1 Gradient Descent

$$C'(w) = \lim_{\epsilon \to 0} \frac{C(w + \epsilon) - C(w_i)}{\epsilon} \tag{1.0.1}$$

1.1 "Twice"

$$C(w) = \frac{1}{n} \sum_{i=1}^{n} (x_i w - y_i)^2 \tag{1.1.1}$$

$$C'(w) = \left(\frac{1}{n} \sum_{i=1}^{n} (x_i w - y_i)^2 \right)' \tag{1.1.2}$$

$$= \frac{1}{n} \left(\sum_{i=1}^{n} (x_i w - y_i)^2 \right)' \tag{1.1.3}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left((x_i w - y_i)^2 \right)' \tag{1.1.4}$$

$$= \frac{1}{n} \sum_{i=1}^{n} 2(x_i w - y_i) \left(x_i w - y_i \right)' \tag{1.1.5}$$

$$= \frac{1}{n} \sum_{i=1}^{n} 2(x_i w - y_i) x_i \tag{1.1.6}$$

1.2 One Neuron Model with 2 inputs - sigmoid

$$y = \sigma(xw + b) \tag{1.2.1}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{1.2.2}$$

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \tag{1.2.3}$$

1.3 Cost

$$a_i = \sigma(wx + b) \tag{1.3.1}$$

$$C = \frac{1}{n}\sum_{i=0}^{n}(a_i - z_i)^2 \tag{1.3.2}$$

$$\partial_{w1}C = \partial_w\left(\frac{1}{n}\sum_{i=0}^{n}(a_i - z_i)^2\right) \tag{1.3.3}$$

$$= \frac{1}{n}\sum_{i=0}^{n}2(a_i - z_i)\partial_{w1}a_i \tag{1.3.4}$$

$$= \frac{1}{n}\sum_{i=0}^{n}2(a_i - z_i)a_i(1 - a_i)x_i \tag{1.3.5}$$

$$\partial_{w2}C = \frac{1}{n}\sum_{i=0}^{n}2(a_i - z_i)a_i(1 - a_i)y_i \tag{1.3.6}$$

$$\partial_b C = \frac{1}{n}\sum_{i=0}^{n}2(a_i - z_i)a_i(1 - a_i) \tag{1.3.7}$$

1.4 2 Neuron Model with 1 input - Forward Pass

$$a_i^{(1)} = \sigma(x_i w^{(1)} + b^{(1)}) \tag{1.4.1}$$

$$\partial_{w^{(1)}}a_i^{(i)} = 2(\partial_{a^{(1)}}C^{(2)})x_i \tag{1.4.2}$$

$$\partial_{b^{(1)}}a_i^{(i)} = 2(\partial_{a^{(1)}}C^{(2)}) \tag{1.4.3}$$

$$a^{(2)} = \sigma(a^{(1)}w^{(2)} + b^{(2)}) \tag{1.4.4}$$

$$\partial_{w^{(2)}}a_i^{(2)} = a_i^{(2)}(1 - a_i^{(2)})a_i^{(1)} \tag{1.4.5}$$

$$\partial_{b^{(2)}}a_i^{(2)} = a_i^{(2)}(1 - a_i^{(2)}) \tag{1.4.6}$$

$$\partial_{a_i^{(1)}}a_i^{(2)} = a_i^{(2)}(1 - a_i^{(2)})w^{(2)} \tag{1.4.7}$$

1.5 2 Neuron Model with 1 input - Back-Propagation

$$C^{(2)} = \frac{1}{n}\sum_{i=1}^{n}(a_i^{(2)} - y_i)^2 \tag{1.5.1}$$

$$\partial_{w^{(2)}} C^{(2)} = \frac{1}{n}\sum_{i=1}^{n}\partial_{w^{(2)}}((a_i^{(2)} - y_i)^2) \tag{1.5.2}$$

$$= \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(2)} - y_i)\partial_{w^{(2)}} a_i^{(2)} \tag{1.5.3}$$

$$= \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)})a_i^{(1)} \tag{1.5.4}$$

$$\partial_{b^{(2)}} C^{(2)} = \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)}) \tag{1.5.5}$$

$$\partial_{a^{(1)}} C^{(2)} = \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)})w^{(2)} \tag{1.5.6}$$

$$e_i = a_i^{(1)} - (a_i^{(1)} - \partial_{a^{(1)}} C^{(2)}) \tag{1.5.7}$$

$$C^{(1)} = \frac{1}{n}\sum_{i=1}^{n}(a_i^{(1)} - e_i)^2 \tag{1.5.8}$$

$$\partial_{w^{(1)}} C^{(1)} = \partial_{w^{(1)}}\left(\frac{1}{n}\sum_{i=1}^{n}(a_i^{(1)} - e_i)^2\right) \tag{1.5.9}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\partial_{w^{(1)}}\left((a_i^{(1)} - e_i)^2\right) \tag{1.5.10}$$

$$= \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(1)} - e_i)\partial_{w^{(1)}} a_i^{(i)} \tag{1.5.11}$$

$$= \frac{1}{n}\sum_{i=1}^{n}2(\partial_{a_i^{(1)}} C^{(2)})x_i \tag{1.5.12}$$

$$\partial_{b^{(1)}} C^{(1)} = \frac{1}{n}\sum_{i=1}^{n}2(\partial_{a_i^{(1)}} C^{(2)}) \tag{1.5.13}$$

1.6 Forward Pass (general)

Let $a_i^{(0)}$ equal x_i

$$a^{(l)} = \sigma(a^{(l-1)}w^{(l)} + b^{(l)}) \tag{1.6.1}$$

$$\partial_{w^{(l)}} a_i^{(l)} = a_i^{(l)}(1 - a_i^{(l)})a_i^{(l-1)} \tag{1.6.2}$$

$$\partial_{b^{(l)}} a_i^{(l)} = a_i^{(l)}(1 - a_i^{(l)}) \tag{1.6.3}$$

$$\partial_{a_i^{(l-1)}} a_i^{(l)} = a_i^{(l)}(1 - a_i^{(l)})w^{(l)} \tag{1.6.4}$$

1.7 Back-Propagation (general)

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

$$C^{(l)} = \frac{1}{n}\sum_{i=1}^{n}(a_i^{(l)} - e_i)^2 \tag{1.7.1}$$

$$\partial_{w^{(l)}} C^{(l)} = \frac{1}{n}\sum_{i=1}^{n} 2(\partial_{a_i^{(l)}} C^{(l+1)})a_i^{(l)}(1 - a_i^{(l)})x_i \tag{1.7.2}$$

$$\partial_{b^{(l)}} C^{(l)} = \frac{1}{n}\sum_{i=1}^{n} 2(\partial_{a_i^{(l)}} C^{(l+1)})a_i^{(l)}(1 - a_i^{(l)}) \tag{1.7.3}$$

$$\partial_{a_i^{(l-1)}} C^{(l)} = \frac{1}{n}\sum_{i=1}^{n} 2(a_i^{(l)} - y_i)a_i^{(l)}(1 - a_i^{(l)})w^{(l)} \tag{1.7.4}$$