3. Linearna regresija

Strojno učenje 1, UNIZG FER, ak. god. 2022./2023.

Jan Šnajder, vježbe, v1.12

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. (T) Rješenje najmanjih kvadrata za vektor težina $\mathbf{w}$ jest:

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Pod kojim uvjetima ćemo težine moći izračunati na ovaj način, i o čemu dominantno ovisi složenost tog postupka?

A Ako je matrica $\mathbf{X}^T\mathbf{X}$ kvadratna i punog ranga, a složenost izračuna dominantno ovisi o N

B Ako je rang matrice $\mathbf{X}$ jednak $N+1$, a složenost izračuna dominantno ovisi o N

C Ako je rang matrice $\mathbf{X}^T\mathbf{X}$ jednak N, a složenost izračuna dominantno ovisi o n

D Ako je rang matrice $\mathbf{X}$ jednak $n+1$, a složenost izračuna dominantno ovisi o n

3. (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 $(0, 0)$

4. (T) Model linearne regresije je poopćeni linearni model i ima probabilističku interpretaciju. Prisjetite se, tu smo interpretaciju upotrijebili smo kako bismo opravdali empirijsku funkciju pogreške definiranu na temelju kvadratnog gubitka. **Kako formalno glasi probabilistička pretpostavka modela linearne regresije?**

A $p(\mathbf{x}|y) = \mathcal{N}(f(\mathbf{x}), \sigma^2)$

B $p(y|\mathbf{x}) = \mathcal{N}(0, \sigma^2)$

C $p(y|\mathbf{x}) = \mathcal{N}(h(\mathbf{x}), \sigma^2)$

D $p(y) = \mathcal{N}(0, \sigma^2)$

5. (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Induktivna pristranost sastoji se od pristranosti jezika i pristranosti preferencije. **Kako glasi induktivna pristranost preferencije (neregulariziranog) linearnog modela regresije?**

   A Težine $\mathbf{w}$ maksimiziraju iznos $\|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2$

   B Hipoteza h je linearna kombinacija težina $\mathbf{w}$ i značajki $\mathbf{x}$

   C Težine $\mathbf{w}$ minimiziraju iznos $\|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2$

   D Hipoteza h je funkcija koja preslikava iz $\mathbb{R}^n$ u $\mathbb{R}$