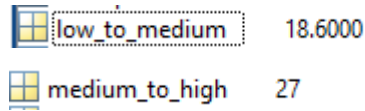


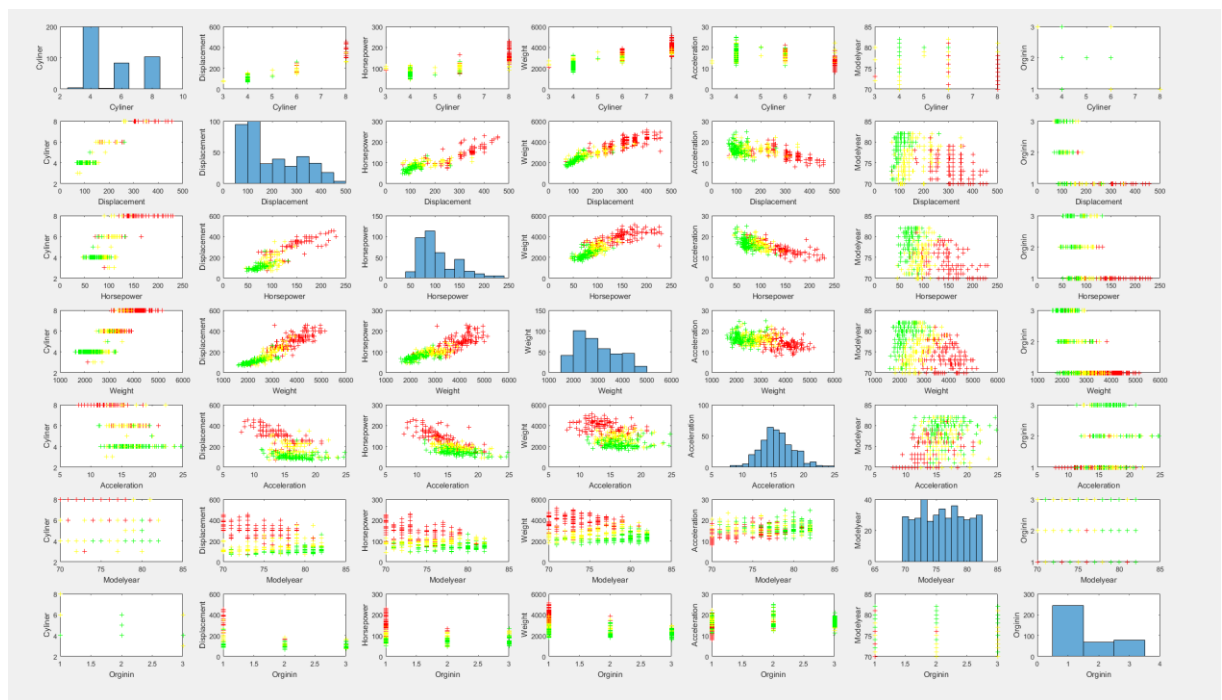
ECS 171 Homework Set 1

Haozhe Gu 999200555

- The threshold for three categories are:



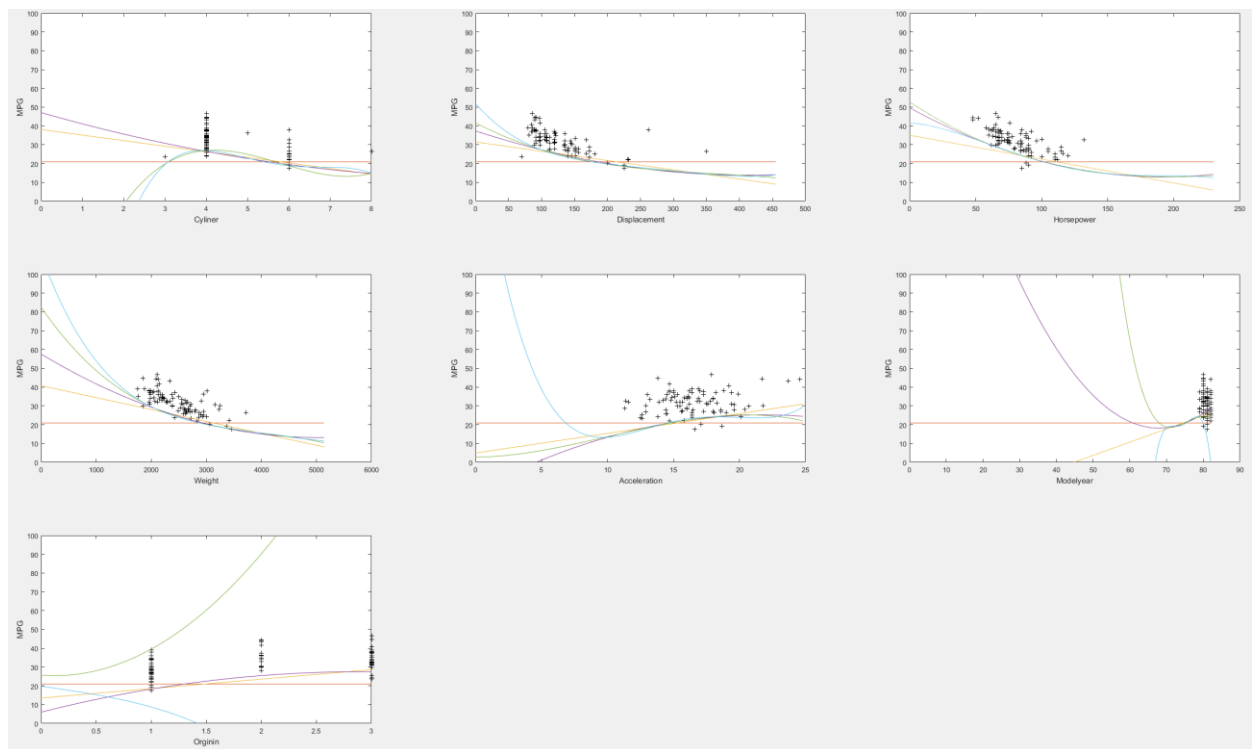
- From the plots below, we can see the plot of Weight Vs Model year (6,4) or (4,6) might have the best separation regard to three categories.



- The file ECS171Q3.m is my solver that integrated OLS and Logistic Linear Regression. The part below respond to OLS.

```
if(mode == -1)
    y = y(1:size(x, 1));
    w = inv(x.' * x) * x.' * y;
```

4. The following seven graphs are for predict lines and testing data.



Training MSE (rows correspond to 0th order, 1st order... 4th order; columns correspond to 7 individual features: "Cylinder" "Displacement" "Horsepower" "Weight" "Acceleration" "Modelyear" "Origin")

	1	2	3	4	5	6	7
1	136.5338	136.5338	136.5338	136.5338	136.5338	136.5338	136.5338
2	45.9019	39.5422	48.8628	31.2209	108.5269	123.4435	89.9703
3	45.0785	33.2435	36.7479	25.9554	105.3313	122.9541	87.0884
4	40.2613	32.7103	36.6578	25.1115	105.0950	122.7115	1.5007e+04
5	40.1616	32.0810	36.4829	25.0208	102.7818	122.0681	7.7543e+03

Testing MSE (Same format as training MSE)

	1	2	3	4	5	6	7
1	155.8924	155.8924	155.8924	155.8924	155.8924	155.8924	155.8924
2	73.5274	69.5172	73.1724	66.2068	131.3637	77.9547	108.5153
3	73.1248	64.1460	60.3076	64.7402	130.9554	59.0355	109.5944
4	68.2405	64.8623	60.1029	67.2976	132.2631	93.4689	8.3636e+03
5	68.5275	66.7271	59.8016	67.9022	128.8050	504.4618	6.0138e+03

Best order:

Cylinder	Displacement	Horsepower	Weight	Acceleration	Modelyear	Origin
3 rd	2 nd	4 th	2 nd	4 th	2 nd	1 st

The most informative feature should be horsepower because it gives the best prediction.

5.

3x1 double		
	1	2
1	40.6889	
2	7.2103	
3	4.3717	

3x1 double		
	1	
1	155.8924	
2	32.7617	
3	19.5974	

6. Solver is in file ECS171Q3.m

q6training_error_percentage 0.2053


q6testing_error_percentage 0.3556

Also, detail information can be found in part1(training) & parttest(testing) matrix
Part1 consist of

- 1- MPG
- 2- Value of first logistic regression for separating low and (median+high)
- 3- Result category of logistic regression (0 means low, 1.5 means above below)
- 4- Value of second logistic regression for separating median and high
- 5- Result category of logistic regression (1 means medium, 2 means high)
- 6- Logical True (1) or False (0) for correctness of categorization



	1	2	3	4	5	6
16	22	0.5122	1.5000	6.8298e-04	1	1
17	18	0.5432	1.5000	0.0010	1	1
18	21	0.7814	1.5000	0.0043	1	1
19	27	0.9703	1.5000	0.0089	1	1
20	26	0.9968	1.5000	0.0777	1	1
21	25	0.7715	1.5000	1.6683e-04	1	1
22	24	0.9049	1.5000	9.6788e-04	1	1
23	25	0.8845	1.5000	0.0013	1	1
24	26	0.8887	1.5000	0.0035	1	1
25	21	0.7207	1.5000	0.0027	1	1
26	10	4.8588e-06	0	0	0	1
27	10	3.2678e-05	0	0	0	1
28	11	2.1832e-05	0	0	0	1
29	9	9.1642e-06	0	0	0	1
30	27	0.9814	1.5000	0.0184	1	1
31	28	0.9503	1.5000	0.0064	1	0
32	25	0.9613	1.5000	0.0092	1	1
33	19	0.7143	1.5000	0.0066	1	1
34	16	0.0885	0	0	0	1
35	17	0.1206	0	0	0	1
36	19	0.1957	0	0	0	0
37	18	0.1600	0	0	0	1
38	14	3.2158e-04	0	0	0	1
39	14	5.4389e-05	0	0	0	1
40	14	5.5167e-04	0	0	0	1

7. Prediction using 2nd order, multivariate linear regression

 Q7polypredict

21.3321

Prediction using 2nd order, multivariate Logistic regression (actually only need first classification, because result is low, don't need to categorize between medium and high. Value of sigmoid is less than 0.5, thus categorize into low)

 Q7sig1
 Q7sig2

0

0

8. For a horse, the intuitive answer to its miles per gallon should be 0 because horse don't consume gasoline. But if we are considering from an energy standpoint, then the energy required by the horse can be transformed in to miles per gallon.

From the online source:

- A standard gallon of gas has 35,249 Joules
- 1000 lb. performance moderate work horse consume 21,200 cal/day

Thus, a horse uses 2.52 gallons gas equivalent energy per day. Assume the horse dragging the car 25 miles per day, then the MPG of this horse should be approximately 9.93.