

Name: Connor Damato
Github Username: Connor-Damato
Purdue Username: 0033498677 (damato0)
Instructor: Sahay

Problem1_writeup.

(1) Estimated Functions:

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

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

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

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

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

(2) Data Visualization:

(insert plot obtained from data in poly.txt)

Figure 1

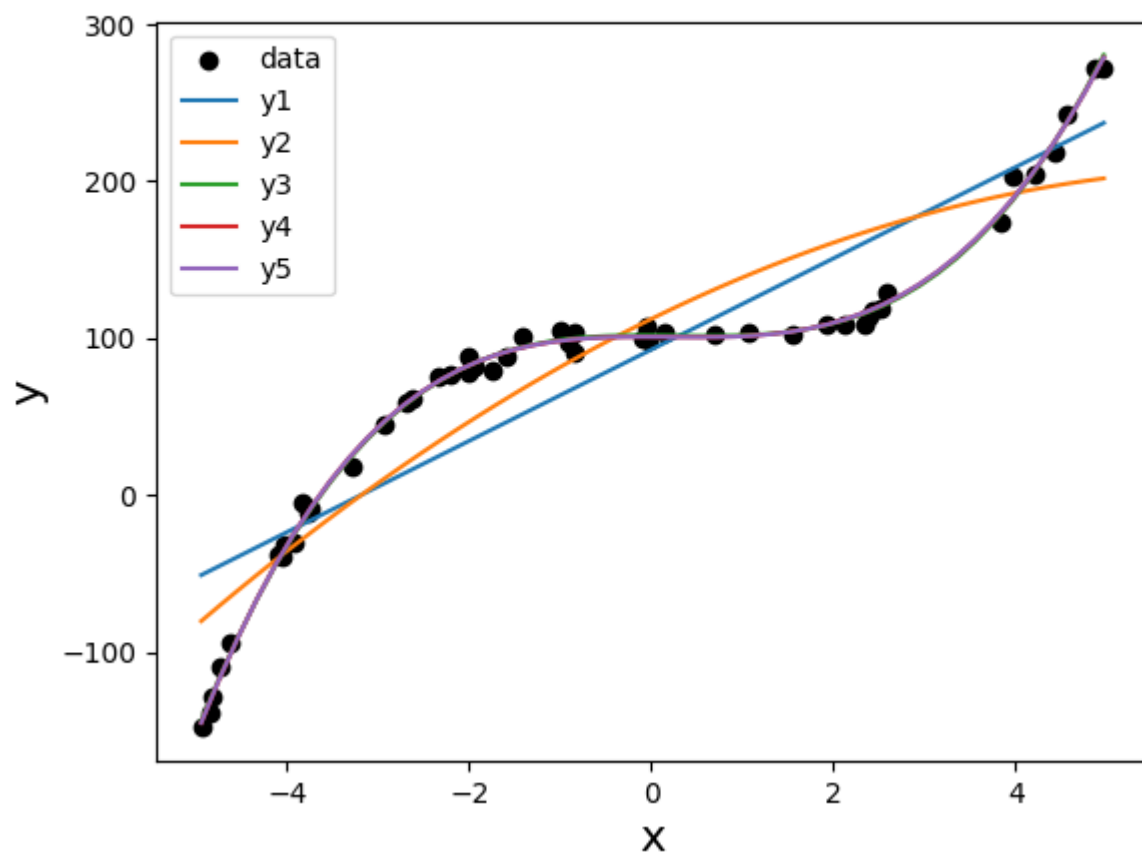
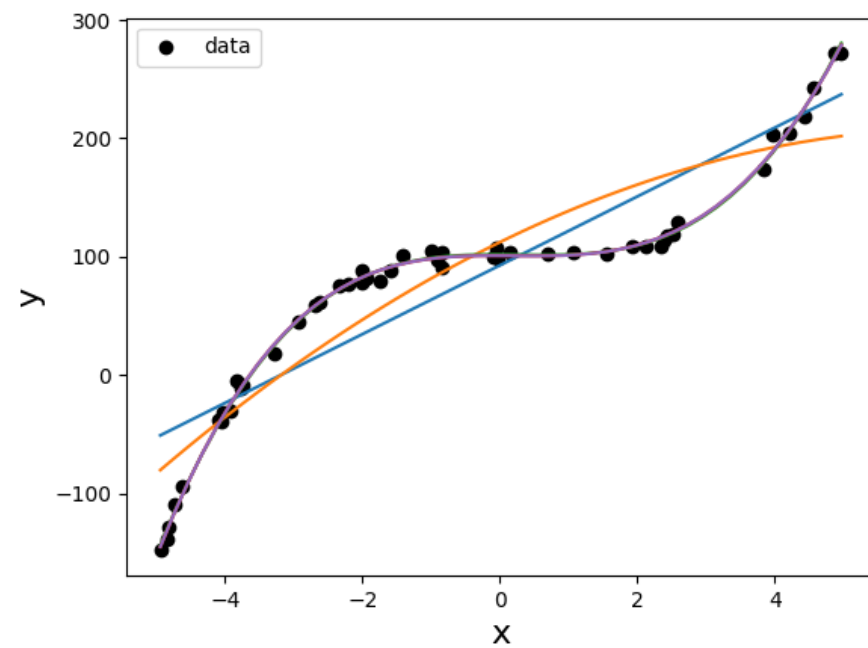


Figure 1



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

(Discuss relationship of data and insert numerical value of c calculated from best regression)

Sample answer:

The data seems to follow a third order polynomial which can be seen from the data plot above. Y3, Y4, and Y5 all have the same equation, meaning the equation with the best fit without overfitting would be Y3, a third order polynomial.

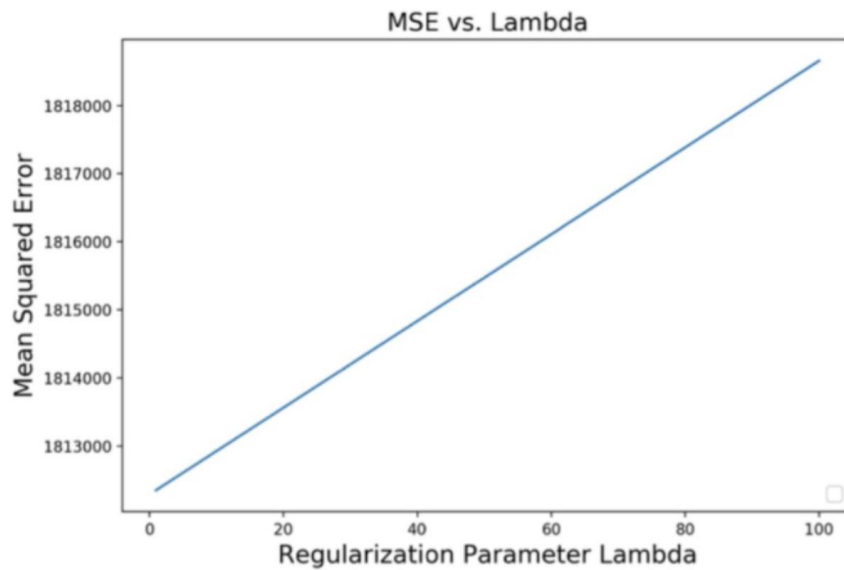
(4) If we measured a new data point, $x = 3$, what would be the predicted value of y , based on the polynomial identified as the best fit in Question (3)?

If we measured a new data point, $x=3$, the corresponding predicted value would be, $\widehat{y_3(3)} = 146.0894734$.

Problem2_writeup.

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

*(Insert plot obtained by completing the **main** function)*



(2) Find best lambda:

Sample answer:

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

(3) Find equation of the best fitted model:

(Insert numerical values for a_i 's and b)

$$\hat{y}(x) = a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + a_5 x_5 + a_6 x_6 + b$$

(4) Draw a prediction plot using Google data

(Note that the plot below is not the solution)

