Adam Levin
998408921

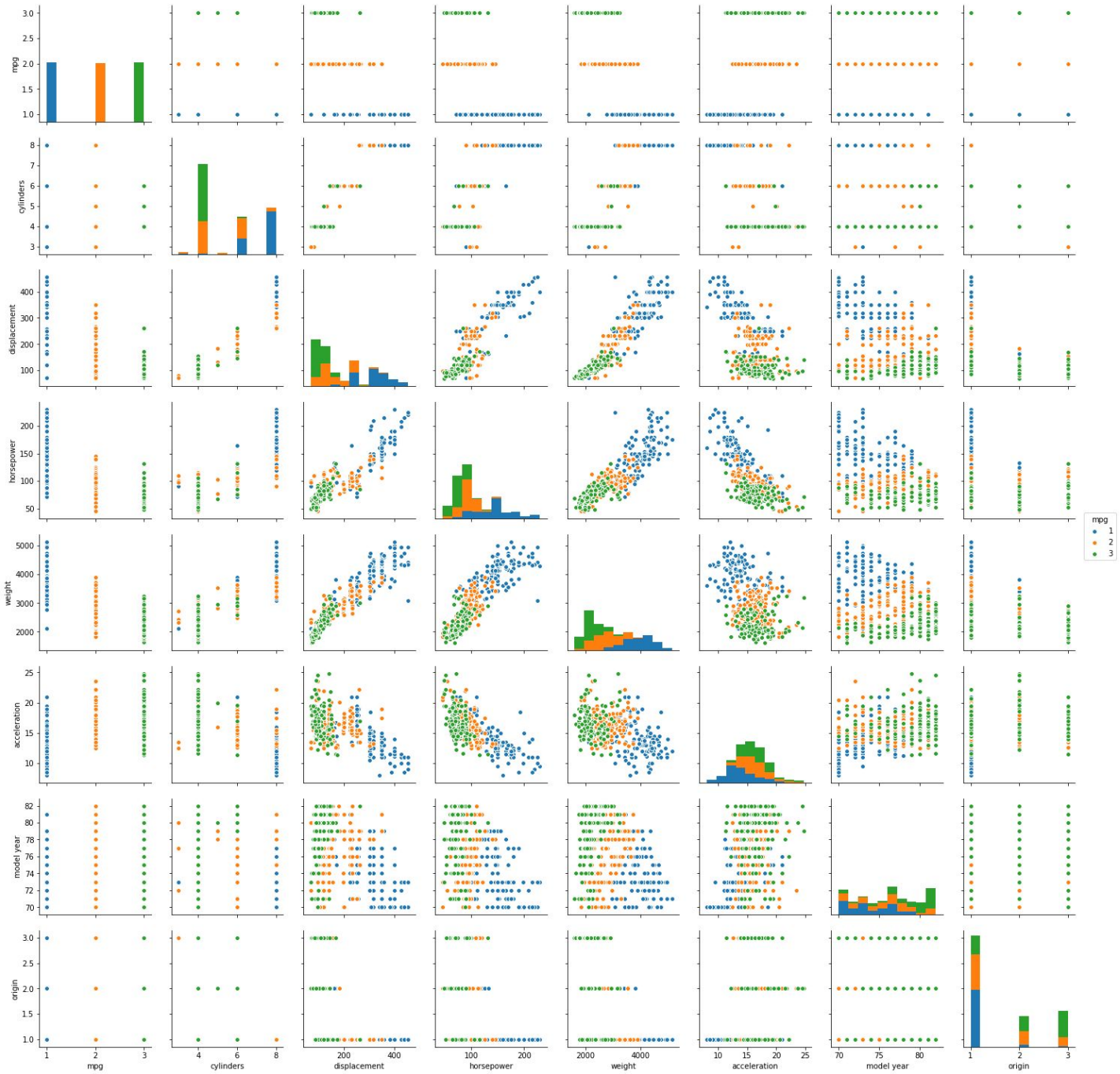
Assignment 1

Stipulations on running code:

1. All problems other than number 2 of the assignment can be compiled and run in all_other_problems.py
2. Due to superimposing of matplotlib upon the later graphs, I had to use a separate file for the seaborn pairplot function

Assignment Answers:

1. Using a sorted list of MPG and index access of 130 and 260, the 3 rules for low, medium and high mpg are as follows:
    a. Low <= 18.6 MPG
    b. 18.6 MPG < Medium <= 26.8 MPG
    c. High > 26.8 MPG
2. Based on looking at the graphs of all pairwise combinations, I hypothesize that the most influential pair of features upon MPG are weight and horsepower. (see figure below)



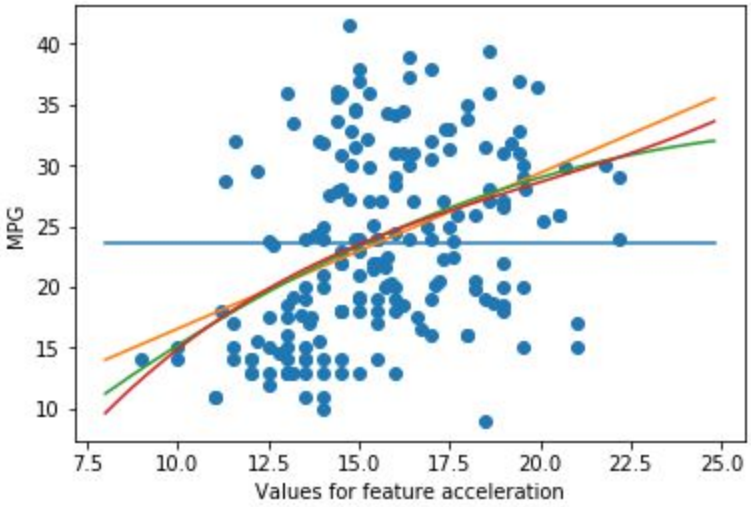
3. See the solver class in all_other_problems.py
4. On different runs the data fluctuates a little bit due to random shuffling of the dataset, here is the mean squared error for the training and testing sets this particular run:
    a. Mean squared train error

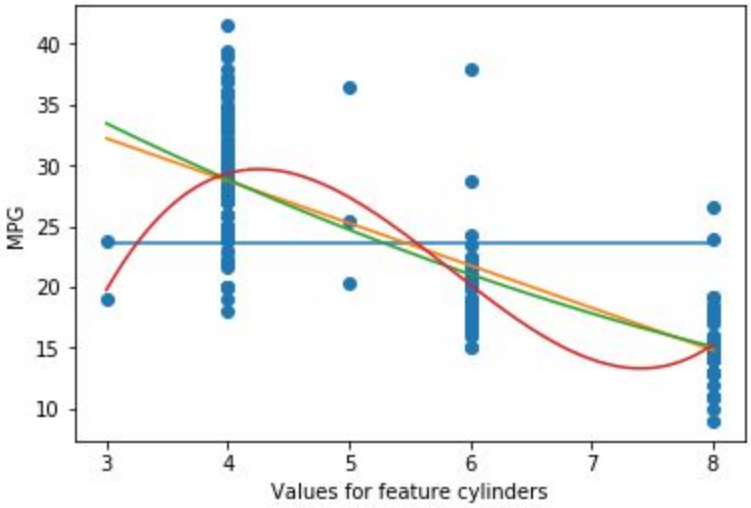
Index	0th order	1st order	2nd order	3rd order
cylinders	61.9854	25.9238	25.7739	23.6128
displaceme…	61.9854	22.0556	19.1901	34.7229
horsepower	61.9854	25.0613	18.3155	30.3774
weight	61.9854	20.1389	38.8644	92.1247
accelerati…	61.9854	47.9421	47.4435	47.3431
model year	61.9854	43.0625	39.7121	39.712
origin	61.9854	38.3494	37.4848	37.4848

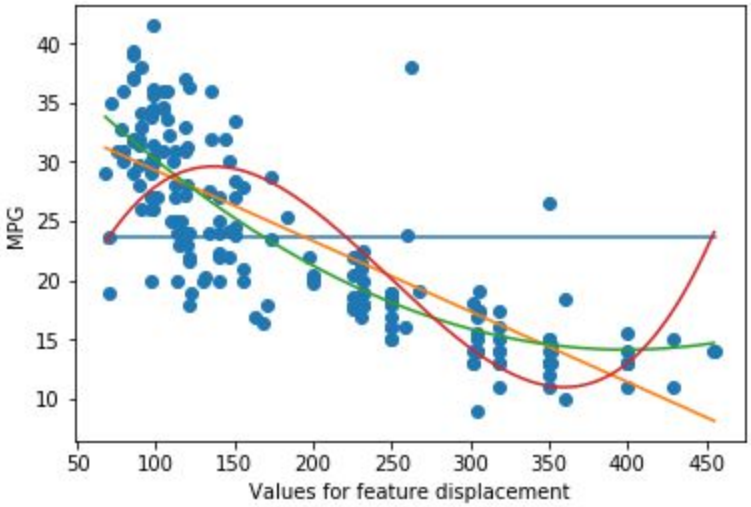
b. Mean squared test error

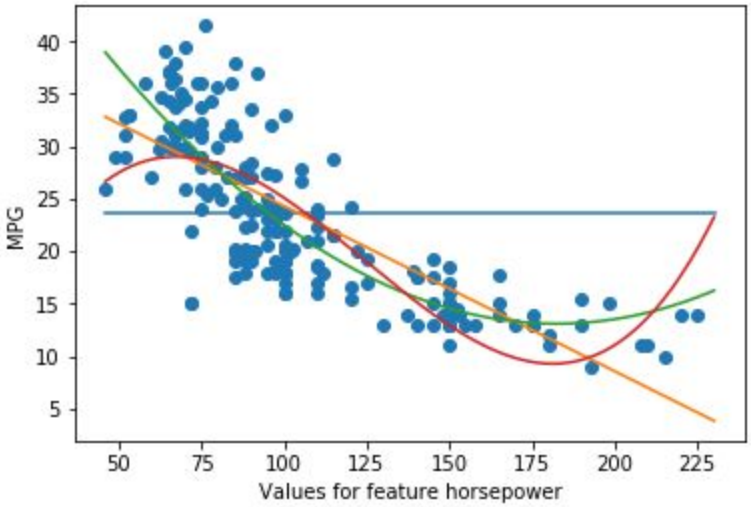
Index	0th order	1st order	2nd order	3rd order
cylinders	59.5489	22.124	21.9256	19.9932
displaceme…	59.5489	20.7805	18.7046	31.5986
horsepower	59.5489	22.8298	19.8953	24.6852
weight	59.5489	17.2003	34.2541	87.9142
accelerati…	59.5489	52.0958	51.1055	50.9259
model year	59.5489	37.6091	37.5973	37.6962
origin	59.5489	45.1673	44.4279	44.4279

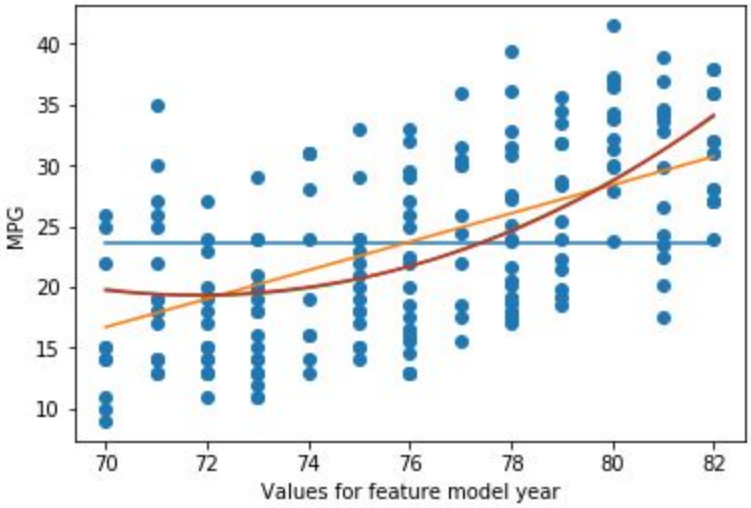
c. Plots of lines and data for testing set

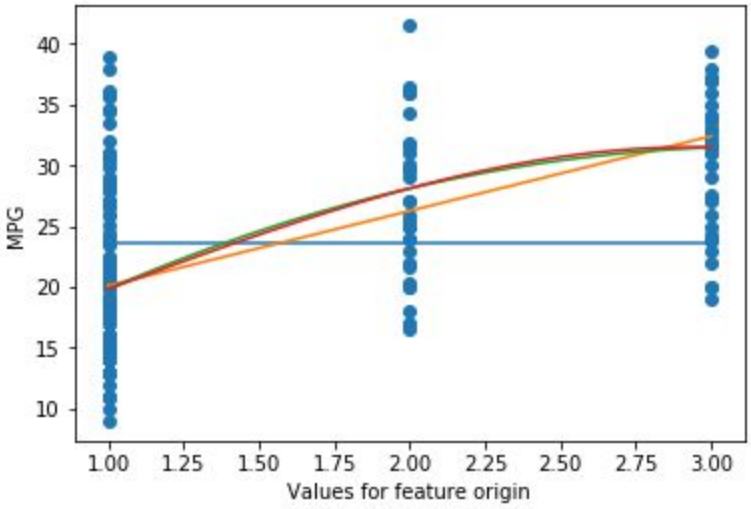

MPG
40
35
30
25
20
15
10
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
Values for feature acceleration

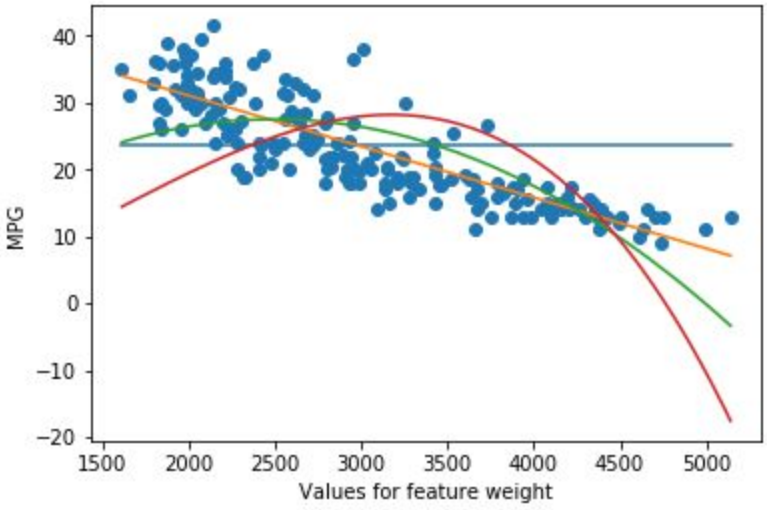

MPG
40
35
30
25
20
15
10
3
4
5
6
7
8
Values for feature cylinders


MPG
40
35
30
25
20
15
10
50
100
150
200
250
300
350
400
450
Values for feature displacement


MPG
40
35
30
25
20
15
10
5
50
75
100
125
150
175
200
225
Values for feature horsepower


MPG
40
35
30
25
20
15
10
70
72
74
76
78
80
82
Values for feature model year





It appears that the second order polynomial performs the best on the data given, and given that particular order, horsepower seems to be the most informative feature.

5. Multivariate polynomial regression MSE report

Index	0th Order	1st Order	2nd Order
Training MSE	64.215	10.756	7.63445
Testing MSE	58.0371	11.3857	9.45447

6. Logistic Regression Precision
    a. Precision score noted

prec_score	float64	1	0.7817370537958773

7. MPG prediction

mpg_class_pred	float64	(1,)	[1.]
mpg_pred	float64	1	18.19042134103937

    a. The 1 for class prediction indicates a MPG of a low class
    b. The prediction mpg in this run is roughly 18.2 MPG, but varies on multiple runs of the program depending on the random shuffling of the dataset

8. Bonus Question -
    a. Assumptions:
        i. the Pony has a method of ingesting gasoline that doesn't poison and kill it
        ii. The gasoline can serve as caloric intake as well as the water intake for the horse
        iii. The Pony can run 50 miles in a day
        iv. The Pony can consume 5 gallons of liquid
    b. By these assumptions, the pony can operate at 10 miles per gallon