

Extra credits PSTAT 126

Deadline: Monday 27th of March 11:59pm

2023-03-13

The data set “cars.txt” includes fuel consumption and 10 aspects of automotive design and performance for 32 automobiles:

- id: Car id
- mpg: Miles/(US) gallon (with a gallon \approx 3.79 liters)
- cyl: Number of cylinders
- disp: Displacement (cu.in.)
- hp: Gross horsepower
- drat: Rear axle ratio
- wt: Weight (1000 lbs, with 1000 lbs \approx 453.59 kg)
- qsec: 1/4 mile time (with 1/4 mile \approx 402.34 meters)
- vs: Engine (0 = V-shaped, 1 = straight)
- am: Transmission (0 = automatic, 1 = manual)
- gear: Number of forward gears
- carb: Number of carburetors

```
cars<-read.table(file="cars.txt", head=T)
```

Our goal is to model the response mpg in terms of the rest of the variables (except name).

- Partition the data set into two sets a training data and a test data. Remove every fifth observation from the data for use as a test sample.
- Perform an exploratory analysis. Comment on your findings.
- Perform a regression analysis and come up with the best multiple linear regression model that explains the response mpg in terms of the rest (except name). Comment on your findings and explain the methods and strategies that you employed in order to select the model you picked. Things you have to include in this part:
 - Model diagnostics
 - Justification on whether it is necessary or not to do any transformation on the response or the predictors
 - Variable selection
- Assess the prediction performance by using the test sample.