

$$f(w_0, w_1, w_2) = w_0 + w_1 x_1 + w_2 x_2 \leftarrow \text{Linear function}$$

$$\frac{\partial f}{\partial w_0} = 1, \quad \frac{\partial f}{\partial w_1} = x_1, \quad \frac{\partial f}{\partial w_2} = x_2$$

$$h = w^T x \quad w, x \text{ are both column vector}$$

Binary Cross Entropy Loss

$$z = \sigma(h) = \frac{1}{1+e^{-h}} \leftarrow \text{Sigmoid function}$$

$$L(w) = -[y \log(z) + (1-y) \log(1-z)] \leftarrow \text{Loss function}$$

$$\text{By chain rule: } \frac{\partial L(w)}{\partial w} = \frac{\partial L(w)}{\partial z} \cdot \frac{\partial z}{\partial h} \cdot \frac{\partial h}{\partial w}$$

$$\frac{\partial L}{\partial z} = -\left[\frac{y}{z} - \frac{1-y}{1-z}\right] = \boxed{\frac{z-y}{z(1-z)}}$$

$$\begin{cases} \text{If } y=1, \frac{\partial L}{\partial z} = -\frac{y}{z} \\ \text{If } y=0, \frac{\partial L}{\partial z} = \frac{1}{1-z} \end{cases}$$

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

$$\frac{dz}{dh} = \frac{e^{-h}}{(1+e^{-h})^2} = \left(\frac{1+e^{-h}-1}{1+e^{-h}}\right) \left(\frac{1}{1+e^{-h}}\right) = \sigma(h)(1-\sigma(h)) = \boxed{z(1-z)}$$

$$\frac{\partial h}{\partial w} = x \quad \leftarrow \text{by Quotient Rule}$$

$$\frac{\partial L}{\partial w} = x^T (z-y)$$

$$\text{Gradient descent: } w = w - \underset{\substack{\uparrow \\ \text{Learning Rate}}}{\ell} \frac{\partial L(w)}{\partial w}$$