1 Gradient Descent

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

1.1 "Twice"

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

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

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

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

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

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

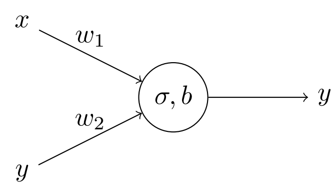
$$= \frac{1}{n}\sum_{i=1}^{n}2(x_i w - y_i)(x_i w - y_i)' = \tag{8}$$

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

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

$$\tag{11}$$

1.2 One Neuron Model with 2 inputs

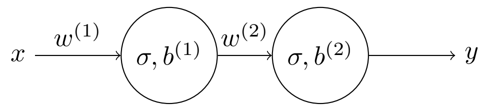


$$
\begin{aligned}
z &= \sigma(xw_1 + yw_2 + b) && (12)\\
\sigma(x) &= \frac{1}{1+e^{-x}} && (13)\\
\sigma'(x) &= \sigma(x)(1-\sigma(x)) && (14)\\
& && (15)
\end{aligned}
$$

1.3 Cost

$$
\begin{aligned}
a_i &= \sigma(x_i w + y_i w + b) && (16)\\
\partial_{w_1} a_i &= \partial_{w_1}(\sigma(x_i w + y_i w + b)) = && (17)\\
&= \partial_{w_1}\sigma(x_i w + y_i w + b)\partial_w(x_i w + y_i w + b) = && (18)\\
&= \sigma(x_i w + y_i w + b)(1 - \sigma(x_i w + y_i w + b))\partial_w(x_i w + y_i w + b) = && (19)\\
&= a_i(1 - a_i)\partial_{w_1}(x_i w + y_i w + b) = && (20)\\
&= a_i(1 - a_i)x_i && (21)\\
\partial_{w_2} a_i &= a_i(1 - a_i)y_i && (22)\\
\partial_b a_i &= a_i(1 - a_i) && (23)\\
C &= \frac{1}{n}\sum_{i=1}^{n}(\sigma(x_i w + b) - z_i)^2 && (24)\\
\partial_{w_1} C &= \frac{1}{n}\sum_{i=1}^{n}\partial_{w_1}\left((a_1 - z_i)^2\right) = && (25)\\
&= \frac{1}{n}\sum_{i=1}^{n}\partial_{w_1}(a_1 - z_i)^2\partial_{w_1}(a_1 - z_i) = && (26)\\
&= \frac{1}{n}\sum_{i=1}^{n}2(a_1 - z_i)\partial_{w_1}(a_1 - z_i) = && (27)\\
&= \frac{1}{n}\sum_{i=1}^{n}2(a_1 - z_i)\partial_{w_1} a_1 = && (28)\\
&= \frac{1}{n}\sum_{i=1}^{n}2(a_1 - z_i)a_1(1 - a_i)x_i && (29)\\
\partial_{w_2} C &= \frac{1}{n}\sum_{i=1}^{n}2(a_1 - z_i)a_1(1 - a_i)y_i && (30)\\
\partial_b C &= \frac{1}{n}\sum_{i=1}^{n}2(a_1 - z_i)a_i(1 - a_i) && (31)\\
& && (32)
\end{aligned}
$$

1.4 Two Neurons Model with 1 input



$$a^{(1)} = \sigma(xw^{(1)} + b^{(1)}) \tag{33}$$

$$y = \sigma(a^{(1)}w^{(2)} + b^{(2)}) \tag{34}$$

$$\tag{35}$$

1.5 Cost

$$C^{(2)} = \frac{1}{n}\sum_{i=1}^{n}(a_i^{(2)} - y_i)^2 \tag{36}$$

$$a_i^{(1)} = \sigma(x_i w^{(1)} + b^{(1)}) \tag{37}$$

$$\partial_{w^{(1)}} a_i^{(1)} = a_i^{(1)}(1 - a^{(1)})x_i \tag{38}$$

$$\partial_{b^{(1)}} a_i^{(1)} = a_i^{(1)}(1 - a^{(1)}) \tag{39}$$

$$a_i^{(2)} = \sigma(a_i^{(1)} w^{(2)} + b^{(2)}) \tag{40}$$

$$\partial_{a_i^{(1)}} a_i^{(2)} = a_i^{(2)}(1 - a^{(2)})w^{(2)} \tag{41}$$

$$\partial_{w^{(2)}} a_i^{(2)} = a_i^{(2)}(1 - a_i^{(2)})a_i^{(1)} \tag{42}$$

$$\partial_{w^{(2)}} a_i^{(2)} = a_i^{(2)}(1 - a_i^{(2)}) \tag{43}$$

$$\partial_{w^{(2)}} C^{(2)} = \frac{1}{n}\sum_{i=1}^{n} 2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)})a_i^{(1)} \tag{44}$$

$$\partial_{b^{(2)}} C^{(2)} = \frac{1}{n}\sum_{i=1}^{n} 2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)}) \tag{45}$$

$$\partial_{a_i^{(1)}} C^{(2)} = \frac{1}{n}\sum_{i=1}^{n} 2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)})w^{(1)} \tag{46}$$

$$e_i = a_i^{(1)} - \partial_{a_i^{(1)}} C^{(2)} \tag{47}$$

$$C^{(1)} = \frac{1}{n}\sum_{i=1}^{n}(a_i^{(1)} - e_i)^2 \tag{48}$$

$$\partial_{w^{(1)}} C^{(1)} = \partial_{w^{(1)}}\left(\frac{1}{n}\sum_{i=1}^{n}(a_i^{(1)} - e_i)^2\right) = \tag{49}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \partial_{w^{(1)}}\left((a_i^{(1)} - e_i)^2\right) = \tag{50}$$

$$= \frac{1}{n}\sum_{i=1}^{n} 2(a_i^{(1)} - e_i)\partial_{w^{(1)}}(a_i^{(1)} - e_i) = \tag{51}$$

$$= \frac{1}{n}\sum_{i=1}^{n} 2(a_i^{(1)} - e_i)\partial_{w^{(1)}} a_i^{(1)} = \tag{52}$$

$$= \frac{1}{n}\sum_{i=1}^{n} 2(a_i^{(1)} - (a_i^{(1)} - \partial_{a_i^{(1)}} C^{(2)}))\partial_{w^{(1)}} a_i^{(1)} = \tag{53}$$

$$= \frac{1}{n}\sum_{i=1}^{n} 2(a_i^{(1)} - a_i^{(1)} + \partial_{a_i^{(1)}} C^{(2)})\partial_{w^{(1)}} a_i^{(1)} = \tag{54}$$

$$= \frac{1}{n}\sum_{i=1}^{n} 2\partial_{a_i^{(1)}} C^{(2)}\partial_{w^{(1)}} a_i^{(1)} = \tag{55}$$

$$= \frac{1}{n}\sum_{i=1}^{n} 2\partial_{a_i^{(1)}} C^{(2)} a_i^{(1)}(1 - a^{(1)})x_i \tag{56}$$

$$\partial_{b^{(1)}} C^{(1)} = \frac{1}{n}\sum_{i=1}^{n} 2(\partial_{a_i^{(1)}} C^{(2)})\partial_{b^{(1)}} a_i^{(1)} = \tag{57}$$

$$= \frac{1}{n}\sum_{i=1}^{n} 2(\partial_{a_i^{(1)}} C^{(2)})a_i^{(1)}(1 - a^{(1)}) \tag{58}$$

1.6 Arbitrary Neurons Model with 1 input

Let's assume that we have m layers.

1.6.1 Feed-Forward

Let's assume that $a_i^{(0)}$ is x_i.

$$\begin{aligned}
a_i^{(l)} &= \sigma(a_i^{(l-1)} w^{(l)} + b^{(l)}) && (60)\\
\partial_{w^{(l)}} a_i^{(l)} &= a_i^{(l)}(1 - a_i^{(l)}) a_i^{(l-1)} && (61)\\
\partial_{b^{(l)}} a_i^{(l)} &= a_i^{(l)}(1 - a_i^{(l)}) && (62)\\
\partial_{a_i^{(l-1)}} a_i^{(l)} &= a_i^{(l)}(1 - a^{(l)}) w^{(l)} && (63)\\
& && (64)
\end{aligned}$$

1.6.2 Back-Propagation

Let's denote $a_i^{(m)} - y_i$ as $\partial_{a_i^{(m)}} C^{(m-1)}$.

$$\begin{aligned}
C^{(l)} &= \frac{1}{n} \sum_{i=1}^{n} (a_i^{(l)} - y_i)^l + 1 && (65)\\
\partial_{w^{(l)}} C^{(l)} &= \frac{1}{n} \sum_{i=1}^{n} 2(\partial_{a_i^{(l)}} C^{(l+1)}) a_i^{(l)} (1 - a^{(l)}) a_i^{(l-1)} && (66)\\
\partial_{b^{(l)}} C^{(l)} &= \frac{1}{n} \sum_{i=1}^{n} 2(\partial_{a_i^{(l)}} C^{(l+1)}) a_i^{(l)} (1 - a^{(l)}) && (67)\\
\partial_{a_i^{(l-1)}} C^{(l)} &= \frac{1}{n} \sum_{i=1}^{n} 2(\partial_{a_i^{(l)}} C^{(l+1)}) a_i^{(l)} (1 - a_i^{(l)}) w^{(l-1)} && (68)\\
& && (69)
\end{aligned}$$