Homework 1

ECS 171

Ryan Gosiaco 912819444

1. I found the threshold mpg values by first arranging the data in ascending order and then splitting it up into 3 bins: the first two contained 131 rows and the last one contained 130 rows. So the low mpg ranges from 9 mpg to 18.6 mpg, the med mpg is from 19.0 mpg to 27.0 mpg, and the high mpg is from 27 to 46.6 mpg.

2. The pair-wise feature combination that is most informative about the mpg categories to me is the HP and mpg combination. The three categories are nicely spread out and the few stray values are still close to the main group of respective values. This graph is called “scatter_mat.png” if you would like to look at it closer.

3. Refer to problem_3.py

4. The polynomial order that performs the best in the test set is the second order polynomial. The most informative feature for mpg consumption is disp.

Training MSE (cyl): [63.698136, 25.605089623471795, 25.545248930021828, 23.088687911266405]

Testing MSE (cyl): [58.02533066666666, 22.440159695544242, 22.206442279766843, 20.605025119653565]

Training MSE (disp): [63.698136, 22.554331692043753, 20.178285465322134, 33.11448357296655]

Testing MSE (disp): [58.02533066666666, 20.218485959125672, 17.62820165639822, 33.48298401277784]

Training MSE (hp): [63.698136, 25.62015783198681, 20.298383687109762, 30.792615994349912]

Testing MSE (hp): [58.02533066666666, 22.429042628438236, 17.698141617866632, 24.468516442802052]

Training MSE (wt): [63.698136, 20.863533564720637, 38.93458624648545, 94.61082387901538]

Testing MSE (wt): [58.02533066666666, 16.494969637283763, 34.47343065163894, 86.22374611318334]

Training MSE (accel): [63.698136, 51.8424610564452, 51.45147985265667, 51.279267433862806]

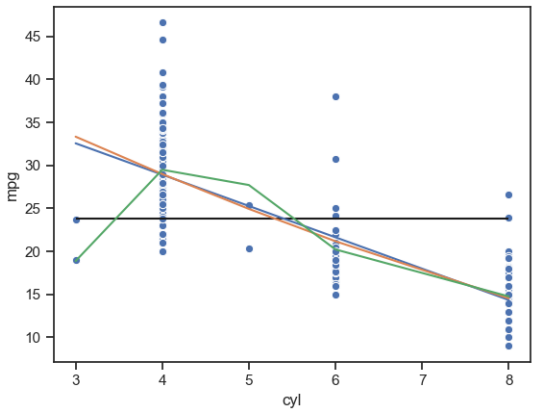
Testing MSE (accel): [58.02533066666666, 48.170212999545875, 47.16872078223728, 47.09455148193545]

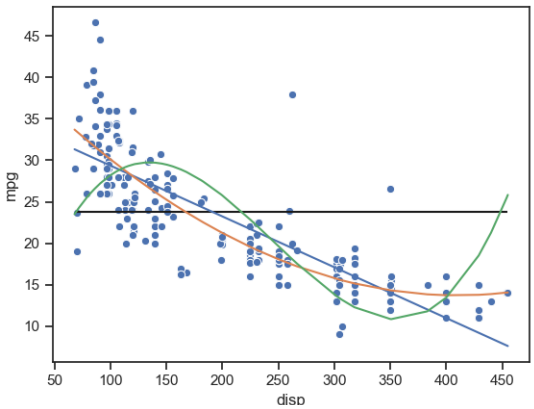
Training MSE (mdl_year): [63.698136, 40.541196645771784, 39.178841897277636, 39.21551882627494]

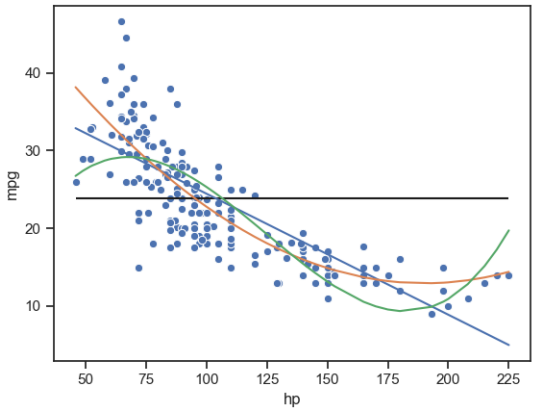
Testing MSE (mdl_year): [58.02533066666666, 40.360324936278424, 37.98214323786516, 38.01030004650418]

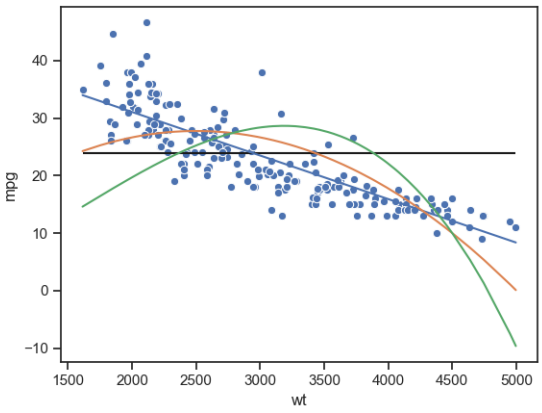
Training MSE (origin): [63.698136, 46.90959932199261, 44.68203621068339, 44.68203621068339]

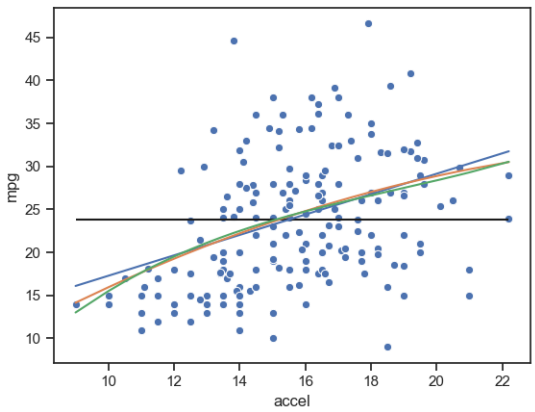
Testing MSE (origin): [58.02533066666666, 36.07267103877044, 37.44670802943601, 37.44670802943447]

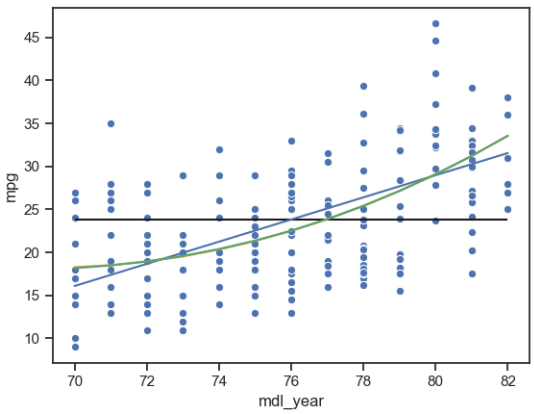

mpg
cyl

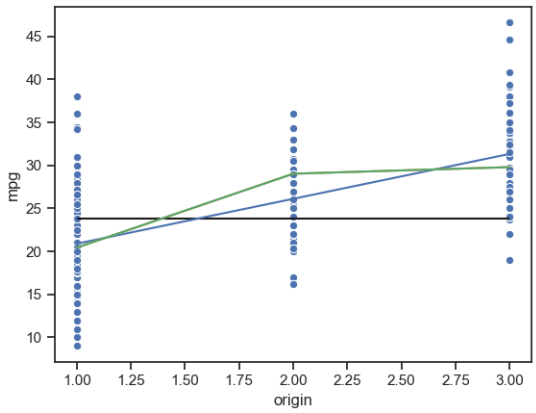

mpg
disp


mpg
hp


mpg
wt


mpg
accel


mpg
mdl_year



5. The equation I used for the 2nd order multi-variate polynomial is $w_0 + w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_7x_7 + w_8x_1^2 + w_9x_2^2 + \cdots + w_{15}x_7^2$.

Training MSE (7 indep): [63.698136, 11.997239747882185, 8.804931157236519]

Testing MSE (7 indep): [58.02533066666666, 10.059978551868028, 8.300906477496087]

6. The training classification precision is 0.945. The testing classification precision is 0.9322916666666666.

7. I would expect this to have a mpg rating of around 9 mpg to 20 mpg and belong in the low mpg category because the characteristics of the car such as 6 cylinders and 3700 lb weight point to a heavier car that is not as efficient. Using the second order multi-variate polynomial and logistic regression, I predicted that the mpg rating of this car would be 18.450269 and the predicted mpg category would be 'low'.

8. Zooming in on the photo shows that this vehicle is being powered by a horse or similar animal. I believe this vehicle would have an mpg rating of around 30 mpg because the amount of energy contained in one gallon of fuel is quite large and the equivalent amount in food for the horse would result in a very full horse.