1 Gradient Descent

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

1.1 "Double"

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

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

1.2 One neuron model

$$\begin{tikzcd} x \arrow[r] & \sigma, w, b \arrow[r] & y \end{tikzcd}$$

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

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

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

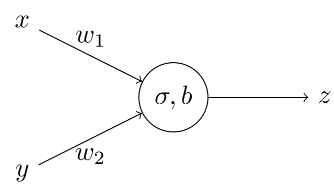
$$\tag{17}$$

1.2.1 Cost

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

$$\tag{31}$$

1.3 One Neuron Model with 2 inputs



$$z = \sigma(x_i w_1 + y_i w_2 + b) \tag{32}$$
$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{33}$$
$$\sigma'(x) = \sigma(x)(1-\sigma(x)) \tag{34}$$
$$\tag{35}$$

1.3.1 Cost

$$a_i = \sigma(x_i w_1 + y_i w_2 + b) \tag{36}$$
$$\partial_{w_1} a_i = \partial_w \left(\sigma(x_i w_1 + y_i w_2 + b)\right) \tag{37}$$
$$= a_i(1-a_i)\partial_w(x_i w_1 + y_i w_2 + b) \tag{38}$$
$$= a_i(1-a_i)x_i \tag{39}$$
$$\partial_{w_2} a_i = a_i(1-a_i)y_i \tag{40}$$
$$\partial_b a_i = a_i(1-a_i) \tag{41}$$
$$C = \frac{1}{n}\sum_{i=1}^{n}(a_i - z_i)^2 \tag{42}$$
$$\partial_{w_1} C = \partial_{w_1}\left(\frac{1}{n}\sum_{i=1}^{n}(a_i - z_i)^2\right) \tag{43}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\partial_{w_1}\left((a_i - z_i)^2\right) \tag{44}$$
$$= \frac{1}{n}\sum_{i=1}^{n}2(a_i - z_i)\partial_{w_1}(a_i - z_i) \tag{45}$$
$$= \frac{1}{n}\sum_{i=1}^{n}2(a_i - z_i)\partial_{w_1} a_i \tag{46}$$
$$= \frac{1}{n}\sum_{i=1}^{n}2(a_i - z_i)(a_i(1-a_i)x_i) \tag{47}$$
$$\partial_{w_2} C = \frac{1}{n}\sum_{i=1}^{n}2(a_i - z_i)(a_i(1-a_i)y_i) \tag{48}$$

$$\partial_b C = \frac{1}{n} \sum_{i=1}^{n} 2(a_i - z_i) \partial_b a_i \tag{49}$$

$$= \frac{1}{n} \sum_{i=1}^{n} 2(a_i - z_i)(a_i(1 - a_i)) \tag{50}$$

$$\tag{51}$$

1.4 Two Neuron Model with 1 input

$$a^{(0)} \xrightarrow{w_1} \boxed{\sigma, b^{(1)}} \xrightarrow{w_2} \boxed{\sigma, b^{(2)}} \longrightarrow a^{(2)}$$

$$\begin{align}
a^{(1)} &= \sigma\left(a^{(0)}w^{(1)} + b^{(1)}\right) \tag{52}\\
a^{(2)} &= \sigma\left(a^{(1)}w^{(2)} + b^{(2)}\right) \tag{53}\\
& \tag{54}
\end{align}$$

1.4.1 Cost

$$\begin{align}
a_i^{(1)} &= \sigma\left(a_i^{(0)}w^{(1)} + b^{(1)}\right) \tag{55}\\
\partial_{w^1} a_i^{(1)} &= a_i^{(1)}(1 - a_i^{(1)})a_i^{(0)} \tag{56}\\
\partial_{b^1} a_i^{(1)} &= a_i^{(1)}(1 - a_i^{(1)}) \tag{57}\\
a_i^{(2)} &= \sigma\left(a_i^{(1)}w^{(2)} + b^{(2)}\right) \tag{58}\\
\partial_{w^{(2)}} a_i^{(2)} &= \partial_{w^{(2)}}\sigma\left(a_i^{(1)}w^{(2)} + b^{(2)}\right) \tag{59}\\
&= a_i^{(2)}(1 - a_i^{(2)})\partial_{w^{(2)}}\left(a_i^{(1)}w^{(2)} + b^{(2)}\right) \tag{60}\\
&= a_i^{(2)}(1 - a_i^{(2)})a_i^{(1)} \tag{61}\\
\partial_{b^{(2)}} a_i^{(2)} &= a_i^{(2)}(1 - a_i^{(2)}) \tag{62}\\
\partial_{a^{(1)}} a_i^{(2)} &= a_i^{(2)}(1 - a_i^{(2)})w^{(2)} \tag{63}\\
C^(2) &= \frac{1}{n}\sum_{i=1}^{n}(a_i^{(2)} - y_i)^2 \tag{64}\\
\partial_{w^{(2)}} C^(2) &= \frac{1}{n}\sum_{i=1}^{n} pd[w^{(2)}](a_i^{(2)} - y_i)^2 \tag{65}\\
&= \frac{1}{n}\sum_{i=1}^{n} 2(a_i^{(2)} - y_i)\partial_{w^{(2)}} a_i^{(2)} \tag{66}\\
&= \frac{1}{n}\sum_{i=1}^{n} 2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)})a_i^{(1)} \tag{67}\\
\partial_{b^{(2)}} C^(2) &= \frac{1}{n}\sum_{i=1}^{n} 2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)}) \tag{68}\\
\partial_{a^{(1)}} C^(2) &= \frac{1}{n}\sum_{i=1}^{n} 2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)})w^{(2)} \tag{69}
\end{align}$$

$$e_i = a_i^{(1)} - \partial_{a_i^{(1)}} C^{(2)} \tag{70}$$

$$C^(1) = \frac{1}{n} \sum_{i=1}^{n} (a_i^{(1)} - e_i)^2 \tag{71}$$

$$\partial_{w^{(1)}} C^(1) = \partial_{w^{(1)}} \left(\frac{1}{n} \sum_{i=1}^{n} (a_i^{(1)} - e_i)^2 \right) \tag{72}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \partial_{w^{(1)}} \left((a_i^{(1)} - e_i)^2 \right) \tag{73}$$

$$= \frac{1}{n} \sum_{i=1}^{n} 2(a_i^{(1)} - e_i) \partial_{w^{(1)}} (a_i^{(1)} - e_i) \tag{74}$$

$$= \frac{1}{n} \sum_{i=1}^{n} 2(a_i^{(1)} - e_i) \partial_{w^{(1)}} a_i^{(1)} \tag{75}$$

$$= \frac{1}{n} \sum_{i=1}^{n} 2(a_i^{(1)} - (a_i^{(1)} - \partial_{a_i^{(1)}} C^{(2)})) \partial_{w^{(1)}} a_i^{(1)} \tag{76}$$

$$= \frac{1}{n} \sum_{i=1}^{n} 2(a_i^{(1)} - a_i^{(1)} + \partial_{a_i^{(1)}} C^{(2)}) \partial_{w^{(1)}} a_i^{(1)} \tag{77}$$

$$= \frac{1}{n} \sum_{i=1}^{n} 2(\partial_{a_i^{(1)}} C^{(2)}) a_i^{(0)} \tag{78}$$

$$\partial_{b^{(1)}} C^(1) = \frac{1}{n} \sum_{i=1}^{n} 2(\partial_{a_i^{(1)}} C^{(2)}) \tag{79}$$

$$\tag{80}$$

1.5 Arbitrary Neurons Model with 1 input

Assume we have m no of layers.

1.5.1 Feed-Forward

Assuming $a_i^{(0)} = x_i$

$$a_i^{(l)} = \sigma\left(a_i^{(l-1)} w^{(l)} + b^{(l)}\right) \tag{81}$$

$$\partial_{w^{(l)}} a_i^{(l)} = a_i^{(l)}(1 - a_i^{(l)}) a_i^{(l-1)} \tag{82}$$

$$\partial_{b^{(l)}} a_i^{(l)} = a_i^{(l)}(1 - a_i^{(l)}) \tag{83}$$

$$\partial_{a^{(l-1)}} a_i^{(l)} = a_i^{(l)}(1 - a_i^{(l)}) w^{(l)} \tag{84}$$

$$\tag{85}$$

1.5.2 Back-Propagation

Assume $a_i^{(m)} - y_i$ is $\partial_{a^{(m)}} C^{(m+1)}$

$$C^{(l)} = \frac{1}{n} \sum_{i=1}^{n} (a_i^{(l)} - (a_i^{(l)} - \partial_{a_i^{(l)}} C^{(l+1)}))^2 \tag{86}$$

$$\partial_{w^{(l)}} C^{(l)} = \frac{1}{n} \sum_{i=1}^{n} 2(\partial_{a_i^{(l)}} C^{(l+1)}) a_i^{(l)} (1 - a_i^{(l)}) a_i^{(l-1)} \tag{87}$$

$$\partial_{b^{(l)}} C^{(l)} = \frac{1}{n} \sum_{i=1}^{n} 2(\partial_{a_i^{(l)}} C^{(l+1)}) a_i^{(l)} (1 - a_i^{(l)}) \tag{88}$$

$$\partial_{a^{(l-1)}} C^{(l)} = \frac{1}{n} \sum_{i=1}^{n} 2(\partial_{a_i^{(l)}} C^{(l+1)}) a_i^{(l)} (1 - a_i^{(l)}) w^{(l)} \tag{89}$$

$$\tag{90}$$