1 Gradient Descent

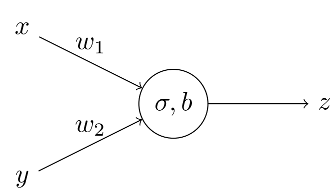
$$C'(w) = \lim_{\epsilon \to 0} \frac{C(w+\epsilon) - C(w)}{\epsilon} \tag{1}$$

1.1 "Twice"

$$\begin{align}
C(w) &= \frac{1}{n}\sum_{i=1}^{n}(x_i w - y_i)^2 \tag{2}\\
C'(w) &= \left(\frac{1}{n}\sum_{i=1}^{n}(x_i w - y_i)^2\right)' = \tag{3}\\
&= \frac{1}{n}\left(\sum_{i=1}^{n}(x_i w - y_i)^2\right)' \tag{4}\\
&= \frac{1}{n}\sum_{i=1}^{n}\left((x_i w - y_i)^2\right)' \tag{5}\\
&= \frac{1}{n}\sum_{i=1}^{n}2(x_i w - y_i)x_i \tag{6}\\
& \tag{7}
\end{align}$$

$$\begin{align}
C(w) &= \frac{1}{n}\sum_{i=1}^{n}(x_i w - y_i)^2 \tag{8}\\
C'(w) &= \frac{1}{n}\sum_{i=1}^{n}2(x_i w - y_i)x_i \tag{9}\\
& \tag{10}
\end{align}$$

1.2 One Neuron Model with 2 inputs

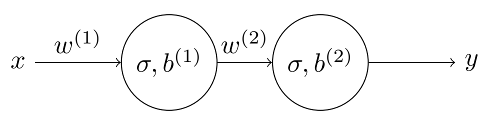


$$\begin{align}
y &= \sigma(xw_1 + yw_2 + b) \tag{11}\\
\sigma(x) &= \frac{1}{1+e^{-x}} \tag{12}\\
\sigma'(x) &= \sigma(x)(1-\sigma(x)) \tag{13}\\
& \tag{14}
\end{align}$$

1.2.1 Cost

$$\begin{aligned}
a_i &= \sigma(x_i w_1 + y_i w_2 + b) && (15)\\
\partial_{w_1} a_i &= \partial_{w_1}(\sigma(x_i w_1 + y_i w_2 + b)) = && (16)\\
&= a_i(1-a_i)\partial_{w_1}(x_i w_1 + y_i w_2 + b) = && (17)\\
&= a_i(1-a_i)x_i && (18)\\
\partial_{w_2} a_i &= a_i(1-a_i)y_i && (19)\\
\partial_b a_i &= a_i(1-a_i) && (20)\\
C &= \frac{1}{n}\sum_{i=1}^{n}(a_i - z_i)^2 && (21)\\
\partial_{w_1} C &= \frac{1}{n}\sum_{i=1}^{n}\partial_{w_1}\left((a_i - z_i)^2\right) = && (22)\\
&= \frac{1}{n}\sum_{i=1}^{n}2(a_i - z_i)\partial_{w_1} a_i = && (23)\\
&= \frac{1}{n}\sum_{i=1}^{n}2(a_i - z_i)a_i(1-a_i)x_i && (24)\\
\partial_{w_2} C &= \frac{1}{n}\sum_{i=1}^{n}2(a_i - z_i)a_i(1-a_i)y_i && (25)\\
\partial_b C &= \frac{1}{n}\sum_{i=1}^{n}2(a_i - z_i)a_i(1-a_i) && (26)\\
& && (27)
\end{aligned}$$

1.3 Two Neurons Model with 1 input



$$\begin{aligned}
a^{(1)} &= \sigma(xw^{(1)} + b^{(1)}) && (28)\\
y &= \sigma(a^{(1)}w^{(2)} + b^{(2)}) && (29)
\end{aligned}$$

1.3.1 Cost

$$a_i^{(1)} = \sigma(x_i w^{(1)} + b^{(1)}) \tag{30}$$

$$\partial_{w^1} a_1^{(i)} = a_i^{(1)}(1 - a_i^{(1)})x_i \tag{31}$$

$$\partial_{b^1} a_1^{(i)} = a_i^{(1)}(1 - a_i^{(1)}) \tag{32}$$

$$a_i^{(2)} = \sigma(a_i^{(1)} w^{(2)} + b^{(2)}) \tag{33}$$

$$\partial_{w^{(2)}} a_i^{(2)} = a_i^{(2)}(1 - a_i^{(2)})a_i^{(1)} \tag{34}$$

$$\partial_{b^{(2)}} a_i^{(2)} = a_i^{(2)}(1 - a_i^{(2)}) \tag{35}$$

$$\partial_{a_i^{(1)}} a_i^{(2)} = a_i^{(2)}(1 - a_i^{(2)})w^{(2)} \tag{36}$$

$$C^{(2)} = \frac{1}{n}\sum_{i=1}^{n}(a_i^{(2)} - y_i)^2 \tag{37}$$

$$\partial_{w^{(2)}} C^{(2)} = \frac{1}{n}\sum_{i=1}^{n}\partial_{w^{(2)}}((a_i^{(2)} - y_i)^2) = \tag{38}$$

$$= \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(2)} - y_i)\partial_{w^{(2)}} a_i^{(2)} = \tag{39}$$

$$= \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)})a_i^{(1)} \tag{40}$$

$$\partial_{b^{(2)}} C^{(2)} = \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)}) \tag{41}$$

$$\partial_{a_i^{(1)}} C^{(2)} = \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)})w^{(2)} \tag{42}$$

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

$$C^{(1)} = \frac{1}{n}\sum_{i=1}^{n}(a_1^{(i)} - e_i)^2 \tag{44}$$

$$\partial_{w^1} C^{(1)} = \partial_{w^1}\left(\frac{1}{n}\sum_{i=1}^{n}(a_1^{(i)} - e_i)^2\right) = \tag{45}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\partial_{w^1}\left((a_1^{(i)} - e_i)^2\right) = \tag{46}$$

$$= \frac{1}{n}\sum_{i=1}^{n}2(a_1^{(i)} - e_i)\partial_{w^1} a_1^{(i)} = \tag{47}$$

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

$$\partial_{b^1} C^{(1)} = \frac{1}{n}\sum_{i=1}^{n}2(\partial_{a_i^{(1)}} C^{(2)}) \tag{49}$$

$$\tag{50}$$

1.4 Arbitrary Neurons Model with 1 input

Let's assume that we have m layers.

1.4.1 Feed-Forward

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

$$
\begin{align}
a_i^{(l)} &= \sigma(a_i^{(l-1)} w^{(l)} + b^{(l)}) \tag{51} \\
\partial_{w^{(l)}} a_i^{(l)} &= a_i^{(l)}(1 - a_i^{(l)}) a_i^{(l-1)} \tag{52} \\
\partial_{b^{(l)}} a_i^{(l)} &= a_i^{(l)}(1 - a_i^{(l)}) \tag{53} \\
\partial_{a_i^{(l-1)}} a_i^{(l)} &= a_i^{(l)}(1 - a_i^{(l)}) w^{(l)} \tag{54} \\
& \tag{55}
\end{align}
$$

1.4.2 Back-Propagation

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

$$
\begin{align}
C^{(l)} &= \frac{1}{n} \sum_{i=1}^{n} (\partial_{a_i^{(l)}} C^{(l+1)})^2 \tag{56} \\
\partial_{w^{(l)}} C^{(l)} &= \frac{1}{n} \sum_{i=1}^{n} 2(\partial_{a_i^{(l)}} C^{(l+1)}) a_i^{(l)}(1 - a_i^{(l)}) a_i^{(l-1)} = \tag{57} \\
\partial_{b^{(l)}} C^{(l)} &= \frac{1}{n} \sum_{i=1}^{n} 2(\partial_{a_i^{(l)}} C^{(l+1)}) a_i^{(l)}(1 - a_i^{(l)}) \tag{58} \\
\partial_{a_i^{(l-1)}} C^{(l)} &= \frac{1}{n} \sum_{i=1}^{n} 2(\partial_{a_i^{(l)}} C^{(l+1)}) a_i^{(l)}(1 - a_i^{(l)}) w^{(l)} \tag{59} \\
& \tag{60}
\end{align}
$$