AI - Assignment - 4

1) a. Linear Regression Equation :-

$$y = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + x_5w_5 + b$$

MSE $$L(y, \hat{y}) = \frac{1}{2N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$$

N is the number of data points,
$\hat{y}_i$ is the predicted value for sample i
y_i is the actual value for sample i
α is the learning rate

Weight update formulas :-

- for w_1.

$$w_1 = w_1 - \alpha \frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial w_1} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - \hat{y}_i) \cdot x_{1i}$$

- for w_2.

$$w_2 = w_2 - \alpha \frac{\partial L}{\partial w_2}$$

$$\frac{\partial L}{\partial w_2} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i) \cdot x_{2i}$$

- for w_3

$$w_3 = w_3 - \alpha \frac{\partial L}{\partial w_3}$$

$$\frac{\partial L}{\partial w_2} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i) \cdot x_{3i}$$

- for w_4.

$$w_4 = w_4 - \alpha \frac{\partial L}{\partial w_4}$$

$$\frac{\partial L}{\partial w_4} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i) \cdot x_{4i}$$

- for w_5

$$w_5 = w_5 - \alpha \frac{\partial L}{\partial w_5}$$

$$\frac{\partial L}{\partial w_5} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i) x_{5i}$$

- for bias b.

$$b = b - \alpha \frac{\partial L}{\partial b}$$

$$\frac{\partial L}{\partial b} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)$$

- This process is repeated for multiple epochs until convergence.
- for SGD Stochastic Gradient descent we use just one sample for weight and bias update therefore in the above formulaes we remove the Σ sign and work with just one sample each time.

e.g

$$w_1 = w_1 - \alpha \frac{\partial L}{\partial w_1} \qquad L = (\hat{y}_i - y_i)$$

$$\frac{\partial L}{\partial w_1} = (\hat{y}_i - y_i) x_{1i}$$

Similary for all.

Vector Form

- $y = W^T X + b$

MSE

$$L(y, \hat{y}) = \frac{1}{2N} \sum (\hat{y}_i - y_i)^2$$

Gradients in Vector Form

$$\frac{\partial L}{\partial W} = \frac{1}{N} \sum (\hat{y}_i - y_i) \cdot X_i$$

$$\frac{\partial L}{\partial b} = \frac{1}{N} \sum (\hat{y}_i - y_i)$$

bias & Weight update

$$W = W - \alpha \frac{\partial L}{\partial W}$$

$$b = b - \alpha \frac{\partial L}{\partial b}$$

where,

$$W = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{bmatrix} \qquad X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix} \qquad \alpha = \text{learning rate}$$

Similarly, for SGD we remove the Σ sign and work with just one sample.