

Name: Calvin Madsen
Github Username: MadSons
Purdue Username: 31132999
Instructor: Qiu

Problem1_writeup.

(1) Estimated Functions:

$$\widehat{y_1}(x) = 29.0587x + 92.7676$$

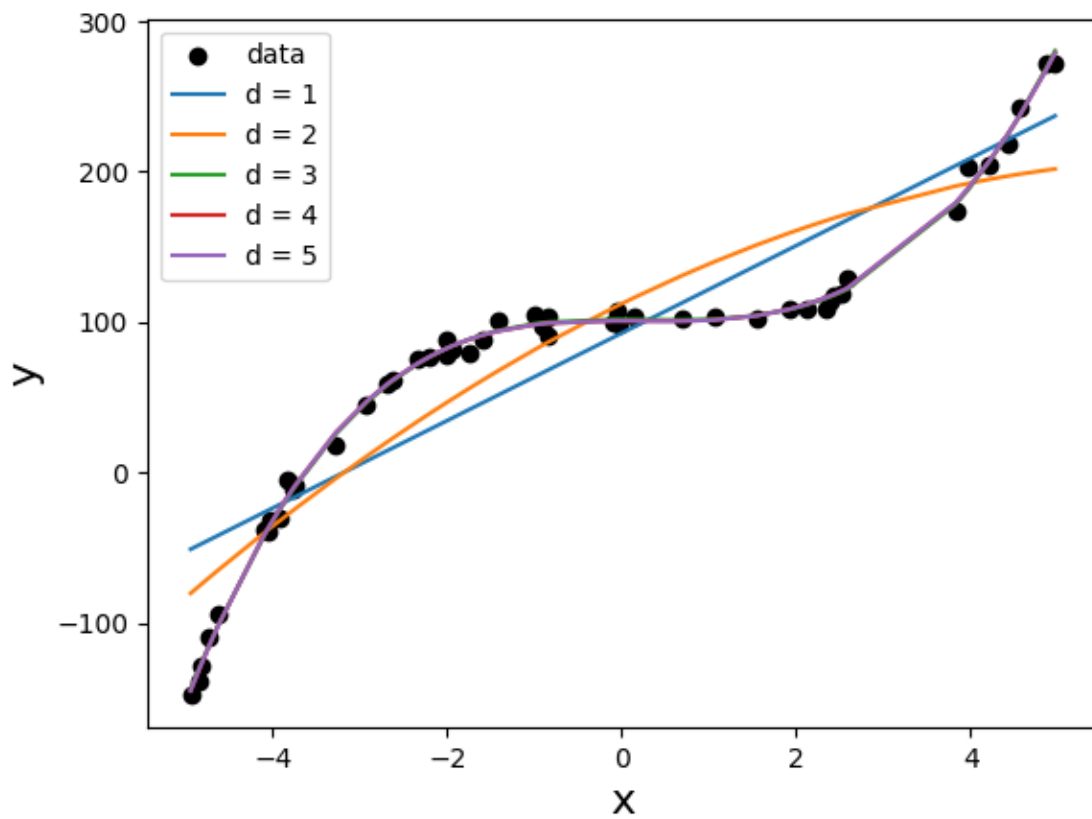
$$\widehat{y_2}(x) = -2.1111x^2 + 28.5066x + 112.3148$$

$$\widehat{y_3}(x) = 1.7574x^3 + -1.4324x^2 + -0.3307x + 101.8661$$

$$\widehat{y_4}(x) = -0.0151x^4 + 1.7541x^3 + -1.0821x^2 + -0.2558x + 100.9145$$

$$\widehat{y_5}(x) = -0.00045x^5 + -0.0154x^4 + 1.7668x^3 + -1.0743x^2 + -0.3227x + 100.8875$$

(2) Data Visualization:



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

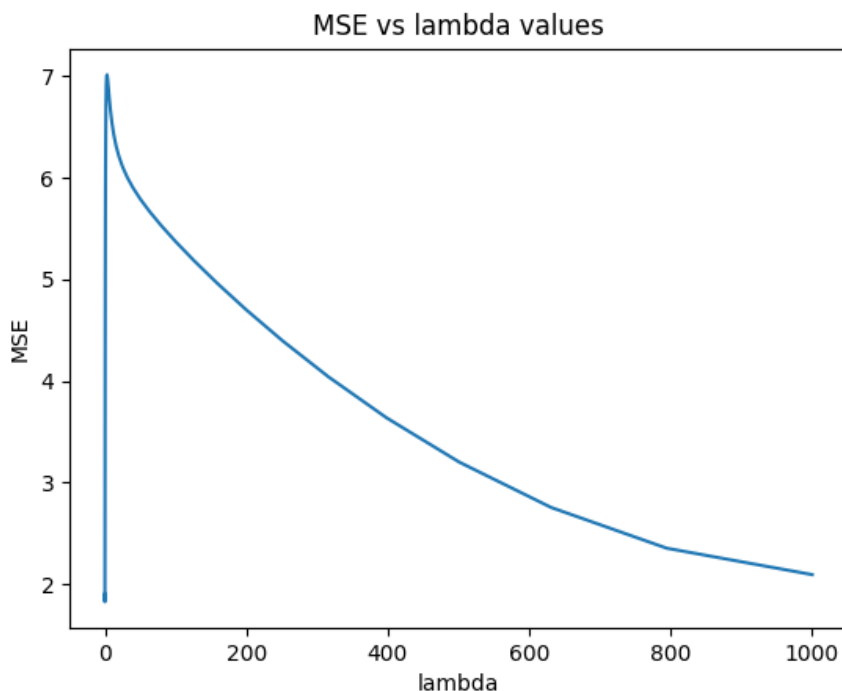
The data seems to best follow a third order polynomial (i.e., a cubic) which can be seen from the low error between the estimated regression function, $\widehat{y_3(x)}$, and the data in the plot above. Degree 4 and degree 5 polynomials show a very similar plot but become no more accurate than the degree 3 polynomial, making the degree 3 polynomial the best option to reduce complexity.

(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, $\widehat{y_3(3)} = 135.43283$

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 0.039810717055349734, which yields an MSE of 1.8240068113322172 as shown on the plot above.

(3) Find equation of the best fitted model:

$$\hat{y}(x) = -0.3861x_1 + 0.6665x_2 + 0.3732x_3 + 4.2754x_4 \\ + 0.0176x_5 + 0.0002x_6 + 2.5618$$

(4) Draw a prediction plot using Google data

