Armand Nasseri
912679383
ECS 171

Homework 1

For this homework assignment, I have created 8 different files that correspond to each of the problems in the assignment description. Each file will have to be run separately to view the output.

1. In my code I sorted all the mpg categories so that I could divide them into 4 cutoff points to make 4 equally sized bins. The values are printed to the screen when the program runs.

Low mpg	Med mpg	High mpg	Very high mpg
mpg <= 17.0	17.0 < mpg <=22.4	22.4 < mpg <= 29.0	29.0 <mpg <=46.6

2. It appears that horsepower vs. weight is the most informative due to the fact that the data is clustered very well between the four mpg categories. It makes sense due to the fact that cars will burn more fuel when more weight is added. Below is a chart with all the features and threshold categories scattered to indicate low, med, high, or very high mpg.

3. Refer to the problem3.py file for function implementation.

4. MPG vs. Cylinder

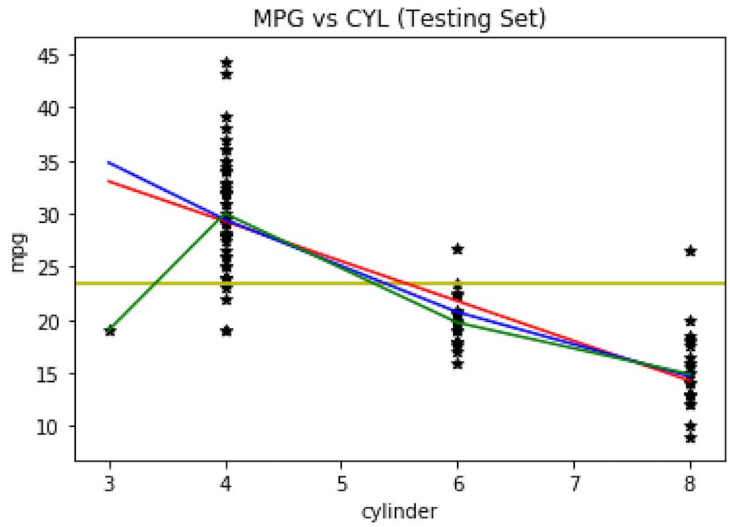
Training Error

0th Order	1st Order	2nd Order	3rd Order
59.497	24.45	24.32	22.51

Testing Error

0th Order (Yellow)	1st Order (Red)	2nd Order (Blue)	3rd Order (Green)
64.45	22.48	22.17	19.30

3rd Order appears to be best for this dataset



MPG vs. Displacement

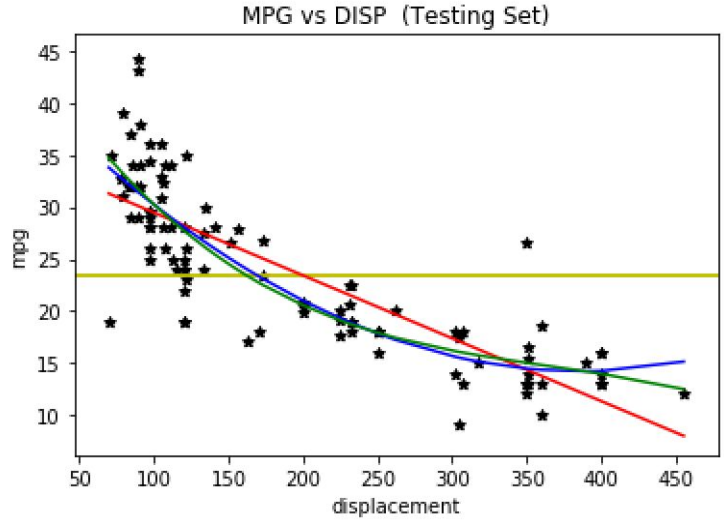
Training Error

0th Order	1st Order	2nd Order	3rd Order
64.45	20.39	17.23	17.00

Testing Error

0th Order (Yellow)	1st Order (Red)	2nd Order (Blue)	3rd Order (Green)
59.49	24.64	20.47	20.47

2nd Order appears to be best for this dataset



MPG vs. Horsepower

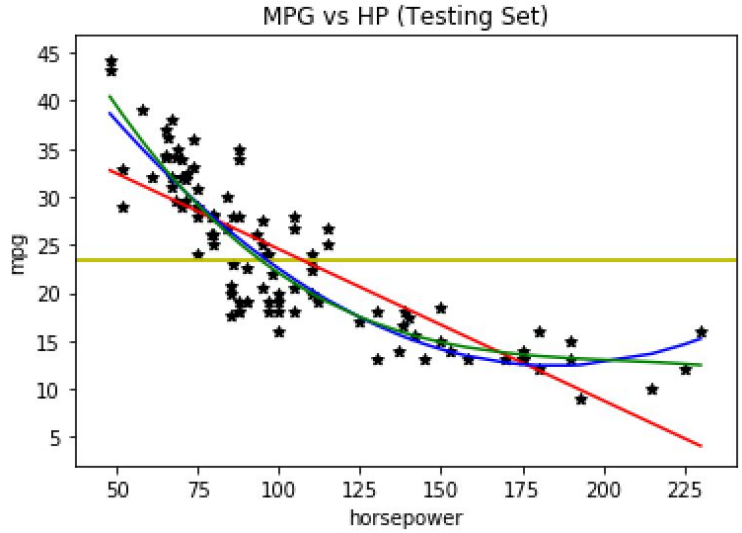
Training Error

0th Order	1st Order	2nd Order	3rd Order
59.49	24.64	20.47	20.47

Testing Error

0th Order (Yellow)	1st Order (Red)	2nd Order (Blue)	3rd Order (Green)
64.45	21.69	14.31	13.87

2nd Order appears to be best for this dataset



MPG vs. Weight

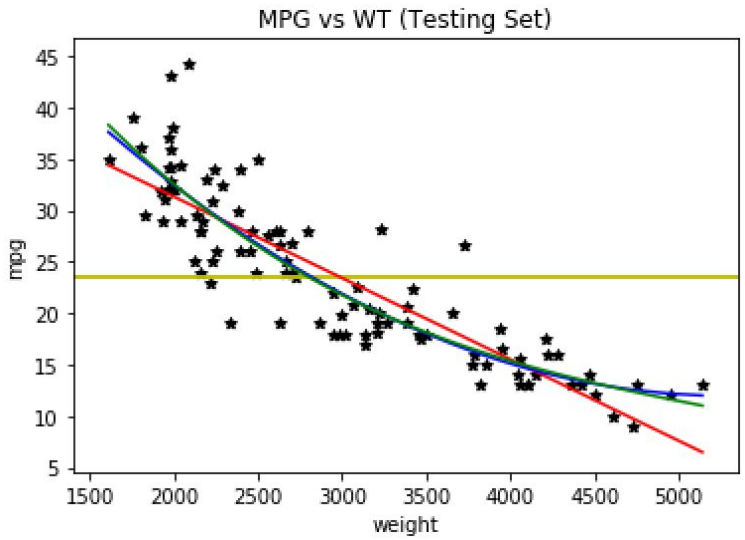
Training Error

0th Order	1st Order	2nd Order	3rd Order
59.49	19.87	18.71	18.70

Testing Error

0th Order (Yellow)	1st Order (Red)	2nd Order (Blue)	3rd Order (Green)
64.45	15.07	13.07	13.02

3rd Order appears to be best for this dataset



MPG vs. Acceleration

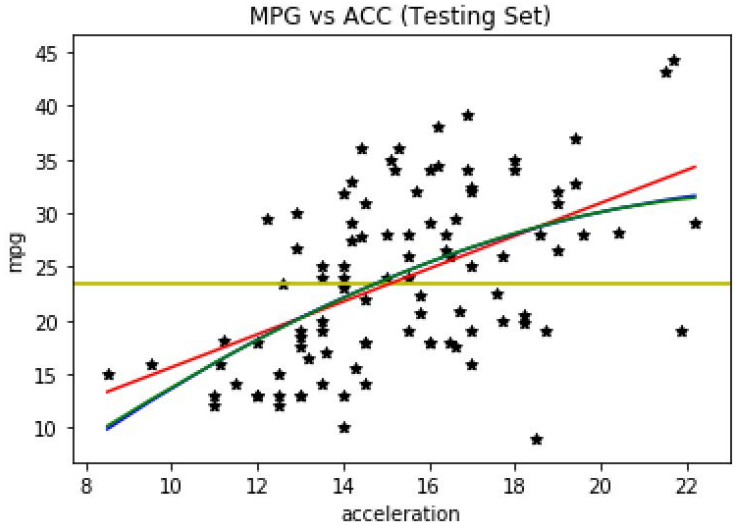
Training Error

0th Order	1st Order	2nd Order	3rd Order
59.49	50.30	49.38	49.16

Testing Error

0th Order (Yellow)	1st Order (Red)	2nd Order (Blue)	3rd Order (Green)
64.45	47.11	46.50	46.49

3rd Order appears to be best for this dataset



MPG vs. Year

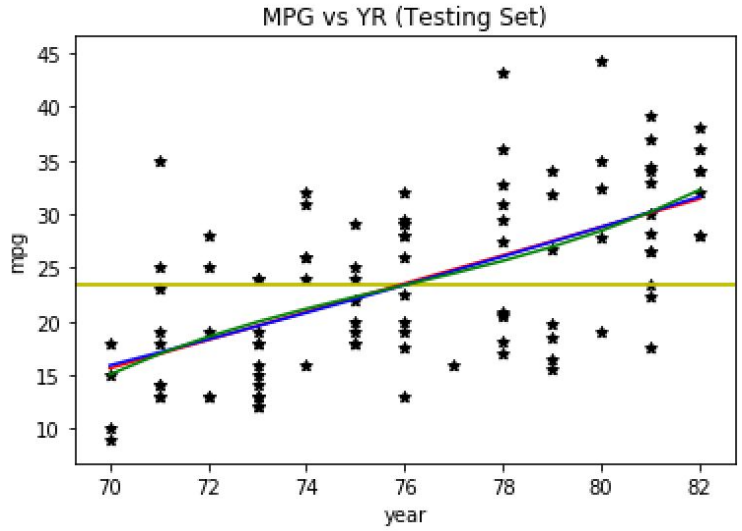
Training Error

0th Order	1st Order	2nd Order	3rd Order
59.49	39.90	36.85	36.67

Testing Error

0th Order (Yellow)	1st Order (Red)	2nd Order (Blue)	3rd Order (Green)
64.45	41.25	41.23	41.10

3rd Order appears to be best for this dataset



MPG vs. Origin

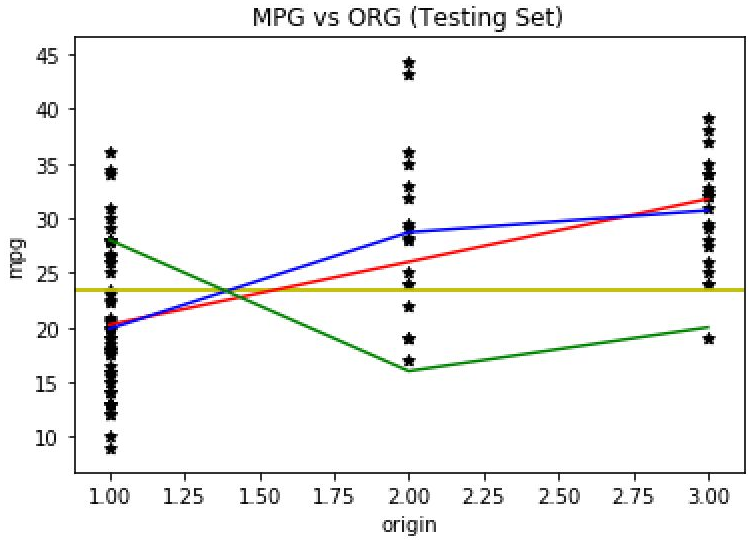
Training Error

0th Order	1st Order	2nd Order	3rd Order
59.49	41.11	40.56	22839.50

Testing Error

0th Order (Yellow)	1st Order (Red)	2nd Order (Blue)	3rd Order (Green)
64.45	41.95	40.38	133.30

1st Order appears to be best for this dataset



It appears that weight is the most informative out of all the features in the dataset.

5.

Training Error

0th Order	1st Order	2nd Order
59.48	11.19	7.49

Testing Error

0th Order	1st Order	2nd Order
64.36	10.26	7.19

6. The training classification precision for the various classes is: **0.73.**The testing classification precision for the various classes is: **0.81**

7. After I applied min-max normalization to the dataset, there was a definite decrease in performance. The new calculated training classification precision decreased to **0.71.** The new calculated testing classification precision decreased to **0.74**

8. The expected MPG rating using multivariate polynomial regression is **23.64**, which would make this car fall into the high mpg category.

The expected MPG rating using logistic regression for these specific features would place this car in the **high** mpg category.