



Zadaci

Zadatak 1 (5).

Pokažite kako je funkcija $J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$ konveksna.

(Pomoć: veza sa Hessijanom.)

Zadatak 2 (10).

Izračunajte gradijent funkcije $J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$ tj. raspišite kako biste dobili izraz dan na predavanju.

Zadatak 3 (15).

Neka su zadani ulazni podaci $X \in \mathbb{R}^{m \times n}$ i pripadne izlazne vrijednosti $y \in \mathbb{R}$. Na ovim podacima možemo provesti proces učenja modela linearne regresije.

Znamo kako se učenje svodi na minimizaciju funkcije

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

te kako do rješenje tog minimizacijskog problema možemo doći gradijentnom metodom.

Implementirajte gradijentnu metodu za učenje modela linearne regresije u općenitom slučaju kada su ulazni podaci dimenzije $X \in \mathbb{R}^{m \times n}$.

Dodajte sljedeću mogućnost svojoj implementaciji gradijentne metode: crtanje grafa ovisnosti funkcije troška $J(\theta)$ o iteracijama. (U svakoj iteraciji spremite vrijednost funkcije troška) .

U datoteci *house_train.csv* i *house_test.csv* su vam dani podaci o nekretninama, koji sadrže različite karakteristike nekretnina poput površine ili broja soba te cijenu.

Cilj narednih zadataka je uspostaviti linearnu regresiju između različitih karakteristika nekretnine u svrhu predviđanja cijene nekretnine.

Zadatak 4. (15)

Kao ulaznu varijablu uzmite *sqft_living*, a kao izlaznu varijablu *price*.

- Učitajte podatke koji su vam dani u datoteci *house_train.csv*. Skalirajte podatke koristeći metodu objašnjenu na ovom link
- Koristeći proizvoljnu biblioteku vizualizirajte odnos varijable ulaznih i izlaznih podataka.



- c) Koristeći *numpy array* napravite matricu dizajna $X \in \mathbb{R}^{m \times (n+1)}$ i vektor izlaznih podataka $y \in \mathbb{R}^m$.
- d) Pokrenite gradijentnu metodu koju ste implementirali u prethodnom zadatku i pomoću nje odredite parametre θ_0 i θ_1 . (Stopu učenja i broj iteracija odredite sami)
- e) Pravac određen izračunatim parametrima $\theta_0 + \theta_1 x$ prikažite na grafu iz podzadatka b).
- f) Nacrtajte graf promjene funkcije troška $J(\theta)$ kroz iteracije.
- g) Učitajte podatke iz datoteke *house_test.csv*, spremite ih u odgovarajuće *numpy arrays* X_{test} i y_{test} . Pazite, X_{test} je ponovno matrica dizajna kao i X . Kao u podzadatku b), podatke prikažite grafički.
- h) Na učitanim podacima za testiranje provedite testiranje vašeg modela tako da izračunate pogrešku kao što smo pokazali na vježbama. Ispišite vrijednost pogreške.
- i) Za podatke X i y učitane iz datoteke *house_train.csv* odredite opet parametre θ_0, θ_1 , ali koristeći gotovu implementaciju iz *sklearn* (*sklearn.LinearRegression*) biblioteke. (Pazite na dimenzije)
- j) Na učitanim podacima za testiranje provedite testiranje ovog modela kao što smo pokazali na vježbama. Ispišite vrijednost pogreške.
- k) Razlikuju li se model kojeg ste dobili koristeći vašu implementaciju gradijentne metode i model kojeg ste dobili koristeći gotovu implementaciju iz *scikit-learn* paketa?
- l) Za podatke X i y učitane iz datoteke *house_train.csv* odredite opet parametre θ_0, θ_1 , ali koristeći sustav normalnih jednadžbi.

Zadatak 5. (15)

U ovom zadatku ulazni podaci će biti višestruki. Izlazna varijabla je i dalje *price*.

- a) Odaberite nekoliko karakteristika koje će vam biti ulazni podaci. Reprezentirajte ih kao *numpy arrays*. (Npr. 3 karakteristike)
- b) Izračunajte parametre θ_0, θ_1 koristeći implementaciju gradijentne metode iz trećeg zadatka.
- c) Učitajte podatke iz datoteke *house_test.csv*, spremite ih u odgovarajuće *numpy arrays* X_{test} i y_{test} . Pazite, X_{test} je ponovno matrica dizajna kao i X .
- d) Na učitanim podacima za testiranje provedite testiranje vašeg modela tako da izračunate pogrešku kao što smo pokazali na vježbama. Ispišite vrijednost pogreške.