1 Gradient Descent

$$\frac{dC(w_i)}{dw_i} = \lim_{\epsilon \to 0} \frac{C(w_i + \epsilon) - C(w_i)}{\epsilon} \tag{1}$$

1.1 "Twice"

$$C(w) = \frac{1}{n} \sum_{i=1}^{n} (x_i w - y_i)^2 \tag{2}$$

$$C'(w) = \left(\frac{1}{n} \sum_{i=1}^{n} (x_i w - y_i)^2 \right)' \tag{3}$$

$$= \frac{1}{n} \left(\sum_{i=1}^{n} (x_i w - y_i)^2 \right)' \tag{4}$$

$$= \frac{1}{n} \left((x_0 w - y_0)^2 + (x_1 w - y_1)^2 + (x_2 w - y_2)^2 + \ldots \right)' \tag{5}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left((x_i w - y_i)^2 \right)' \tag{6}$$

$$= \frac{1}{n} \sum_{i=1}^{n} 2 \left(x_i w - y_i \right) (x_i w - y_i)' \tag{7}$$

$$= \frac{1}{n} \sum_{i=1}^{n} 2 \left(x_i w - y_i \right) ((x_i w)' - (y_i)') \tag{8}$$

$$= \frac{1}{n} \sum_{i=1}^{n} 2 \left(x_i w - y_i \right) (x_i w)' \tag{9}$$

$$= \frac{1}{n} \sum_{i=1}^{n} 2 \left(x_i w - y_i \right) x_i \tag{10}$$

$$\tag{11}$$

$$C(w) = \frac{1}{n} \sum_{i=1}^{n} (x_i w - y_i)^2 \tag{12}$$

$$C'(w) = \frac{1}{n} \sum_{i=1}^{n} 2 \left(x_i w - y_i \right) x_i \tag{13}$$

$$\tag{14}$$

1.2 One Neuron Model

$$y = \sigma(xw + b) \tag{15}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{16}$$

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \tag{17}$$

$$\tag{18}$$

1.3 Cost

$$v_i = (x_i w + b) \tag{19}$$

$$a_i = \sigma(v_i) \tag{20}$$

$$\partial_w a_i = \partial_w(\sigma(v_i)) \tag{21}$$

$$= \sigma(v_i)(1 - \sigma(v_i))\partial_w v_i \tag{22}$$

$$C = \frac{1}{n}\sum_{i=0}^{n}(a_i - y_i)^2 \tag{23}$$

$$\partial_w C = \frac{1}{n}\sum_{i=0}^{n}\partial_w\left((a_i - y_i)^2\right) \tag{24}$$

$$= \frac{1}{n}\sum_{i=0}^{n}2\left(a_i - y_i\right)\partial_w\left(a_i - y_i\right) \tag{25}$$

$$\tag{26}$$