

Jednostruka Linearna Regresija

Tim:

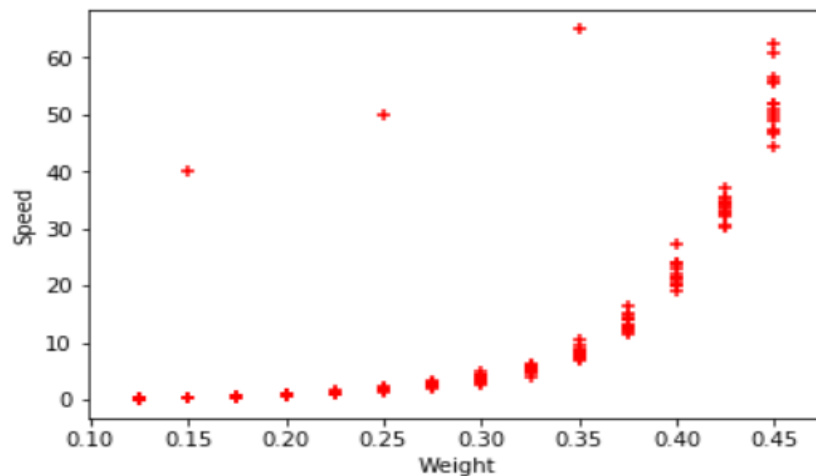
- Matija Aleksić

Zadatak:

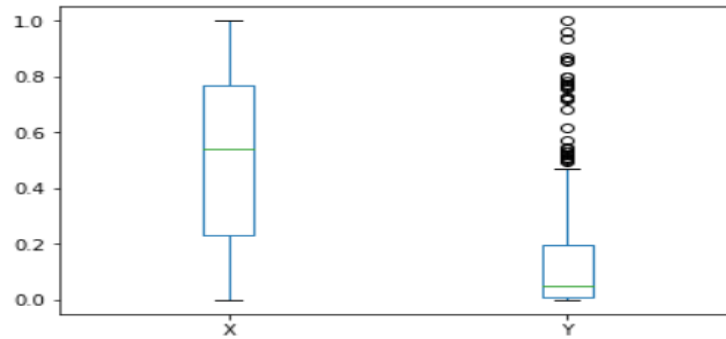
Jednostruka linearna regresija nad skupom podataka koji sadrže zavisnost rezidualne otpornosti broda po jedinici težine tokom pomeranja (Y) u odnosu na njegove dimenzije i brzinu kretanja (X). Pronaći regresioni model koji se najbolje uklapa u dati skup podataka.

Analiza podataka:

Analizom scattera plota podataka, 3 tacke na grafu su vidno outlieri.



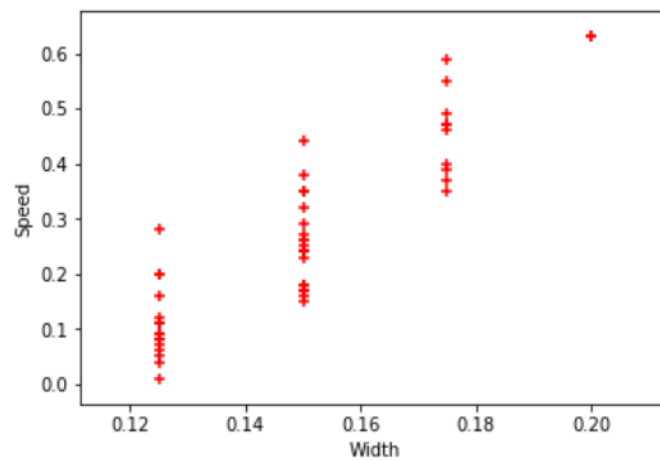
Ali daljim eksperimentisanjem sa metodama **z-scora** i **IQR** izbacivanja outlieri, koji ukljucujuci navedene 3 tacke koji su ocigledni outlieri, putem navedenih metoda dolazimo do zakljucka da bi morali da izbacimo veliku kolicinu podataka koje prelaze prag gornjih viskera BoxPlota pokazanog na slici ispod sto bi dovelo do slabijeg treniranja modela za predikciju testnog skupa. Radi ovih razloga sam odlucio da ostavim ove outliere u training skupu.



Na slici ispod je pokazan graf kako bi podaci izgledali kada bi se izbacili outlieri pomocu gore navedenih metoda. Takodje mozemo da vidimo da od 217 training podataka nakon izbacivanja je ostalo samo 50 elemenata.

Element number after IQR outlier removed: 50

Out[16]: <matplotlib.collections.PathCollection at 0x15321a3cf28>

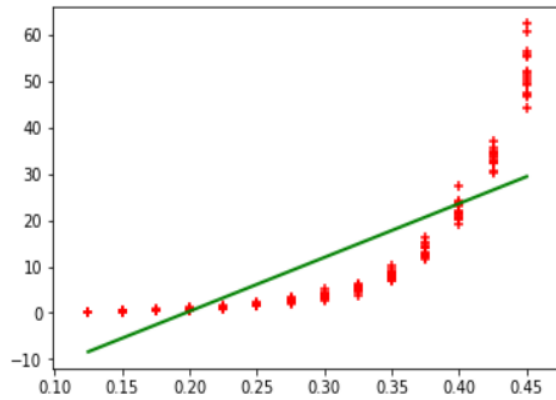


Primjenjene metode:

- Prilikom obucavanja linearne regresije training skup smo podelili na dva dela jedan za training i jedan za testiranje u razmjeri 80%/20% na random.
- Metoda koja je koriscena je **Polinomialna Regresija** jel sam zakljucio sa jednostavnom linearnom regresijom ne mozemo dobiti liniju na grafu koja je pozeljna za predikciju testnih podataka, pa sam se odlucio na ovu tehniku. Dalje slike opisuju rezultate dobijene sa polinomima razlicitih stepena:

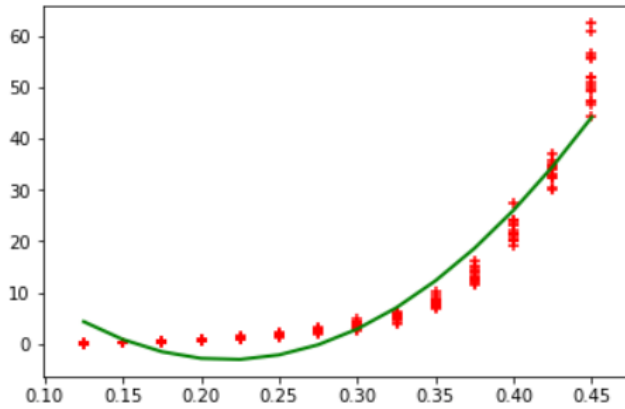
- Polinom prvog stepena $f(x) = a + bx$

Out[38]: 8.063107719700565



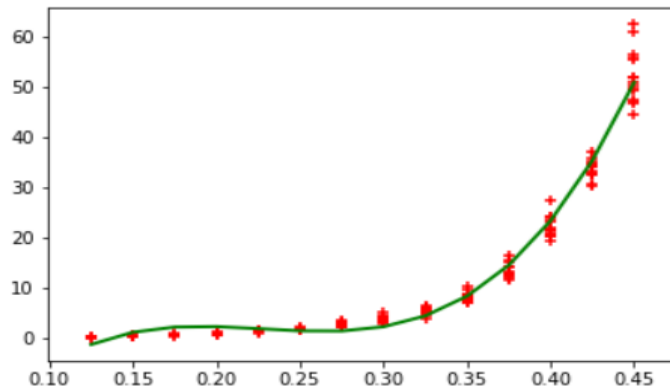
- Polinom drugog stepena $f(x) = a + bx + cx^2$

4.439477780612391



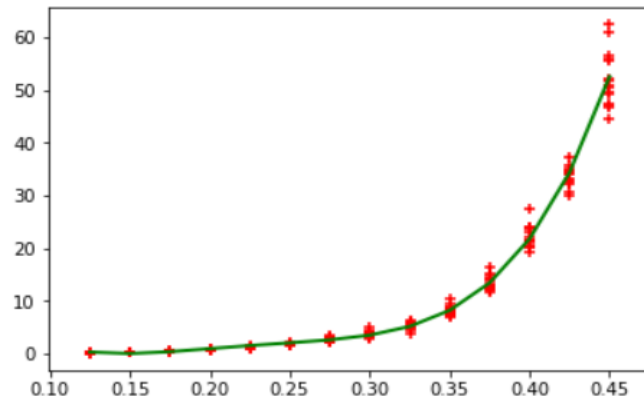
- Polinom treceg stepena $f(x) = a + bx + cx^2 + dx^3$

2.0511895880375715



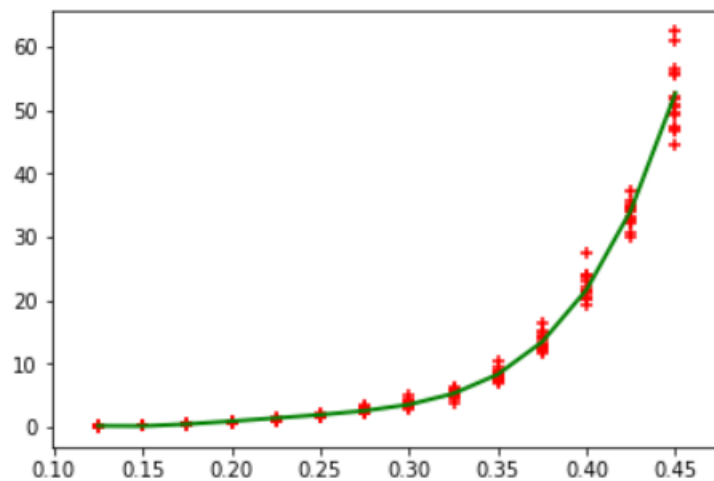
- Polinom cetvrtog stepena $f(x) = a + bx + cx^2 + dx^3 + ex^4$

1.493351811411137



- Polinom cetvrtog stepena $f(x) = a + bx + cx^2 + dx^3 + ex^4 + fx^5$

1.0842493939934825



Izabrana metoda:

Za resenje ovoga modela izabrao sam polinom petog stepena jel dovoljno dobro opisuje date podatke I radi jednostavnosti sam ostavio na petom stepenu. Svaki sledeci stepen polinoma bi davao jos tacnije rezultate ali sa malom promenom RMSE skora, uz veliku kolicinu racunja podataka kroz matrice.

Dobijeno resenje RMSE: 1.78304