Name: Bogomir Glavan
Github Username: mglavan07
Purdue Username: 00360-54905 (bglavan@purdue.edu)
Instructor: Qiu

Problem 1.

Estimated Functions:

$\hat{y}_1(x) = 27.406798636142426x + 76.54862254204512$

$\hat{y}_2(x) = -1.2648866448486866x^2 + 27.02773666903971x +$
88.44135382520403

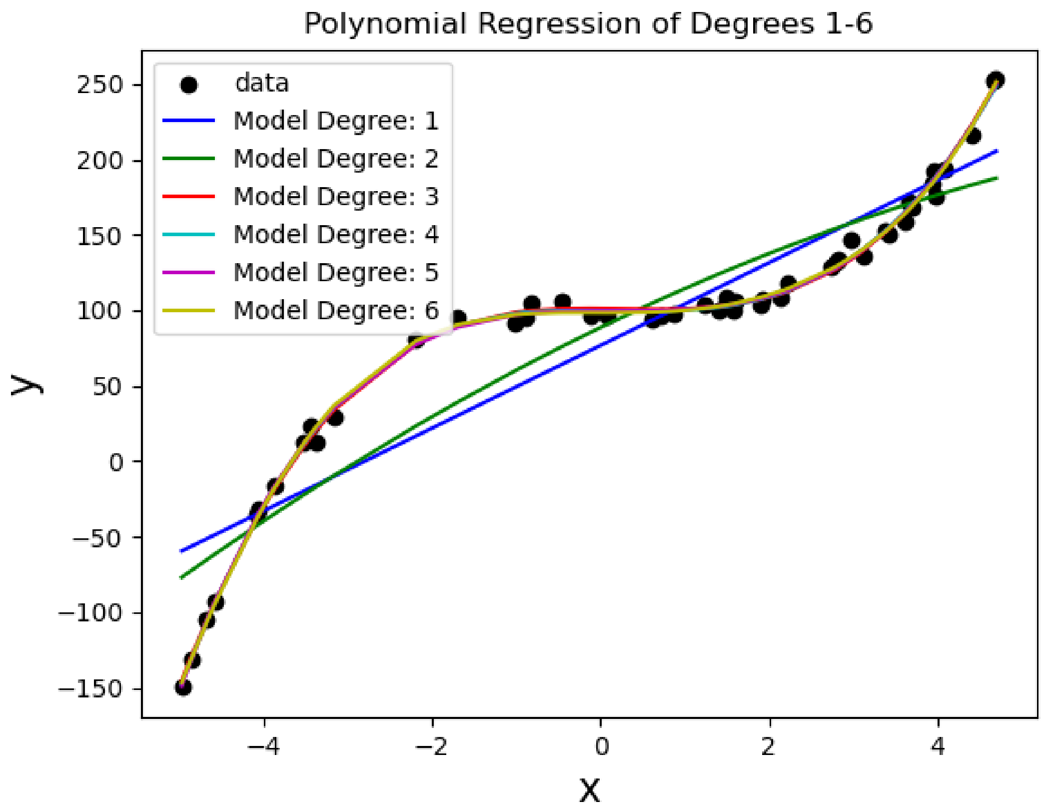
$\hat{y}_3(x) = 1.762383193944788x^3 - 1.3682572662308559x^2 -$
$0.6498400559683155x + 101.16436653650119$

$\hat{y}_4(x) = -0.022514493706442817x^4 + 1.755887097697185x^3$
$-0.8882893402084138x^2 - 0.6518810852880286x + 99.92393813058679$

$\hat{y}_5(x) = 0.005007633185103513x^5 - 0.020435144478004708x^4$
$+1.61451848602943x^3 - 0.899320700489445x^2 + 0.1705612875646505x$
$+99.7076810284182$

$\hat{y}_6(x) = 0.004056621447540954x^6 + 0.005285210041171565x^5 -$
$0.16206024287973864x^4 + 1.6376473758417345x^3 +$
$0.3353859398703367x^2 - 0.23809253411224773x + 98.3073297819154$

(2) <u>Data Visualization:</u>



---

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

(Discuss relationship of data and state a numerical value for the best-fitting polynomial degree)

The data seems to best fit a third-degree (degree = 3) polynomial (a cubic function). This can be demonstrated by the low error, seen visual by the residuals in the scatter and line plot above. However, it can be noticed that the third, fourth, fifth, and sixth polynomials all follow similar trends. This further suggests that $\hat{y}_3(x)$ is the best fit for the data, as the coefficients for x^4, x^5, and x^6 are very close to 0, therefore, just "overfitting" the data that is best left as a cubic regression approximation.

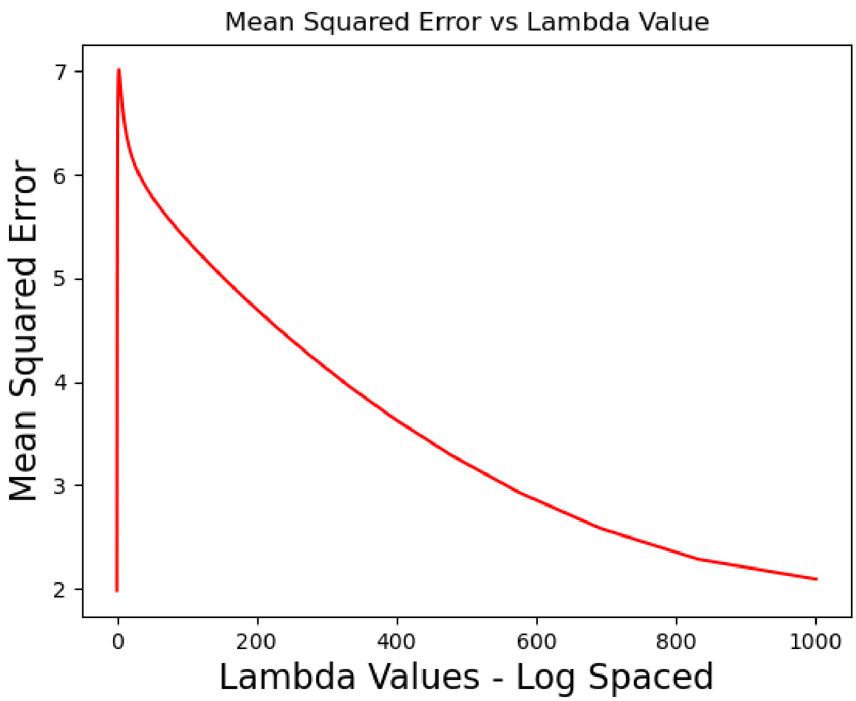
(4) <u>If we measured a new data point, $x = 2$, what would be the predicted value of y, based on the polynomial identified as the best fit in Question (3)?</u>

If we measured a new data point, x = 2, the corresponding predicted value would be $\hat{y}_3(2) = 108.49072291119944$

(Python Output: y_hat(2) at Degree = 3 Regression is: 108.49072291119944)

Problem 2.

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



(2) <u>Find best lambda:</u>

Based on the range of Lambda values tested, the best lambda value is 0.1, which yields an MSE of 1.981514407475458 as shown on the plot above.

(Python Output: Best lambda tested is 0.1, which yields an MSE of 1.981514407475458)

(3) Find equation of the best fitted model:

$$\hat{y}(x) = -0.43399263018209167x_1 + 0.8162047620779096x_2 + 0.519495066516436x_3 + 3.8334219221012993x_4 + 0.21135908932280634x_5 + 0.00045371931035987 44x_6 + 2.5617645349536864$$

(4) Plot the predicted stock prices and actual stock prices using Google data