

13M051MU, 2. domaći zadatak 2023/24

Generalizovani linearni modeli i Generativni algoritmi

Poslednja kolona u datoteci `multiclass_data.csv` je oznaka klase. Ostale kolone sadrže vrednosti prediktora. Problem treba da rešite pomoću 2 od 4 različita klasifikatora, kao što je opisano u nastavku.

1 Generalizovani linearni modeli

Koristite samo originalne prediktore (bez polinomijske ekspanzije prediktora). U ovom slučaju, modeli su dovoljno jednostavni, tako da nije potreba regularizacija. Pošto nema dodatnih hiper-parametara, u ovom domaćem zadatku nema ni potrebe za unakrsnom validacijom.

1.1 Logistička regresija (LR)

Pošto je u pitanju problem sa više od 2 klase, treba primeniti pristup tipa “jedan-protiv-ostalih”:

- projektuje se po jedan klasifikator za svaku klasu, koji treba da prepozna da li je primer iz te klase ili iz neke (bilo koje) od preostalih;
- novi primer se pušta kroz sva 3 klasifikatora, a konačna odluka se donosi poređenjem vrednosti hipoteza.

1.2 Softmax

U nastavku je izvođenje izraza za gradijent kriterijumske f-je. Prirodni parametar multinomijalne raspodele je

$$\boldsymbol{\eta} = [\ln \phi_1 / \phi_k \quad \cdots \quad \ln \phi_{k-1} / \phi_k]^\top$$

gde je $\phi_i = p(y = i | \mathbf{x}; \boldsymbol{\theta})$. Usvajamo $\eta_i = \boldsymbol{\theta}_i^\top \mathbf{x}$, pa je

$$\phi_i = \frac{e^{\boldsymbol{\theta}_i^\top \mathbf{x}}}{1 + \sum_{j=1}^{k-1} e^{\boldsymbol{\theta}_j^\top \mathbf{x}}}, \quad i = 1, \dots, k-1,$$

$$\phi_k = \frac{1}{1 + \sum_{j=1}^{k-1} e^{\boldsymbol{\theta}_j^\top \mathbf{x}}}.$$

Definišimo $\boldsymbol{\theta}_k = \mathbf{0}$ i matricu parametara

$$\boldsymbol{\Theta} = [\boldsymbol{\theta}_1 \quad \cdots \quad \boldsymbol{\theta}_k].$$

Sada je

$$\phi_i = \frac{e^{\boldsymbol{\theta}_i^\top \mathbf{x}}}{\sum_{j=1}^k e^{\boldsymbol{\theta}_j^\top \mathbf{x}}}, \quad i = 1, \dots, k.$$

Izraz za log-verodostojnost je

$$J(\boldsymbol{\Theta}) = l(\boldsymbol{\Theta}) = \sum_{i=1}^m \ln \phi_{y^{(i)}} \\ = \sum_{i=1}^m \boldsymbol{\theta}_{y^{(i)}}^\top \mathbf{x}^{(i)} - \ln \left(\sum_{j=1}^k e^{\boldsymbol{\theta}_j^\top \mathbf{x}^{(i)}} \right)$$

Gradijent po $\boldsymbol{\theta}_l$ je

$$\nabla_{\boldsymbol{\theta}_l} J(\boldsymbol{\Theta}) = \sum_{i=1}^m \left(\mathbb{1}_{y^{(i)}=l} - \frac{e^{\boldsymbol{\theta}_l^\top \mathbf{x}^{(i)}}}{\sum_{j=1}^k e^{\boldsymbol{\theta}_j^\top \mathbf{x}^{(i)}}} \right) \mathbf{x}^{(i)}$$

Jedna iteracija gradijentnog uspona data je sa

$$\boldsymbol{\theta}_l \leftarrow \boldsymbol{\theta}_l + \alpha \nabla_{\boldsymbol{\theta}_l} J(\boldsymbol{\Theta}), \quad l = 1, \dots, k-1$$

($\boldsymbol{\theta}_k$ se ne ažurira i uvek ostaje $\mathbf{0}$).

1.3 Numerička optimizacija

Za numeričku maksimizaciju verodostojnosti koristite stohastički gradijentni spust sa “mini-saržama”. Eksperimentišite sa stopom učenja α i veličinom mini-sarže m_{mb} , i probajte da odredite kombinaciju koja daje najbržu konvergenciju. Opciono, možete eksperimentisati sa naprednim tehnikama (inercija, AdaGrad, RMSProp, Adam). Ilustrujte konvergenciju sa pet grafika:

- jedan sa “optimalnom” kombinacijom vrednosti α^* i m_{mb}^* ,
- dva sa α^* i prevelikim, odnosno premalim m_{mb} ,
- dva sa m_{mb}^* i prevelikim, odnosno premalim α .

Na x -osama ovih grafika treba da bude prikazan broj obučavajućih primera kroz koje je algoritam “prošao”¹: u jednoj iteraciji “konzumirate” jednu mini-saržu, pa će

¹Jedna iteracija sa velikim m_{mb} može biti numerički skuplja od nekoliko iteracija sa manjim m_{mb} ; ako bi x -osa prikazivala redni broj iteracije, ne bismo imali fer poređenje.

x -koordinata prve tačke na grafiku biti m_{mb} itd. Na y -osama grafika treba da bude verodostojnost. Kod logističke regresije imate 3 modela; prikažite verodostojnosti za svaki od njih na istom grafiku (različitim bojama), ili samo za jedan od njih—odlučite sami šta je ilustrativnije.

2 Generativni algoritmi

Implementirajte Gaussovsu diskriminantnu analizu (GDA) i Gaussovski naivni Bayesov (GNB) klasifikator. Kod GDA dozvolite da svaka klasa ima sopstvenu kovarijacionu matricu.

3 Kombinacija modela

Ako vam je broj indeksa $GG/B_1B_2B_3B_4$ radite:

- LR za parno B_3 , odnosno softmax za neparno B_3 ,
- GDA za parno B_4 , odnosno GNB za neparno B_4 .

Napomene

Nije dozvoljeno korišćenje gotovih f-ja za logističku regresiju, softmax, gradijentni spust, GDA i GNB. Kôd (isključivo Python ili Matlab/Octave) i kratak izveštaj sa traženim graphicima i tačnostima u pdf formatu, predati putem Teams-a do datuma naznačenog u Assignments. Kôd i izveštaj možete objediniti u Jupyter Notebook; u tom slučaju pošaljite ipynb i njegovu html verziju sa generisanim rezultatima. Na kraju izveštaja navedite ostvarene tačnosti na trening skupu za oba korišćena modela.