

1 Gradient Descent

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

(2)

1.1 Derivative Of One Parameter Function

Within the *Twice* example we described a model with one parameter - w

The formula had a form like this:

$$f(x) = x \cdot w \quad (3)$$

Function C which takes one parameter w is defined as:

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

Let's compute the derivative C' of our function:

$$C'(w) = (C)' \quad (5)$$

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

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

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

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

$$= \frac{1}{n} \sum_{i=1}^n (2 \cdot (x_i \cdot w - y_i)(x_i \cdot w - y_i)') = \quad (10)$$

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

The final form of our derivative:

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

1.2 One Neuron Model With 2 Inputs

One neuron model is defined as:

$$z = \sigma(x \cdot w_1 + y \cdot w_2 + b) \quad (13)$$

x_1 ... input parameter

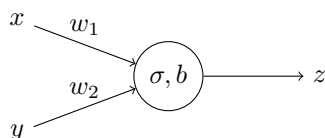
x_2 ... input parameter

w_1 ... weight parameter

w_2 ... weight parameter

b ... bias parameter

σ ... sigmoid activation function



1.2.1 Cost

Let's recall the Sigmoid activation function

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (14)$$

$$\sigma(x)' = \sigma(x) \cdot (1 - \sigma(x)) \quad (15)$$

Let's define the cost function C for our model

$$a_i = \sigma(x_i \cdot w_1 + y_i \cdot w_2 + b) \quad (16)$$

$$C(x) = \frac{1}{n} \sum_{i=1}^n (a_i - y_i)^2 \quad (17)$$

Let's compute the derivative C' for our function

We have to modify TWO parameters w, b

For this we will use PARTIAL DERIVATIVES this means that first we compute a derivative in respect to w_1, w_2 and then we compute another derivative in respect to b

1. Partial Derivative in respect to w_1

$$a_i = \sigma(x_i \cdot w_1 + y_i \cdot w_2 + b) = \quad (18)$$

$$\partial_{w_1} a_i = \partial_{w_1} (\sigma(x_i \cdot w_1 + y_i \cdot w_2 + b)) = \quad (19)$$

$$= a_i(1 - a_i) \partial_{w_1} (x_i \cdot w_1 + y_i \cdot w_2 + b) = \quad (20)$$

$$\partial_{w_1} a_i = a_i(1 - a_i) \cdot x_i \quad (21)$$

$$(22)$$

$$\partial_{w_1} C = \partial_{w_1} \left(\frac{1}{n} \sum_{i=1}^n (a_i - z_i)^2 \right) = \quad (23)$$

$$= \frac{1}{n} \sum_{i=1}^n \partial_{w_1} ((a_i - z_i)^2) = \quad (24)$$

$$= \frac{1}{n} \sum_{i=1}^n 2 \cdot (a_i - z_i) \partial_{w_1} (a_i - z_i) = \quad (25)$$

$$= \frac{1}{n} \sum_{i=1}^n 2 \cdot (a_i - z_i) \partial_{w_1} a_i = \quad (26)$$

$$\partial_{w_1} C = \frac{1}{n} \sum_{i=1}^n 2 \cdot (a_i - z_i) \cdot a_i(1 - a_i) \cdot x_i \quad (27)$$

$$(28)$$

2. Partial Derivative in respect to w_2

$$a_i = \sigma(x_i \cdot w_1 + y_i \cdot w_2 + b) = \quad (29)$$

$$\partial_{w_2} a_i = \partial_{w_2} (\sigma(x_i \cdot w_1 + y_i \cdot w_2 + b)) = \quad (30)$$

$$= a_i(1 - a_i) \partial_{w_2} (x_i \cdot w_1 + y_i \cdot w_2 + b) = \quad (31)$$

$$\partial_{w_2} a_i = a_i(1 - a_i) \cdot y_i \quad (32)$$

$$(33)$$

$$\partial_{w_2} C = \partial_{w_2} \left(\frac{1}{n} \sum_{i=1}^n (a_i - z_i)^2 \right) = \quad (34)$$

$$= \frac{1}{n} \sum_{i=1}^n \partial_{w_2} ((a_i - z_i)^2) = \quad (35)$$

$$= \frac{1}{n} \sum_{i=1}^n 2 \cdot (a_i - z_i) \partial_{w_2} (a_i - z_i) = \quad (36)$$

$$= \frac{1}{n} \sum_{i=1}^n 2 \cdot (a_i - z_i) \partial_{w_2} a_i = \quad (37)$$

$$\partial_{w_2} C = \frac{1}{n} \sum_{i=1}^n 2 \cdot (a_i - z_i) \cdot a_i(1 - a_i) \cdot y_i \quad (38)$$

$$(39)$$

3. Partial Derivative in respect to b

$$a_i = \sigma(x_i \cdot w_1 + y_i \cdot w_2 + b) \quad (40)$$

$$\partial_b a_i = \partial_b (\sigma(x_i \cdot w_1 + y_i \cdot w_2 + b)) = \quad (41)$$

$$= a_i(1 - a_i) \partial_b (x_i \cdot w_1 + y_i \cdot w_2 + b) = \quad (42)$$

$$= a_i(1 - a_i) \cdot 1 = \quad (43)$$

$$\partial_b a_i = a_i(1 - a_i) \quad (44)$$

$$(45)$$

$$\partial_b C = \partial_b \left(\frac{1}{n} \sum_{i=1}^n (a_i - z_i)^2 \right) = \quad (46)$$

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

$$= \frac{1}{n} \sum_{i=1}^n 2 \cdot (a_i - z_i) \partial_b (a_i - z_i) = \quad (48)$$

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

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

$$(51)$$

To summarize the partial derivatives are:

$$a_i = \sigma(x_i \cdot w_1 + y_i \cdot w_2 + b) \quad (52)$$

$$\partial_{w_1} C = \frac{1}{n} \sum_{i=1}^n 2 \cdot (a_i - z_i) \cdot a_i(1 - a_i) \cdot x_i \quad (53)$$

$$\partial_{w_2} C = \frac{1}{n} \sum_{i=1}^n 2 \cdot (a_i - z_i) \cdot a_i(1 - a_i) \cdot y_i \quad (54)$$

$$\partial_b C = \frac{1}{n} \sum_{i=1}^n 2 \cdot (a_i - z_i) \cdot a_i(1 - a_i) \quad (55)$$

$$(56)$$

1.3 Execution Time Comparison

Let's compare computation time difference between **Finite Difference** and **Gradient Descent**

My machine is Lenovo Legion Slim 5:

- All computations are run on the CPU
- CPU: AMD Rayzen 7 7840HS (16) 5.137Ghz

The test:

- Neural network will try to learn the proper configuration for simulating NAND gate
- Comparison of training the model using the *Finite Difference* method and *Gradient Descent*
- 8.000.000 iterations(epochs) of training will be run (overkill I know)

RESULTS:

Finite Difference : ≈ 1.556 *seconds*

Gradient Descent : ≈ 0.473 *seconds*