

6. Logistička regresija

Strojno učenje 1, UNIZG FER, ak. god. 2023./2024.

Jan Šnajder, vježbe, v2.7

1 Zadatci za učenje

1. [Svrha: Znati definirati model logističke regresije. Razumjeti izvod funkcije pogreške unakrsne entropije i pripadne funkcije gubitka. Shvatiti zašto je ta funkcija gubitka unakrsne entropije prikladna za klasifikaciju, dok funkcija kvadratnog gubitka to nije.]

- (a) Definirajte poopćeni linearni model. Koja je svrha aktivacijske funkcije?
- (b) Definirajte model logističke regresije. Zašto je sigmoidna (logistička) funkcija prikladan odabir za aktivacijsku funkciju?
- (c) Izvedite pogrešku unakrsne entropije $E(\mathbf{w}|\mathcal{D})$ kao negativan logaritam vjerojatnosti oznaka svih primjera iz skupa za učenje prema hipotezi s težinama \mathbf{w} .
- (d) Napišite funkciju gubitka unakrsne entropije i nacrtajte njezin graf. Koliki je najveći a koliki najmanji mogući gubitak?
- (e*) Pretpostavimo da su oznake $y \in \{-1, +1\}$ umjesto $y = \{0, 1\}$. Reformulirajte funkciju gubitka unakrsne entropije $L(y, h(\mathbf{x}))$ tako da koristi takve oznake te da vrijedi $L(y, 0) = 1$ (kako bi funkcija bila kompatibilna s ostalim funkcijama gubitka koje smo radili).
- (f) Nacrtajte graf funkcije gubitka $L(y, h(\mathbf{x}))$ u ovisnosti o udjelu pogrešne klasifikacije $y\mathbf{w}^T\phi(\mathbf{x})$, i to za: gubitak 0-1, kvadratni gubitak i logistički gubitak iz (e). Na temelju skice, odgovorite: (i) zašto je logistički gubitak dobar za klasifikaciju, a kvadratni gubitak to nije?; (ii) nanose li ispravno klasificirani primjeri ikakav gubitak?; (iii) možemo li reći da je logistički gubitak konveksni surogat gubitka 0-1, i što to znači?

2. [Svrha: Prisjetiti se definicije konveksnosti funkcije. Razumjeti da konveksnost i unimodalnost nisu jedno te isto.]

- (a*) Formalno definirajte kada je funkcija $f : \mathbb{R}^n \rightarrow \mathbb{R}$ konveksna.
- (b*) Funkcija f je kvazikonveksna (ili unimodalna) akko je njezina domena $\text{dom } f$ konveksna te ako za svaki $\mathbf{x}, \mathbf{y} \in \text{dom } f$ i $0 \leq \alpha \leq 1$ vrijedi

$$f(\alpha\mathbf{x} + (1 - \alpha)\mathbf{y}) \leq \max \{f(\mathbf{x}), f(\mathbf{y})\}.$$

Kvazikonveksnost je poopćenje konveksnosti: svaka je konveksna funkcija unimodalna, ali obrat ne vrijedi. Pokažite primjerom da obrat ne vrijedi.

- (c) Zašto u strojnom učenju preferiramo konveksne funkcije pogreške? Koja je veza između konveksnosti funkcije pogreške i konveksnosti funkcije gubitka?
3. [Svrha: Razumjeti gradijentni spust i potrebu za linijskim pretraživanjem. Znati izvesti gradijentni spust za logističku regresiju. Demonstrirati upoznatost s prednostima i nedostacima optimizacije drugog reda.]
- (a) Objasnite ideju gradijentnog spusta i potrebu za linijskim pretraživanjem.
 - (b) Objasnite razliku između grupnog (*batch*) i stohastičkog gradijentnog spusta. Koja je prednost ovog drugog?
 - (c) Izrazite gradijent funkcije pogreške unakrsne entropije $\nabla E(\mathbf{w}|\mathcal{D})$ i napišite pseudokôd algoritma gradijentnog spusta (grupna i stohastička inačica).

4. [Svrha: Razumjeti kako regularizacija i linearna (ne)odvojivost utječu na gradijenti spust i na izgled funkcije pogreške u prostoru parametara.] Koristimo model L2-regularizirane logističke regresije učene algoritmom gradijentnog spusta. Iskušavamo dvije vrijednosti regularizacijskog faktora: $\lambda = 0$ i $\lambda = 100$. Razmatramo posebno linearno odvojiv i linearno neodvojiv problem.
- Skicirajte pogreške učenja i ispitivanja $E(\mathbf{w}|\mathcal{D})$ u ovisnosti o broju iteracija za $\lambda = 0$ i $\lambda = 100$ te za slučaj (i) linearno odvojivih i (ii) linearno neodvojivih primjera (četiri grafikona sa po dvije krivulje).
 - Načinite skice izokontura funkcije neregularizirane pogreške $E(\mathbf{w}|\mathcal{D})$ i L2-regularizacijskog izraza u ravni w_1 - w_2 . Napravite dvije odvojene skice: za linearno odvojive i linearno neodvojive primjere.
 - Na grafikone iz prethodnoga zadatka do crtajte izokonture L2-regulariziranih funkcija pogreške za $\lambda = 100$ i naznačite gdje se nalazi točka minimuma (w_1^*, w_2^*) . Gdje bi se nalazila točka minimuma za $\lambda = 0$?

2 Zadaci s ispita

1. (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 \mathbf{x} kada bismo njegove značajke pomnožili sa dva i promijenili mu oznaku?**

☐ A 4.03 ☐ B 2.54 ☐ C 7.11 ☐ D 1.19

2. (N) Na skupu \mathcal{D} označenih primjera trenirali smo model binarne logističke regresije. Naknadno smo uočili da jedan primjer iz skupa \mathcal{D} modelu nanosi razmjerno velik gubitak. Konkretno, iznos gubitka za dotični primjer je $L(y, h(\mathbf{x})) = 1.20$. Ispostavilo se da je taj primjer pogrešno označen. **Koliko bi iznosio gubitak na istom ovom primjeru, ako bismo sada naknadno promijenili njegovu oznaku, ali model ostavili nepromijenjenim?**

☐ A 0.70 ☐ B 0.28 ☐ C 0.36 ☐ D 0.52

3. (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijanskog ulaznog prostora preslikali smo u prostor značajki funkcijom

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

U jednoj iteraciji treniranja modela vektor parametara jednak je

$$\mathbf{w} = (0.2, 0.5, -1.1, 2.7)$$

Koliko u toj iteraciji iznosi L_2 -norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((-0.5, 2), 1)$?

☐ A 0.70 ☐ B 2.48 ☐ C 1.28 ☐ D 4.00

4. (P) Na primjerima iz dvodimenzijanskoga 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}_3 ☐ D \mathbf{w}_4

5. (P) Na skupu označenih primjera treniramo tri modela: (1) model neregularizirane logističke regresije (NR), (2) model L2-regularizirane logističke regresije (L2R) i (3) perceptron. Sva tri modela koriste istu funkciju preslikavanja u prostor značajki. Za sva tri algoritma promatramo iznos empirijske pogreške učenja kroz iteracije optimizacijskog postupka. Nakon određenog broja iteracija, algoritam perceptrona uspješno se zaustavio s rješenjem. **Kako se u ovom slučaju ponaša empirijska pogreška učenja kroz iteracije za dva spomenuta modela logističke regresije, NR i L2R?**

- ☐ A Pogreška učenja modela NR nakon određenog broja iteracije doseže nulu, dok pogreška učenja modela L2R najprije pada pa raste
- ☐ B Pogreške učenja modela NR i modela L2R dosežu nulu, ali modelu L2R za to treba više iteracija
- ☐ C Pogreške učenja modela NR i modela L2R obje stagniraju nakon određenog broja iteracija, ali modelu NR za to treba više iteracija
- ☐ D Pogreška učenja modela NR asimptotski teži nuli, dok pogreška učenja modela L2R nakon određenog broja iteracija stagnira

6. (P) Razmotrimo sljedeći skup označenih primjera:

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

Nad ovim skupom treniramo dva modela: perceptron (P) i neregulariziranu logističku regresiju (LR). Pored toga, razmatramo tri funkcije preslikavanja:

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

Ukupno, dakle, isprobavamo šest kombinacija modela i funkcije preslikavanja. **Za koje će algoritme (model+preslikavanje) optimizacijski postupak pronaći minimizator empirijske pogreške?**

- ☐ A $P+\phi_0$ ☐ B $P+\phi_2$ ☐ C $LR+\phi_1$ ☐ D $LR+\phi_2$

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

- ☐ A -12 ☐ B -5 ☐ C -2 ☐ D $+22$