1 Gradient Descent

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

1.1 "Twice"

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

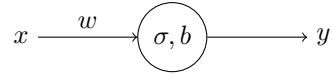$$C'(w) = \left(\frac{1}{n} \sum_{i=1}^{n} (x_i w - y_i)^2 \right)' \tag{3}$$

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

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

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

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

1.2 One Neuron Model / Perceptron

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

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

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

1.2.1 Cost

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

$$(a_i)'_w = (\sigma(x_i w + b))'_w \tag{12}$$

$$= a_i(1 - a_i)(x_i w + b)'_w \tag{13}$$

$$= a_i(1 - a_i)x_i \tag{14}$$

$$(a_i)'_b = a_i(1 - a_i) \tag{15}$$

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

$$C'_w = \left(\frac{1}{n} \sum_{i=0}^{n} (a_i - y_i)^2 \right)'_w \tag{17}$$

$$= \frac{1}{n} \sum_{i=0}^{n} \left((a_i - y_i)^2 \right)'_w \tag{18}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i - y_i) \left(a_i - y_i \right)'_w \tag{19}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i - y_i)(a_i)'_w \tag{20}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i - y_i)a_i(1 - a_i)x_i \tag{21}$$

$$C'_b = \frac{1}{n} \sum_{i=0}^{n} 2(a_i - y_i)a_i(1 - a_i) \tag{22}$$

1.3 One Neuron Model with Two Inputs
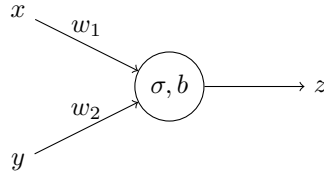
$$y = \sigma(xw_1 + yw_2 + b) \tag{23}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{24}$$

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \tag{25}$$

$$\tag{26}$$

1.3.1 Cost

$$a_i = \sigma(x_i w_1 + y_i w_2 + b) \tag{27}$$

$$(a_i)'_{w_1} = (\sigma(x_i w_1 + y_i w_2 + b))'_{w_1} \tag{28}$$

$$= a_i(1 - a_i)(x_i w_1 + y_i w_2 + b)'_{w_1} \tag{29}$$

$$= a_i(1 - a_i)x_i \tag{30}$$

$$(a_i)'_{w_2} = a_i(1 - a_i)y_i \tag{31}$$

$$(a_i)'_b = a_i(1 - a_i) \tag{32}$$

$$C = \frac{1}{n} \sum_{i=0}^{n} (a_i - z_i)^2 \tag{33}$$

$$C'_{w_1} = \left(\frac{1}{n} \sum_{i=0}^{n} (a_i - z_i)^2 \right)'_{w_1} \tag{34}$$

$$= \frac{1}{n} \sum_{i=0}^{n} \left((a_i - z_i)^2 \right)'_{w_1} \tag{35}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i - z_i) \left(a_i - z_i \right)'_{w_1} \tag{36}$$
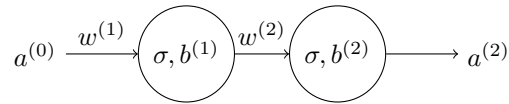
$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i - z_i)(a_i)'_{w_1} \tag{37}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i - z_i)a_i(1 - a_i)x_i \tag{38}$$

$$C'_{w_2} = \frac{1}{n} \sum_{i=0}^{n} 2(a_i - z_i)a_i(1 - a_i)y_i \tag{39}$$

$$C'_b = \frac{1}{n} \sum_{i=0}^{n} 2(a_i - z_i)a_i(1 - a_i) \tag{40}$$

1.4 Two Neuron Model with One Input

$$a^{(1)} = \sigma(a^{(0)} w^{(1)} + b^{(1)}) \tag{41}$$

$$a^{(2)} = \sigma(a^{(1)} w^{(2)} + b^{(2)}) \tag{42}$$

1.4.1 Cost

$$a_i^{(1)} = \sigma(a_i^{(0)} w^{(1)} + b^{(1)}) \tag{43}$$

$$a_i^{(2)} = \sigma(a_i^{(1)} w^{(2)} + b^{(2)}) \tag{44}$$

$$(a_i^{(2)})'_{w^{(2)}} = \left(\sigma(a_i^{(1)} w^{(2)} + b^{(2)})\right)'_{w^{(2)}} \tag{45}$$

$$= a_i^{(2)}(1 - a_i^{(2)})(a_i^{(1)} w^{(2)} + b^{(2)})'_{w^{(2)}} \tag{46}$$

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

$$(a_i^{(2)})'_{b^{(2)}} = \left(\sigma(a_i^{(1)} w^{(2)} + b^{(2)})\right)'_{b^{(2)}} \tag{48}$$

$$= a_i^{(2)}(1 - a_i^{(2)}) \tag{49}$$

$$(a_i^{(2)})'_{a^{(1)}} = \left(\sigma(a_i^{(1)} w^{(2)} + b^{(2)})\right)'_{a^{(1)}} \tag{50}$$

$$= a_i^{(2)}(1 - a_i^{(2)})(a_i^{(1)} w^{(2)} + b^{(2)})'_{a^{(1)}} \tag{51}$$

$$= a_i^{(2)}(1 - a_i^{(2)}) w^{(2)} \tag{52}$$

$$(a_i^{(1)})'_{w^{(1)}} = \sigma(a_i^{(0)} w^{(1)} + b^{(1)})'_{w^{(1)}} \tag{53}$$

$$= a_i^{(1)}(1 - a_i^{(1)})(a_i^{(0)} w^{(1)} + b^{(1)})'_{w^{(1)}} \tag{54}$$

$$= a_i^{(1)}(1 - a_i^{(1)}) a_i^{(0)} \tag{55}$$

$$(a_i^{(1)})'_{b^{(1)}} = \sigma(a_i^{(0)} w^{(1)} + b^{(1)})'_{b^{(1)}} \tag{56}$$

$$= a_i^{(1)}(1 - a_i^{(1)}) \tag{57}$$

$$C^{(2)} = \frac{1}{n} \sum_{i=0}^{n} (a_i^{(2)} - y_i)^2 \tag{58}$$

$$C_{w^{(2)}}'^{(2)} = \left(\frac{1}{n} \sum_{i=0}^{n} (a_i^{(2)} - y_i)^2 \right)_{w^{(2)}}' \tag{59}$$

$$= \frac{1}{n} \sum_{i=0}^{n} ((a_i^{(2)} - y_i)^2)_{w^{(2)}}' \tag{60}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i^{(2)} - y_i)(a_i^{(2)} - y_i)_{w^{(2)}}' \tag{61}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i^{(2)} - y_i)(a_i^{(2)})_{w^{(2)}}' \tag{62}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)})a_i^{(1)} \tag{63}$$

$$C_{b^{(2)}}'^{(2)} = \left(\frac{1}{n} \sum_{i=0}^{n} (a_i^{(2)} - y_i)^2 \right)_{b^{(2)}}' \tag{64}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i^{(2)} - y_i)(a_i^{(2)})_{b^{(2)}}' \tag{65}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)}) \tag{66}$$

$$C_{a_i^{(1)}}'^{(2)} = \left(\frac{1}{n} \sum_{i=0}^{n} (a_i^{(2)} - y_i)^2 \right)_{a_i^{(1)}}' \tag{67}$$

$$= \frac{1}{n} \sum_{i=0}^{n} ((a_i^{(2)} - y_i)^2)_{a_i^{(1)}}' \tag{68}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i^{(2)} - y_i)(a_i^{(2)} - y_i)_{a_i^{(1)}}' \tag{69}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i^{(2)} - y_i)(a_i^{(2)})_{a_i^{(1)}}' \tag{70}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i^{(2)} - y_i)a_i^{(2)}(1 - a_i^{(2)})w^{(2)} \tag{71}$$

$$e_i^{(1)} = a_i^{(1)} - C_{a_i^{(1)}}'^{(2)} \tag{72}$$

$$C^{(1)} = \frac{1}{n} \sum_{i=0}^{n} (a_i^{(1)} - e_i^{(1)})^2 \tag{73}$$

$$C_{w^{(1)}}^{(1)} = \left(\frac{1}{n} \sum_{i=0}^{n} (a_i^{(1)} - e_i^{(1)})^2 \right)_{w^{(1)}}' \tag{74}$$

$$= \frac{1}{n} \sum_{i=0}^{n} \left((a_i^{(1)} - e_i^{(1)})^2 \right)_{w^{(1)}}' \tag{75}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i^{(1)} - e_i^{(1)})(a_i^{(1)})_{w^{(1)}}' \tag{76}$$

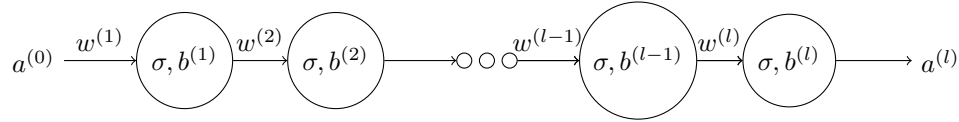$$= \frac{1}{n} \sum_{i=0}^{n} 2C_{a_i^{(1)}}'^{(2)} a_i^{(1)}(1 - a_i^{(1)})a_i^{(0)} \tag{77}$$

$$.C^{(1)}_{b^{(1)}} = \left(\frac{1}{n} \sum_{i=0}^{n} (a_i^{(1)} - e_i^{(1)})^2 \right)'_{b^{(1)}} \tag{78}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2(a_i^{(1)} - e_i^{(1)})(a_i^{(1)})'_{b^{(1)}} \tag{79}$$

$$= \frac{1}{n} \sum_{i=0}^{n} 2C'^{(2)}_{a_i^{(1)}} a_i^{(1)}(1 - a_i^{(1)}) \tag{80}$$

1.5 Arbitrary Neuron Model with One Input

Let's assume that we have m layers.

$$a^{(1)} = \sigma(a^{(0)} w^{(1)} + b^{(1)}) \tag{81}$$
$$a^{(2)} = \sigma(a^{(1)} w^{(2)} + b^{(2)}) \tag{82}$$

1.5.1 Feed-Forward

$$a_i^{(l)} = \sigma(a_i^{(l-1)} w^{(l)} + b^{(l)}) \tag{83}$$

$$(a_i^{(l)})'_{w^{(l)}} = \left(\sigma(a_i^{(l-1)} w^{(l)} + b^{(l)}) \right)'_{w^{(l)}} \tag{84}$$

$$= a_i^{(l)}(1 - a_i^{(l)})(a_i^{(l-1)} w^{(l)} + b^{(l)})'_{w^{(l)}} \tag{85}$$

$$= a_i^{(l)}(1 - a_i^{(l)})a_i^{(l-1)} \tag{86}$$

$$(a_i^{(l)})'_{b^{(l)}} = \left(\sigma(a_i^{(l-1)} w^{(l)} + b^{(l)}) \right)'_{b^{(l)}} \tag{87}$$

$$= a_i^{(l)}(1 - a_i^{(l)}) \tag{88}$$

$$(a_i^{(l)})'_{a^{(l-1)}} = \left(\sigma(a_i^{(l-1)} w^{(l)} + b^{(l)}) \right)'_{a^{(l-1)}} \tag{89}$$

$$= a_i^{(l)}(1 - a_i^{(l)})(a_i^{(l-1)} w^{(l)} + b^{(l)})'_{a^{(l-1)}} \tag{90}$$

$$= a_i^{(l)}(1 - a_i^{(l)})w^{(l)} \tag{91}$$

1.5.2   Back-Propagation

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

$$C^{(l)} = \frac{1}{n}\sum_{i=0}^{n}(C'^{(l+1)}_{a_i^{(l)}})^2 \tag{92}$$

$$C^{(l)}_{w^{(l)}} = \left(\frac{1}{n}\sum_{i=0}^{n}(C'^{(l+1)}_{a_i^{(l)}})^2\right)'_{w^{(l)}} \tag{93}$$

$$= \frac{1}{n}\sum_{i=0}^{n}\left((C'^{(l+1)}_{a_i^{(l)}})^2\right)'_{w^{(l)}} \tag{94}$$

$$= \frac{1}{n}\sum_{i=0}^{n}2(C'^{(l+1)}_{a_i^{(l)}})(a_i^{(l)})'_{w^{(l)}} \tag{95}$$

$$= \frac{1}{n}\sum_{i=0}^{n}2C'^{(l+1)}_{a_i^{(l)}}a_i^{(l)}(1 - a_i^{(l)})a_i^{(0)} \tag{96}$$

$$C^{(l)}_{b^{(l)}} = \left(\frac{1}{n}\sum_{i=0}^{n}(C'^{(l+1)}_{a_i^{(l)}})^2\right)'_{b^{(l)}} \tag{97}$$

$$= \frac{1}{n}\sum_{i=0}^{n}2(C'^{(l+1)}_{a_i^{(l)}})(a_i^{(l)})'_{b^{(l)}} \tag{98}$$

$$= \frac{1}{n}\sum_{i=0}^{n}2C'^{(l+1)}_{a_i^{(l)}}a_i^{(l)}(1 - a_i^{(l)}) \tag{99}$$

$$C'^{l+1}_{a_i^{(l)}} = \left(\frac{1}{n}\sum_{i=0}^{n}(C'^{(l+1)}_{a_i^{(l)}})^2\right)'_{a_i^{(l)}} \tag{100}$$

$$= \frac{1}{n}\sum_{i=0}^{n}((C'^{(l+1)}_{a_i^{(l)}})^2)'_{a_i^{(l)}} \tag{101}$$

$$= \frac{1}{n}\sum_{i=0}^{n}2(C'^{(l+1)}_{a_i^{(l)}})(C'^{(l+1)}_{a_i^{(l)}})'_{a_i^{(l)}} \tag{102}$$

$$= \frac{1}{n}\sum_{i=0}^{n}2(C'^{(l+1)}_{a_i^{(l)}})(a_i^{l+1})'_{a_i^{(l)}} \tag{103}$$

$$= \frac{1}{n}\sum_{i=0}^{n}2(C'^{(l+1)}_{a_i^{(l)}})a_i^{l+1}(1 - a_i^{l+1})w^{l+1} \tag{104}$$

$$\tag{105}$$