

# Mašinsko učenje – Domaći 1

## LASSO regresija

Nemanja Saveski 2023/3163

28. oktobar 2023.

## 1 Zadatak

Potrebno je napraviti model LASSO regresije za koji se usvaja polinom 2. reda i za taj model naći optimalnu vrednost hiperparametra  $\lambda$  kao i procenjenu vrednost korena srednje-kvadratne greške modela (**RMSE**).

## 2 Rešenje

Hipoteza (ciljna promenljiva), za polinom 2. reda, a koji ima 5 prediktora, ima 21 član:

$$\hat{y} = h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \dots + \theta_i x_1^2 + \dots + \theta_j x_1 x_2 + \dots + \theta_{20} x_4 x_5$$

Kada smo definisali polinom, moramo izvršiti standardizaciju promenljivih:

$$x_i = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$$

Dalje je potrebno definisati kriterijumsku funkciju LASSO regresije:

$$J(\theta) = \sum_{i=1}^m (y^{(i)} - \hat{y}^{(i)})^2 + \lambda R(\theta), \quad R = \sum_{i=1}^n |\theta_i|$$

U matričnom obliku izgleda ovako:

$$J(\theta) = \|X\theta - y\|_2^2 + \lambda \|\theta\|_1$$

Kada uradimo izvod kriterijuma po  $\theta$ , dobijamo sledeći oblik:

$$\frac{dJ(\theta)}{d\theta} = \frac{1}{2} X^T e + \lambda \text{sgn}(\theta), \quad e = X\theta - y$$

Ovde je u redu da se uzme  $\text{sgn}$  funkcija kao rezultat izvoda apsolutne vrednosti, iako funkcija apsolutne vrednosti nije diferencijabilna u 0. Razlog ovoga je to što su male šanse da neko  $\theta$  dobije u nekom trenutku baš vrednost 0 (može se dosta približiti 0, ali ne biti 0). Sama LASSO regresija nam smanjuje broj relevantnih prediktora sa njihovim  $\theta$  parametrima, ali ti parametri u praksi nikada ne postaju tačno 0.

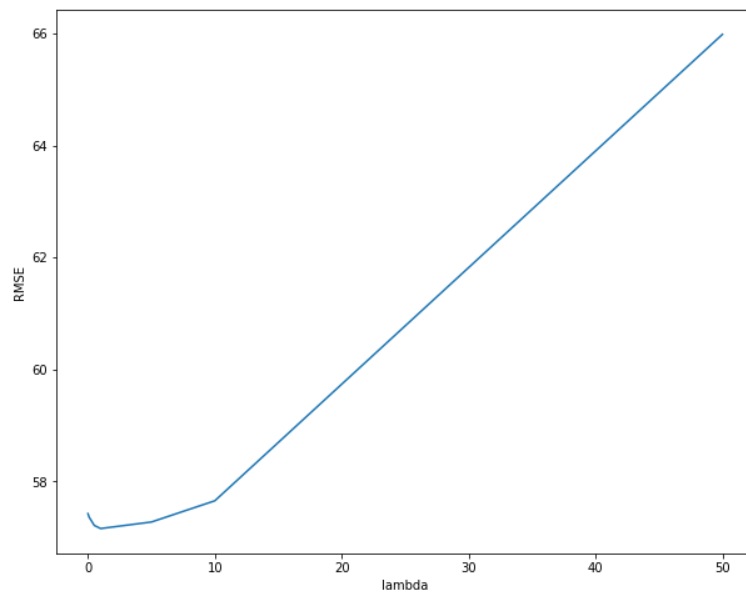
Na kraju ćemo iskoristiti ovaj izvod kriterijuma u gradijentnom metodu:

$$\theta = \theta + \eta \frac{dJ(\theta)}{d\theta}$$

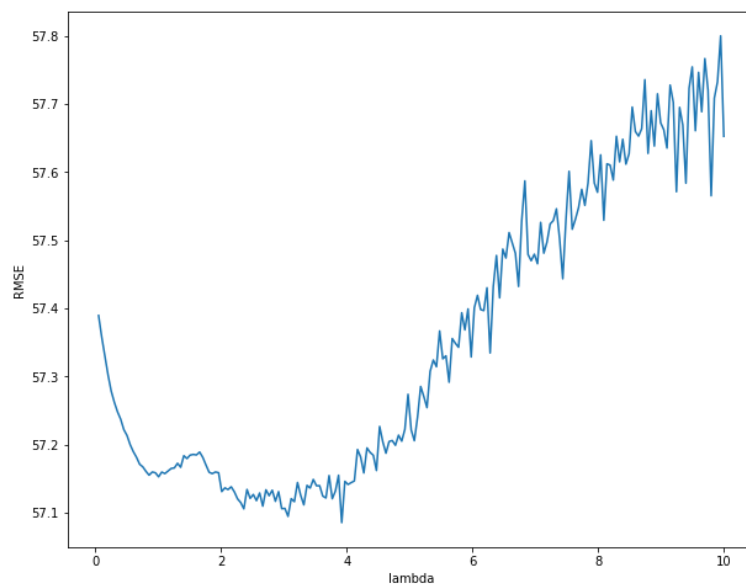
Za biranje optimalnog parametra  $\lambda$  koristila se *k-fold* krosvalidaciona metoda, gde je ceo skup podataka podeljen na 5 delova, a u svakoj epizodi se jedan od tih delova koristi kao validacioni skup, a ostalih 4 kao trening skup.

Optimalno  $\theta$  za svako  $\lambda$  dobijamo posle 1000 iteracija, gde je  $\eta = 0.0005$ .

Prvo je pokrenuta simulacija za  $\theta \in [0.00001, 50]$ :



Sa grafika se vidi da se optimalno  $\theta$  nalazi u opsegu  $[1, 10]$ , pa ćemo taj deo uveličati (pustiti simulaciju samo na tom delu sa većom rezolucijom):



Dobija se traženo  $\lambda = 3.92$ , za koje se dobija  $\text{RMSE} = 57.08$ .