Name: Andrew Wu
Github Username: AndrewWu20
Purdue Username: wu1795
Instructor: Qiu

Problem1_writeup.

(1) Estimated Functions:

$$\hat{y}_1(x) = 29.05867495x + 92.76756053$$

$$\hat{y}_2(x) = -2.11108454x^2 + 28.50662487x + 112.31481224$$

$$\hat{y}_3(x) = 1.75743661x^3 + -1.43242754x^2 + -0.3307411x + 101.86611055$$

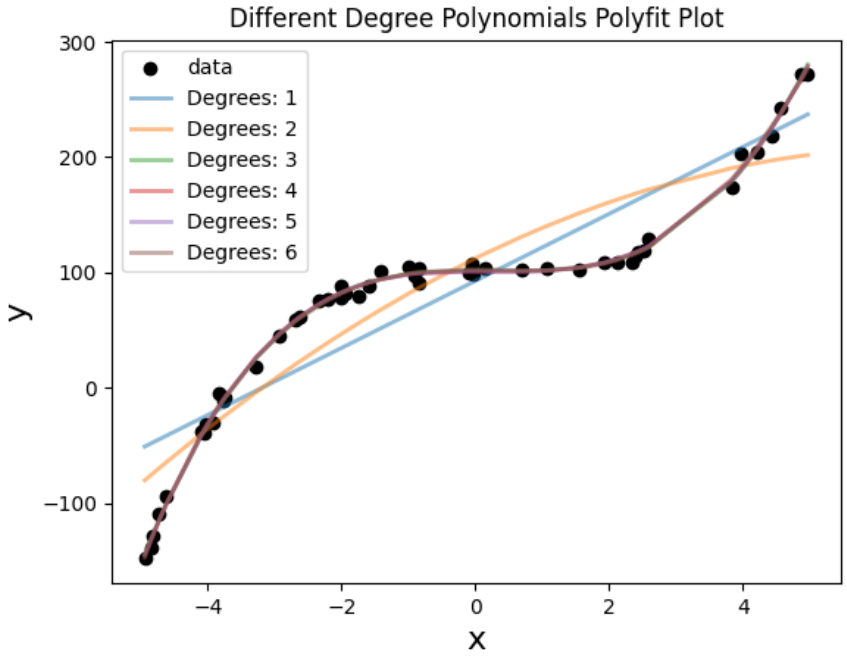
$$\hat{y}_4(x) = -1.51249835e-02x^4 + 1.75412364e+00x^3 + -1.08212257e+00x^2 + -2.55843975e-01x + 1.00914532e+02$$

$$\hat{y}_5(x) = -4.45092599e-04x^5 + -1.54226284e-02x^4 + 1.76681929e+00x^3 + -1.07434416e+00x^2 + -3.22742703e-01x + 1.00887487e+02$$

$$\hat{y}_6(x) = -3.98888854e-03x^6 + -1.29035900e-03x^5 + 1.26647021e-01x^4 + 1.79048310e+00x^3 + -2.33277278e+00x^2 + -4.25298673e-01x + 1.02701524e+02$$

(2) Data Visualization:

(insert plot obtained from data in poly.txt)



(3) What degree polynomial does the relationship seem to follow? Please explain your answer.

The data seems to best follow a third order polynomial which can be seen from the low error between the estimated regression function, $\widehat{y_3}(x)$ and the data in the plot above. The reason why the third order polynomial was chosen instead of the fourth, fifth, or sixth order polynomial was because we want to minimize complexity of the model and maintain the best fit possible to choose the best polyfit for the data.

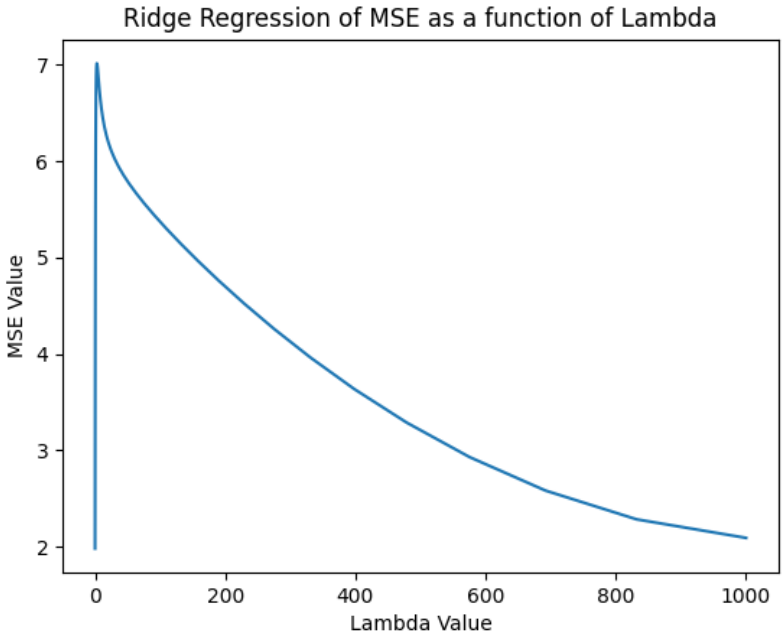
(4) If we measured a new data point, $x = 2$, what would be the predicted value of y, based on the polynomial identified as the best fit in Question (3)?

Sample answer:

If we measured a new data point, x=2, the corresponding predicted value would be $\widehat{y_3}(2) = 109.5344110539675$.

Problem2_writeup.

(1) Plot the mean squared error as a function of lambda in Ridge Regression:



(2) Find best lambda:

Based on the range of Lambda values tested, the best lambda value is 0.1, which yields an MSE of 1.9815144074968696 as shown on the plot above.

(3) Find equation of the best fitted model:

(*Insert numerical values for a_i's and b)*

$$\hat{y}(x) = -4.33992630e-01x_1 + 8.16204762e-01x_2 + 5.19495066e-01x_3 + 3.83342192e+00x_4 + 2.11359089e-01x_5 + 4.53719310e-04x_6 + 2.5617645349536864$$

(4) Draw a prediction plot using Google data

(Note that the plot below is not the solution)