Logistic Regression

Let $\vec{w}$ be the weight vector
b be the bias and
$\vec{x}$ be the input vector
y be the true label

$$z = w^T x + b$$

prediction

$$\hat{y} = \sigma(z) \rightarrow 5$$

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Error function is defined as

$$L = -y \cdot \log \hat{y} - (1-y) \cdot \log(1-\hat{y})$$

$$
\begin{tikzcd}
b \arrow[dr] & & & \\
x \arrow[r] & z \arrow[r] & \hat{y} \arrow[r] & L \\
w \arrow[ur] & & &
\end{tikzcd}
$$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial w} \longrightarrow (1)$$

Calculating each term separately

$$z = w^T x + b$$

$$\boxed{\frac{\partial z}{\partial w} = \vec{x}} \longrightarrow (2)$$

$$\boxed{\frac{\partial \hat{y}}{\partial z} = \sigma(z)\left[1 - \sigma(z)\right]} \longrightarrow (3)$$

$$\frac{\partial L}{\partial \hat{y}} = -\left[\frac{\partial}{\partial \hat{y}} y \log \hat{y} + \frac{\partial}{\partial \hat{y}} (1-y) \log(1-\hat{y})\right]$$

$$= -\left[\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}(-1)\right]$$

$$\boxed{\frac{\partial L}{\partial \hat{y}} = \frac{1-y}{1-\hat{y}} - \frac{y}{\hat{y}}} \longrightarrow (4)$$

putting ② ③ ④ in ①

$$\frac{\partial L}{\partial W} = \left[\frac{1-y}{1-\hat{y}} - \frac{y}{\hat{y}}\right] \sigma(z)\left[1-\sigma(z)\right] \vec{x}$$

by eq ⑤

$$= \left[\frac{1-y}{1-\hat{y}} - \frac{y}{\hat{y}}\right] \hat{y}\left[1-\hat{y}\right] \vec{x}$$

$$= \left[\hat{y}\left[1-y\right] - \left[1-\hat{y}\right]y\right] \vec{x}$$

$$= \left[\hat{y} - y\hat{y} - y + y\hat{y}\right] \vec{x}$$

$$\boxed{\frac{\partial L}{\partial W} = \left[\hat{y} - y\right] x} \rightarrow ⑥$$

Gradient Step is defined by

Learning Rate

$$\boxed{W_1 = W_0 - \alpha \frac{\partial L}{\partial W}}$$

Now Calculating for bias b.

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z} \frac{\partial z}{\partial b}$$

$$z = w^T x + b$$

$$\frac{\partial z}{\partial b} = 1$$

$$\frac{\partial L}{\partial b} = [\hat{y} - y] \cdot 1$$

gradient step for bias.

$$\boxed{b_1 = b_0 - \alpha \frac{\partial L}{\partial b}}$$