1 Gradient Descent

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

1.1 "Double"

$$\begin{aligned}
C(w) &= \frac{1}{n} \sum_{i=1}^{n} (x_i w - y_i)^2 && (2)\\
C'(w) &= \left(\frac{1}{n} \sum_{i=1}^{n} (x_i w - y_i)^2 \right)' && (3)\\
&= \frac{1}{n} \left(\sum_{i=1}^{n} (x_i w - y_i)^2 \right)' && (4)\\
&= \frac{1}{n} \left((x_0 w - y_0)^2 + (x_1 w - y_1)^2 + \ldots + (x_n w - y_n)^2 \right)' && (5)\\
&= \frac{1}{n} \sum_{i=1}^{n} \left((x_i w - y_i)^2 \right)' && (6)\\
&= \frac{1}{n} \sum_{i=1}^{n} 2(x_i w - y_i) \left(x_i w - y_i \right)' && (7)\\
&= \frac{1}{n} \sum_{i=1}^{n} 2(x_i w - y_i) \left(x_i w \right)' && (8)\\
&= \frac{1}{n} \sum_{i=1}^{n} 2(x_i w - y_i) x_i' && (9)\\
& && (10)
\end{aligned}$$

$$\begin{aligned}
C(w) &= \frac{1}{n} \sum_{1=1}^{n} (x_i w - y_i)^2 && (11)\\
C'(w) &= \frac{1}{n} \sum_{1=1}^{n} 2(x_i w - y_i) x_i' && (12)\\
& && (13)
\end{aligned}$$

1.2 One neuron model

$$x \longrightarrow \sigma, w, b \longrightarrow y$$

$$y = \sigma(xw + b) \tag{14}$$

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

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \tag{16}$$

$$\tag{17}$$

1.2.1 Cost

$$
\begin{align}
a_i &= \sigma(x_i w + b) \tag{18}\\
\partial_w a_i &= \partial_w \left(\sigma(x_i w + b)\right) \tag{19}\\
&= a_i(1 - a_i)\partial_w(x_i w + b) \tag{20}\\
&= a_i(1 - a_i)x_i \tag{21}\\
\partial_b a_i &= a_i(1 - a_i) \tag{22}\\
C &= \frac{1}{n}\sum_{i=1}^{n}(a_i - y_i)^2 \tag{23}\\
\partial_w C &= \partial_w \left(\frac{1}{n}\sum_{i=1}^{n}(a_i - y_i)^2\right) \tag{24}\\
&= \frac{1}{n}\sum_{i=1}^{n}\partial_w \left((a_i - y_i)^2\right) \tag{25}\\
&= \frac{1}{n}\sum_{i=1}^{n}2(a_i - y_i)\partial_w \left(a_i - y_i\right) \tag{26}\\
&= \frac{1}{n}\sum_{i=1}^{n}2(a_i - y_i)\partial_w a_i \tag{27}\\
&= \frac{1}{n}\sum_{i=1}^{n}2(a_i - y_i)(a_i(1 - a_i)x_i) \tag{28}\\
\partial_b C &= \frac{1}{n}\sum_{i=1}^{n}2(a_i - y_i)\partial_b a_i \tag{29}\\
&= \frac{1}{n}\sum_{i=1}^{n}2(a_i - y_i)(a_i(1 - a_i)) \tag{30}\\
& \tag{31}
\end{align}
$$