Strojno učenje – međuispit

UNIZG FER, ak. god. 2016./2017.

28. studenog 2016.

Ispit traje 180 minuta i nosi 35 bodova. Svaki zadatak rješavajte na zasebnoj stranici. Pišite uredno i čitko. Nemojte pretpostavljati da je nešto očito; Vaše znanje može se ocijeniti samo na temelju onog što napišete. Kod skica grafikona, označite osi, budite uredni i precizni te označite ekstreme krivulja, ako postoje.

1. (*6 bodova*) Uvod i osnovni koncepti.

   (a) Formalno definirajte induktivnu pristranost i povežite vrste pristranosti s trima osnovnim komponentima algoritma strojnog učenja.

   (b) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α koji određuje složenost (veći α daje složeniji model). Model je nelinearan, ali je funkcija pogreške nekonveksna i optimizacija može zaglaviti u lokalnom optimumu. Model treniramo dvama algoritmima, L_1 i L_2, gdje je L_2 općenito bolji optimizacijski algoritam u smislu da češće nalazi globalni optimum od algoritma L_1. Skicirajte pogreške učenja i ispitivanja kao funkcije od α za oba algoritma (četiri krivulje na jednom grafikonu).

   (c) Primjere $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((0,0),1),(0,1),1),((0,-1),0)\}$ iz prostora $\mathcal{X} = \{-1,0,1\}^2$ želimo klasificirati modelom $h(x_1,x_2;\boldsymbol{\theta}) = \mathbf{1}\{\theta_1 x_1 + \theta_2 x_2 \geqslant 0\}$, gdje $(\theta_1,\theta_2) \in \mathbb{R}^2$. Odredite $|VS_{\mathcal{H},\mathcal{D}}|$ i $|\mathcal{H}|$.

2. (*6 bodova*) Regresija.

   (a) L2-regularizirana empirijska pogreška linearnog modela regresije jest $E(\mathbf{w}|\mathcal{D}) = \frac{1}{2}\sum_{i=1}^N \big(y^{(i)} - \mathbf{w}^\mathrm{T}\mathbf{x}^{(i)}\big)^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2$. Krenuvši od tog izraza, izvedite rješenje u matričnoj formi za težine $\mathbf{w}^*$ koje minimiziraju regulariziranu pogrešku.

   (b) Model regresije treniramo na podatcima koji su generirani funkcijom $f(x) = 3\cdot(x-2)^2 + 1$. Koristimo funkciju preslikavanja $\phi(x) = (1, x, x^2)$. Skicirajte izokonture neregularizirane funkcije pogreške u ravnini $\mathbb{R}^2$ koju definiraju parametri w_1 i w_2 i izokonture L2-regularizacijskog izraza. Ako je faktor λ odabran tako da se jednak značaj pridaje složenosti modela i minimizaciji pogreške, skicirajte (otprilike) vektor optimalnih težina (w_1^*, w_2^*).

   (c) Linearnom višestrukom regresijom modeliramo ovisnost prihoda (zavisna varijabla) o dobi, godinama radnog staža i broju djece (nezavisne varijable). Na ovom primjeru ukratko objasnite problem multikolinearnosti i kako ga regularizacija rješava.

3. (*4 boda*) Linearni diskriminativni modeli.

   (a) Definirajte poopćeni linearan model i navedite varijante modela koje su (i) nelinearne u parametrima i (ii) nelinearne u granici između klasa.

(b) Raspolažemo sljedećim primjerima za učenje

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

Za klasifikaciju koristimo regresiju i shemu *jedan-naspram-ostali* (OVR). Definirajte dizajn-matricu i vektor oznaka $\mathbf{y}$ za svaki od triju modela te skicirajte (otprilike) granice između klasa u ulaznome prostoru.

(c) Napišite funkciju pogreške perceptrona te navedite nedostatke perceptrona.

4. (*6 bodova*) Logistička regresija.

(a) Izvedite pogrešku unakrsne entropije $E(\mathbf{w}|\mathcal{D})$ kao negativnu log-izglednost na skupu označenih primjera.

(b) Izrazite gradijent funkcije pogreške unakrsne entropije $\nabla E(\mathbf{w}|\mathcal{D})$ te napišite pseudokôd algoritma gradijentnog spusta (*batch* i stohastičku izvedbu) za L2-regulariziranu logističku regresiju. Ukratko objasnite potrebu za linijskim pretraživanjem.

(c) Nacrtajte graf funkcije gubitka $L(\mathbf{x}, y)$ u ovisnosti o točnosti klasifikacije za: (i) gubitak 0-1, (ii) kvadratni gubitak i (iii) logistički gubitak. Na temelju skice odgovorite zašto je logistički gubitak dobar za klasifikaciju, a kvadratni gubitak to nije.

5. (*6 bodova*) Stroj potpornih vektora.

(a) Formulirajte, korak po korak, problem maksimalne margine i zatim izvedite, korak po korak, dualni kvadratni problem (tvrde) maksimalne margine s pripadnim uvjetima KKT.

(b) Neka su potporni vektori linearnog SVM-a $\mathbf{x}^{(1)} = (-2,3,5,5)$ i $\mathbf{x}^{(2)} = (6,4,3,1)$. Prvi primjer je negativan, a drugi pozitivan. Dualni parametri su $\alpha_1 = 0.2$ i $\alpha_2 = 0.5$, a pomak je $w_0 = -2$. Napišite izraz za gubitak zglobnice i odredite gubitak hipoteze za primjer $\mathbf{x}^{(3)} = (1,1,1,1)$, ako $y^{(3)} = -1$.

(c) Koristimo SVM s Gaussovom jezgrenom funkcijom. Skicirajte pogrešku učenja i ispitnu pogrešku kao funkcije hiperparametra γ, i to za $C = \{1, 100, 1000\}$ (ukupno šest krivulja).

6. (*7 bodova*) Jezgrene i neparametarske metode.

(a) Raspolažemo skupom primjera

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

Rabimo jezgreni stroj s dvije značajke i Gaussovom jezgrom $\kappa(\mathbf{x}, \boldsymbol{\mu}_j) = \exp(-\gamma_j \|\mathbf{x} - \boldsymbol{\mu}_j\|^2)$. Parametri jezgara su $\boldsymbol{\mu}_1 = (0,0)$, $\boldsymbol{\mu}_2 = (-3,3)$, $\gamma_1 = \gamma_2 = 1$. Skicirajte (otprilike) primjere u prostoru značajki i granicu koju biste dobili linearnom regresijom (nakon preslikavanja).

(b) Raspolažemo s 1000 primjera za učenje u 100-dimenzijskome ulaznom prostoru. Odredite (i) broj parametara i broj hiperparametara jezgrenog stroja sa Gaussovom jezgrom koji preslikava u 10-dimenzijski prostor, (ii) interval mogućeg broja parametara i hiperparametara rijetkog jezgrenog stroja s Gaussovom jezgrom i (iii) interval broja (hiper)parametara za rijetki stroj s linearnom jezgrom.

(c) Koristimo polinomijalnu jezgrenu funkciju $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^{\mathrm{T}}\mathbf{z} + 1)^2$. Pokažite da je za $n = 2$ jezgra κ Mercerova jezgra. Što to znači i zašto je to bitno?

(d) Nabrojite prednosti inverznog oblikovanja.

(e) Definirajte model k-NN i težinski k-NN. Kako biste odredili optimalnu vrijednost za k?