4. Linearna regresija II

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

Jan Šnajder, vježbe, v1.11

1 Zadatci za učenje

1. [***Svrha: Shvatiti kako se nelinarna funkcija u ulaznom prostoru funkcija preslikava u linearnu funkciju odnosno (hiper)ravninu u prostoru značajki.***]

   (a) Regresijom želimo aproksimirati funkciju jedne varijable $y = 3 \cdot (x-2)^2 + 1$. Skicirajte graf te funkcije. Definirajte linearan model $h(x)$ uz funkciju preslikavanja u prostor značajki $\boldsymbol{\phi}(x) = (1, x, x^2)$. Odredite vektor težina $\mathbf{w} = (w_0, w_1, w_2)$ tog modela.

   (b) Skicirajte u prostoru sa dimenzijama x_1 i x_2 (dakle u prostoru značajki) izokonture funkcije y. Naznačite u tom prostoru točke u koje se preslikavaju primjeri $x^{(1)} = 1$, $x^{(2)} = 2$ i $x^{(3)} = 3$. Koja je vrijednost od $h(x)$ za navedene primjere?

2. [***Svrha: Razumjeti matrično rješenje za L2-regulariziranu regresiju. Razumjeti kako regularizacija popravlja lošu kondiciju matrice.***]

   (a) Izvedite u matričnom obliku rješenje za vektor $\mathbf{w}$ za hrbatnu (L2-regulariziranu) regresiju.

   (b) Raspolažemo sljedećim skupom primjera za učenje:

   $$\mathcal{D} = \big\{(x^{(i)}, y^{(i)})\big\}_{i=1}^4 = \big\{(0,4),(1,1),(2,2),(4,5)\big\}.$$

   Podatke želimo modelirati polinomijalnom regresijskom funkcijom $h(x) = w_0 + w_1 x + w_2 x^2$. Napišite kako bi u ovome konkretnom slučaju izgleda jednadžba iz zadatka (a), ako se koristi regularizacijski faktor $\lambda = 10$. (Ne morate ju izračunavati, samo ju napišite da se vide konkretni brojevi.)

   (c) Komentirajte na koji način L2-regularizacija rješava problem numeričke nestabilnosti rješenja za $\mathbf{w}$.

   (d) Koristimo regresiju za predviđanje cijene nekretnine na temelju površine, starosti i udaljenosti od glavne prometnice. Koliko primjera nam je minimalno potrebno a da bi rješenje bilo izračunljivo jednadžbom iz (a), ako pritom ne koristimo preslikavanje. Koliko primjera nam je potrebno ako koristimo preslikavanja s polinomom drugog stupnja i interakcijskim značajkama? Što bi se dogodilo da kao značajku dodamo godinu izgradnje nekretnine? Obrazložite.

3. [***Svrha: Isprobati izračun regresijskog modela s različitim funkcijama preslikavanja u prostor značajki te razviti intuiciju kako o tome kako ta funkcija određuje složenost hipoteze u ulaznome prostoru.***] Linearnim modelom univarijatne regresije želimo aproksimirati jednu periodu funkcije $f(x) = \sin(\pi x)$. Raspolažemo sljedećim skupom primjera za učenje:

   $$\mathcal{D} = \big\{(0.25, 0.707), (0.5, 1), (1, 0), (1.5, -1), (2, 0)\big\}.$$

   (a) Izračunajte parametre linearnog modela regresije u ulaznom prostoru primjera, tj. s funkcijom preslikavanja u prostor značajki definiranom kao $\boldsymbol{\phi}(x) = (1, x)$. Skicirajte dobivenu regresijsku funkciju.

   (b) Izračunajte parametre modela polinomijalne regresije drugog stupnja, tj. modela koji koristi funkciju preslikavanja u prostor značajki definiranu kao $\boldsymbol{\phi}(x) = (1, x, x^2)$. Skicirajte dobivenu regresijsku funkciju.

(c) Izračunajte parametre modela polinomijalne regresije četvrtog stupnja, tj. modela koji koristi funkciju preslikavanja u prostor značajki definiranu kao $\boldsymbol{\phi}(x) = (1, x, x^2, x^3, x^4)$, uz L2-regularizaciju ($\lambda = 1$). Skicirajte dobivenu regresijsku funkciju.

(d) Koji je model u ovom slučaju najprikladniji? Zašto?

Napomena: Izračun možete načiniti u nekom alatu koji podržava izračun matričnih operacija. Skicu također možete načiniti u nekom alatu, ili je možete napraviti ručno, izračunom vrijednosti regresijske funkcije u nekoliko odabranih točaka.

4. [***Svrha: Razumjeti vezu između faktora regularizacije i složenosti modela.***] Neka $\mathcal{H}_{d,\lambda}$ označava model polinomijalne regresije stupnja d s L2-regularizacijskim faktorom λ. Razmatramo četiri modela: $\mathcal{H}_{2,0}$, $\mathcal{H}_{5,0}$, $\mathcal{H}_{5,100}$, $\mathcal{H}_{5,1000}$ u ulaznome prostoru $\mathcal{X} = \mathbb{R}$. Pretpostavimo da su podatci u stvarnosti generirani funkcijom koja je polinom trećeg stupnja ($d = 3$). Pretpostavite da imamo razmjerno malo podataka i da je šum u podatcima razmjerno velik. Na dva odvojena crteža skicirajte

   (a) regresijsku funkciju $h(x)$ za sva četiri modela te

   (b) pogrešku učenja i ispitnu pogrešku za sva četiri modela.

5. [***Svrha: Shvatiti kako regularizacija utječe na optimizaciju. Shvatiti geometrijski argument zašto L1-regularizacija rezultira rijetkim modelima, a L2-regularizacije ne.***]

   (a) Objasnite koja je svrha regularizacije i na kojoj se pretpostavci temelji.

   (b) Koja je prednost regulariziranog modela u odnosu na neregularizirani? Dolazi li ta prednost više do izražaja u slučajevima kada imamo puno primjera za učenje ili kada ih imamo malo?

   (c) Razmatramo višestruku regresiju, $h(\mathbf{x}; \mathbf{w}) = w_0 + w_1 x_1 + w_2 x_2$. Skicirajte izokonture neregularizirane funkcije pogreške u ravnini $\mathbb{R}^2$ koju definiraju parametri w_1 i w_2 (napomena: funkcija pogreške je konveksna). Zatim skicirajte izokonture regularizacijskog izraza definiranog L2-normom vektora težina (i ova je funkcija konveksna). Pomoću ove skice objasnite na koji način regularizacija utječe na izbor optimalnih parametara (w_1^*, w_2^*). Skicirajte krivulju mogućih rješenja za $\lambda \in [0, \infty)$.

   (d) Ponovite prethodnu skicu, ali ovog puta sa L1-regularizacijom. Na temelju ove skice pokušajte odgovoriti na pitanje zašto L1-regularizacija daje rjeđe modele od L2-regularizacije.

6. [***Svrha: Shvatiti vezu između težine značajki, važnosti značajki i složenosti modela.***] Treniramo model regresije uz nelinearno preslikavanje u prostor značajki $\boldsymbol{\phi} : \mathbb{R}^n \to \mathbb{R}^{m+1}$, gdje $m > n$, uz L2-regularizaciju.

   (a) Kako biste odredili optimalan regularizacijski faktor λ?

   (b) Kako, nakon treniranja modela, možemo provjeriti (1) koje su značajke nebitne i (2) je li izvorni (neregularizirani) model presložen?

   (c) Kako bi se u ovom slučaju ponašao L1-regularizirani model?

   (d) Pretpostavite da u podatcima postoji skup multikolinearnih značajki koje su, osim što su redundantne, također i irelevantne, odnosno zavisna varijabla u stvarnosti uopće ne ovisi o tim varijablama. Ako model nije regulariziran, koje su očekivane težine tih značajki?

2 Zadatci s ispita

1. (N) Raspolažemo sljedećim skupom primjera u dvodimenzijskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{((1,0),1),((2,-3),2),((3,5),-1),((5,0),-4)\}$$

Na ovom skupu gradijentnim spustom trenirali smo L_1-regularizirani model linearne regresije sa $\lambda = 1$. Dobili smo težine $\mathbf{w} = (2.12, -0.94, -0.08)$. **Koliko iznosi L_1-regularizirana pogreška $E(\mathbf{w}|\mathcal{D})$?**

A 7.10   B 2.69   C 1.58   D 0.29

2. (P) Raspolažemo skupom označenih primjera $\mathcal{D} \subset \mathbb{R}^n \times \mathbb{R}$ koji su u stvarnosti generirani funkcijom koja je polinom trećeg stupnja. Podataka imamo razmjerno malo, a šum u podatcima je velik. Skup $\mathcal{D}$ dijelimo na skup za učenje i skup za ispitivanje. Neka je $\mathcal{H}_{d,\lambda}$ familija modela polinomijalne regresije stupnja d s L2-regularizacijskim faktorom λ. Na skupu za učenje postupkom najmanjih kvadrata treniramo četiri modela iz te familije: $\mathcal{H}_{2,0}$, $\mathcal{H}_{5,0}$, $\mathcal{H}_{5,100}$ i $\mathcal{H}_{5,1000}$. Zatim izračunavamo empirijsku pogrešku (očekivanje kvadratnog gubitka) ovih modela na skupu za ispitivanje. **Što možemo zaključiti o ponašanju hipoteza iz ovih modela naučenih na skupu primjera $\mathcal{D}$?**

   A Najbolje će generalizirati hipoteza iz $\mathcal{H}_{5,100}$ ili hipoteza iz $\mathcal{H}_{5,1000}$, ovisno o količini šuma u podatcima

   B Hipoteza iz $\mathcal{H}_{2,0}$ imati će veću pogrešku na skupu za učenje od hipoteze $\mathcal{H}_{5,0}$, ali mogu podjednako loše generalizirati

   C Hipoteza iz $\mathcal{H}_{5,1000}$ će generalizirati bolje od hipoteze iz $\mathcal{H}_{5,0}$, ali će imati veću pogrešku na skupu za učenje

   D Hipoteza iz $\mathcal{H}_{5,100}$ će bolje generalizirati od hipoteze iz $\mathcal{H}_{2,0}$ i imat će manju pogrešku na skupu za učenje

3. (P) Koristimo regresiju za predviđanje uspjeha na studiju. Kao značajke možemo koristiti ocjene u četiri razreda srednje škole (značajke x_1–x_4), prosjek ocjena sva četiri razreda (x_5) te uspjeh iz matematike (x_6) i fizike (x_7) na državnoj maturi (ukupno 7 značajki). Ne moramo iskoristiti sve značajke, ali ih želimo iskoristiti što više. Za preslikavanje u prostor značajki koristimo preslikavanje s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki (npr. x_1x_2) i interakcije trojki (npr. $x_1x_2x_3$) između svih značajki koje koristimo. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

   A 75 B 38 C 48 D 63