

1 Gradient Descent

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

1.1 “Twice”

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

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

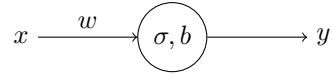
$$= \frac{1}{n} \left(\sum_{i=1}^n (x_i w - y_i)^2 \right)' \quad (4)$$

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

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

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

1.2 One Neuron Model / Perceptron



$$y = \sigma(xw + b) \quad (8)$$

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

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \quad (10)$$

1.2.1 Cost

$$a_i = \sigma(x_i w + b) \quad (11)$$

$$(a_i)'_w = (\sigma(x_i w + b))'_w \quad (12)$$

$$= a_i(1 - a_i)(x_i w + b)'_w \quad (13)$$

$$= a_i(1 - a_i)x_i \quad (14)$$

$$(a_i)'_b = a_i(1 - a_i) \quad (15)$$

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

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

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

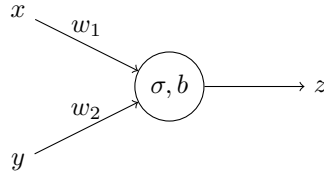
$$= \frac{1}{n} \sum_{i=0}^n 2(a_i - y_i)(a_i - y_i)'_w \quad (19)$$

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

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

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

1.3 One Neuron Model with Two Inputs



$$y = \sigma(xw_1 + yw_2 + b) \quad (23)$$

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

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \quad (25)$$

$$(26)$$

1.3.1 Cost

$$a_i = \sigma(x_i w_1 + y_i w_2 + b) \quad (27)$$

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

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

$$= a_i(1 - a_i)x_i \quad (30)$$

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

$$(a_i)'_b = a_i(1 - a_i) \quad (32)$$

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

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

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

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

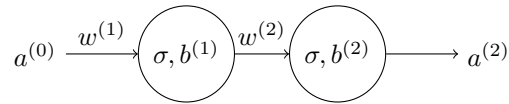
$$= \frac{1}{n} \sum_{i=0}^n 2(a_i - z_i)(a_i)'_{w_1} \quad (37)$$

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

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

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

1.4 Two Neuron Model with One Input



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

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

1.4.1 Cost

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

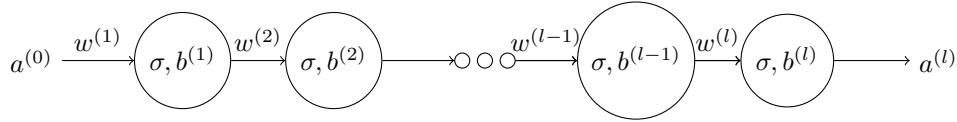
$$C_{b^{(1)}}^{(1)} = \left(\frac{1}{n} \sum_{i=0}^n (a_i^{(1)} - e_i^{(1)})^2 \right)'_{b^{(1)}} \quad (78)$$

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

$$= \frac{1}{n} \sum_{i=0}^n 2C_{a_i^{(1)}}'^{(2)} a_i^{(1)} (1 - a_i^{(1)}) \quad (80)$$

1.5 Arbitrary Neuron Model with One Input

Let's assume that we have m layers.



$$a^{(l+1)} = \sigma(a^{(l)} w^{(l+1)} + b^{(l+1)}) \quad (81)$$

1.5.1 Feed-Forward

$$a_i^{(l)} = \sigma(a_i^{(l-1)} w^{(l)} + b^{(l)}) \quad (82)$$

$$(a_i^{(l)})'_{w^{(l)}} = \left(\sigma(a_i^{(l-1)} w^{(l)} + b^{(l)}) \right)'_{w^{(l)}} \quad (83)$$

$$= a_i^{(l)} (1 - a_i^{(l)}) (a_i^{(l-1)} w^{(l)} + b^{(l)})'_{w^{(l)}} \quad (84)$$

$$= a_i^{(l)} (1 - a_i^{(l)}) a_i^{(l-1)} \quad (85)$$

$$(a_i^{(l)})'_{b^{(l)}} = \left(\sigma(a_i^{(l-1)} w^{(l)} + b^{(l)}) \right)'_{b^{(l)}} \quad (86)$$

$$= a_i^{(l)} (1 - a_i^{(l)}) \quad (87)$$

$$(a_i^{(l)})'_{a^{(l-1)}} = \left(\sigma(a_i^{(l-1)} w^{(l)} + b^{(l)}) \right)'_{a^{(l-1)}} \quad (88)$$

$$= a_i^{(l)} (1 - a_i^{(l)}) (a_i^{(l-1)} w^{(l)} + b^{(l)})'_{a^{(l-1)}} \quad (89)$$

$$= a_i^{(l)} (1 - a_i^{(l)}) w^{(l)} \quad (90)$$

1.5.2 Back-Propagation

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

$$C^{(l)} = \frac{1}{n} \sum_{i=0}^n (C'_{a_i^{(l)}})^2 \quad (91)$$

$$C'_{w^{(l)}}^{(l)} = \left(\frac{1}{n} \sum_{i=0}^n (C'_{a_i^{(l)}})^2 \right)'_{w^{(l)}} \quad (92)$$

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

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

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

$$C'_{b^{(l)}}^{(l)} = \left(\frac{1}{n} \sum_{i=0}^n (C'_{a_i^{(l)}})^2 \right)'_{b^{(l)}} \quad (96)$$

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

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

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

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

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

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

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

$$(104)$$