CPEN 355 Homework 1

Aloysio Kouzak Campos da Paz
Student ID 58687526

Collaboration statement I discussed the assignment with the student Leonardo Dayrell, and I completed the assignment with the help of Matlab for matrix multiplication, Python for plotting and verifying results and ChatGPT for understanding the concepts behind the assignment and for generating Python code.

Here is a link to the chat I had with ChatGPT:
https://chat.openai.com/share/58b549e7-808f-4786-9d9f-2d574b9f6815

1 Linear Regression

We will use the following equation for solving part a, b and c of question 1:

$$\Theta^* = (X^T X)^{-1} X^T Y \tag{1}$$

This equation allows us to find the optimal parameters of the linear functions we want to fit in part a, b, and c.

1.1 Question 1 a

We will fit the training data with the model:

$$f_\theta(x) = \theta_0 + \theta_1 x \tag{2}$$

We begin by populating the matrix $\mathbf{X}$, $\mathbf{X^T}$, and $\mathbf{Y}$. We will use the training data in $\mathbf{Y}$.

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1 & 1.65 \\ 1 & 2.15 \\ 1 & 1.81 \\ 1 & 1.63 \\ 1 & 1.27 \\ 1 & 1.94 \\ 1 & 1.31 \\ 1 & 2.68 \\ 1 & 2.89 \\ 1 & 1.15 \end{bmatrix}$$

$$\mathbf{X^T} = \begin{bmatrix} 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\ 1.6500 & 2.1500 & 1.8100 & 1.6300 & 1.2700 & 1.9400 & 1.3100 & 2.6800 & 2.8900 & 1.1500 \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} 3.78 \\ 6.70 \\ 4.69 \\ 3.72 \\ 2.40 \\ 5.25 \\ 2.80 \\ 9.85 \\ 11.67 \\ 1.70 \end{bmatrix}$$

$$(\mathbf{X^T X})^{-1} = \begin{bmatrix} 1.2100 & -0.6007 \\ -0.6007 & 0.3250 \end{bmatrix}$$

$$\mathbf{X^T Y} = \begin{bmatrix} 52.5600 \\ 114.1748 \end{bmatrix}$$

$$\mathbf{\Theta}^* = (\mathbf{X^T X})^{-1}\mathbf{X^T Y} = \begin{bmatrix} -4.9818 \\ 5.5399 \end{bmatrix}$$

Therefore we obtain the model:

$$f_\theta(x) = -4.9818 + 5.5399 * x \tag{3}$$

Now we plot the model vs the training data we used to train it:

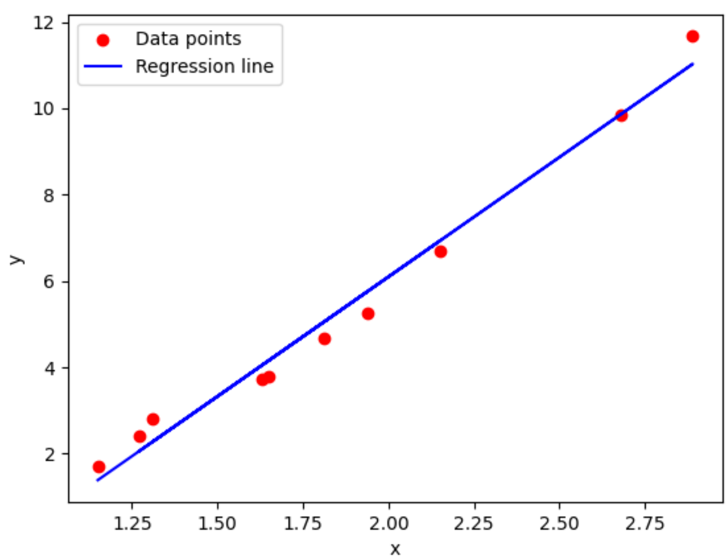


Figure 1: Plotting model $f_\theta(x) = -4.9818 + 5.5399 * x$ against training data

1.2 Question 1 b

We want to increase the model complexity, by considering y as a linear function of both x and $\mathbf{x^2}$. Namely :

$$f_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 \tag{4}$$

We begin by populating the matrix $\mathbf{X}$, $\mathbf{X^T}$, and $\mathbf{Y}$:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1 & 1.65 & 1.65^2 \\ 1 & 2.15 & 2.15^2 \\ 1 & 1.81 & 1.81^2 \\ 1 & 1.63 & 1.63^2 \\ 1 & 1.27 & 1.27^2 \\ 1 & 1.94 & 1.94^2 \\ 1 & 1.31 & 1.31^2 \\ 1 & 2.68 & 2.68^2 \\ 1 & 2.89 & 2.89^2 \\ 1 & 1.15 & 1.15^2 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1.0000 & 1.6500 & 2.7225 \\ 1.0000 & 2.1500 & 4.6225 \\ 1.0000 & 1.8100 & 3.2761 \\ 1.0000 & 1.6300 & 2.6569 \\ 1.0000 & 1.2700 & 1.6129 \\ 1.0000 & 1.9400 & 3.7636 \\ 1.0000 & 1.3100 & 1.7161 \\ 1.0000 & 2.6800 & 7.1824 \\ 1.0000 & 2.8900 & 8.3521 \\ 1.0000 & 1.1500 & 1.3325 \end{bmatrix}$$

$$\mathbf{X^T} = \begin{bmatrix} 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\ 1.6500 & 2.1500 & 1.8100 & 1.6300 & 1.2700 & 1.9400 & 1.3100 & 2.6800 & 2.8900 & 1.1500 \\ 2.7225 & 4.6225 & 3.2761 & 2.6569 & 1.6129 & 3.7636 & 1.7161 & 7.1824 & 8.3521 & 1.3325 \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} 3.78 \\ 6.70 \\ 4.69 \\ 3.72 \\ 2.40 \\ 5.25 \\ 2.80 \\ 9.85 \\ 11.67 \\ 1.70 \end{bmatrix}$$

Now we multiply the matrices to find the optimal parameters :

$$\mathbf{X^T X} = \begin{bmatrix} 10.0000 & 18.4800 & 37.2276 \\ 18.4800 & 37.2276 & 81.1961 \\ 37.2276 & 81.1961 & 189.3760 \end{bmatrix}$$

$$(\mathbf{X^T X})^{-\mathbf{1}} = \begin{bmatrix} 18.2877 & -19.0768 & 4.5843 \\ -19.0768 & 20.3141 & -4.9597 \\ 4.5843 & -4.9597 & 1.2306 \end{bmatrix}$$

$$\mathbf{X^T Y} = \begin{bmatrix} 52.5600 \\ 114.1748 \\ 265.4092 \end{bmatrix}$$

$$\mathbf{\Theta}^* = (\mathbf{X^TX})^{-1}\mathbf{X^TY} = \begin{bmatrix} -0.1751 \\ 0.3396 \\ 1.2903 \end{bmatrix}$$

Therefore we obtain the model:

$$f_\theta(x) = -0.1751 + 0.3396 * x + 1.2903 * x^2 \tag{5}$$

Now we use the optimal theta parameters to plot the function to the data:

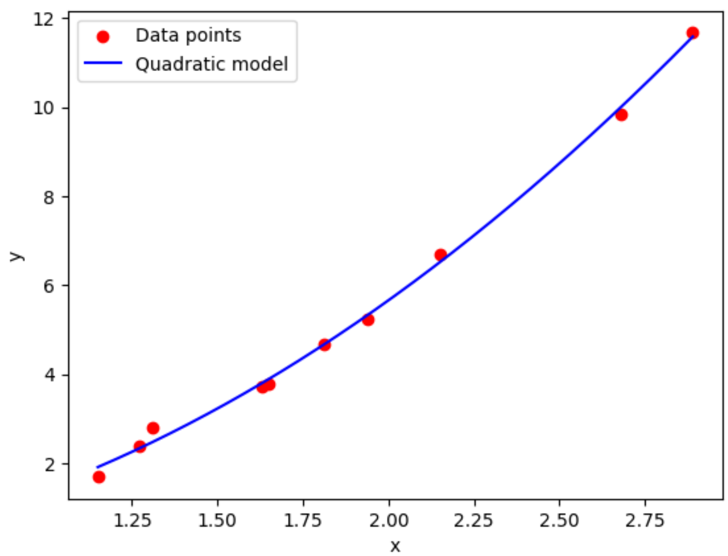


Figure 2: Fitting $f_\theta(x) = -0.1751 + 0.3396 * x + 1.2903 * x^2$ to training data

1.3 Question 1 c

We want to increase the model complexity further, by considering y as a linear function of $\mathbf{x}$, $\mathbf{x^2}$, $\mathbf{x^3}$ and $\mathbf{x^4}$.

Namely,

$$f_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4 \tag{6}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & x_n^4 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1 & 1.65 & 1.65^2 & 1.65^3 & 1.65^4 \\ 1 & 2.15 & 2.15^2 & 2.15^3 & 2.15^4 \\ 1 & 1.81 & 1.81^2 & 1.81^3 & 1.81^4 \\ 1 & 1.63 & 1.63^2 & 1.63^3 & 1.63^4 \\ 1 & 1.27 & 1.27^2 & 1.27^3 & 1.27^4 \\ 1 & 1.94 & 1.94^2 & 1.94^3 & 1.94^4 \\ 1 & 1.31 & 1.31^2 & 1.31^3 & 1.31^4 \\ 1 & 2.68 & 2.68^2 & 2.68^3 & 2.68^4 \\ 1 & 2.89 & 2.89^2 & 2.89^3 & 2.89^4 \\ 1 & 1.15 & 1.15^2 & 1.15^3 & 1.15^4 \end{bmatrix}$$

$$
\mathbf{X} = \begin{bmatrix}
1.0000 & 1.6500 & 2.7225 & 4.4921 & 7.4120 \\
1.0000 & 2.1500 & 4.6225 & 9.9384 & 21.3675 \\
1.0000 & 1.8100 & 3.2761 & 5.9297 & 10.7328 \\
1.0000 & 1.6300 & 2.6569 & 4.3307 & 7.0591 \\
1.0000 & 1.2700 & 1.6129 & 2.0484 & 2.6014 \\
1.0000 & 1.9400 & 3.7636 & 7.3014 & 14.1647 \\
1.0000 & 1.3100 & 1.7161 & 2.2481 & 2.9450 \\
1.0000 & 2.6800 & 7.1824 & 19.2488 & 51.5869 \\
1.0000 & 2.8900 & 8.3521 & 24.1376 & 69.7576 \\
1.0000 & 1.1500 & 1.3225 & 1.5209 & 1.7490
\end{bmatrix}
$$

$$
\mathbf{X}^T = \begin{bmatrix}
1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\
1.6500 & 2.1500 & 1.8100 & 1.6300 & 1.2700 & 1.9400 & 1.3100 & 2.6800 & 2.8900 & 1.1500 \\
2.7225 & 4.6225 & 3.2761 & 2.6569 & 1.6129 & 3.7636 & 1.7161 & 7.1824 & 8.3521 & 1.3225 \\
4.4921 & 9.9384 & 5.9297 & 4.3307 & 2.0484 & 7.3014 & 2.2481 & 19.2488 & 24.1376 & 1.5209 \\
7.4120 & 21.3675 & 10.7328 & 7.0591 & 2.6014 & 14.1647 & 2.9460 & 51.5869 & 69.7576 & 1.7490
\end{bmatrix}
$$

$$
(\mathbf{X^T X})^{-\mathbf{1}} = \begin{bmatrix}
6.9953e+03 & -1.5169e+04 & 1.1897e+04 & -4.0120e+03 & 492.1785 \\
-1.5169e+04 & 3.3020e+04 & -2.6002e+04 & 8.8024e+03 & -1.0838e+03 \\
1.1897e+04 & -2.6002e+04 & 2.0561e+04 & -6.9890e+03 & 863.8780 \\
-4.0120e+03 & 8.8024e+03 & -6.9890e+03 & 2.3856e+03 & -296.0494 \\
492.1785 & -1.0838e+03 & 863.8780 & -296.0494 & 36.8812
\end{bmatrix}
$$

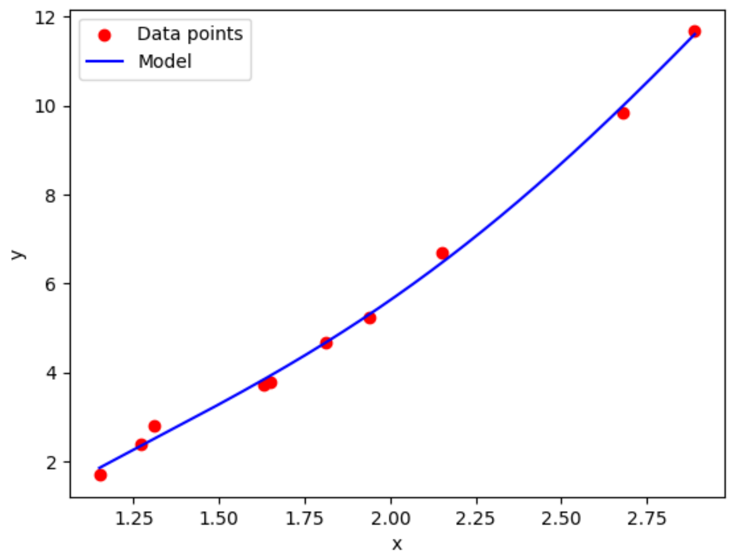
$$
\mathbf{X^T Y} = \begin{bmatrix}
52.5600 \\
114.1748 \\
265.4092 \\
650.9032 \\
1.6618e+03
\end{bmatrix}
$$

$$
\mathbf{\Theta}^* = (\mathbf{X^T X})^{-\mathbf{1}} \mathbf{X^T Y} = \begin{bmatrix}
-6.3545 \\
12.9433 \\
-7.8985 \\
2.8502 \\
-0.3193
\end{bmatrix}
$$

Therefore we obtain the model:

$$
f_\theta(x) = -6.3545 + 12.9433 * x - 7.8985 * x^2 + 2.8502 * x^3 - 0.3193 * x^4 \tag{7}
$$

Now we use the optimal theta parameters to plot the function to the training data:



Graph of the function $f_\theta(x) = -6.3545 + 12.9433 * x - 7.8985 * x^2 + 2.8502 * x^3 - 0.3193 * x^4$.

1.4 Question 1 d

Observing the functions of part a,b and c, we can guess that the model in part a has underfitted and the model in part c has overfitted. We can confirm this by using the testing data to calculate the mean absolute error (MAE), mean squared error (MSE) and root mean squared error (RMSE) for each of the functions we found in part a, b and c. I used Python to calculate each error.

Table 1: Error measures for the models in part a, b and c using the testing data

	Mean absolute error	**Mean squared error**	**Root mean squared error**
function 1a	1.2796	4.3308	2.0810
function 1b	0.1878	0.0669	0.2587
function 1c	1.0958	5.3193	2.3063

The models $f_\theta(x) = -4.9818 + 5.5399 * x$ and $f_\theta(x) = -6.3545 + 12.9433 * x - 7.8985 * x^2 + 2.8502 * x^3 - 0.3193 * x^4$ have the largest error. $f_\theta(x) = -0.1751 + 0.3396 * x + 1.2903 * x^2$ has the smallest error.

Since $f_\theta(x) = -4.9818 + 5.5399 * x$ has a higher error and is a less complex model, we call it underfitted.

Since $f_\theta(x) = -6.3545 + 12.9433 * x - 7.8985 * x^2 + 2.8502 * x^3 - 0.3193 * x^4$ has a higher error and is a more complex model, we call it overfitted.

2 Machine Learning Question

2.1 Question 1

Answer: (B)

2.2 Question 2

Answer: (A),(B) and (C)