

# 1 Gradient Descent

$$C'(w) = \lim_{\epsilon \rightarrow 0} \frac{C(w + \epsilon) - C(w)}{\epsilon} \quad (1)$$

## 1.1 “Double“

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

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

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

$$= \frac{1}{n} ((x_0 w - y_0)^2 + (x_1 w - y_1)^2 + \dots + (x_n w - y_n)^2)' \quad (5)$$

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

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

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

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

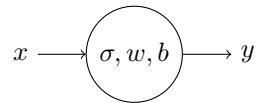
$$(10)$$

$$C(w) = \frac{1}{n} \sum_{i=1}^n (x_i w - y_i)^2 \quad (11)$$

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

$$(13)$$

## 1.2 One neuron model



$$y = \sigma(xw + b) \quad (14)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (15)$$

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \quad (16)$$

$$(17)$$

### 1.2.1 Cost

$$a_i = \sigma(x_i w + b) \quad (18)$$

$$\partial_w a_i = \partial_w (\sigma(x_i w + b)) \quad (19)$$

$$= a_i(1 - a_i)\partial_w(x_i w + b) \quad (20)$$

$$= a_i(1 - a_i)x_i \quad (21)$$

$$\partial_b a_i = a_i(1 - a_i) \quad (22)$$

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

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

$$= \frac{1}{n} \sum_{i=1}^n \partial_w ((a_i - y_i)^2) \quad (25)$$

$$= \frac{1}{n} \sum_{i=1}^n 2(a_i - y_i) \partial_w (a_i - y_i) \quad (26)$$

$$= \frac{1}{n} \sum_{i=1}^n 2(a_i - y_i) \partial_w a_i \quad (27)$$

$$= \frac{1}{n} \sum_{i=1}^n 2(a_i - y_i)(a_i(1 - a_i)x_i) \quad (28)$$

$$\partial_b C = \frac{1}{n} \sum_{i=1}^n 2(a_i - y_i) \partial_b a_i \quad (29)$$

$$= \frac{1}{n} \sum_{i=1}^n 2(a_i - y_i)(a_i(1 - a_i)) \quad (30)$$

$$(31)$$