Test Project:Gas Mileage Prediction

Xiaochen Zhuo

Table Details: car

Schema

mpg	FLOAT	NULLABLE	Describe this field...
cylinders	FLOAT	NULLABLE	Describe this field...
displacement	FLOAT	NULLABLE	Describe this field...
horsepower	FLOAT	NULLABLE	Describe this field...
weight	FLOAT	NULLABLE	Describe this field...
acceleration	FLOAT	NULLABLE	Describe this field...
year	FLOAT	NULLABLE	Describe this field...
origin	FLOAT	NULLABLE	Describe this field...
name	STRING	NULLABLE	Describe this field...

The column I'm interested in is mpg.I want to make a prediction about whether a car's mpg will be high or low(compare to the median) based on other columns, a classification problem.

Procedure

- Add a new column into the data indicate whether the mpg of the row is higher or lower than average, I call it mpg01.
- Remove the mpg column.
- Explore the data graphically and numerically .
- Split the data into training and test.
- Throw different classification methods on the training data and build models
- Test the model on test data

Result(error rate):

Linear Discriminant Analysis: 0.10
Quadratic Discriminant Analysis: 0.11
Logistic Regression: 0.33
K nearest neighbour: 0.09
Please see detail in R script.