Name : Deep Pawar
CWID : A20545137
Sub : CSP 571 DPA Assignment I (Writing)

Date
Page

Q. 1.

→ Given,

| X | Y |
|---|---|
| -1 | 0 |
| 0 | 2 |
| 1 | 4 |
| 2 | 5 |

- Step 1 : X & Y vectors

$X = [-1, 0, 1, 2]$
$Y = [0, 2, 4, 5]$

- Step 2 : Matrix Formulation

Here, we need to create Xd to include the intercept in our linear regression model.
Xd matrix contains a column of 1's (for the intercept) with a column of X vector values.

$$\therefore Xd = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$$

The equation is $y = (1 \; x) \begin{pmatrix} b \\ m \end{pmatrix}$

$$\therefore y = (Xd) \begin{pmatrix} b \\ m \end{pmatrix}$$

$$\therefore Xd = (1 \; x)$$

Date
Page

And the target vector $Y = \begin{bmatrix} 0 \\ 2 \\ 4 \\ 5 \end{bmatrix}$

- Step 3: Closed form equation:

The closed form solution for the weight vector $W$ will be given by:

$$\boxed{W = (X_d^T X_d)^{-1} X_d^T Y}$$

where,

$X_d$ is a matrix
$T$ is matrix transpose
$Y$ is vector of the value of the response variable

$$X_d^T X_d = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\therefore X_d^T X_d = \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}$$

$$\therefore X_d^T Y = \begin{bmatrix} 11 \\ 14 \end{bmatrix}$$

$$(X_d^T X_d)^{-1} = \frac{1}{\Delta [X_d^T X_d]} \times adj\,(X_d^T X_d)$$

$$= \frac{1}{(4 \times 6) - (2 \times 2)} \begin{bmatrix} 6 & -2 \\ -2 & 4 \end{bmatrix}$$

Date
Page

$$\therefore (X_d^T X_d)^{-1} = \begin{bmatrix} 0.3 & -0.1 \\ -0.1 & 0.2 \end{bmatrix}$$

$$\therefore W = (X_d^T X_d)^{-1} X_d^T Y$$

$$= \begin{bmatrix} 0.3 & -0.1 \\ -0.1 & 0.2 \end{bmatrix} \begin{bmatrix} 11 \\ 14 \end{bmatrix}$$

$$= \begin{bmatrix} 0.3 \times 11 + -0.1(14) \\ -0.1 \times 11 + 0.2(14) \end{bmatrix}$$

$$= \begin{bmatrix} 3.3 & -1.4 \\ -1.1 & 2.8 \end{bmatrix}$$

$$\therefore W = \begin{bmatrix} 1.9 \\ 1.7 \end{bmatrix}$$

$\therefore$ Optimal Linear Regression model is

$$y = b + wx$$

$$\therefore y = 1.9 + 1.7x$$

- Step 4: Plotting the Dataset

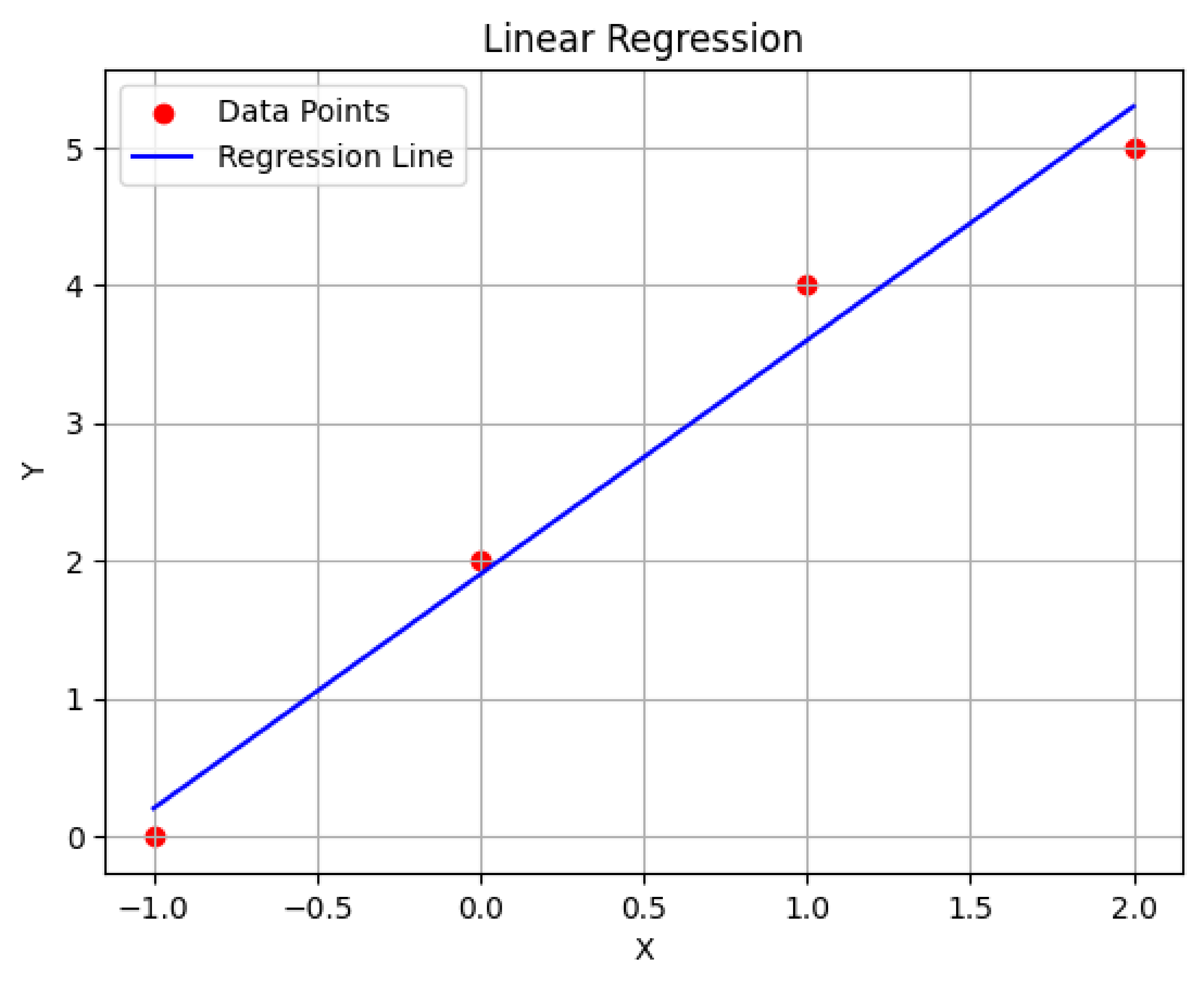
Here,

$$Y = 1.9 + 1.7x$$

$$x = [-1, 0, 1, 2]$$

After putting values of $x$ is $Y$ equation we get the values of $Y$ as

$$Y = [0.2, 1.9, 3.6, 5.3]$$



Linear Regression
Data Points
Regression Line
Y
X
0
1
2
3
4
5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0

Date
Page

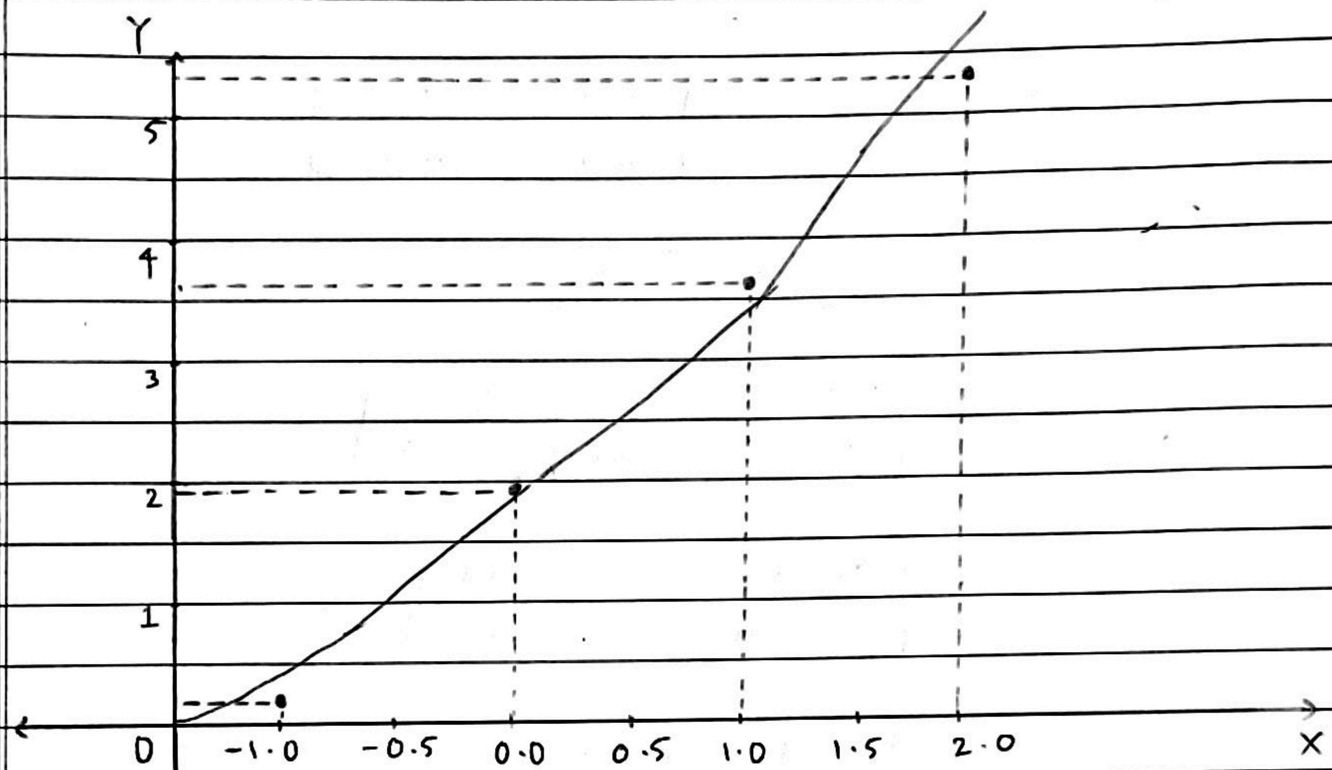



Fig. Linear Regression Plot

Plot points : [(-1, 0.2), (0, 1.9), (1, 3.6), (2, 5.3)]

Q. 2.

→ Here,

$Y = [0, 2, 4, 5]$

Predicted $\hat{Y} = [0.2, 1.9, 3.6, 5.3]$

Model $Y = 1.9 + 1.7x$

• Step I : Mean Squared Error (MSE) :

$$\boxed{MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}$$

Here,

$Y_i$ is the actual values

$\hat{Y}_1$ is the predicted values

First, let's calculate each value of $[Y_i - \hat{Y}_i]^2$

$$[0 - 0.2]^2 = 0.04$$
$$[2 - 1.9]^2 = 0.01$$
$$[4 - 3.6]^2 = 0.16$$
$$[5 - 5.3]^2 = 0.09$$

$$\therefore MSE = \frac{0.04 + 0.01 + 0.16 + 0.09}{4}$$

$$= \frac{0.3}{4}$$

$$\boxed{\therefore MSE = 0.075 = 0.08}$$

- Step 2: Mean Absolute Error [MAE]:

$$\boxed{MAE = \frac{1}{n} \sum_{i=1}^{n} |Y_i - \hat{Y}_i|}$$

First, lets calculate the $|Y_i - \hat{Y}_i|$ values:

$$|0 - 0.2| = 0.2$$
$$|2 - 1.9| = 0.1$$
$$|4 - 3.6| = 0.4$$
$$|5 - 5.3| = 0.3$$

$$\therefore MAE = \frac{0.2 + 0.1 + 0.4 + 3}{4}$$

$$= \frac{1}{4}$$

$$\boxed{\therefore MAE = 0.25}$$

```python
mse = mean_squared_error(Y, Y_pred)
mae = mean_absolute_error(Y, Y_pred)

print(f"Mean Squared Error (MSE) : {mse:.2f}\n \nMean Absolute Error (MAE) : {mae:.2f}")
```

```
Mean Squared Error (MSE) : 0.08

Mean Absolute Error (MAE) : 0.25
```

Date________
Page________

Q. 3.

→

- Step 1 : Calculate Bias :

$$\boxed{Bias = \frac{1}{n} \sum_{i=1}^{n} (\hat{Y}_i - Y_i)}$$

Here,

$Y_i$ = actual values = [0, 2, 4, 5]
$\hat{Y}_i$ = Predicted values = [0.2, 1.9, 3.6, 5.3]

First, lets calculate $(\hat{Y}_i - Y_i)$

$(0.2 - 0) = 0.2$
$(1.9 - 2) = -0.1$
$(3.6 - 4) = -0.4$
$(5.3 - 5) = 0.3$

$\therefore Bias = \frac{0.2 + (-0.1) + (-0.4) + 0.3}{4}$

$\therefore Bias = \frac{0}{4}$

$$\boxed{\therefore Bias = 0}$$

- Step 2 : Calculate Variance :

$$\boxed{Variance = \frac{1}{n} \sum_{i=1}^{n} (\hat{Y}_i - mean(\hat{Y}))^2}$$

$mean(\hat{Y}) = \frac{0.2 + 1.9 + 3.6 + 5.3}{4} = \frac{11}{4}$

$\therefore mean(\hat{Y}) = 2.75$

Data
Page

$$(0.2 - 2.75)^2 = (-2.55)^2 = 6.5025$$
$$(1.9 - 2.75)^2 = (-0.85)^2 = 0.7225$$
$$(3.6 - 2.75)^2 = (0.85)^2 = 0.7225$$
$$(5.3 - 2.75)^2 = (2.55)^2 = 6.5025$$

$$\therefore \text{Variance} = \frac{6.5025 + 0.7225 + 0.7225 + 6.5025}{4}$$

$$= \frac{14.45}{4}$$

$$\boxed{\therefore \text{Variance} = 3.6125}$$

```python
bias = np.mean(Y_pred - Y)

variance = np.var(Y_pred)

print(f"The value of Bias : {bias:.2f} \n\nThe value of Variance : {variance:.2f}")
```

```
The value of Bias : 0.00

The value of Variance : 3.61
```