Name - Sairamvinay Vijayaraghavan
Student ID - 913603345
ECS 171 Fall 2019

HW1 Report

1. The data set was cleaned initially by removing all the missing value records and the size of the data set was 392 samples and 9 features. The bins were separated equally into 4 equally sized, each bin of size 98 samples. The thresholds are recorded as follows:

   minimum value is 9.000000
   first quartile maximum threshold: 17.0
   second quartile maximum threshold: 22.5
   third quartile maximum threshold: 29.0
   maximum value is 46.600000

   Therefore,

   First bin has mpg values between 9 and 17
   Second bin has mpg values between 17 and 22.5
   Third bin has mpg values between 22.5 and 29
   Fourth bin has mpg values between 29 and 46.6

   The bin size was checked and it showed each bin to be of equal size = 98 samples.

2. The scatter matrix plot was obtained and illustrated as follows. From the graph obtained, the most informative graph can be **weight vs horsepower.** This is because we can see how a regressive curve could be used to separate the classification.
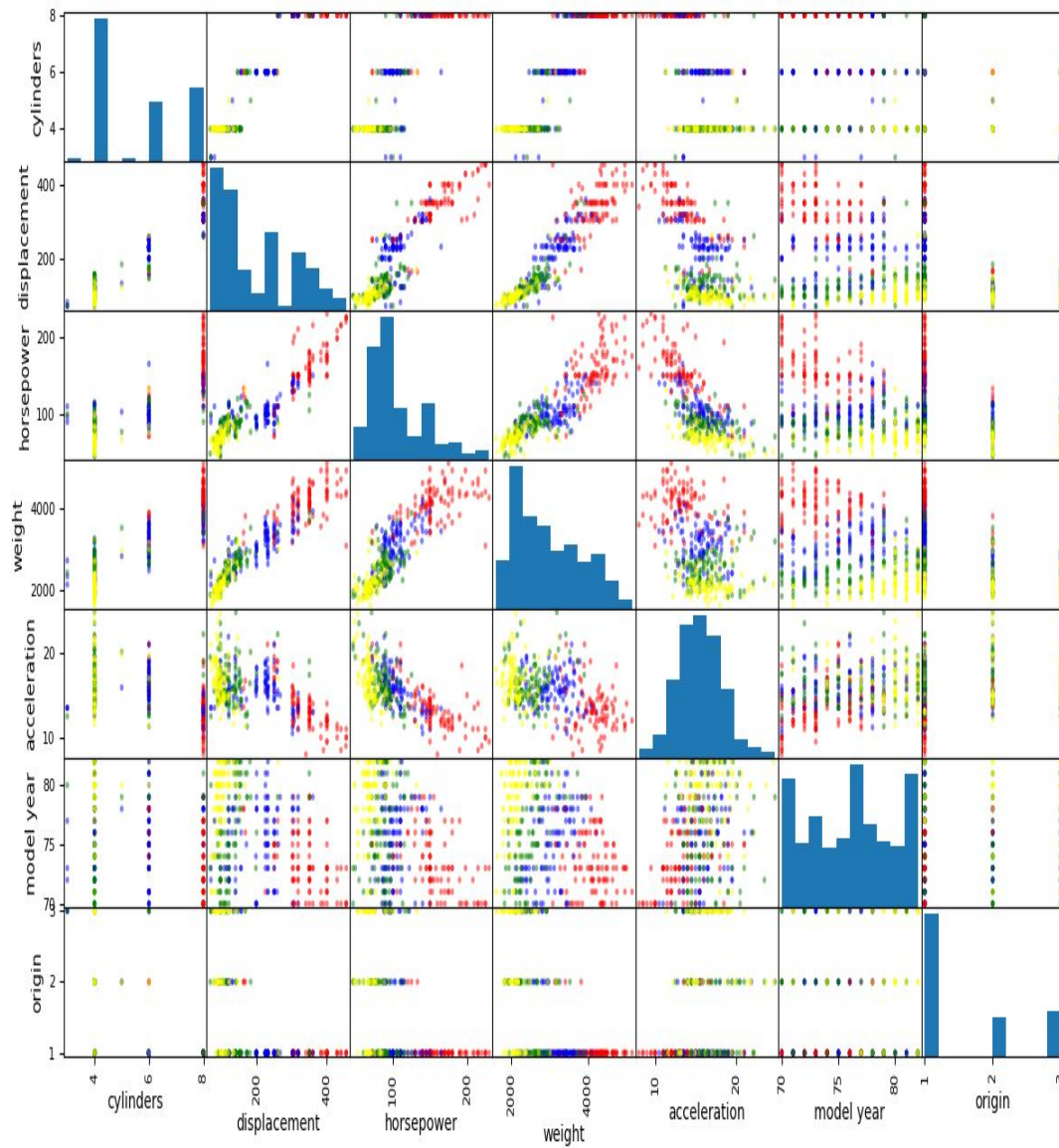
3. Please refer to the code.
4. Here are the reports for each of the feature and degree independently regressed:

Feature: Cylinders

The training Mean square error for the 0th order solver is 59.48581910302119
The testing Mean square error for the 0th order solver is 64.53554802026645

The training Mean square error for the 1th order solver is 24.451476497376085
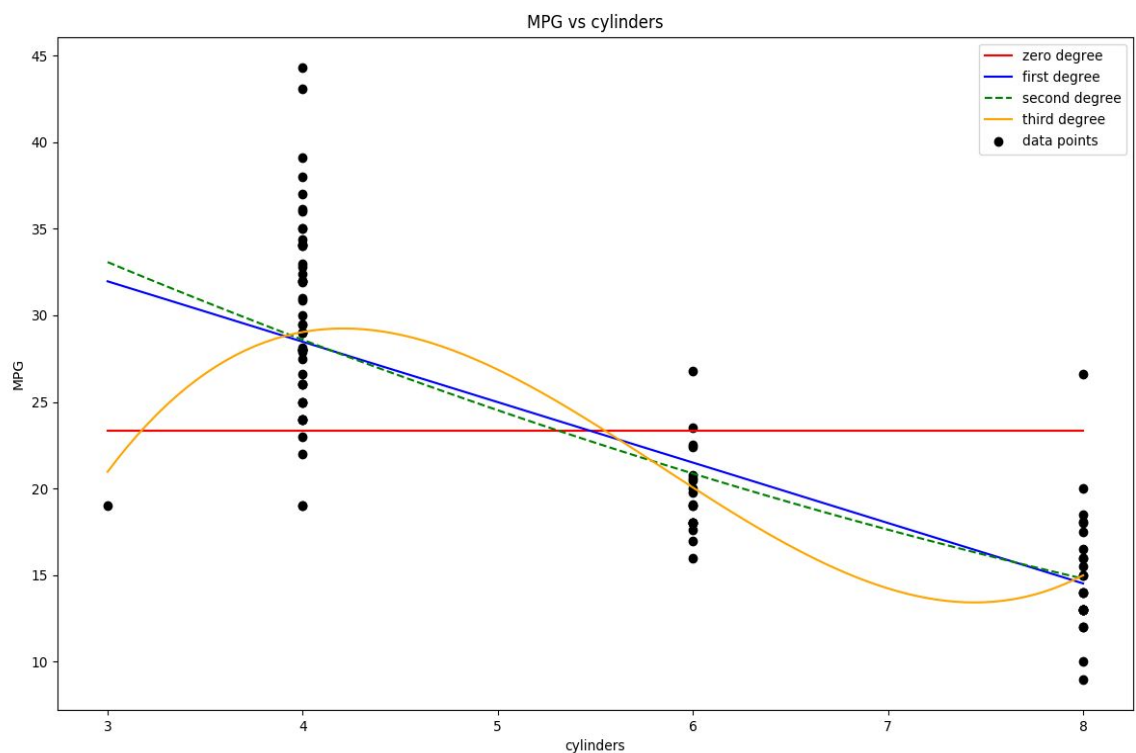The testing Mean square error for the 1th order solver is 22.85869953189948

The training Mean square error for the 2th order solver is 24.323460057053637
The testing Mean square error for the 2th order solver is 22.583307491756468

The training Mean square error for the 3th order solver is 22.511843946250337
The testing Mean square error for the 3th order solver is 19.8543220440778

BEST: The best polynomial solver is degree 3 (least test MSE).

MPG vs cylinders

Feature: displacement

The training Mean square error for the 0th order solver is 59.48581910302119
The testing Mean square error for the 0th order solver is 64.53554802026645

The training Mean square error for the 1th order solver is 21.662820868102283
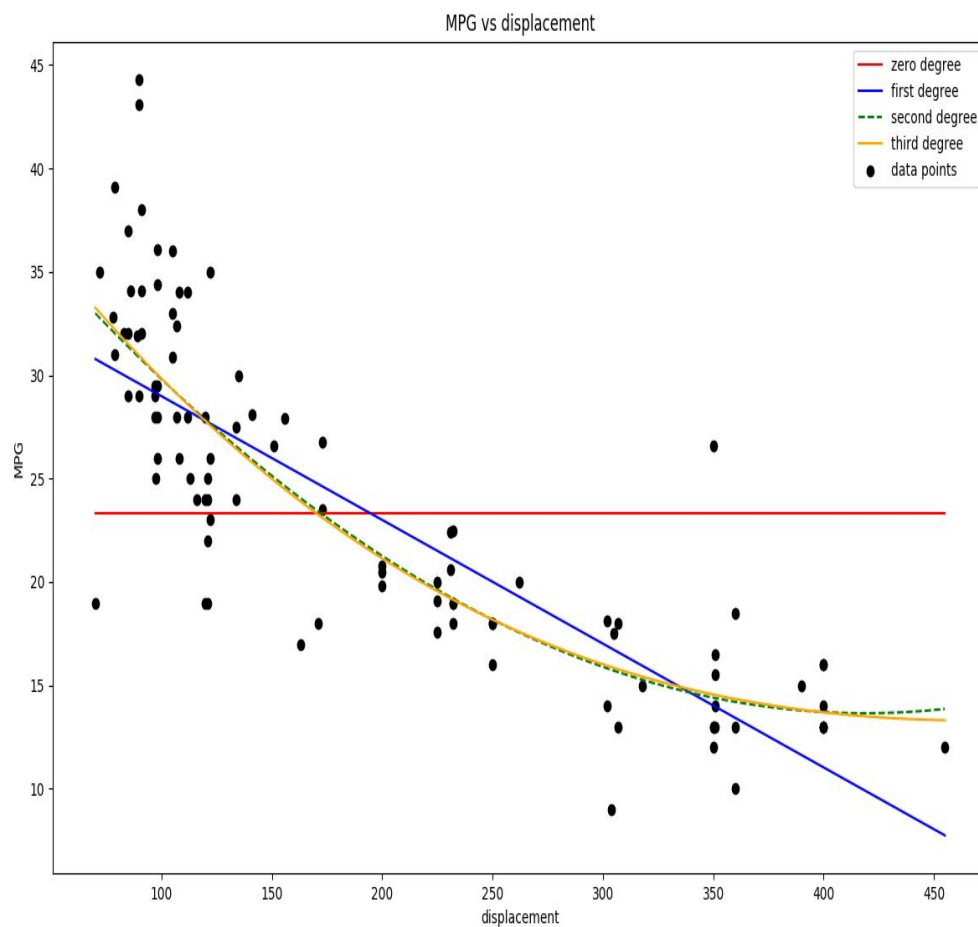The testing Mean square error for the 1th order solver is 20.58072315410922

The training Mean square error for the 2th order solver is 19.440143603799203
The testing Mean square error for the 2th order solver is 17.40263593794041

The training Mean square error for the 3th order solver is 19.423418449094815
The testing Mean square error for the 3th order solver is 17.28472652167644

BEST: The best polynomial solver is degree 3 (least test MSE).

MPG vs displacement

Feature: horsepower

The training Mean square error for the 0th order solver is 59.48581910302119
The testing Mean square error for the 0th order solver is 64.53554802026645

The training Mean square error for the 1th order solver is 24.640751142237306
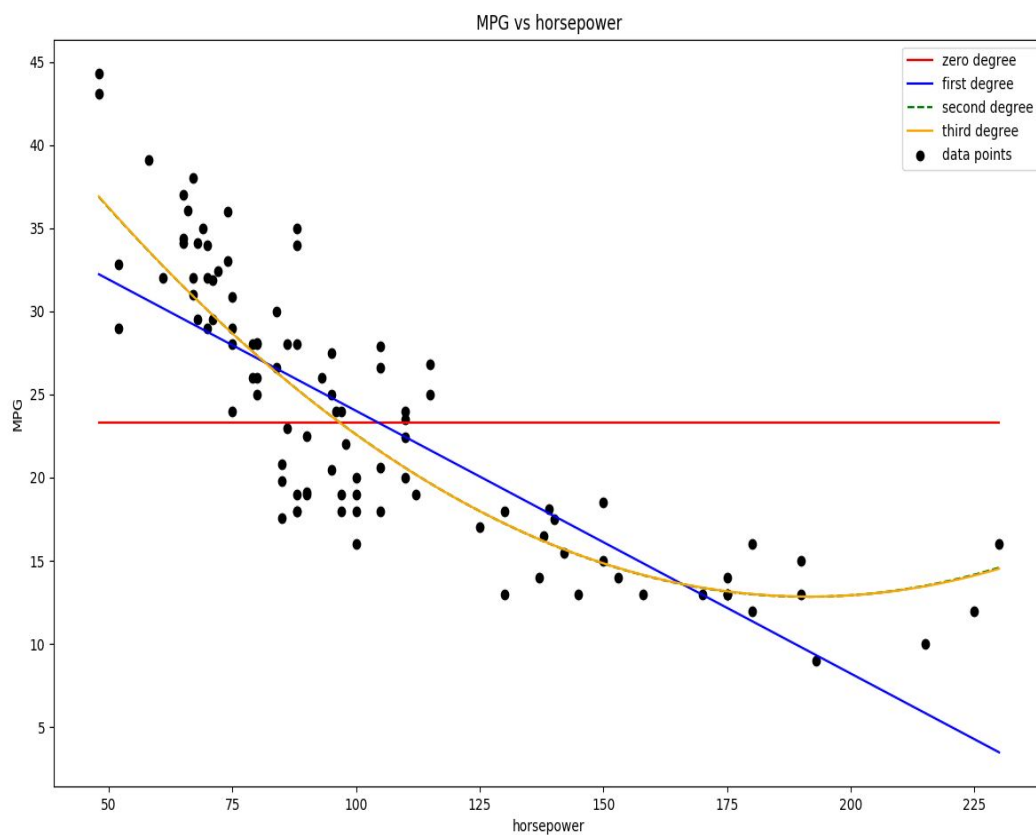The testing Mean square error for the 1th order solver is 21.981890488957607

The training Mean square error for the 2th order solver is 20.471138611034814
The testing Mean square error for the 2th order solver is 14.762714224610585

The training Mean square error for the 3th order solver is 20.471021018060934
The testing Mean square error for the 3th order solver is 14.745719977817302

BEST: The best polynomial solver is degree 3 (least test MSE).

Feature: weight

The training Mean square error for the 0th order solver is 59.48581910302119
The testing Mean square error for the 0th order solver is 64.53554802026645

The training Mean square error for the 1th order solver is 19.87003215547306
The testing Mean square error for the 1th order solver is 15.23557094967358
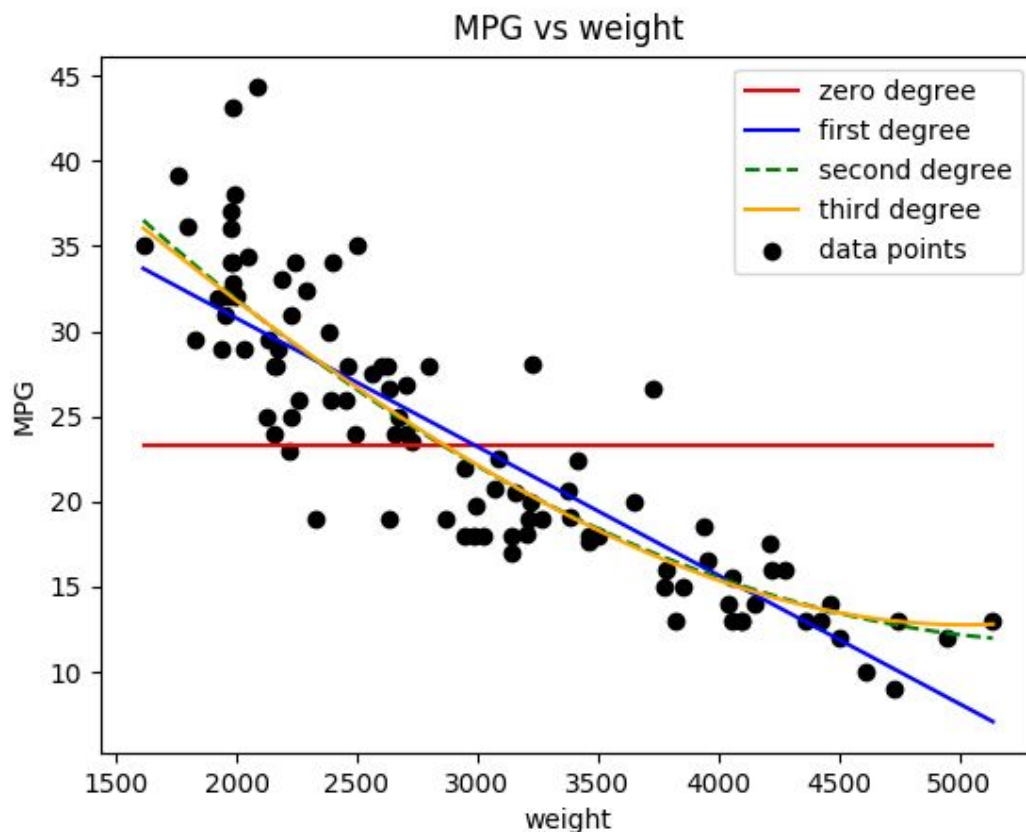
The training Mean square error for the 2th order solver is 18.715920961618533
The testing Mean square error for the 2th order solver is 13.249529723819727

The training Mean square error for the 3th order solver is 18.7021451059892
The testing Mean square error for the 3th order solver is 13.344201423438053

BEST: The best polynomial solver is degree 2 (least test MSE)

MPG vs weight

Feature: acceleration

The training Mean square error for the 0th order solver is 59.48581910302119
The testing Mean square error for the 0th order solver is 64.53554802026645

The training Mean square error for the 1th order solver is 50.30389289665868
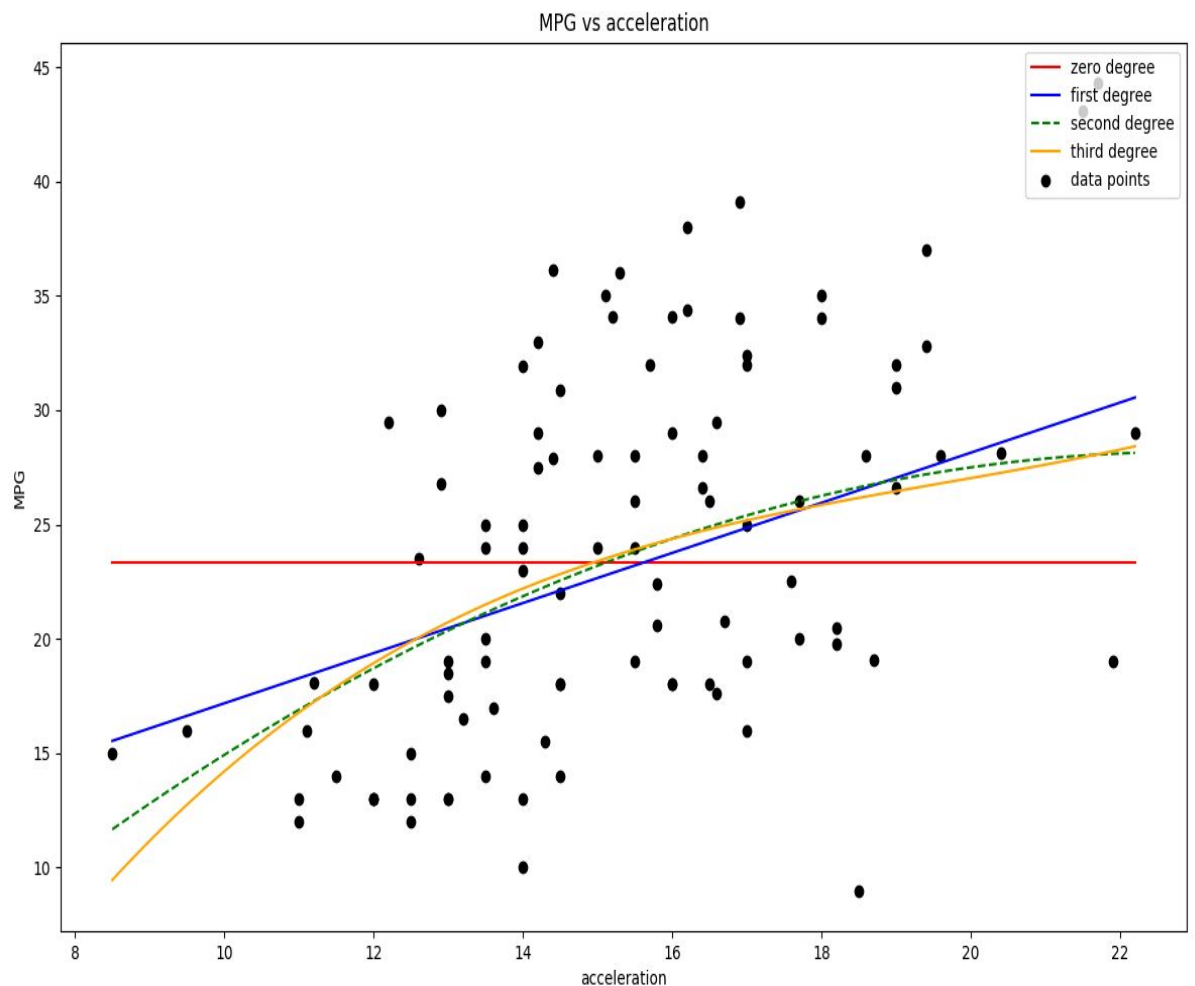The testing Mean square error for the 1th order solver is 49.06910136524679

The training Mean square error for the 2th order solver is 49.38576545575227
The testing Mean square error for the 2th order solver is 48.17149999939908

The training Mean square error for the 3th order solver is 49.169369038745586
The testing Mean square error for the 3th order solver is 48.55602952086114

BEST: The best polynomial solver is degree 2 (least test MSE)

MPG vs acceleration

Feature: model year

The training Mean square error for the 0th order solver is 59.48581910302119
The testing Mean square error for the 0th order solver is 64.53554802026645

The training Mean square error for the 1th order solver is 39.90508369380596
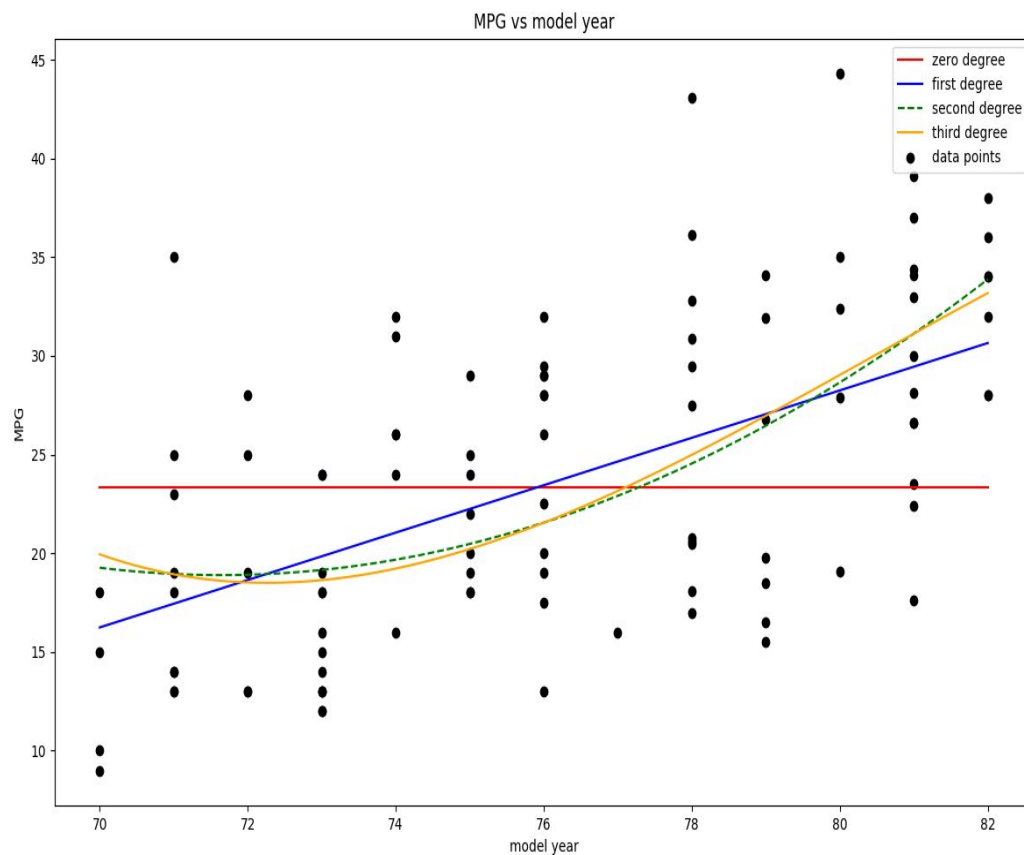The testing Mean square error for the 1th order solver is 41.43822914897107

The training Mean square error for the 2th order solver is 36.85847814251237
The testing Mean square error for the 2th order solver is 43.53799465432915
The training Mean square error for the 3th order solver is 36.672689817101315
The testing Mean square error for the 3th order solver is 43.87281689956256

BEST: The best polynomial solver is degree 1 (least test MSE)

Feature: origin

The training Mean square error for the 0th order solver is 59.48581910302119
The testing Mean square error for the 0th order solver is 64.53554802026645

The training Mean square error for the 1th order solver is 41.114905063187244
The testing Mean square error for the 1th order solver is 42.076645237015
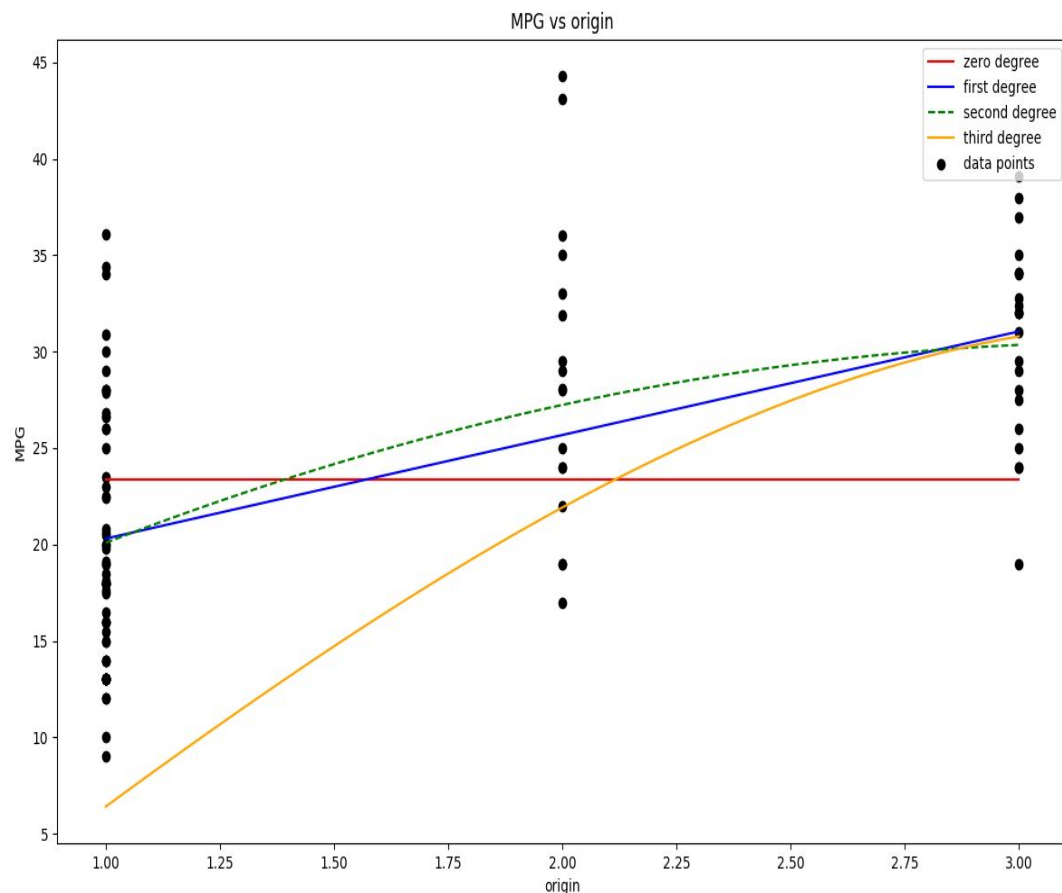
The training Mean square error for the 2th order solver is 40.565102796580845
The testing Mean square error for the 2th order solver is 40.802454638704134

The training Mean square error for the 3th order solver is 163.53842113032866
The testing Mean square error for the 3th order solver is 158.8688145056166

BEST: The best polynomial solver is degree 2 (least test MSE)

Most informative feature for mpg consumption is least error feature: which is **weight.**

5. The results obtained from regressing against all the features at once are:

   For regression against all features

   The training Mean square error for the 0th order solver is 59.48581910302119
   The testing Mean square error for the 0th order solver is 64.53554802026645

   The training Mean square error for the 1th order solver is 11.199890587155481
   The testing Mean square error for the 1th order solver is 10.263709230784107

   The training Mean square error for the 2th order solver is 7.495914671960289
   The testing Mean square error for the 2th order solver is 7.197244443775177

6. The accuracy (the precision_score as well as the score() method for Logistic
   Regression) for training and testing data are recorded as follows:

   Logistic Regression Algorithm results
   Training Accuracy is 74.657534%
   Testing Accuracy is 83.000000%

7. After min-max normalization using sklearn's MinMaxScaler(), the results obtained
   are:

   Logistic Regression Algorithm results
   Training Accuracy is 71.575342%
   Testing Accuracy is 73.000000%

   Yes, there is a decrease in both train and test data accuracies.

8. For the sample data with features of 4 cylinders, 400 cc displacement, 150
   Horsepower, 3500 lb weight, 8 m/s^2 acceleration and 1981 year (= 81) and origin 1
   model is tested using both Linear Regression solver and Logistic Regression trained
   on the entire 392 samples of the data set, the following results are obtained:

   MPG Pred is **23.802693**
   Cat Pred is **'high'**