

# 1 Gradient Descent

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

## 1.1 “Double“

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

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

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

$$= \frac{1}{n} \left( (x_0 w - y_0)^2 + (x_1 w - y_1)^2 + \dots + (x_n w - y_n)^2 \right)' \quad (5)$$

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

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

$$= \frac{1}{n} \sum_{i=1}^n 2(x_i w - y_i) (x_i)' \quad (8)$$

$$= \frac{1}{n} \sum_{i=1}^n 2(x_i w - y_i) x_i' \quad (9)$$

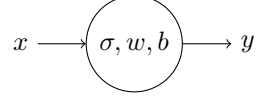
$$(10)$$

$$C(w) = \frac{1}{n} \sum_{i=1}^n (x_i w - y_i)^2 \quad (11)$$

$$C'(w) = \frac{1}{n} \sum_{i=1}^n 2(x_i w - y_i) x_i' \quad (12)$$

$$(13)$$

## 1.2 One neuron model



$$y = \sigma(xw + b) \quad (14)$$

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

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \quad (16)$$

$$(17)$$

### 1.2.1 Cost

$$a_i = \sigma(x_i w + b) \quad (18)$$

$$\partial_w a_i = \partial_w (\sigma(x_i w + b)) \quad (19)$$

$$= a_i(1 - a_i)\partial_w(x_i w + b) \quad (20)$$

$$= a_i(1 - a_i)x_i \quad (21)$$

$$\partial_b a_i = a_i(1 - a_i) \quad (22)$$

$$C = \frac{1}{n} \sum_{i=1}^n (a_i - y_i)^2 \quad (23)$$

$$\partial_w C = \partial_w \left( \frac{1}{n} \sum_{i=1}^n (a_i - y_i)^2 \right) \quad (24)$$

$$= \frac{1}{n} \sum_{i=1}^n \partial_w ((a_i - y_i)^2) \quad (25)$$

$$= \frac{1}{n} \sum_{i=1}^n 2(a_i - y_i) \partial_w (a_i - y_i) \quad (26)$$

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

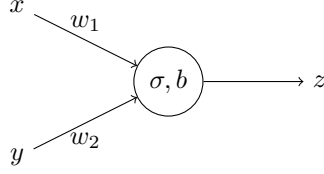
$$= \frac{1}{n} \sum_{i=1}^n 2(a_i - y_i)(a_i(1 - a_i)x_i) \quad (28)$$

$$\partial_b C = \frac{1}{n} \sum_{i=1}^n 2(a_i - y_i) \partial_b a_i \quad (29)$$

$$= \frac{1}{n} \sum_{i=1}^n 2(a_i - y_i)(a_i(1 - a_i)) \quad (30)$$

$$(31)$$

### 1.3 One Neuron Model with 2 inputs



$$z = \sigma(x_i w_1 + y_i w_2 + b) \quad (32)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (33)$$

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \quad (34)$$

$$(35)$$

#### 1.3.1 Cost

$$a_i = \sigma(x_i w_1 + y_i w_2 + b) \quad (36)$$

$$\partial_{w_1} a_i = \partial_w (\sigma(x_i w_1 + y_i w_2 + b)) \quad (37)$$

$$= a_i(1 - a_i) \partial_w (x_i w_1 + y_i w_2 + b) \quad (38)$$

$$= a_i(1 - a_i) x_i \quad (39)$$

$$\partial_{w_2} a_i = a_i(1 - a_i) y_i \quad (40)$$

$$\partial_b a_i = a_i(1 - a_i) \quad (41)$$

$$C = \frac{1}{n} \sum_{i=1}^n (a_i - z_i)^2 \quad (42)$$

$$\partial_{w_1} C = \partial_{w_1} \left( \frac{1}{n} \sum_{i=1}^n (a_i - z_i)^2 \right) \quad (43)$$

$$= \frac{1}{n} \sum_{i=1}^n \partial_{w_1} ((a_i - z_i)^2) \quad (44)$$

$$= \frac{1}{n} \sum_{i=1}^n 2(a_i - z_i) \partial_{w_1} (a_i - z_i) \quad (45)$$

$$= \frac{1}{n} \sum_{i=1}^n 2(a_i - z_i) \partial_{w_1} a_i \quad (46)$$

$$= \frac{1}{n} \sum_{i=1}^n 2(a_i - z_i) (a_i(1 - a_i) x_i) \quad (47)$$

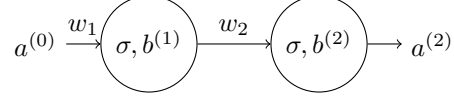
$$\partial_{w_2} C = \frac{1}{n} \sum_{i=1}^n 2(a_i - z_i) (a_i(1 - a_i) y_i) \quad (48)$$

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

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

$$(51)$$

## 1.4 Two Neuron Model with 1 input



$$a^{(1)} = \sigma \left( a^{(0)} w^{(1)} + b^{(1)} \right) \quad (52)$$

$$a^{(2)} = \sigma \left( a^{(1)} w^{(2)} + b^{(2)} \right) \quad (53)$$

$$(54)$$

### 1.4.1 Cost

$$a_i^{(1)} = \sigma \left( a_i^{(0)} w^{(1)} + b^{(1)} \right) \quad (55)$$

$$\partial_{w^{(1)}} a_i^{(1)} = a_i^{(1)} (1 - a_i^{(1)}) a_i^{(0)} \quad (56)$$

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

$$a_i^{(2)} = \sigma \left( a_i^{(1)} w^{(2)} + b^{(2)} \right) \quad (58)$$

$$\partial_{w^{(2)}} a_i^{(2)} = \partial_{w^{(2)}} \sigma \left( a_i^{(1)} w^{(2)} + b^{(2)} \right) \quad (59)$$

$$= a_i^{(2)} (1 - a_i^{(2)}) \partial_{w^{(2)}} \left( a_i^{(1)} w^{(2)} + b^{(2)} \right) \quad (60)$$

$$= a_i^{(2)} (1 - a_i^{(2)}) a_i^{(1)} \quad (61)$$

$$\partial_{b^{(2)}} a_i^{(2)} = a_i^{(2)} (1 - a_i^{(2)}) \quad (62)$$

$$\partial_{a^{(1)}} a_i^{(2)} = a_i^{(2)} (1 - a_i^{(2)}) w^{(2)} \quad (63)$$

$$C^{(2)} = \frac{1}{n} \sum_{i=1}^n (a_i^{(2)} - y_i)^2 \quad (64)$$

$$\partial_{w^{(2)}} C^{(2)} = \frac{1}{n} \sum_{i=1}^n 2(a_i^{(2)} - y_i) \partial_{w^{(2)}} a_i^{(2)} \quad (65)$$

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

$$= \frac{1}{n} \sum_{i=1}^n 2(a_i^{(2)} - y_i) a_i^{(2)} (1 - a_i^{(2)}) a_i^{(1)} \quad (67)$$

$$\partial_{b^{(2)}} C^{(2)} = \frac{1}{n} \sum_{i=1}^n 2(a_i^{(2)} - y_i) a_i^{(2)} (1 - a_i^{(2)}) \quad (68)$$

$$\partial_{a^{(1)}} C^{(2)} = \frac{1}{n} \sum_{i=1}^n 2(a_i^{(2)} - y_i) a_i^{(2)} (1 - a_i^{(2)}) w^{(2)} \quad (69)$$

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

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

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

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

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

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

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

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

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

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

$$(80)$$

## 1.5 Arbitrary Neurons Model with 1 input

### 1.5.1 Feed-Forward

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

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

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

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

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

$$(85)$$

### 1.5.2 Back-Propagation

$$C^{(2)} = \frac{1}{n} \sum_{i=1}^n (a_i^{(2)} - y_i)^2 \quad (86)$$

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

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

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

$$C^{(1)} = \frac{1}{n} \sum_{i=1}^n (a_i^{(1)} - (a_i^{(1)} - \partial_{a_i^{(1)}} C^{(2)}))^2 \quad (90)$$

$$\partial_{w^{(1)}} C^{(1)} = \frac{1}{n} \sum_{i=1}^n 2(\partial_{a_i^{(1)}} C^{(2)}) a_i^{(0)} \quad (91)$$

$$\partial_{b^{(1)}} C^{(1)} = \frac{1}{n} \sum_{i=1}^n 2(\partial_{a_i^{(1)}} C^{(2)}) \quad (92)$$

$$(93)$$