AI Assignment-4

Q1.(a)

Expanded equation form:-

Let's assume a linear regression model with 5 features x_1, x_2, x_3, x_4, x_5 and target variable $\hat{y}$

Then $\hat{y} = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + w_5 x_5$

where $w_0, w_1, w_2, w_3, w_4, w_5$ are weights of model

MSE loss function for n data points is:-

$$L = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$ where y_i = Actual value
$\hat{y}_i$ = predicted value.

for n-data points

$$\hat{y}_i = w_0 + w_1 x_1 + \cdots w_j x_j + \cdots w_n x_n$$

$$\Rightarrow \frac{\partial L}{\partial w_i} = \frac{2}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i) \times \frac{\partial \hat{y}_i}{\partial w_i}$$

$$\Rightarrow \frac{\partial L}{\partial w_i} = \frac{2}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i) \times x_{ij} \quad \{\text{for } w_0,\ x_0 = 1\}$$

for each $i = 1$ to $i = n$

$$w_i = w_i - \alpha \frac{\partial L}{\partial w_i} \quad \{\alpha = \text{Learning rate}\}$$

$$\Rightarrow \boxed{w_i = w_i - \alpha \frac{2}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i) \times x_{ij}}$$

(x_{ij} represents value of j^{th} feature in i^{th} data point)

Hence all 6 weights would be updated as:-

$$w_0 = w_0 - \alpha \frac{2}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)$$

$$w_1 = w_1 - \alpha \frac{2}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)\, x_{i1}$$

$$w_2 = w_2 - \frac{\alpha}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)\, x_{i2}$$

$$w_3 = w_3 - \frac{\alpha}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)\, x_{i3}$$

$$w_4 = w_4 - \frac{\alpha}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)\, x_{i4}$$

$$w_5 = w_5 - \frac{\alpha}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)\, x_{i5}$$

Vector form:-

Target value vector:- & Model parameter vector

$$y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ y^{(4)} \\ y^{(5)} \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{bmatrix}$$

Instance's feature vector:-

$$X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$

Here, Hypothesis function

$$h(x) = \theta^T \cdot x = \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$

$$= \theta_0 + x_1\theta_1 + x_2\theta_2 + x_3\theta_3 + x_4\theta_4 + x_5\theta_5$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(\theta^T x^{(i)} - y^{(i)} \right)^2 = J(\theta)$$

So, $$\nabla_\theta J(\theta) = \begin{bmatrix} \frac{\partial}{\partial \theta_0} J(\theta) \\ \frac{\partial}{\partial \theta_1} J(\theta) \\ \vdots \\ \frac{\partial J(\theta)}{\partial \theta_5} \end{bmatrix} = \frac{2}{n} X^T (X \cdot \theta - y)$$

$$X = \begin{bmatrix} (x_0)^T \\ (x_1)^T \\ \vdots \\ (x_5)^T \end{bmatrix} \quad x_0 = 1,$$

Hence, simultaneous update of all weights is:-

$$\boxed{\theta = \theta - \alpha \frac{2}{n} X^T (X \cdot \theta - y)}$$ (α = Learning rate)

Example output of gradient descent:-

```
Weight = 23.007157833712 32, Bias = 47.1718185131469
```

Example output of linear regression:-

```
Full dataset train and eval R2 score: 0.59
70-15-15 Cross validation boxplot: mean=0.56, std=0.03
```

Box plot: