Monday, February 6, 2023 5:43 PM

b) For the weight vector :-

As 1) we can calculated the gradient of loss function

$$\nabla_w f_{loss} = x(\hat{y} - y)$$

$$\nabla_w f_{loss} = x[\hat{y} - y] = x[\hat{y} - 1]$$

$$\nabla_w f_{loss} = x[\sigma(x^T w + b) - 1]$$

Assume $b = 0$ for simplicity.

$$\nabla_w f_{loss} = x[\sigma(x^T w) - 1]$$

Let $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ ; $w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

$$\nabla_w f_{loss} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \left[ \sigma\left( [x_1 \; x_2] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \right) - 1 \right]$$

$$\nabla_w f_{loss} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} [\sigma(w_1 x_1 + w_2 x_2) - 1]$$

Update for $w$

$$w = w_0 - \alpha \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} [\sigma(x_1 w_1 + w_2 x_2) - 1]$$

Now value of $\sigma(x_1 w_1 + w_2 x_2) - 1 < 0$

$\alpha_{new} = \alpha \times a$ ; $a = [\sigma(x_1 w_1 + x_2 w_2) - 1]$

$$w = w_0 - \alpha_{new} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

This equation denotes that the convergence of $w$ is dependent on the exact values of $x$.

If $x = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ it will never converge.

If $x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ or some suitable value it might converge.

$\therefore$ Convergence of $w$ depends on the value of training examples.