

Motivation



Use auto dataset from ISLR library in R



Model mpg based on car features



Features:

Cylinders
Displacement
Horsepower
Weight
Acceleration
Year
Origin

Code Snippet

```
import numpy as np
from pandas import *

# Create Python DataFrame
cylinders = df_test.rdd.map(lambda p: (p.features[0])).collect()
displacement = df_test.rdd.map(lambda p: (p.features[1])).collect()
horsepower = df_test.rdd.map(lambda p: p.features[2]).collect()
mpg = df_test.rdd.map(lambda p: p.mpg).collect()
pred = lr_predictions.rdd.map(lambda r: r.prediction).collect()
```

- Code to collect datapoints into lists, which can in turn be read into a pandas dataframe, which can in turn be used for plotting with matplotlib

Performance of Model on Horsepower Variable

