

7. Logistička regresija II

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

Jan Šnajder, vježbe, v1.11

1 Zadatci za učenje

- [*Svrha: Znati izvesti algoritam multinomijalne logističke regresije.*]
 - Definirajte funkciju *softmax*. Izračunajte $\text{softmax}(\boldsymbol{\alpha})$ za ulazni vektor $\boldsymbol{\alpha} = (2, 8, 1, 5)$. Koja su dva efekta funkcije softmax?
 - Definirajte model multinomijalne logističke regresije.
 - Izvedite pogrešku modela multinomijalne logističke regresije kao negativan logaritam vjerojatnosti oznaka koje model dodjeljuje primjerima iz skupa označenih primjera.
- [*Svrha: Znati izvesti algoritam LMS poopćenih linearnih modela. Razumjeti prednosti tog algoritma.*]
 - Izvedite pravilo za ažuriranje težina algoritma LMS (engl. *least mean squares*) kao gradijent funkcije gubitka, i to za (i) model linearne regresije i (ii) model logističke regresije.
 - Objasnite prednost algoritma LMS (odnosno stohastičkog gradijentnog spusta) nad grupnim (*batch*) gradijentnim spustom.
- [*Svrha: Uočiti zajedničkosti poopćenih linearnih modela.*]
 - Opišite veze između (i) modela linearne regresije, logističke regresije i multinomijalne logističke regresije, (ii) distribucija zavisne varijable y i (iii) aktivacijskih funkcija f . Što je zajedničko svim distribucijama s kojima smo dosada radili?
 - Objasnite riječima ovaj izraz:

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmax}} \ln P(\mathcal{D}|\mathbf{w}) = \underset{\mathbf{w}}{\operatorname{argmin}} E(\mathbf{w}|\mathcal{D})$$

- [*Svrha: Razumjeti motivaciju za adaptivne bazne funkcije i vezu između poopćenih linearnih modela i modela neuronske mreže.*]
 - Objasnite što su bazne funkcije i koji je problem s fiksnim baznim funkcijama.
 - Definirajte poopćeni linearni model s proizvoljnom aktivacijskom funkcijom f koji kao bazne funkcije koristi poopćene linearne modele s istom takvom aktivacijskom funkcijom. Načinite skicu takvog modela, odnosno dvoslojne neuronske mreže. Na skici naznačite komponente ulaznog vektora, težine modela, i bazne funkcije.
 - Koja je prednost ovakvog modela u odnosu na (i) poopćeni model bez baznih funkcija i (ii) poopćeni model s fiksnim baznim funkcijama? Koji je nedostatak takvog modela u odnosu na poopćeni model s fiksnim baznim funkcijama?

2 Zadatci s ispita

- (N) Raspoložemo 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):

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

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

- ☐ A 7 ☐ B 11 ☐ C 23 ☐ D 35

2. (P) Poopćeni linearni modeli mogu koristiti adaptivne bazne funkcije. Prednost toga je da ne moramo ručno definirati preslikavanje ϕ u prostor značajki, već se to preslikavanje može naučiti na temelju podataka. Rasplažemo podacima iz $K = 3$ klase u 10-dimenzijskome ulaznom prostoru. Za taj višeklasni problem koristimo multinomijalnu logističku regresiju, ali s adaptivnim baznim funkcijama. Svaka adaptivna bazna funkcija ϕ_j parametrizirana je kao skalarni produkt vektora značajki i vektora primjera, kao što smo radili na predavanjima. Naš model definirali smo ovako:

$$h_k(\mathbf{x}) = \text{softmax}_k \left(\sum_{j=0}^3 w_{j,k} \phi_j(\mathbf{x}) \right)$$

Ovime je definirana hipoteza za klasu k . Svaka klasa ima svoju hipotezu h_k . Svaka klasa ima i svoje težine $w_{j,k}$. Međutim, bazne funkcije ϕ_j zajedničke su za sve klase (dakle, ti parametri su dijeljeni između klasa). **Koliko ukupno parametara ima ovaj model?**

- ☐ A 45 ☐ B 49 ☐ C 136 ☐ D 142