3. Linearna regresija

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

Jan Šnajder, vježbe, v2.1

1 Zadatci za učenje

1. [***Svrha: Razumjeti osnovne komponente algoritma regresije te motivaciju za kvadratni gubitak i za postupak najmanjih kvadrata.***]
   (a) Definirajte tri komponente algoritma linearnog modela regresije.
   (b) Objasnite zašto koristimo kvadratnu funkciju gubitka, a ne gubitak 0-1.
   (c) Objasnite zašto težine ne možemo izračunati kao rješenje sustava jednadžbi $\mathbf{X}\mathbf{w} = \mathbf{y}$.

2. [***Svrha: Razumjeti matrično rješenje za regulariziranu regresiju i izvježbati potrebnu matematku. Razumjeti kako je rang matrice povezan sa postojanjem i stabilnošću rješenja. Razumijeti algoritamsku složenost postupka.***]
   (a) Izvedite u matričnom obliku rješenje za vektor $\mathbf{w}$ za linearan model regresije uz kvadratnu funkciju gubitka.
   (b) Što minimizira rješenje $\mathbf{w}$ izvedeno pseudoinverzom? Što ako takvih rješenja ima više?
   (c) Raspolažemo sljedećim skupom primjera za učenje:

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

   Podatke želimo modelirati modelom jednostavne regresije: $h(x) = w_0 + w_1 x$. Napišite kako bi u ovome konkretnom slučaju izgleda jednadžba iz zadatka (a) (Ne morate ju izračunavati, samo ju napišite da se vide konkretni brojevi.)
   (d) Jednadžba iz zadatka (a) daje rješenje u zatvorenoj formi, međutim rješenje nije uvijek izračunljivo na taj način. Što predstavlja problem? Pod kojim uvjetom je rješenje izračunljivo pomoću jednadžbe iz (a)? Možemo li rješenje izračunati i kada taj uvjet nije ispunjen? Kako?
   (e) U situacijama kada je rješenje izračunljivo jednadžbom iz zadatka (a), izračun ponekad može biti računalno zahtjevan. Što predstavlja problem? Je li problem izražen kada imamo mnogo primjera za učenje ili kada imamo mnogo značajki? Obrazložite odgovor.

3. [***Svrha: Uvjeriti se da, uz određene pretpostavke, funkcija kvadratne pogreške ima probabilističko tumačenje i opravdanje.***] Kod postupka najmanjih kvadrata empirijska je pogreška definirana kao:

$$E(\mathbf{w}|\mathcal{D}) = \frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(i)} - y^{(i)}\right)^2.$$

   Pokažite da je minimizacija gornjeg izraza istovjetna maksimizaciji log-vjerojatnosti $\ln P(\mathbf{y}|\mathbf{X},\mathbf{w})$ (odnosno minimizaciji negativne log-vjerojatnosti) uz pretpostavku normalno distribuiranog šuma $\mathcal{N}(h(\mathbf{x};\mathbf{w}),\sigma^2)$.

2 Zadatci s ispita

1. (P) Funkcija kvadratne pogreške definirana je kao:

$$E(\mathbf{w}|\mathcal{D}) = \frac{1}{2}\sum_{i=1}^{N}(\mathbf{w}^T\mathbf{x}^{(i)} - y^{(i)})^2$$

Izvedite matrični zapis ove funkcije. **Kako glasi matrični zapis ove funkcije, nakon sređivanja izraza, a prije deriviranja?**

- A $\frac{1}{2}(\mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w} - 2\mathbf{y}^T\mathbf{X} + \mathbf{w}^T\mathbf{w})$
- B $\frac{1}{2}(\mathbf{w}\mathbf{X}^T\mathbf{X}\mathbf{w}^T - 2\mathbf{y}^T\mathbf{w} + \mathbf{y}^T\mathbf{y})$
- C $\frac{1}{2}(\mathbf{w}^T\mathbf{X}^T - 2\mathbf{w}^T\mathbf{X}\mathbf{y} + \mathbf{y}^T\mathbf{y})$
- D $\frac{1}{2}(\mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w} - 2\mathbf{y}^T\mathbf{X}\mathbf{w} + \mathbf{y}^T\mathbf{y})$

2. (P) Razmatramo model jednostavne regresije:

$$h(x; w_0, w_1) = w_0 + w_1 x$$

Model linearne regresije inače koristi funkciju kvadratnog gubitka:

$$L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$$

Međutim, u našoj implementaciji greškom smo funkciju gubitka definirali ovako:

$$L(y, h(\mathbf{x})) = (y + h(\mathbf{x}))^2$$

S tako pogrešno definiranom funkcijom gubitka, postupkom najmanjih kvadrata treniramo naš model na skupu primjera čije su oznake uzorkovane iz distribucije $\mathcal{N}(-1 + 2x, \sigma^2)$, gdje je varijanca σ^2 razmjerno malena (tj. nema mnogo šuma). **Koji vektor težina** (w_0, w_1) **očekujemo (približno) dobiti kao rezultat najmanjih kvadrata?**

A $(1, -2)$ B $(2, -1)$ C $(-1, 2)$ D $(0, 0)$

3. (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable y o zavisnoj varijabli x. Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_i = \{(0,0), (2,0), (3,2), (5,2)\}$. Neka je h hipoteza koju dobivamo treniranjem modela te neka je L^i gubitak hipoteze h na primjeru $x^{(i)}$, tj. $L^i = L(y^{(i)}, h(x^{(i)}))$. **Što vrijedi za gubitke hipoteze na pojedinim primjerima?**

- A $L^1 = L^2 = 1 < L^3 < L^4$
- B $L^1 = L^4 = 1,\ L^2 < L^3$
- C $L^1 = L^3 = 0,\ L^2 = L^4 < 1$
- D $L^1 = L^4 < L^2 = L^3$

4. (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\mathbf{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$. Treniranjem modela na skupu $(\mathbf{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (0.28, -0.58, 1.79, -0.75)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x}; \mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu, \sigma^2) = (\sqrt{2\pi}\sigma)^{-1}\exp(-\frac{1}{2}\sigma^{-2}(x - \mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\mathbf{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln \mathcal{L}(\mathbf{w}|\mathbf{\Phi}, \mathbf{y})$?

A -12.63 B -5.69 C -4.73 D -10.64