Q1.)

Assume that we want to classify the cars into 3 categories: low, medium and high mpg.
Find what the threshold for each category should be, so that all samples are divided into
three equally-sized bins.

First of all, we know use ‘size’ to know how many data we have, then use ‘sort’ to sort data from low to high. Then use totally number of data divide 3 to split them two three groups, which are low, medium and high MPG .

Low from 9 to 18.5

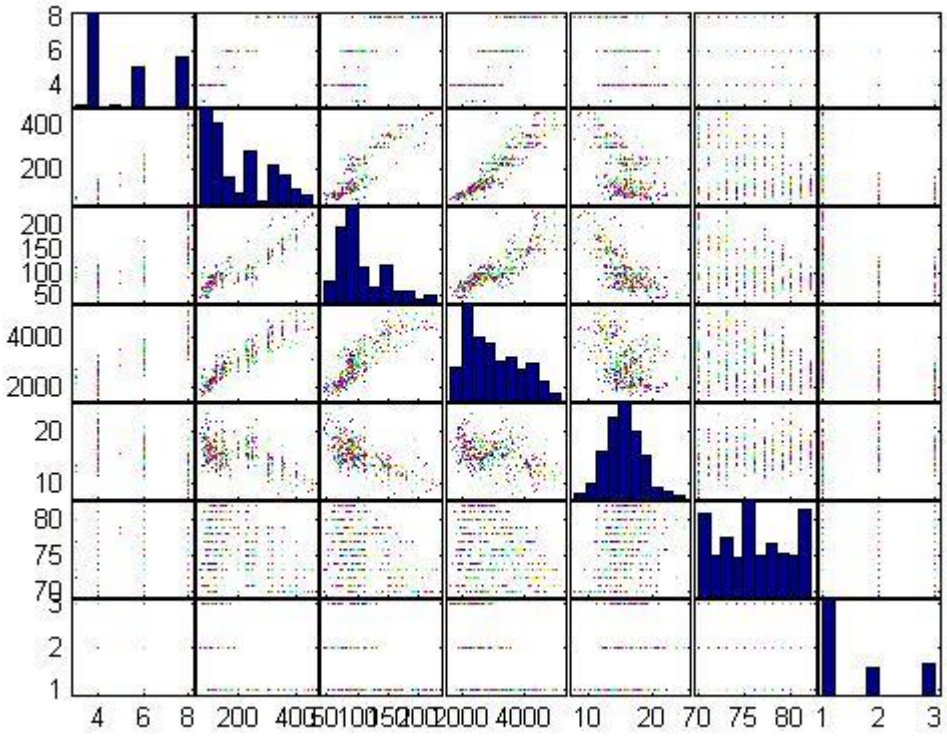
Medium from 19 to 26.8

High from 26.8 to 46.6

Q2

Create a 2D scatterplot matrix

Use gplotmatrix function to plot them, since we want to all the feature compare with MPG, so each plot have to has MPG, then pair from all pairs feature combinations, so that we have 49 plots. Acceleration and weight most informative regarding MPG.



Q3

Write a linear regression solver that can accomodate polynomial basis functions on a single
variable. Your code should use the Ordinary Least Squares (OLS) estimator which is
also the Maximum-likelihood estimator

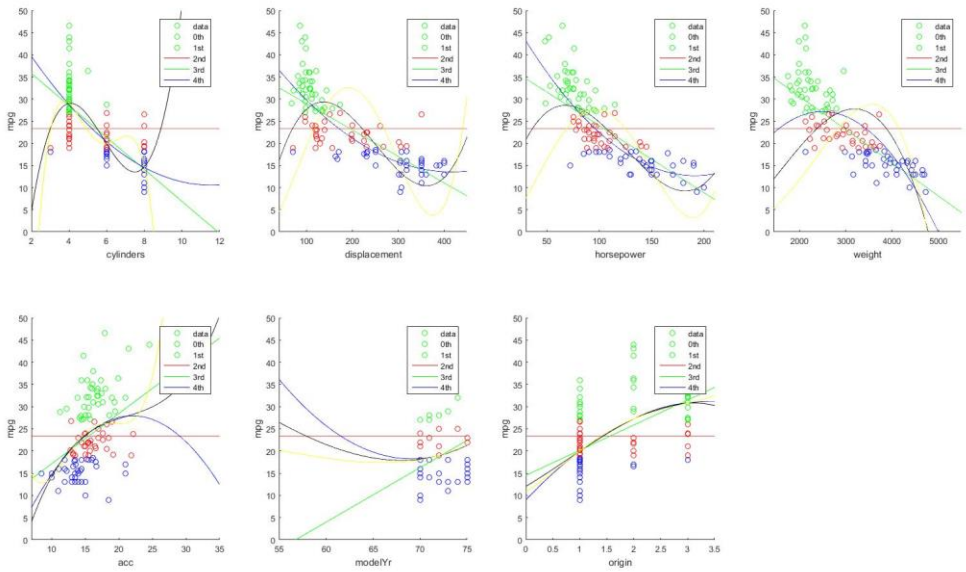
```
function w = LRS3(Y,X,n)
    if n==0,
        X=ones(length(X),1);
    elseif n==1,
        X=[ones(length(X),1),X];
    elseif n>1,
        single=X;

        indices=1:(n-1);
        for i = indices,
            X=[X,single.^(i+1)];
        end
        X=[ones(length(X),1),X];
    end
    w=pinv((X'*X))*X'*Y;
```

Q4

Split the dataset in the first 280 samples for training and the rest 112 samples for testing. Use your solver to regress for 0th to 4th order polynomial on a single independent variable (feature) each time by using mpg as the dependent variable. Report (a) the training and (b) the testing mean squared errors for each variable individually (except the "car name" string variable, so a total of 7 features that are independent variables). Plot the lines and data for the testing set, one plot per variable (so 5 lines in each plot, 7 plots total).

Use the slover (q3) to calculated 0^{th} to 4^{th} order polynomial



cylinders-mpg: training errors for five functions are

127.26

81.48

81.40

78.25
78.25.
cylinders-mpg: test errors for five functions are
80.72
98.29
98.18
98.43
98.42.
disp-mpg: training errors for five functions are
127.26
76.05
72.26
93.97
176.00.
disp-mpg: test errors for five functions are
80.72
103.11
102.36
105.57
128.10.
horsepower-mpg: training errors for five functions are
127.26
79.39
71.28
84.88
129.20.
horsepower-mpg: test errors for five functions are
80.72
110.23
106.28
105.30
114.38.
weight-mpg: training errors for five functions are
127.26
71.38
98.43
153.83
217.67.
weight-mpg: test errors for five functions are
80.72
102.37
104.12
116.15
139.81.

acc-mpg: training errors for five functions are
127.26
117.82
116.58
115.76
113.38.
acc-mpg: test errors for five functions are
80.72
88.31
90.41
91.56
90.97.
modelYr-mpg: training errors for five functions are
127.26
103.72
101.70
101.73
101.75.
modelYr-mpg: test errors for five functions are
80.72
98.11
98.53
98.59
98.63.
origin-mpg: training errors for five functions are
127.26
106.20
105.46
105.46
105.46.
disp-mpg: test errors for five functions are
80.72
89.04
88.90
88.90
88.90.
cylinders-mpg: training errors for five functions are
126.30
81.57
81.32
78.58
78.39.
cylinders-mpg: test errors for five functions are
81.62

102.59
102.57
104.75
104.91.
disp-mpg: training errors for five functions are
126.30
76.46
72.09
95.91
177.15.
disp-mpg: test errors for five functions are
81.62
105.01
105.97
109.43
125.52.
horsepower-mpg: training errors for five functions are
126.30
81.15
71.48
87.17
131.58.
horsepower-mpg: test errors for five functions are
81.62
102.31
106.19
106.09
112.07.
weight-mpg: training errors for five functions are
126.30
71.35
100.01
156.01
218.57.
weight-mpg: test errors for five functions are
81.62
105.60
102.34
106.20
124.91.
acc-mpg: training errors for five functions are
126.30
113.50
112.94

112.94
112.52.
acc-mpg: test errors for five functions are
81.62
88.08
87.82
87.84
88.38.
modelYr-mpg: training errors for five functions are
126.30
103.65
101.48
101.48
101.49.
modelYr-mpg: test errors for five functions are
81.62
92.36
91.94
91.90
91.88.
origin-mpg: training errors for five functions are
126.30
103.55
102.35
102.35
102.35.
disp-mpg: test errors for five functions are
81.62
85.92
88.14
88.14
88.14.
cylinders-mpg: training errors for five functions are
128.04
78.01
77.48
75.80
75.63.
cylinders-mpg: test errors for five functions are
79.78
103.52
103.73
103.21
103.06.

disp-mpg: training errors for five functions are
128.04
73.34
68.51
93.20
177.37.
disp-mpg: test errors for five functions are
79.78
101.54
104.10
105.24
115.92.
horsepower-mpg: training errors for five functions are
128.04
80.99
72.33
86.52
132.28.
horsepower-mpg: test errors for five functions are
79.78
103.93
107.09
107.99
109.66.
weight-mpg: training errors for five functions are
128.04
70.06
100.55
159.15
224.29.
weight-mpg: test errors for five functions are
79.78
104.75
101.90
103.63
117.07.
acc-mpg: training errors for five functions are
128.04
116.73
115.39
115.26
114.54.
acc-mpg: test errors for five functions are
79.78

86.00
86.98
86.88
86.73.
modelYr-mpg: training errors for five functions are
128.04
104.11
102.57
102.64
102.70.
modelYr-mpg: test errors for five functions are
79.78
94.50
94.95
95.03
95.10.
origin-mpg: training errors for five functions are
128.04
104.17
103.51
103.51
103.51.
disp-mpg: test errors for five functions are
79.78
95.88
95.55
95.55
95.55.
M=3 polynomial order performs the best in the test set
Which feature is the most informative regarding mpg consumption in that case? Horsepower

5.
Modify solver to be able to handle second order polynomials of all 8 independent variables simultaneously (i.e. 15 terms). Regress with 0th, 1st and 2nd order and report (a) the training and (b) the testing mean squared error. Use the same 280/112 split as before.
Training and testing mean squared errors are :
84.75
51.46
51.53
6.
Modify your solver to allowfor logistic regression (1st order) and report the training/testing mean squared error, as before.

I spited this question to 3 parts. Low medium and high MPG.
I used training data to compute cost and gradient, then optimizing to [1 0] range
I make all low mpg data equal to 0, otherwise equal to 1. Then use incorrectly predict value divide to totally case.

testing low MPG our predict data MSE is : 0.009524 = 0.95%
traning low MPG our predict data MSE is : 0.025478=2.5%

testing medium MPG our predict data MSE is : 0.318182=31%
traning low MPG our predict data MSE is : 0.114943=11%

testing High MPG our predict data MSE is : 0.054422=5.4%
training High MPG our predict data MSE is : 0.004367=4%
7.
considered to introduce amodel in 1980 with the following
characteristics: 6 cylinders, 300 cc displacement, 170 horsepower, 3600 lb weight,
9 m/sec^2 acceleration, what is the MPG rating that we should have expected? In which
mpg category (low,medium,high mpg) would it belong? Use second-order, multi-variate
polynomial and logistic regression.
I make the cylinders displacement horsepower wight acceleration data as training data as X, and Mpg data as Y . X and Y are training data, then use them to predict the specific data. This question I used feature normalization, since feature normalization could make leaning fast, make the plot convert fast. First of all I calculated sigma/mu for each row feature , then normalization scale , after that we have x' = (x-\mu)/\sigma. ， then we have the function of feature normalization .then use multiple variable linear regression

Predicted 6 cylinders, 300 cc displacement, 170 horsepower, 3600 lb weight,9 m/sec2 acceleration, (using gradient descent) mpg :16.254682

According to Q1, this car belong to low mpg rate category.