

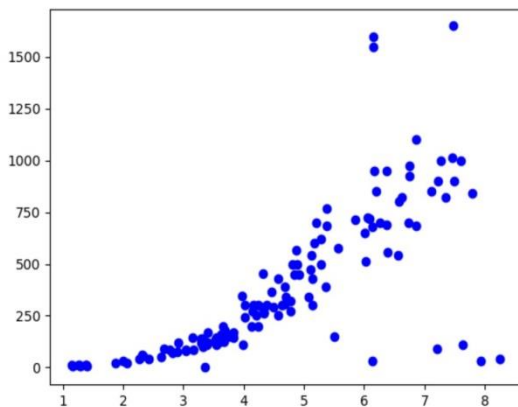
Jednostruka linearna regresija

Tim 6_23

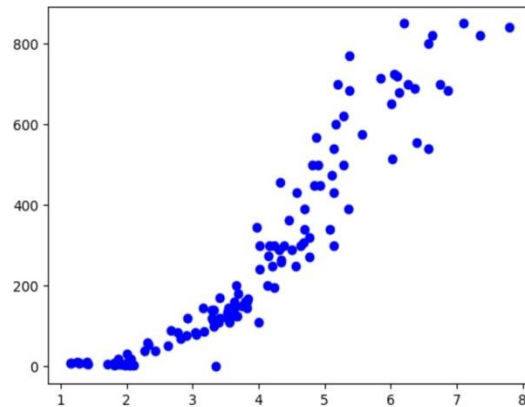
- Anastasija Samčović, SW44/2019
- Strahinja Popović, SW51/2019
- Srđan Đurić, SW63/2019

Pristup problemu

Prvi korak u rešavanju datog problema bila je vizualizacija i analiza datog trening skupa podataka. Na taj način smo uvideli trend koji bi trebalo podaci da prate i uočili koji podaci predstavljaju outlier-e. Uklanjanje outlier-a ne bi pomoglo, zato smo se odlučili za njihovu logaritamsku transformaciju.



Slika br. 1 – Podaci pre transformacije

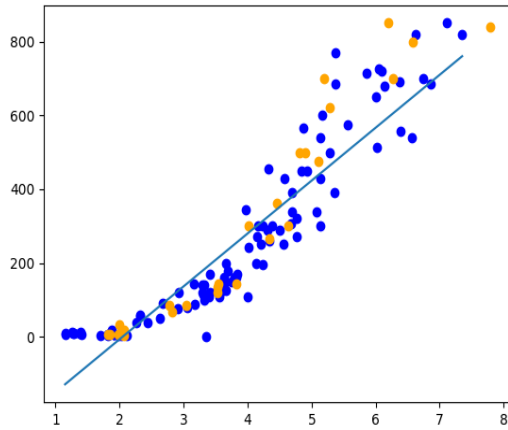


Slika br. 2 – Podaci nakon transformacije

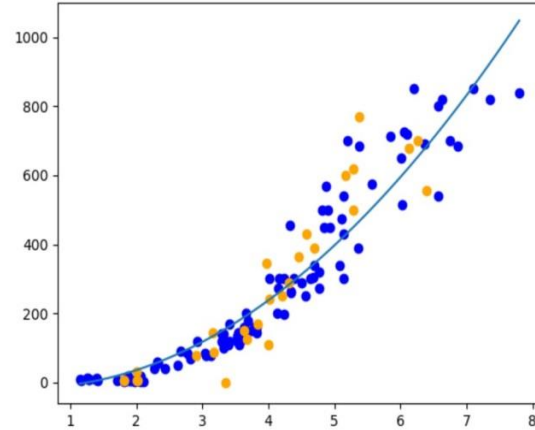
Skup podataka korišćen za testiranje predstavlja 20% datog trening skupa, a skup podataka korišćen za treniranje modela predstavlja 80% datog trening skupa.

U zavisnosti od podataka koji su se nalazili u testnom skupu, rezultat je varirao u dole navedenom opsegu(Tabela 1).

Nakon transformacije outlier-a, primenili smo batch gradient descent algoritam sa jednačinom prvog stepena.



Slika br. 3 – Jednačina prvog stepena, batch gradient descent



Slika br. 4 – Jednačina drugog stepena, batch gradient descent

Slika br. 3 nam pokazuje da prava linija ne prolazi kroz gužvu podataka, dok se na Slika br. 4 vidi da jednačina drugog stepena bolje prati trend podataka. Nakon toga smo primenili Normal Equation i Stochastic Gradient Descent metode nad jednačinom drugog stepena i došli smo do zaključka da Batch Gradient Descent daje bolje rezultate.

Eksperimentalnim putem smo utvrdili da jednačina petog stepena daje najbolje rezultate i da najbolje prati trend podataka.

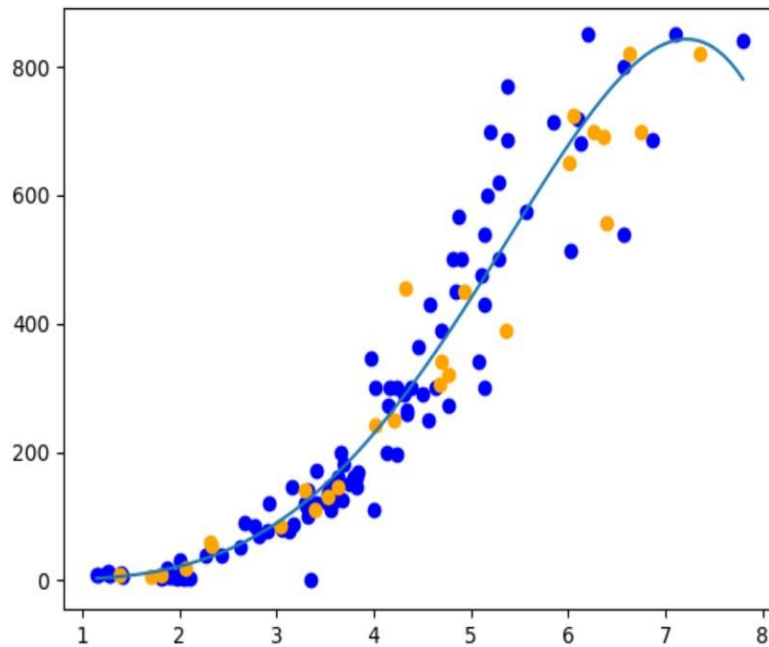
Pristup rešenju	RMSE(80% trening, 20% test)
Jednačina prvog stepena	[87, 102]
Jednačina drugog stepena	[61, 91]
Jednačina petog stepena	[54, 86]

Tabela 1 – Prikaz dobijenih rezultata

Izabrano rešenje

Naše konačno rešenje je Batch Gradient Descent algoritam sa jednačinom petog stepena. Parametri za koje je naš model pokazao najbolji odnos rezultata i brzine izvršavanja su:

- EPOCHES(Maksimalan broj iteracija) = 9000
- LR(Learning rate) = 0.000000031



Slika br. 5 – Jednačina petog stepena, batch gradient descent