2 czerwca 2021

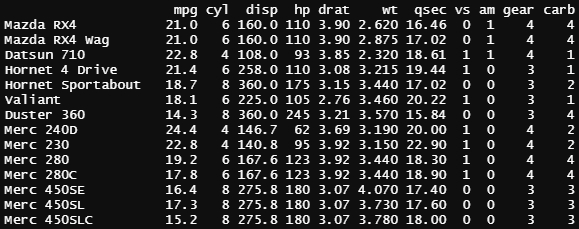
Kinga Kalisz & Marek Pałdyna

Porównanie Metod Regresji

Pakiet R

# MTCars

## Opis danych



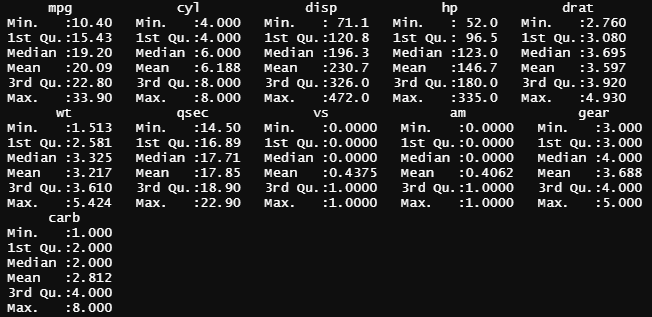
Zbiór danych MTCars zawarty jest w zbiorach danych zawartych w środowisku R.   
Składa się on z 11 kolumn i przedstawia dane na temat 32 samochodów z lat 1973-74. Zaciągnięty został z „Motor Trend US magazine”

Za cel przewidywań na tym zbiorze została obrana kolumna „qsec” która opisuje czas na ćwierć mili.

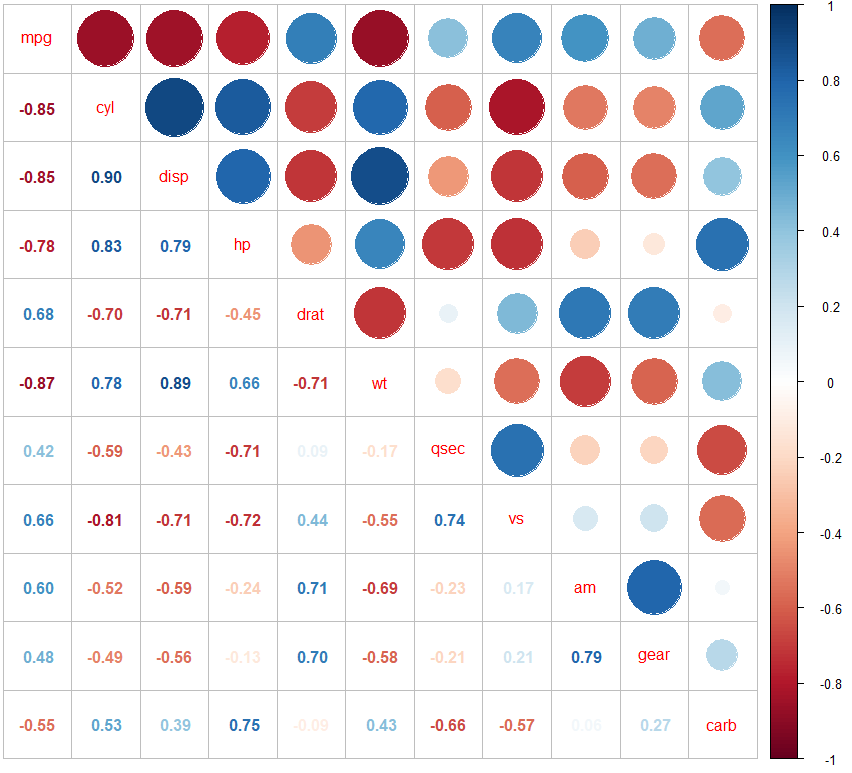
Kolumny:

* „mpg” – spalanie paliwa (mil/galon)
* „cyl” – Liczba cylindrów
* „disp” – pojemność silnika (cu.in.)
* „hp” – moc silnika (KM )
* „drat” – przełożenie tylnej osi
* „wt” – waga (1000 lbs)
* „qsec” – czas na ¼ mili
* „vs” – typ silnika (0 – w kształcie V, 1 - prosty)
* „am” – skrzynia biegów (0 – automatyczna, 1 -manualna)
* „gear” – liczba biegów do jazdy do przodu
* „carb” – liczba gaźników

## Podsumowanie danych:



## Korelacja zmiennych:



W tym zbiorze danych istnieją duże korelacje „pozytywne” i „negatywne”.

Korelacje negatywne:

* pomiędzy spalaniem i ilością cylindrów silnika,
* pomiędzy spalaniem i wagą samochodu
* pomiędzy spalaniem i pojemnością silnika
* pomiędzy liczbą cylindrów i typem silnika

Korelacje pozytywne:

* pomiędzy pojemnością silnika, a wagą samochodu
* pomiędzy pojemnością silnika, a liczbą cylindrów
* pomiędzy rodzajem silnika a liczbą biegów do jazdy do przodu
* pomiędzy liczbą cylindrów, a mocą silnika

## Obróbka danych

Do najlepszego dopasowania danych do zadanego problemu użyto funkcji step. Dobrała ona kolumny tak aby regresja wychodziła jak najlepsza -> najlepsze R-squared. Funkcja ta ma 3 „metody” – forward, backward i both.

Wartości squared dla danych:

* forward r.squared = 0.8523863
* backward r.squared = 0.8453189
* **both r.squared = 0.8642928**

Wybrany został więc model “both” –> „qsec ~ cyl + disp + wt + vs + am + carb”

## Modele regresji

Zostały zaimplementowane 3 modele regresji:

* LM
* GLM
  + Gaussian
  + Poisson
  + …
  + …
* SVM z kernelami:
  + Linear
  + Polynomial
  + Radial
  + Sigmoid

## Porównanie poszczególnych modeli i ich błędów RMSE

|  |  |  |
| --- | --- | --- |
| Model | Typ | RMSE |
| LM |  | ??? |
| SVM | Linear | 0,7227685 |
| Polynomial | 0,8880503 |
| Radial | 0,7340955 |
| Sigmoid | 1,173061 |
| GLM | Gaussian | ??? |
| Gamma | ??? |
|  | … | … |

## Podsumowanie

Najlepszym modelem jest : etc.

# Imports-85

## Dane

Plik imports-85.csv to plik zawierający dane o samochodach.

Na podstawie danych będziemy przeprowadzać przewidywanie mocy silnika (horsepower) na podstawie innych parametrów.

# Machine

## Dane

Plik machine.data zawiera dane na temat Komputerów PC.

Na podstawie danych będziemy przeprowadzać przewidywanie „mocy względna” (PRP) komputera.