1 Gradient Descent

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

1.1 “Twice”

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

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

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

$$= \frac{1}{n} \sum_{i=1}^{n} \left((x_i w - y_i)^2 \right)' \tag{5}$$

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

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

1.2 One Neuron Model with 2 inputs - sigmoid

$$y = \sigma(xw + b) \tag{8}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{9}$$

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \tag{10}$$

1.3 Cost

$$a_i = \sigma(wx + b) \tag{11}$$

$$C = \frac{1}{n} \sum_{i=0}^{n} (a_i - z_i)^2 \tag{12}$$

$$\partial_{w1} C = \partial_w \left(\frac{1}{n} \sum_{i=0}^{n} (a_i - z_i)^2 \right) \tag{13}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i - z_i) \partial_{w1} a_i \tag{14}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i - z_i) a_i (1 - a_i) x_i \tag{15}$$

$$\partial_{w2} C = \frac{1}{n} \sum_{i=0}^{n} 2(a_i - z_i) a_i (1 - a_i) y_i \tag{16}$$

$$\partial_b C = \frac{1}{n} \sum_{i=0}^{n} 2(a_i - z_i) a_i (1 - a_i) \tag{17}$$

1.4 Two Neuron Model with 1 input

$$a_i^{(1)} = \sigma(x_i w^{(1)} + b^{(1)}) \tag{18}$$

$$a^{(1)} = \sigma(x w^{(1)} + b^{(1)}) \tag{19}$$

$$a^{(2)} = \sigma(a^{(1)} w^{(2)} + b^{(2)}) \tag{20}$$

$$\partial_{w^{(2)}} a_i^{(2)} = a_i^{(2)} (1 - a_i^{(2)}) a_i^{(1)} \tag{21}$$

$$\partial_{b^{(2)}} a_i^{(2)} = a_i^{(2)} (1 - a_i^{(2)}) \tag{22}$$

$$\partial_{a_i^{(1)}} a_i^{(2)} = a_i^{(2)} (1 - a_i^{(2)}) w^{(2)} \tag{23}$$

1.5 Back-Propagation

$$C^{(2)} = \frac{1}{n}\sum_{i=1}^{n}(a_i^{(2)} - y_i)^2 \tag{24}$$

$$\partial_{w^{(2)}} C^{(2)} = \frac{1}{n}\sum_{i=1}^{n}\partial_{w^{(2)}}((a_i^{(2)} - y_i)^2) \tag{25}$$

$$= \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(2)} - y_i)\partial_{w^{(2)}} a_i^{(2)} \tag{26}$$

$$= \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)})a_i^{(1)} \tag{27}$$

$$\partial_{b^{(2)}} C^{(2)} = \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)}) \tag{28}$$

$$\partial_{a^{(1)}} C^{(2)} = \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)})w^{(2)} \tag{29}$$

$$e_i = a_i^{(1)} - (a_i^{(1)} - \partial_{a^{(1)}} C^{(2)}) \tag{30}$$

$$C^{(1)} = \frac{1}{n}\sum_{i=1}^{n}(a_i^{(1)} - e_i)^2 \tag{31}$$

$$\partial_{w^{(1)}} C^{(1)} = \partial_{w^{(1)}}\left(\frac{1}{n}\sum_{i=1}^{n}(a_i^{(1)} - e_i)^2\right) \tag{32}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\partial_{w^{(1)}}\left((a_i^{(1)} - e_i)^2\right) \tag{33}$$

$$= \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(1)} - e_i)\partial_{w^{(1)}} a_i^{(i)} \tag{34}$$

$$= \frac{1}{n}\sum_{i=1}^{n}2(\partial_{a^{(1)}} C^{(2)})x_i \tag{35}$$

$$\partial_{b^{(1)}} C^{(1)} = \frac{1}{n}\sum_{i=1}^{n}2(\partial_{a^{(1)}} C^{(2)})x_i \tag{36}$$