
- C Razlika je to veća što je izlaz modela $h(\mathbf{x})$ bliži nuli
- D Razlika je to veća što je stopa učenja η bliža jedinici

11 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.1, 0.5, 2, -0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1, -1), 1)$?**

A 0.50 B 0.09 C 1.42 D 0.22

12 (P) Skup za učenje čine sljedeći označeni primjeri:

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

Na skupu $\mathcal{D}$ treniramo logističku regresiju (LR) i stroj potpornih vektora s tvrdom marginom (SVM). Dodatno, treniramo model linearne regresije (LINR), gdje izlaz tog modela koristimo za klasifikaciju, tj. $h(\mathbf{x}) = \mathbf{1}\{\mathbf{w}^{\mathrm{T}}\mathbf{x} \geq 0\}$. Za modele SVM i LINR umjesto oznake $y = 0$ koristimo oznaku $y = -1$. Za treniranje modela LR koristimo dovoljan broj iteracija tako da možemo pretpostaviti da je dobivena pogreška unakrsne entropije praktički jednaka nuli. Razmotrite primjer $\mathbf{x}^{(7)} = (2, 0)$. Neka je $d(m)$ udaljenost primjera $\mathbf{x}^{(7)}$ od granice između klasa dobivene modelom m. **Što od navedenog vrijedi za tu udaljenost?**

A $d(\mathrm{SVM}) < d(\mathrm{LR}) < d(\mathrm{LINR})$ C $d(\mathrm{LINR}) < d(\mathrm{LR}) < d(\mathrm{SVM})$
B $d(\mathrm{SVM}) < d(\mathrm{LINR}) < d(\mathrm{LR})$ D $d(\mathrm{LR}) < d(\mathrm{SVM}) < d(\mathrm{LINR})$

13 (T) Višeklasni problem može se riješiti binarnim klasifikatorom uz primjenu sheme OVO ili sheme OVR. Obje sheme imaju svoje prednosti i nedostatke. Pretpostavite da raspolažemo sa K klasa i da svaka klasa ima N/K primjera, gdje je N ukupan broj primjera u skupu za učenje. **Što su prednosti odnosno nedostatci OVO i OVR sheme u takvom slučaju?**

- A OVO svaki klasifikator trenira s $K/2$ puta manje primjera nego OVR, ali pozitivne klase kod OVR imaju K puta manje primjera nego kod OVO
- B OVO iziskuje $(K-1)/2$ puta više parametara nego OVR, ali svaki OVR klasifikator ima $K-1$ puta manje pozitivnih primjera nego negativnih
- C OVR treba K puta više klasifikatora nego OVO, ali su kod OVO pozitivne klase $K/2$ puta manje zastupljene nego OVR
- D OVR iziskuje $K-1$ puta više klasifikatora od sheme OVO, ali kod OVO pozitivne klase imaju $K-1$ puta manje primjera nego kod OVR

14 (P) Na primjerima iz dvodimenzijskoga ulaznog prostora treniramo L_2-regulariziranu logističku regresiju. Neka su $\mathbf{w}_0 = (1, -4, 4)$, $\mathbf{w}_1 = (1, -4, 6)$, $\mathbf{w}_2 = (1, -1, 7)$, $\mathbf{w}_3 = (1, -7, 1)$ i $\mathbf{w}_4 = (1, -7, -3)$ vektori u prostoru parametara. Neka je $E(\mathbf{w}|\mathcal{D})$ neregularizirana pogreška unakrsne entropije na skupu za učenje $\mathcal{D}$. Pritom je $\mathbf{w}_0$ minimizator funkcije $E(\mathbf{w}|\mathcal{D})$ te vrijedi $E(\mathbf{w}_1|\mathcal{D}) = E(\mathbf{w}_2|\mathcal{D}) = E(\mathbf{w}_3|\mathcal{D})$. Napravite skicu izokontura funkcije pogreške u potprostoru $w_1 \times w_2$. Za treniranje modela koristimo gradijentni spust s linijskim pretraživanjem uz regularizacijski faktor $\lambda = 100$. Za tako naučen model vrijednost regularizacijskog izraza $\frac{\lambda}{2}\|\mathbf{w}\|^2$ jednaka je 400. Međutim, broj koraka gradijentnog spusta (broj poziva linijskog pretraživanja) ovisi o tome koliko će spust krivudati, a to ovisi o odabiru inicijalnih parametara. Kao moguće inicijalne parametre razmotrite vektore $\mathbf{w}_1$–$\mathbf{w}_4$. **S kojim inicijalnim parametarima će algoritam gradijentnog spusta konvergirati u najmanjem broju koraka?**

A $\mathbf{w}_4$ B $\mathbf{w}_3$ C $\mathbf{w}_1$ D $\mathbf{w}_2$

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

15 (T) Optimizacijski problem algoritma SVM može se postaviti u formulaciji meke ili tvrde margine te u primarnoj ili dualnoj formulaciji. Ovisno o formulaciji, kvadratni program sadrži različit broj varijabli po kojima optimiramo (optimizacijske varijable). **Ako matrica dizajna ima više redaka nego stupaca, koja formulacija ima najmanje optimizacijskih varijabli?**

A Dualni problem tvrde margine
C Dualni problem meke margine
B Primarni problem meke margine
D Primarni problem tvrde margine

16 (T) Algoritmi strojnog učenja mogu biti parametarski ili neparametarski. **Što je karakteristika neparametarskih algoritama strojnog učenja?**

A Pretpostavljaju teorijsku distribuciju podataka
B Broj parametara ovisi o broju primjera
C Hiperparametri nemaju utjecaja na složenost modela
D Broj značajki ne ovisi o dimenziji ulaznog prostora

17 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 5$ označenih primjera. Vektor oznaka je $\mathbf{y} = (+1, +1, -1, -1, +1)$. Euklidske udaljenosti između primjera dane su sljedećom matricom udaljenosti:

$$\mathbf{D} = \begin{pmatrix} 0.0 & 7.48 & 6.16 & 13.42 & 12.21 \\ 7.48 & 0.0 & 12.73 & 20.1 & 14.18 \\ 6.16 & 12.73 & 0.0 & 10.49 & 9.95 \\ 13.42 & 20.1 & 10.49 & 0.0 & 20.02 \\ 12.21 & 14.18 & 9.95 & 20.02 & 0.0 \end{pmatrix}$$

Treniranjem uz $C = 10$ i $\gamma = 0.001$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (10, 0, 10, 2.779, 2.779)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za prvi primjer, $L(y^{(1)}, h(\mathbf{x}^{(1)}))$?**

A 0.24 B 1.18 C 0.03 D 1.64

18 (P) Na 900 primjera sa 50 značajki treniramo rijetki jezgreni stroj s Gaussovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 52 prototipa. **Koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

A Optimiramo 51 parametar, a naučeni model ima 2653 parametara
B Optimiramo 901 parametar, a naučeni model ima 2653 parametara
C Optimiramo 51 parametar, a naučeni model ima 2600 parametara
D Optimiramo 901 parametar, a naučeni model ima 3501 parametar

19 (N) U ulaznome prostoru dimenzije $n = 3$ trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri $((2, -5, 15), -1)$, $((1, 8, -305), -1)$ i $((1, -6, 225), +1)$, a njima odgovarajući dualni koeficijenti su $\alpha_1 = 0.5$, $\alpha_2 = 0.8$ i $\alpha_3 = 0.9$. Treniranje smo proveli na skaliranim značajkama: svaku smo značajku x_j standardizirali primjenom transformacije $\frac{x_j - \mu_j}{\sigma_j}$, gdje su μ_j i σ_j srednja vrijednost odnosno varijanca značajke x_j u skupu označenih podataka $\mathcal{D}$. Parametri skaliranja su $\boldsymbol{\mu} = (15, -2, 100)$ i $\boldsymbol{\sigma} = (4, 1, 12)$. Model SVM-a koristimo za predikciju klase primjera $\mathbf{x} = (1, 2, -30)$. **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera x?**

A -373.22 B $+907.43$ C $+541.53$ D -739.13

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

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

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

A 0.359 B 0.456 C 0.552 D 0.795

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

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

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

A 10 B 25 C 68 D 96

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

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

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

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

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