

Into to Machine Learning - Sheet 3

Ibrahim Mohamed

ID: 7790

November 3, 2024

Note

This assignment was written using LaTeX. The source code for the document and the code used in this assignment can be found in the following GitHub link: <https://github.com/H1MAA/Intro-To-Machine-Assignments>

1 Linear Regression

1.1 Question 1

Problem: Find the least square regression line for the following set of data:

$$\{(-1, 0), (0, 2), (1, 4), (2, 5)\}$$

Solution:

1. For Linear regression I will use the equation:

$$a = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{1}{n}(\sum y - a \sum x)$$

2. I defined the following equations in the function leastsquares in python:

```
def leastSquares(x, y):  
    n = len(x)  
  
    sum_x = sum(x)  
    sum_y = sum(y)  
    sum_x2 = sum(x_i**2 for x_i in x)  
    sum_xy = sum(x_i*y_i for x_i, y_i in zip(x, y))  
  
    m = (n*sum_xy - sum_x*sum_y) / (n*sum_x2 - sum_x**2)  
    b = (sum_y - m*sum_x) / n  
    return m, b  
✓ 0.0s
```

Figure 1: Least Squares Function

3. when entering the data above i get:

$$a = 1.7, \quad b = 1.9$$

The equation of the line is:

$$y = 1.7x + 1.9$$

4. The plot of the data and the line is shown to the right

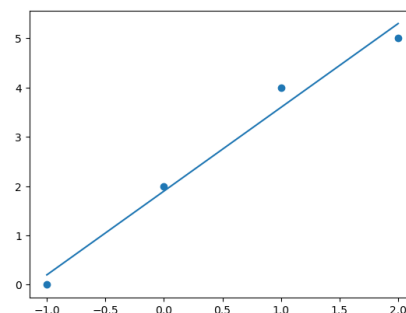


Figure 2: Least Squares Plot

1.2 Question 2

Problem: 2. The value of x and their corresponding values of y are shown in the table below.

x	0	1	2	3	4
y	2	3	5	4	6

Table 1: Problem Givens

- a) Find the least square regression line $y = ax + b$.
- b) Estimate the value of y when $x = 10$.

Solution:

- 1. Using the same function as in the previous question to find the least squares regression line for the data in the table above. The output of the function is shown below:
- 2. the output of the function is:

$$a = 0.9, b = 2.2$$

The equation of the line is:

$$y = 0.9x + 2.2$$

- 3. The value of y given $x = 10$ is:

$$y = 0.9(10) + 2.2 = 11.2$$

1.3 Question 3

Problem: Apply linear regression analytic form, use matrix inverse, to find the parameters of the best-fit line through the 6 points $\{(2, 2), (0, 0), (-1, 1), (1, -1), (-2, 0), (2, 0)\}$ shown in the figure below.

Solution:

a)

$$\mathbf{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{t}$$

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 1 & -1 \\ 1 & 1 \\ 1 & -2 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{t} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$$

$$(\mathbf{X}^T \mathbf{X})^{-1} = \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 0 & -1 & 1 & -2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 1 & -1 \\ 1 & 1 \\ 1 & -2 \\ 1 & 2 \end{bmatrix} \right)^{-1} = \begin{bmatrix} 6 & 2 \\ 2 & 14 \end{bmatrix}^{-1}$$

$$(\mathbf{X}^T \mathbf{X})^{-1} = \frac{1}{80} \begin{bmatrix} 14 & -2 \\ -2 & 6 \end{bmatrix} = \begin{bmatrix} 0.175 & -0.025 \\ -0.025 & 0.075 \end{bmatrix}$$

$$\mathbf{X}^T \mathbf{t} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 0 & -1 & 1 & -2 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\mathbf{w} = \begin{bmatrix} 0.175 & -0.025 \\ -0.025 & 0.075 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.1 \end{bmatrix}$$

$w = 0.1$, $b = 0.3$, Therefore, the equation of the best-fit line is:

$$y = 0.1x + 0.3$$

b) The plot of the data points and the best-fit line is shown below:

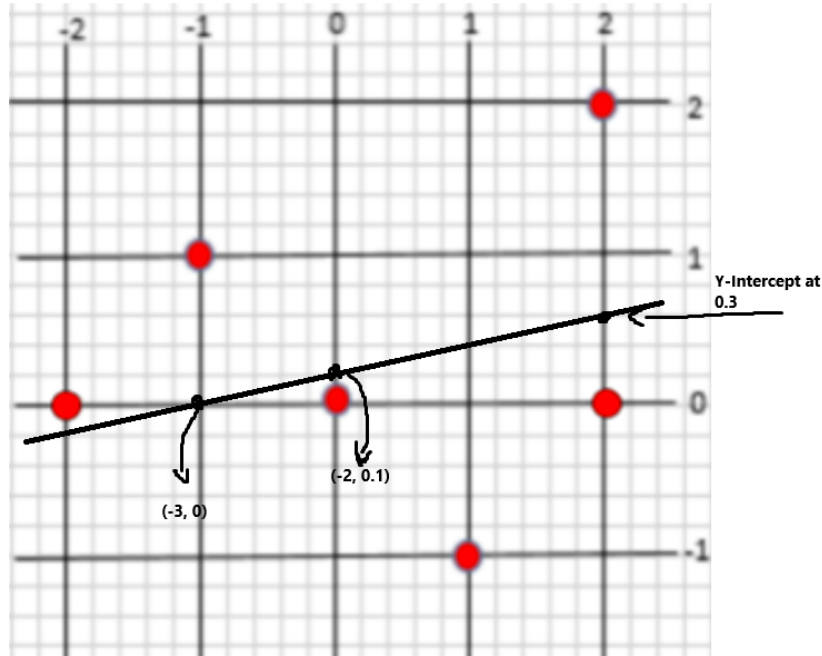


Figure 3: Best-Fit Line

c) The sum of the squared loss is calculated as follows:

$$\text{Loss} = \sum_{i=1}^n (y_i - t)^2$$

$$\begin{aligned} (2, 2) &\rightarrow y = 0.5, \quad t = 2, & \text{Loss} &= (2 - 0.5)^2 = 2.25 \\ (0, 0) &\rightarrow y = 0.3, \quad t = 0, & \text{Loss} &= (0 - 0.3)^2 = 0.09 \\ (-1, 1) &\rightarrow y = 0.2, \quad t = 1, & \text{Loss} &= (1 - 0.2)^2 = 0.64 \\ (1, -1) &\rightarrow y = 0.4, \quad t = -1, & \text{Loss} &= (-1 - 0.4)^2 = 1.96 \\ (-2, 0) &\rightarrow y = 0.1, \quad t = 0, & \text{Loss} &= (0 - 0.1)^2 = 0.01 \\ (2, 0) &\rightarrow y = 0.5, \quad t = 0, & \text{Loss} &= (0 - 0.5)^2 = 0.25 \end{aligned}$$

$$\sum_{i=1}^n (y_i - t)^2 = 2.25 + 0.09 + 0.64 + 1.96 + 0.01 + 0.25 = 5.2$$

- d) Linear regression is sensitive to noise. To illustrate this, we can remove the outlier and recalculate the best-fit line. The new data points are:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & 1 \\ 1 & -2 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$$

Recalculating $W = (X^T X)^{-1} X^T y$:

$$W = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.1 \end{bmatrix} \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

$$W = \begin{bmatrix} 0 \\ -0.2 \end{bmatrix}, b = 0, w = -0.2, \quad \text{The new equation is: } y = -0.2x$$

Comparing this to the previous equation, we can calculate loss:
old line, $y = 0.1x + 0.3$, loss without outlier:

$$\sum_{i=1}^n (y_i - t)^2 = 0.09 + 0.64 + 1.96 + 0.01 + 0.25 = 2.95$$

new line, $y = -0.2x$, loss:

$$\begin{array}{llll} (0, 0) & \rightarrow & y = 0.0, & t = 0, & \text{Loss} = (0 - 0)^2 = 0 \\ (-1, 1) & \rightarrow & y = 0.2, & t = 1, & \text{Loss} = (1 - 0.2)^2 = 0.64 \\ (1, -1) & \rightarrow & y = -0.2, & t = -1, & \text{Loss} = (-1 - (-0.2))^2 = 0.64 \\ (-2, 0) & \rightarrow & y = 0.4, & t = 0, & \text{Loss} = (0 - 0.4)^2 = 0.16 \\ (2, 0) & \rightarrow & y = -0.4, & t = 0, & \text{Loss} = (0 - (-0.4))^2 = 0.16 \end{array}$$

$$\sum_{i=1}^n (y_i - t)^2 = 0 + 0.64 + 0.64 + 0.16 + 0.16 = 1.6$$

as we can see, $1.6 < 2.95$, the loss is smaller without the outlier.

- e) Estimating y for $x = -0.5$, $x = 0.5$, and $x = 1.5$:
using old line, $y = 0.1x + 0.3$:

$$y_1 = 0.1(-0.5) + 0.3 = 0.25, \quad y_2 = 0.1(0.5) + 0.3 = 0.35$$

$$y_3 = 0.1(1.5) + 0.3 = 0.45$$

using new line, $y = -0.2x$:

$$y_1 = -0.2(-0.5) = 0.1, \quad y_2 = -0.2(0.5) = -0.1$$

$$y_3 = -0.2(1.5) = -0.3$$