* Gradient Descent Example ⇒

Consider, a two independent features $x_1 = 3, x_2 = 2$ & third dependent feature $y = 1$.

Assume, cost function as, (Error function)

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - y_i) \right]$$

Find the values of θ upto two descent.

⇒ Here, no. of parameters will be 3 (two coeffient weights & one bias)

$\boxed{\text{First Descent}}$

∴ Assume, $\theta_1 = \theta_2 = \theta_0 = 0$ (Initial)

& $\eta = 0.1$

$$\frac{\partial(J(\theta))}{\partial(\theta_j)} = -\frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i) x_i$$

Here, $\hat{y} = 6(\theta_1 x_1 + \theta_2 x_2 + \theta_0)$

$$\therefore \frac{\partial(J(\theta))}{\partial(\theta_j)} = -\frac{1}{m} \sum_{i=1}^{m} \left[6(\theta_1 x_1 + \theta_2 x_2 + \theta_0) - y \right] x_j$$

$$\nabla J(\theta) = \begin{bmatrix} \dfrac{\partial J(\theta)}{\partial \theta_1} \\ \dfrac{\partial J(\theta)}{\partial \theta_2} \\ \dfrac{\partial J(\theta)}{\partial \theta_0} \end{bmatrix} = \begin{bmatrix} (\sigma(\theta_1 x_1 + \theta_2 x_2 + \theta_0) - y) \cdot x_1 \\ (\sigma(\theta_1 x_1 + \theta_2 x_2 + \theta_0) - y) \cdot x_2 \\ (\sigma(\theta_1 x_1 + \theta_2 x_2 + \theta_0) - y) \end{bmatrix}$$

$$\boxed{\sim (\mathrm{I})}$$

$$\nabla J(\theta) = \begin{bmatrix} (\sigma(0) - 1)3 \\ (\sigma(0) - 1)2 \\ (\sigma(0) - 1) \end{bmatrix} = \begin{bmatrix} (-0.5)3 \\ (-0.5)2 \\ -0.5 \end{bmatrix}$$

$$\nabla J(\theta) = \begin{bmatrix} -1.5 \\ -1.0 \\ -0.5 \end{bmatrix}$$

Now we have a gradient vector $\nabla J(\theta)$.

$\therefore$ We can compute θ_{new},

$$\theta_{new_1} = \theta_{prv} - \eta \nabla J(\theta)$$

$$= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} - 0.1 \begin{bmatrix} -1.5 \\ -1.0 \\ -0.5 \end{bmatrix}$$

$$\boxed{\theta_{new_1} = \begin{bmatrix} 0.15 \\ 0.1 \\ 0.05 \end{bmatrix}}$$

econd Descent

Now we will calculate new values of $J(\theta)$ w.r.t. θ_{new_1} from first descent.

$$\therefore \sigma(\theta_1 x_1 + \theta_2 x_2 + \theta_0) = \frac{1}{1 + e^{-(\theta_1 x_1 + \theta_2 x_2 + \theta_0)}}$$

$$= \frac{1}{1 + e^{-(3(0.15) + 2(0.1) + 0.05)}}$$

$$= \frac{1}{1 + e^{-0.7}}$$

$$\therefore \boxed{\sigma(\theta_1 x_1 + \theta_2 x_2 + \theta_0) = 0.67}$$

$$\nabla J(\theta) = \begin{bmatrix} (0.67-1)\,x_1 \\ (0.67-1)\,x_2 \\ (0.67-1) \end{bmatrix} = \begin{bmatrix} -0.33x_1 \\ -0.33x_1 \\ -0.33 \end{bmatrix}$$

$$= \begin{bmatrix} -0.99 \\ -0.66 \\ -0.33 \end{bmatrix}$$

$$\theta_{new_2} = \theta_{prv} - \eta \nabla J(\theta)$$

$$= \begin{bmatrix} 0.15 \\ 0.1 \\ 0.05 \end{bmatrix} - 0.1 \begin{bmatrix} -0.99 \\ -0.66 \\ -0.33 \end{bmatrix}$$

$$\theta_{new_2} = \begin{bmatrix} 0.249 \\ 0.166 \\ 0.083 \end{bmatrix}$$