5. Linearni diskriminativni modeli

Strojno učenje 1, UNIZG FER, ak. god. 2021./2022.

Jan Šnajder, vježbe, v2.3

1 Zadatci za učenje

1. [***Svrha: Razumjeti geometriju linearnog modela.***]

   (a) Dokažite da je $\mathbf{w}$ normala (hiper)ravnine.

   (b) Izvedite izraz za predznačenu udaljenost primjera $\mathbf{x}$ od (hiper)ravnine.

2. [***Svrha: Isprobati na konkretnom kako se regresija može upotrijebiti za klasifikaciju. Razumjeti kako ostvariti višeklasnu klasifikaciju pomoću više binarnih modela. Razumjeti zašto je korištenje linearne regresije za klasifikaciju loša ideja.***] Na predavanjimo smo pokazali kako se linearan model regresije može (pokušati) koristiti za klasifikaciju. Pokažite to na sljedećim primjerima iz triju ($K = 3$) klasa:

$$
\begin{aligned}
\mathcal{D} &= \left\{(\mathbf{x}^{(i)}, y^{(i)})\right\}_{i=1}^{6} \\
&= \left\{\big((-3,1),0\big), \big((-3,3),0\big), \big((1,2),1\big), \big((2,1),1\big), \big((1,-2),2\big), \big((2,-3),2\big)\right\}.
\end{aligned}
$$

   (a) Primijenite pristup *jedan-naspram-ostali* (OVR), definirajte matricu dizajna i vektor oznaka $\mathbf{y}$ za svaki od triju modela te izračunajte hipoteze $h_j(\mathbf{x})$ za svaku od triju klasa. Izračun možete napraviti ručno ili u nekom alatu.

   (b) Izračunajte diskriminacijske funkcije $h_{01}(\mathbf{x})$, $h_{12}(\mathbf{x})$ i $h_{02}(\mathbf{x})$ između parova susjednih klasa. Skicirajte primjere i dobivene granice u prostoru $\mathbb{R}^2$.

   (c) U koju bi klasu bio klasificiran primjer $\mathbf{x} = (-1, 3)$? Obrazložite odgovor.

   (d) Možete li reći koja je vjerojatnost da primjer pripada toj klasi? Obrazložite odgovor.

   (e) Objasnite koja je prednost pristupa OVR nad pristupom *jedan-naspram-jedan* (OVO), a što je nedostatak.

   (f) U praksi linearnu regresiju ne bismo željeli koristiti za klasifikaciju. Zašto? Pokažite na gornjem primjeru u čemu je problem (možete modificirati primjer).

3. [***Svrha: Razumjeti kriterij perceptrona i ograničenja koja proizlaze iz toga što ta funkcija nije derivabilna.***]

   Algoritam perceptrona minimizira pogrešku $E_p(\mathbf{w}|\mathcal{D})$ koju nazivamo *kriterij perceptrona*. Ta je funkcija aproksimacija udjela pogrešnih klasifikacija (engl. *misclassification ratio*), odnosno očekivanja gubitka 0-1, $E_m(\mathbf{w}|D)$, koju bismo idealno htjeli minimizirati, ali to ne možemo. Pogledajte (u natuknicama s predavanja) kako izgleda pogreška perceptrona u prostoru parametara.

   (a) Objasnite zašto ne možemo izravno minimizirati $E_m(\mathbf{w}|\mathcal{D})$.

   (b) Je li pogreška perceptrona $E_p(\mathbf{w}|\mathcal{D})$ gornja ograda za pogrešku $E_m(\mathbf{w}|\mathcal{D})$? Objasnite.

   (c) Jedan nedostatak perceptrona jest da rješenje $\mathbf{w}^*$ (a time i položaj granice) ovisi o početnim težinama i redoslijedu predočavanja primjera. Pozivajući se na sliku površine pogreške u prostoru parametara, objasnite zbog čega je to tako.

   (d) Drugi nedostatak perceptrona jest da postupak ne konvergira ako primjeri nisu linearno odvojivi. Pozivajući se opet na sliku površine pogreške u prostoru parametara, objasnite zašto je to tako.

4. [***Svrha: Razumjeti odnose između funkcija gubitaka različitih modela. Razumjeti kako funkcija gubitka određuje dobra i loša svojstva modela.***]

   (a) Skicirajte na jednome grafikonu sljedeće tri funkcije gubitka: (1) kvadratni gubitak regresije, (2) gubitak perceptrona i (3) gubitak 0-1.

   (b) Odgovorite čemu odgovara desna strana grafikona (x-os veća od nule), a čemu lijeva (x-os manja od nule).

   (c) Pozivajući se na skicu, odgovorite zašto kvadratni gubitak nije prikladan gubitak u slučajevima kada želimo minimizirati broj pogrešnih klasifikacija.

   (d) Pozivajući se na skicu, odgovorite za koje će modele očekivanje gubitka (empirijska pogreška) biti veće od udjela pogrešnih klasifikacija.

2 Zadatci s ispita

1. (P) Treniramo linearni diskriminativni model u dvodimenzijskome prostoru primjera. Skup za učenje čine samo dva primjera, $(\mathbf{x}_1, y_1) = ((1,0), +1)$ i $(\mathbf{x}_2, y_2) = ((0,1), -1)$. Na tom skupu treniramo model koji ima induktivnu pristranost takvu da rješenje maksimizira minimalnu udaljenost primjera od hiperravnine. Naučen model ispravno klasificira oba primjera, pri čemu za oba primjera vrijedi $y \cdot h(\mathbf{x}) = 5$. **Koliko iznosi težina w_2 tako naučenog modela?**

   A -1 B 5 C -5 D 1

2. (P) Razvijamo sustav za automatsku klasifikaciju novinskih članaka u jednu od pet kategorija. Tih pet kategorija su "sport", "politika", "kriminal", "znanost" i "lifestyle". Najveća razlika u veličini klasa je između kategorija "politika" i "znanost". Očekivano, u kategoriji "politika" ima najviše članaka, dok ih u kategoriji "znanost" ima $5\times$ manje, što je u redu jer to ionako nitko ne čita. Svaki novinski članak prikazujemo kao vektor riječi, gdje su komponente vektora broj pojavljivanja pojedine riječi. Problem rješavamo algoritmom perceptrona. Budući da je perceptron binaran klasifikator, odlučili smo primijeniti shemu OVR ili shemu OVO za dekompoziciju višeklasnog klasifikacijskog problema u skup binarnih klasifikacijskih problema. **Što možemo očekivati?**

   A OVO će imati $2\times$ puta manje značajki od OVR, ali bi mogao raditi bolje na člancima iz kategorije "znanost"

   B OVR će imati $2\times$ manje značajki od OVO, ali bi mogao raditi lošije na člancima iz kategorije "znanost"

   C OVO će imati $5\times$ puta manje značajki od OVR, ali bi mogao raditi lošije na člancima iz kategorije "znanost"

   D OVR će imati $5\times$ manje značajki od OVO, ali bi mogao raditi bolje na člancima iz kategorije "znanost"

3. (T) Na predavanjima smo za klasifikaciju pokušali upotrijebiti algoritam regresije. Zaključili smo da to ne funkcionira, tj. da algoritam linearne regesije jednostavno nije klasifikacijski algoritam. **Koje bismo minimalne preinake trebale učiniti u algoritmu linearne regresije, a da on dobro funkcionira kao klasifikacijski algoritam?**

   A Promijeniti model, funkciju gubitka i optimizacijski postupak

   B Promijeniti funkciju gubitka i optimizacijski postupak

   C Promijeniti funkciju gubitka

   D Promijeniti model i funkciju gubitka

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

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

   Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijski prostor značajki, koja je definirana 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 sljedeće:

$$\mathbf{w} = (1, 0, -1, 2, -2, 0)$$

Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?

A 8 B 9 C 16 D 25

5. (P) Razmotrimo sljedeći skup označenih primjera:

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

Ovaj skup nije linearno odvojiv i algoritam perceptrona neće konvergirati. Linearna neodvojivost podataka je konceptualni razlog zašto algoritam ne konvergira. **Koji je tehnički razlog zašto algoritam perceptrona na ovom skupu primjera neće konvergirati?**

A U svakoj točki prostora parametara postoji barem jedan primjer za koji je gradijent gubitka veći od nule

B Premda je empirijska pogreška na ovom skupu primjera derivabilna, ona je uglavnom konstantna

C U prostoru parametara ne postoji točka u kojoj je gradijent empirijske pogreške jednak nuli

D U prostoru parametara postoji više točaka za koje je empirijska pogreška jednaka nuli