TRAINING & BACKPROPAGATION

Presented by the Mac AI Team!

Sarah Simionescu, 2024

LEVELS OF UNDERSTANDING

LEVELS OF ABSTRACTION

“WOW, A PREDICTION MACHINE!”

Input Health Data, Receive Prediction!

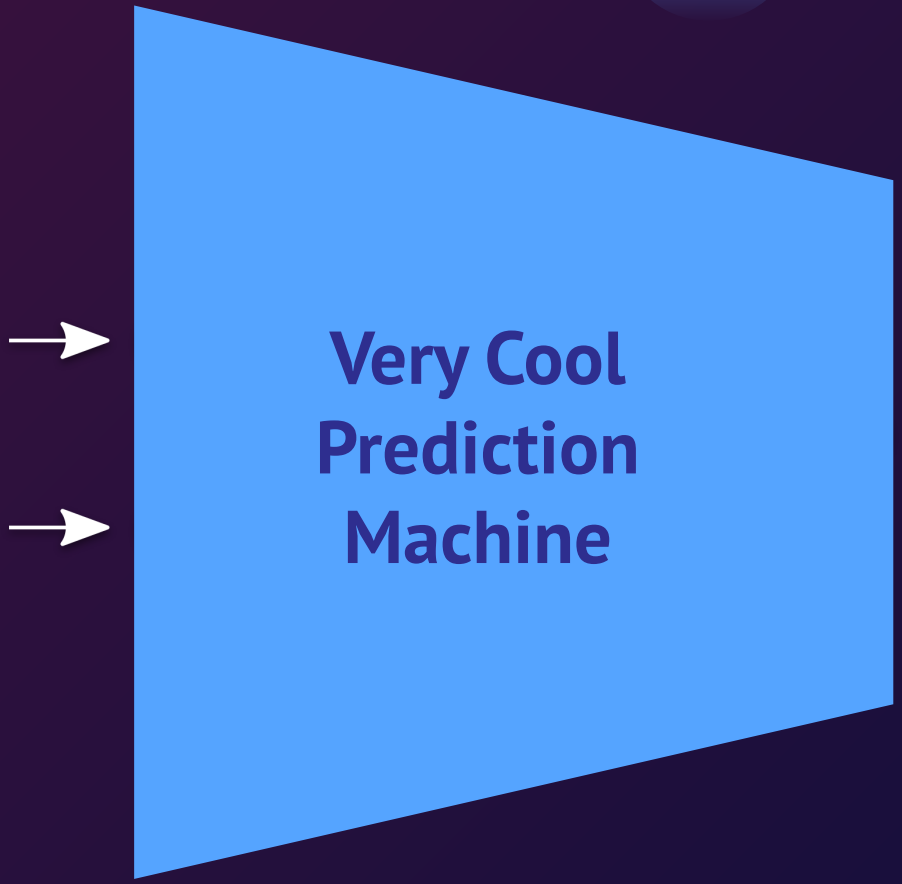
“WOAH, IT’S A BRAIN?”

From Raw Data to Refine Insight, Layer by Layer.

“OH... IT’S MATH.”

Breaking Down to Weights and Biases.

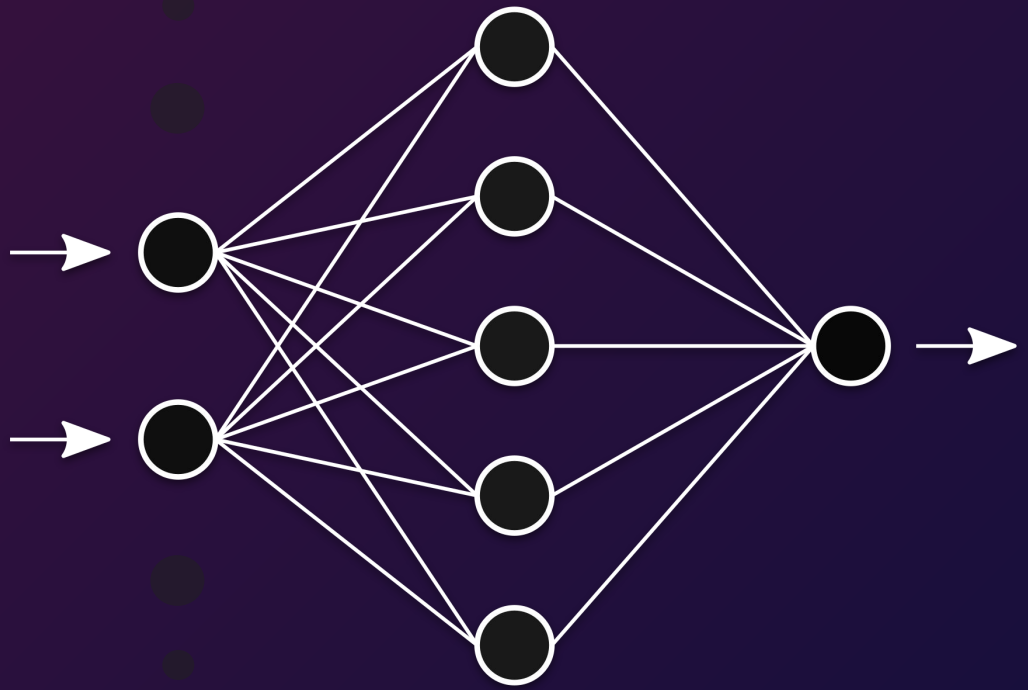
radius1	17.99
texture1	10.38
…	

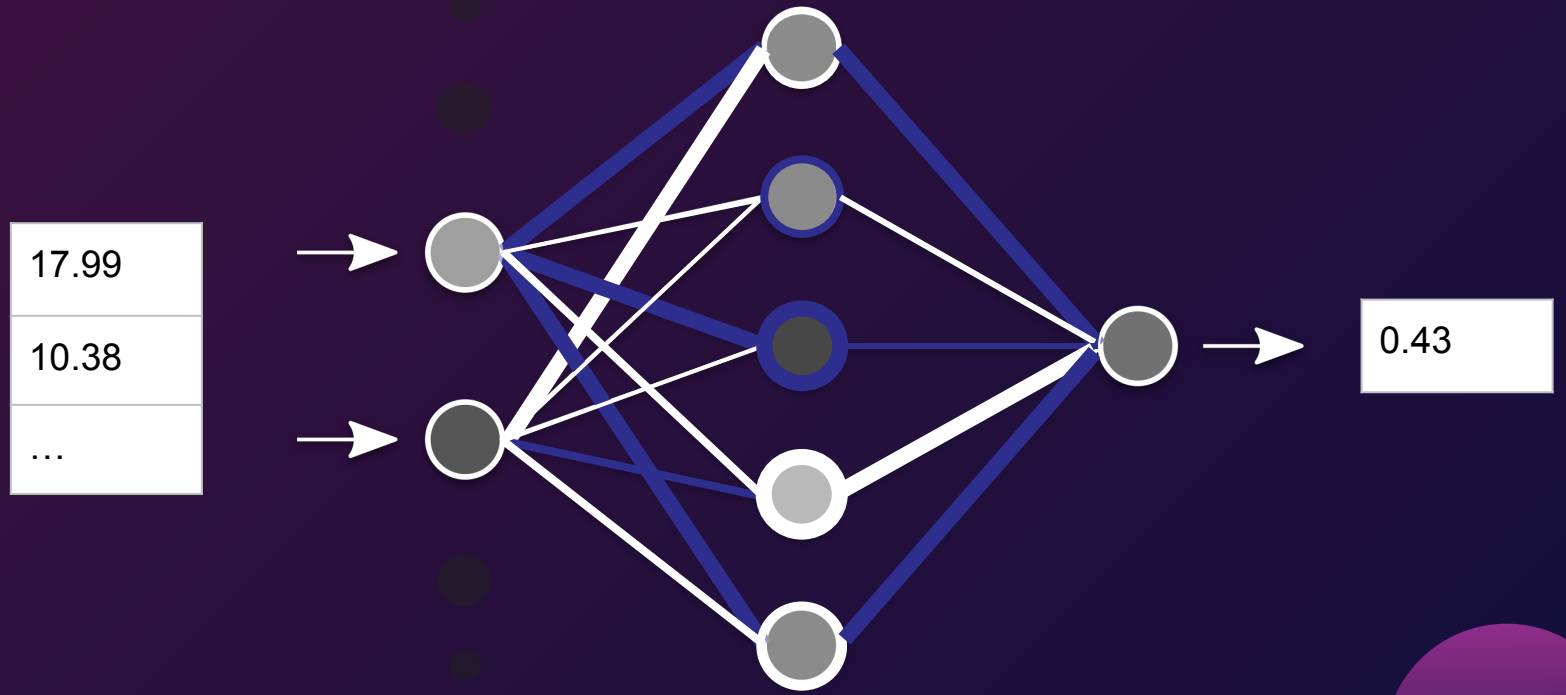


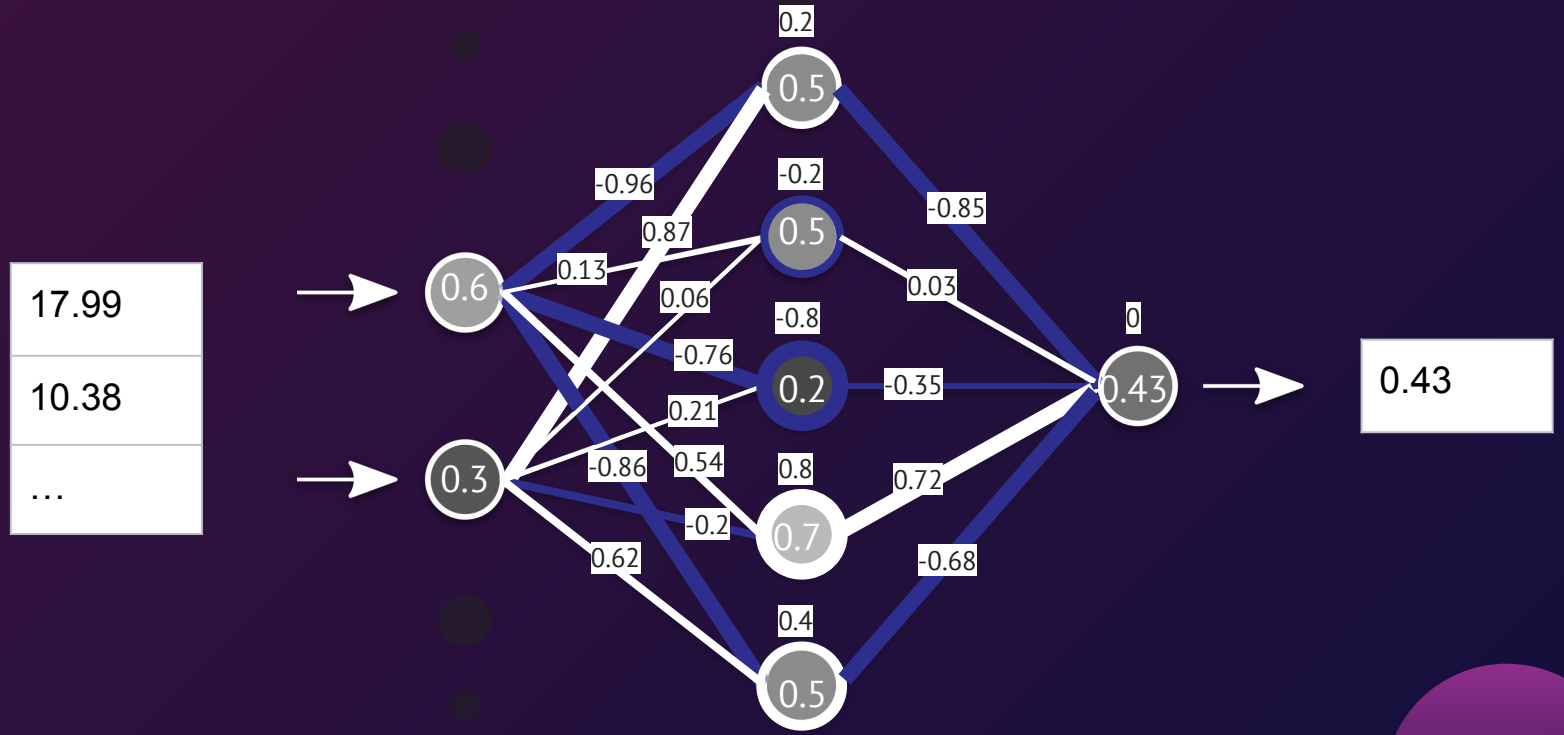
Likelihood Tumor is Malignant

43%

17.99
10.38
…




17.99
10.38
…
0.43


17.99
10.38
...
0.6
0.3
0.2
0.5
-0.96
0.87
0.13
0.06
-0.2
0.5
-0.85
0.03
-0.8
0.2
-0.76
0.21
-0.35
0
0.43
-0.86
0.54
0.8
0.72
-0.2
0.7
0.62
-0.68
0.4
0.5
0.43

BASIC TRAINING PROCESS

STEP 1

Initialize the weights and biases

STEP 2

Test it on a bunch of training data!

STEP 3

Calculate the score (the difference)

STEP 4

Calculate the differences for each weight and bias

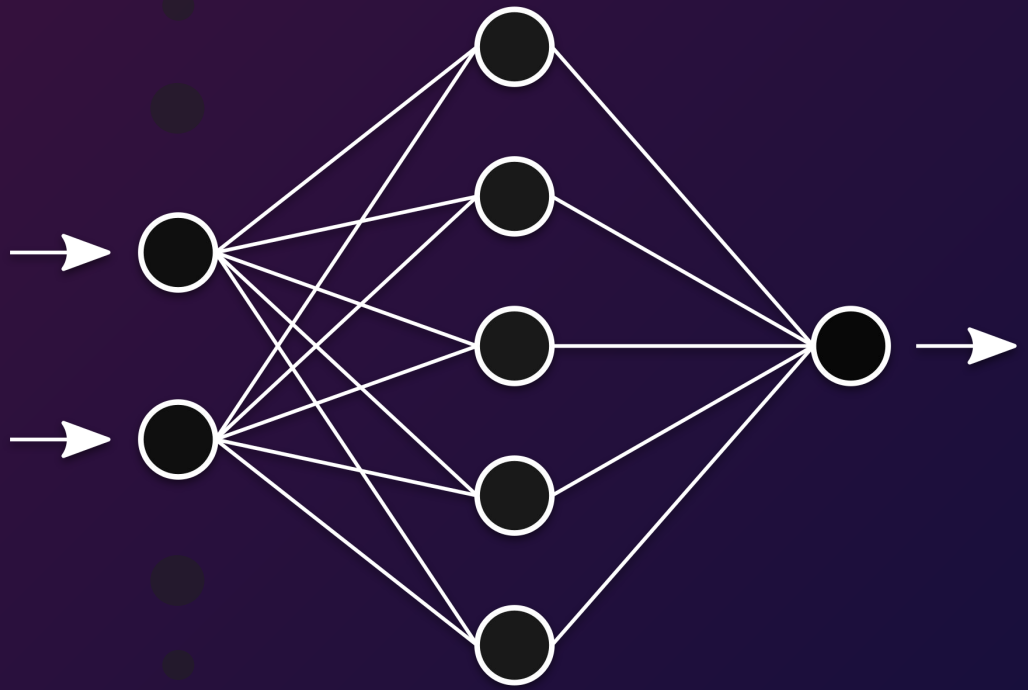
STEP 5

Updates the weights and biases towards a better score

STEP 6

REPEAT!

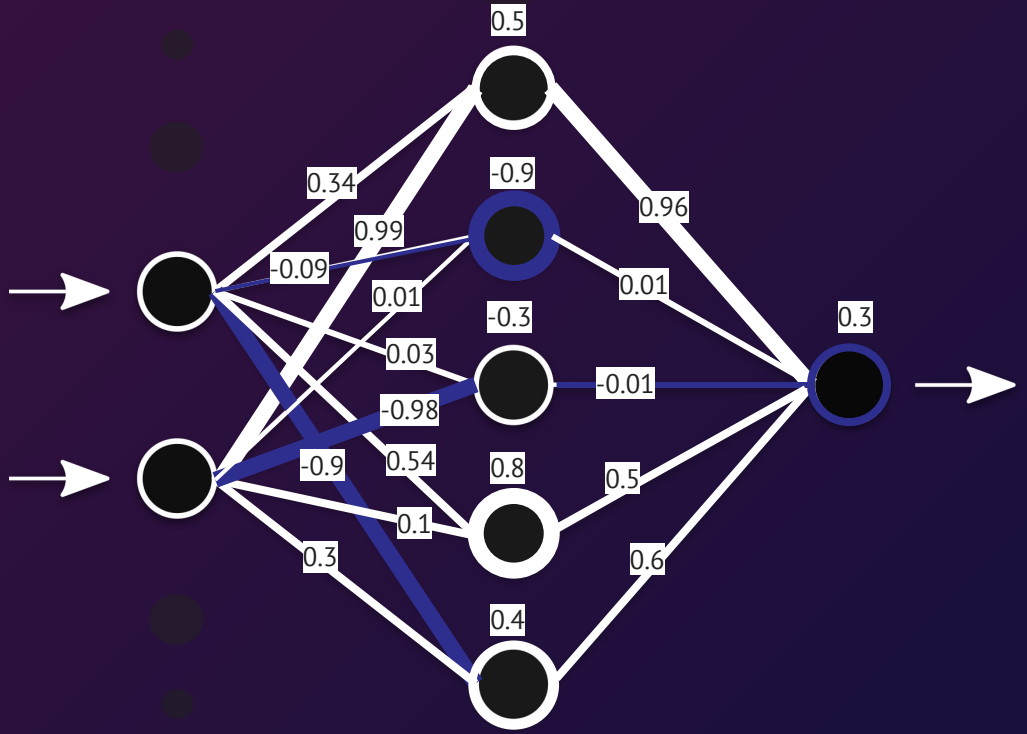
17.99
10.38
…



STEP 1

Initialize the weights and biases

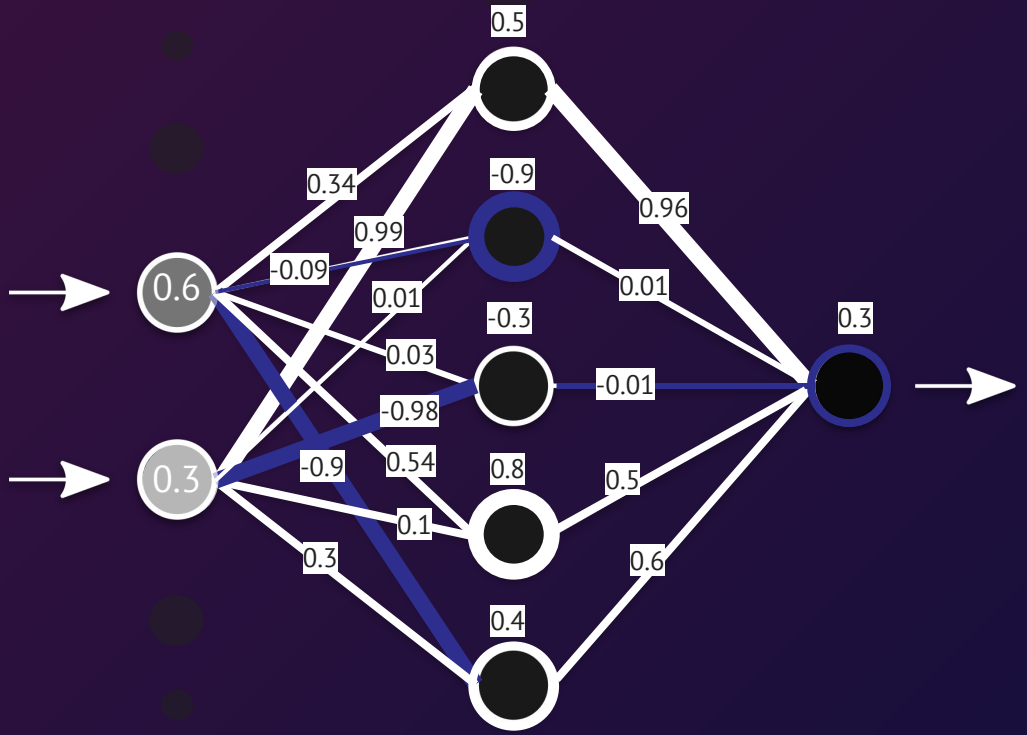
17.99
10.38
…



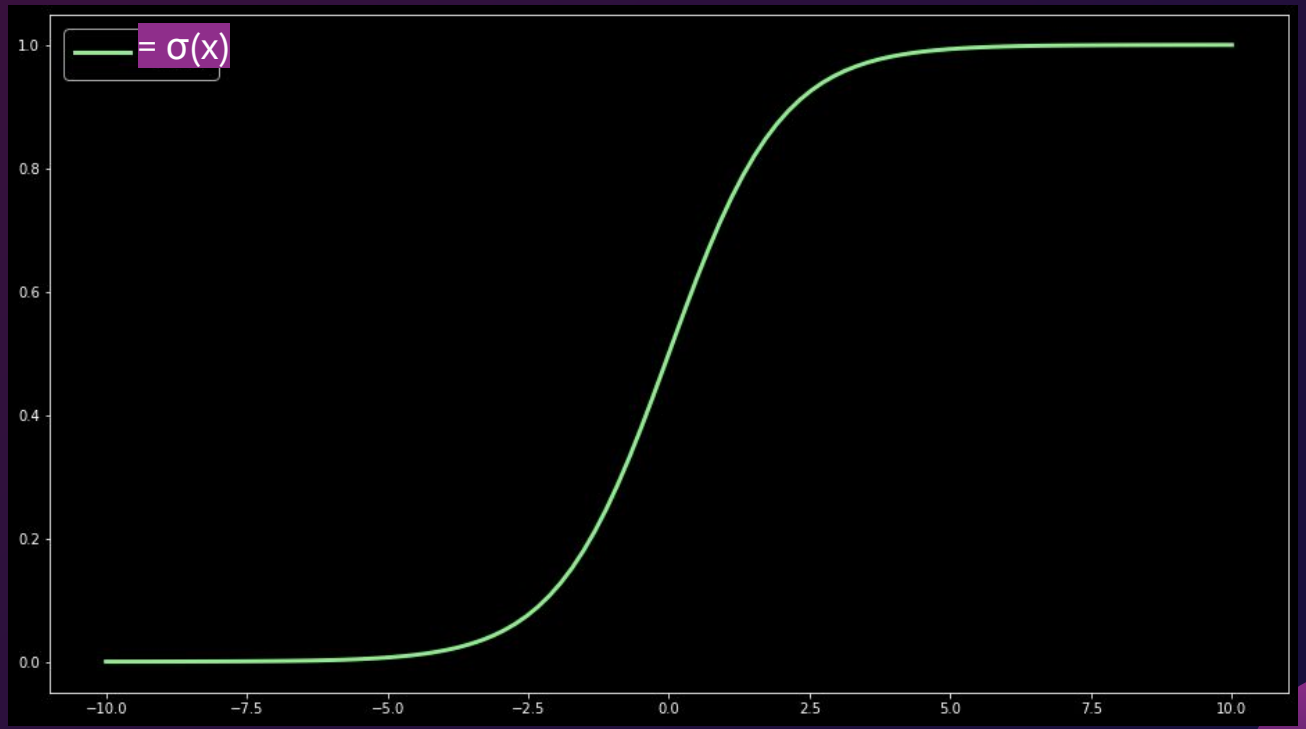
STEP 2

Test it on a ~~bunch~~ of training data!

17.99
10.38
…



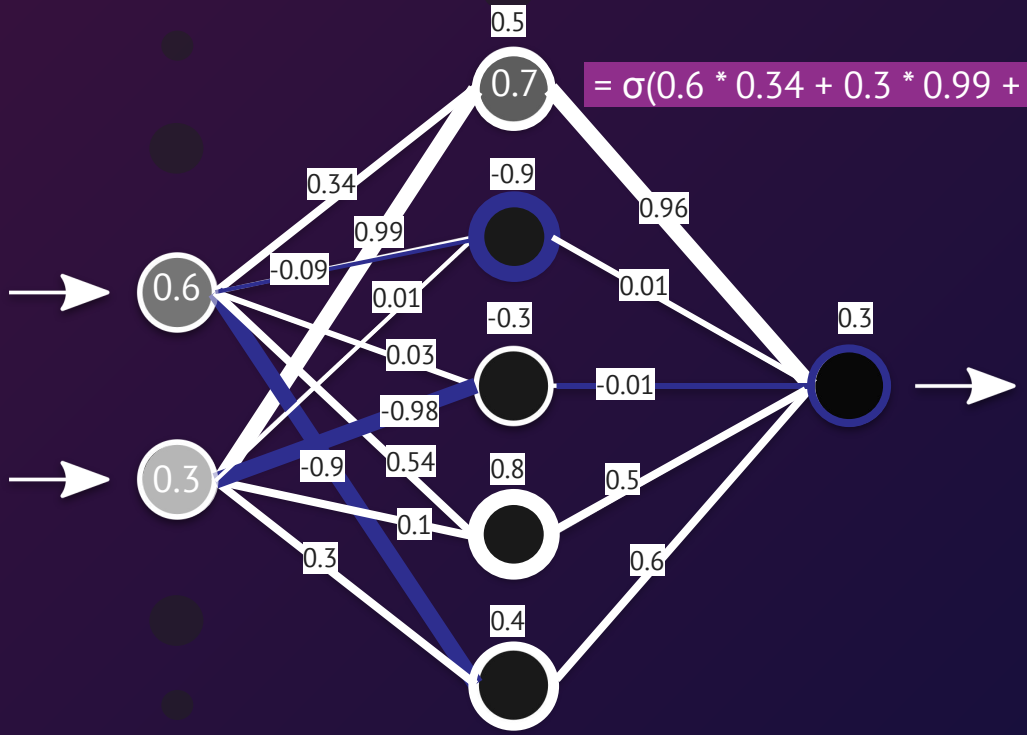
$$\text{Output} = \sigma(\text{Input} \times W + B)$$


= σ(x)
1.0
0.8
0.6
0.4
0.2
0.0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0

STEP 2

Test it on a ~~bunch~~ of training data!

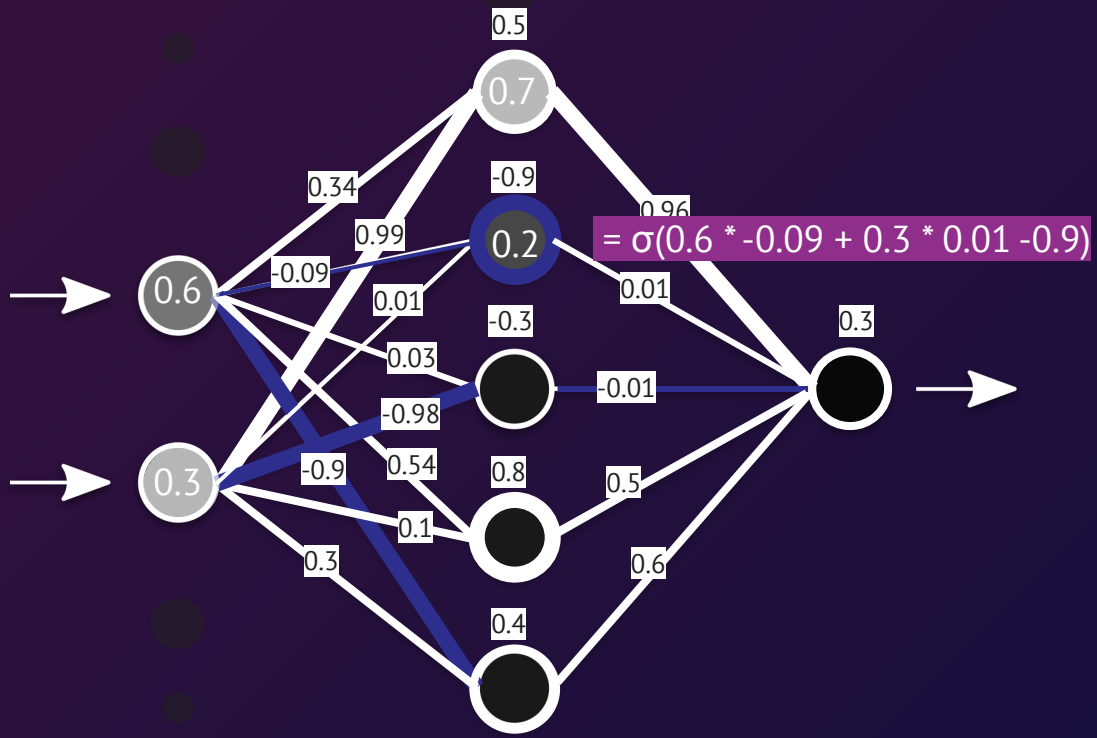
17.99
10.38
…



STEP 2

Test it on a ~~bunch~~ of training data!

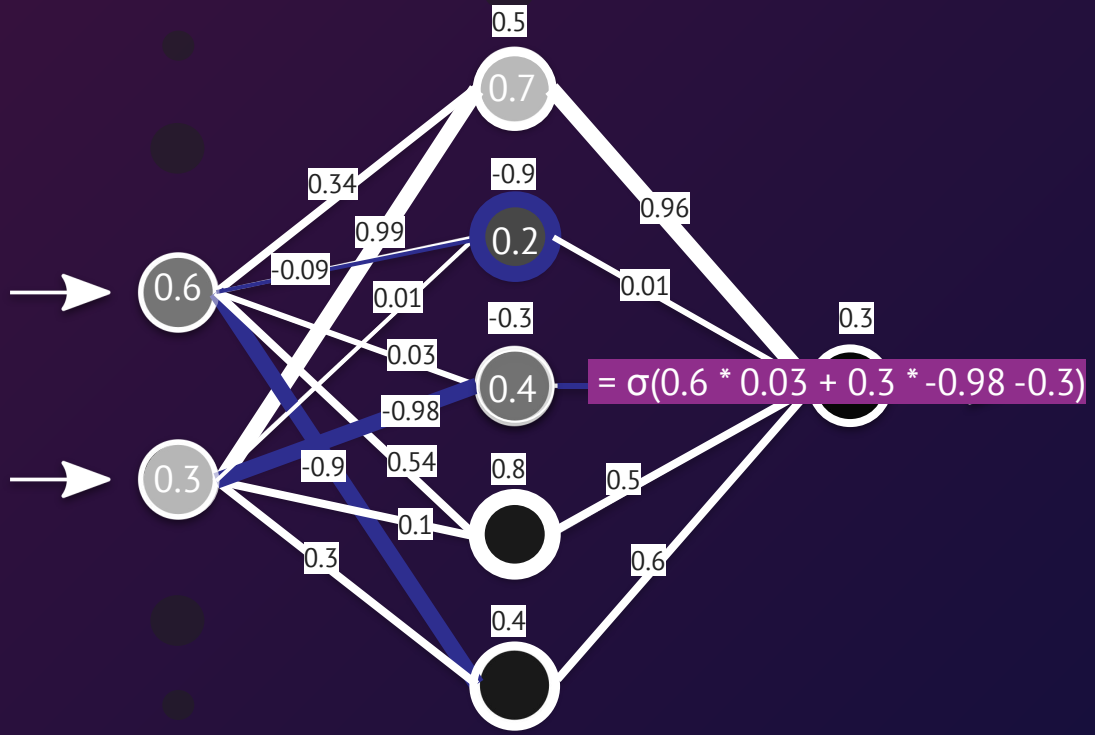
17.99
10.38
…



STEP 2

Test it on a ~~bunch~~ of training data!

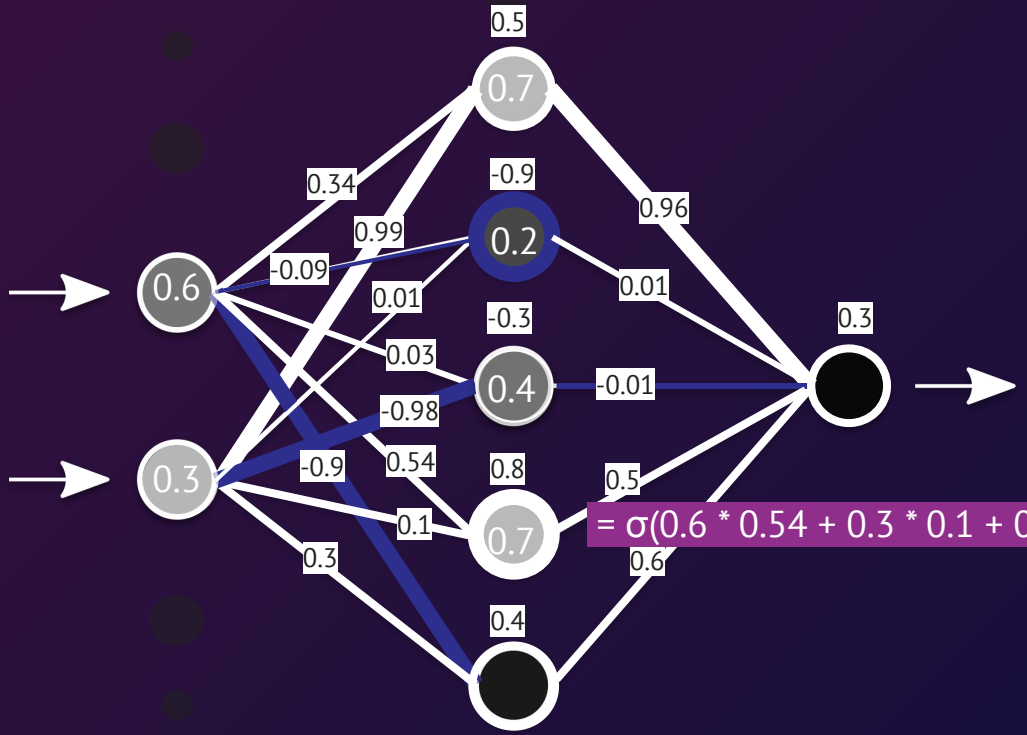
17.99
10.38
…



STEP 2

Test it on a ~~bunch~~ of training data!

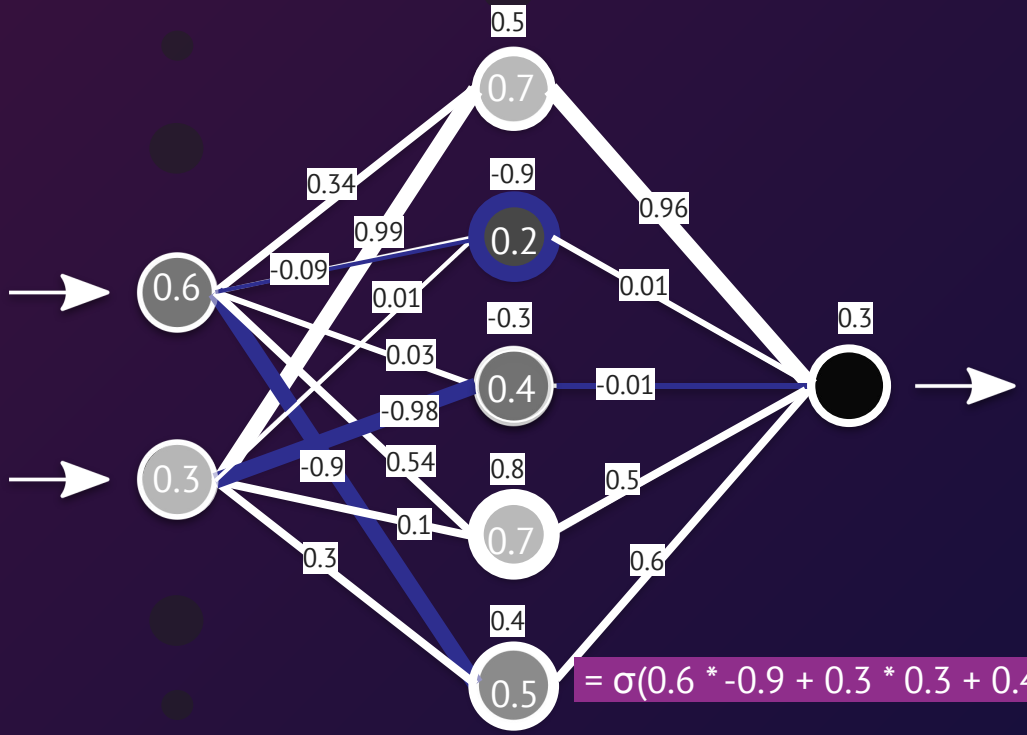
17.99
10.38
…



STEP 2

Test it on a ~~bunch~~ of training data!

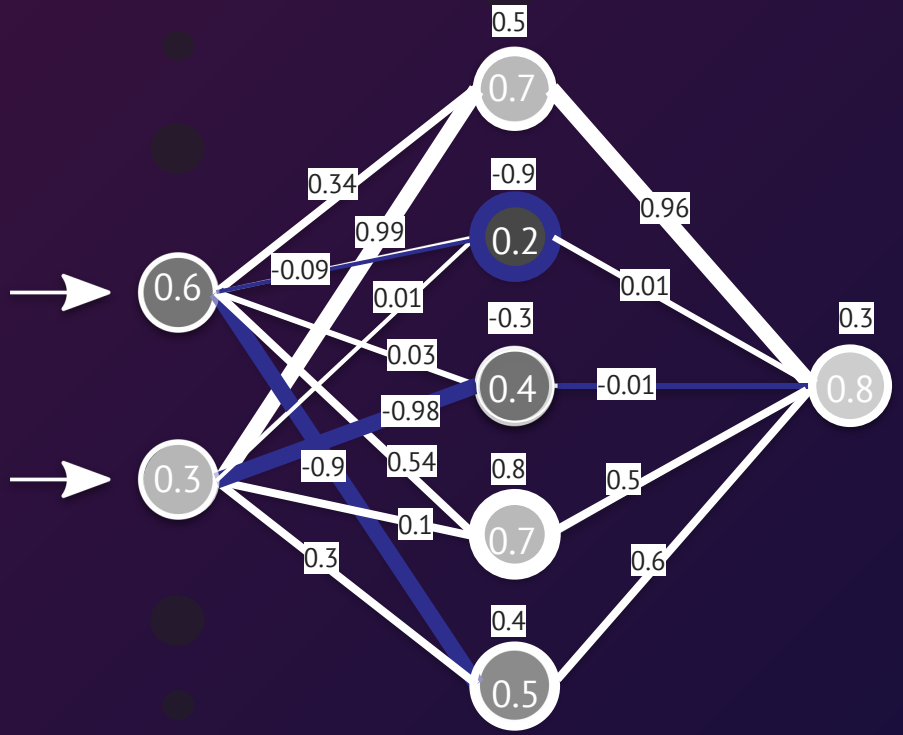
17.99
10.38
…



STEP 2

Test it on a ~~bunch~~ of training data!

17.99
10.38
…

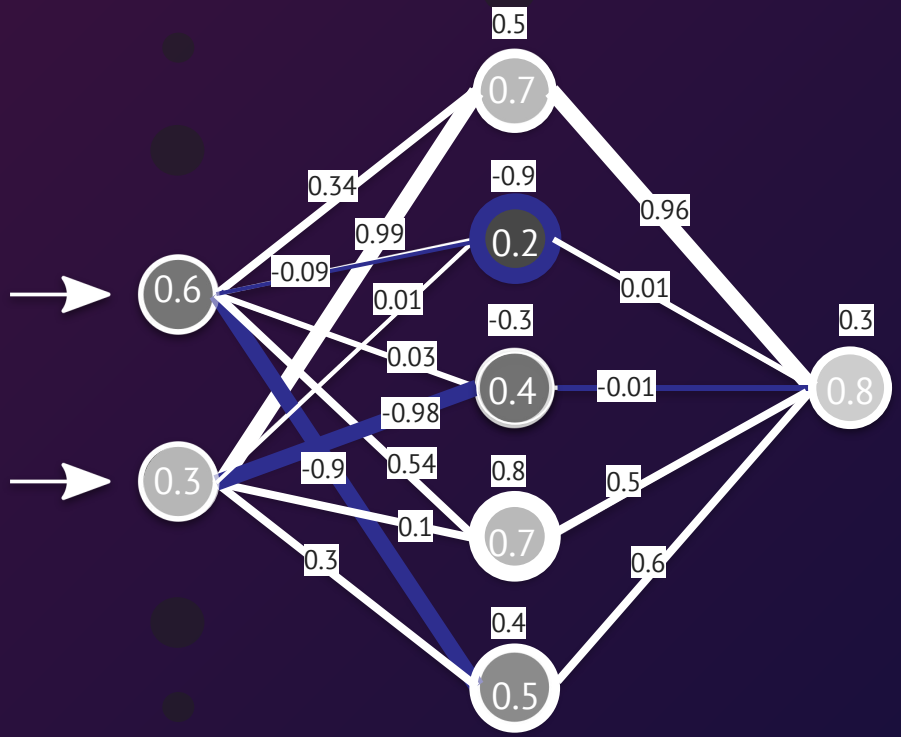


= σ(0.7 * 0.96 + 0.2 * 0.01 + 0.4 * -0.01 + 0.7 * 0.5+ 0.5 * 0.6 + 0.3)

STEP 2

Test it on a ~~bunch~~ of training data!

17.99
10.38
…

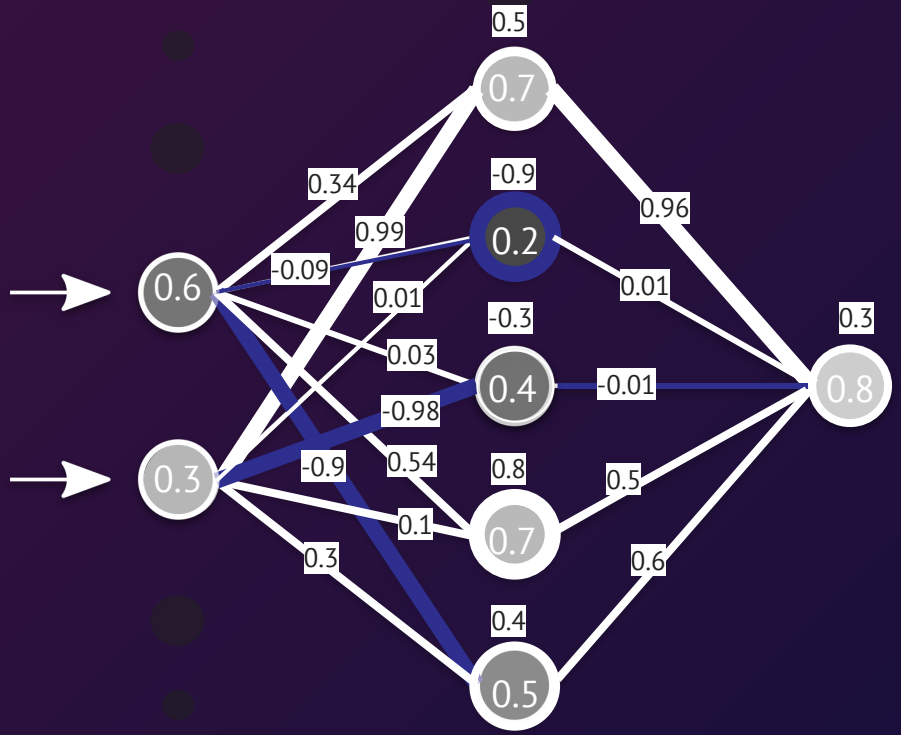


0.83

STEP 3

Calculate the score (the difference)

17.99
10.38
…



We expected:

0.43

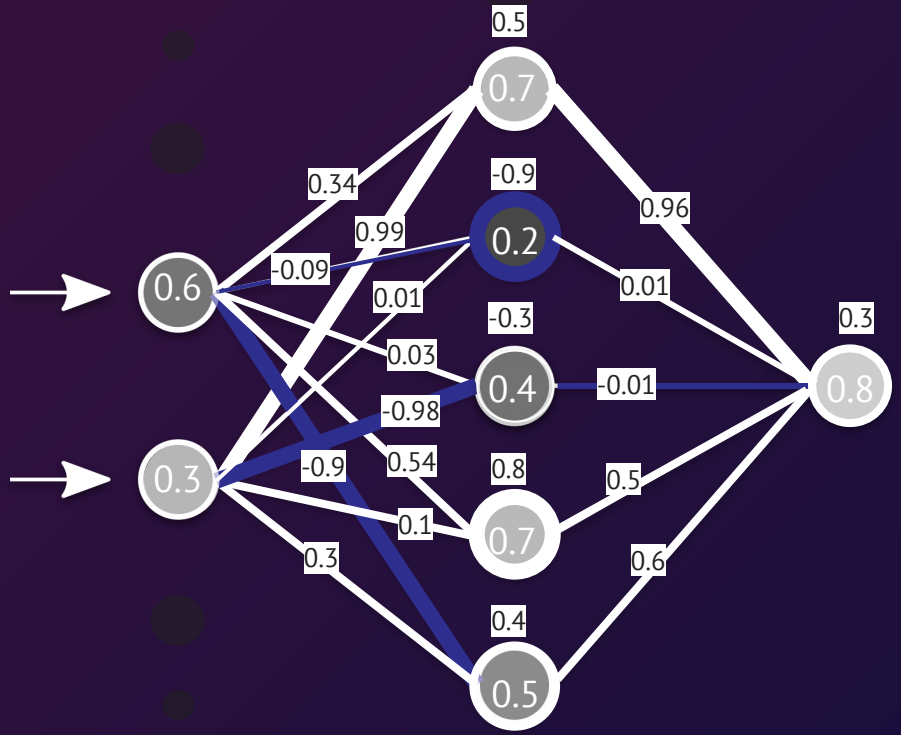
We got:

0.83

STEP 3

Calculate the score (the difference)

17.99
10.38
…



Difference:

$(0.83 - 0.43)^2$

Difference = (Output - Expected Output)^2

Larger when the model made a worse prediction.

Smaller when the model made a more accurate prediction.

$\text{Difference} = \text{Avg}((\text{Output} - \text{Expected Output})^2)$

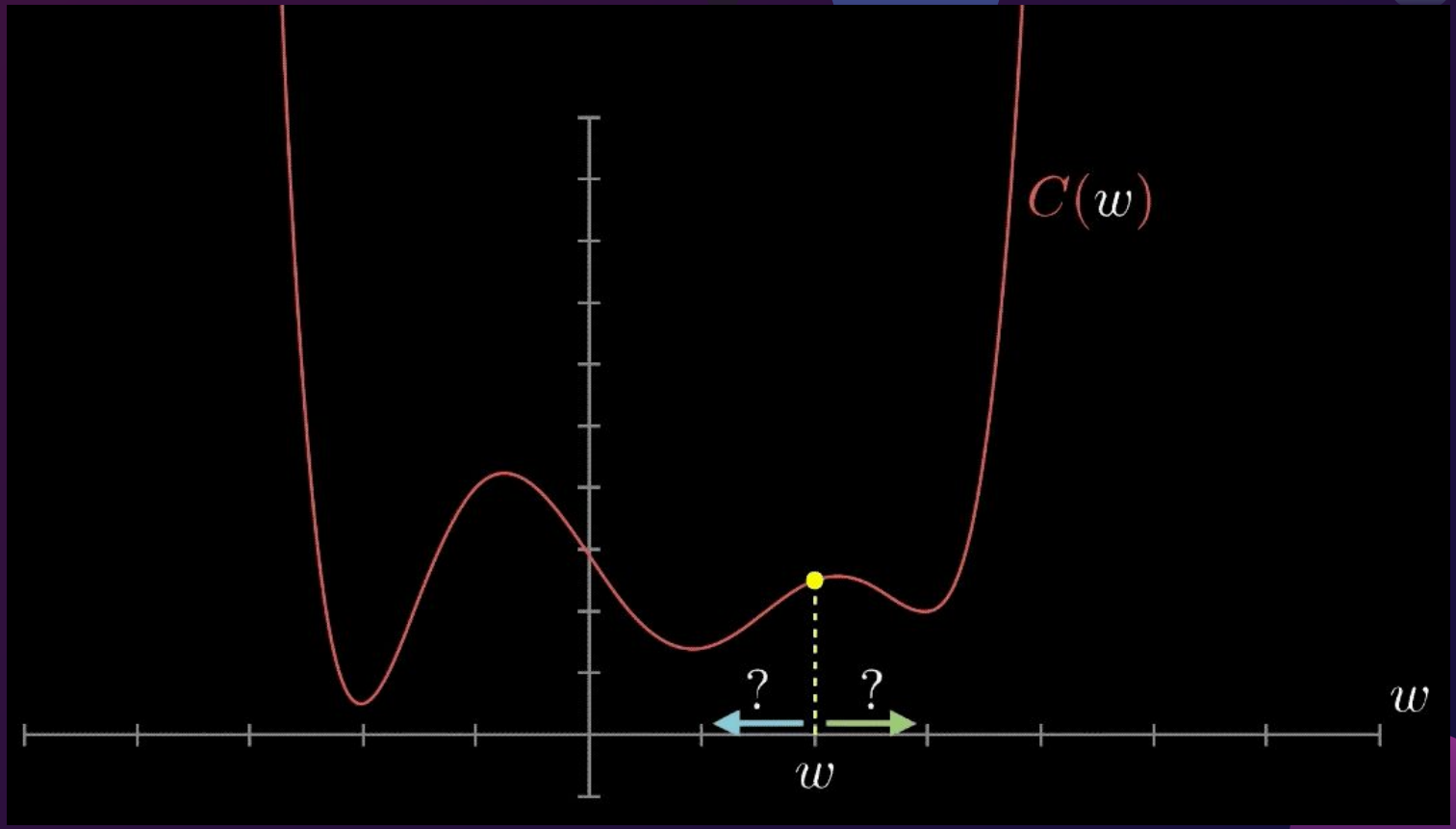
Larger when the model is more inaccurate over the entire training set.

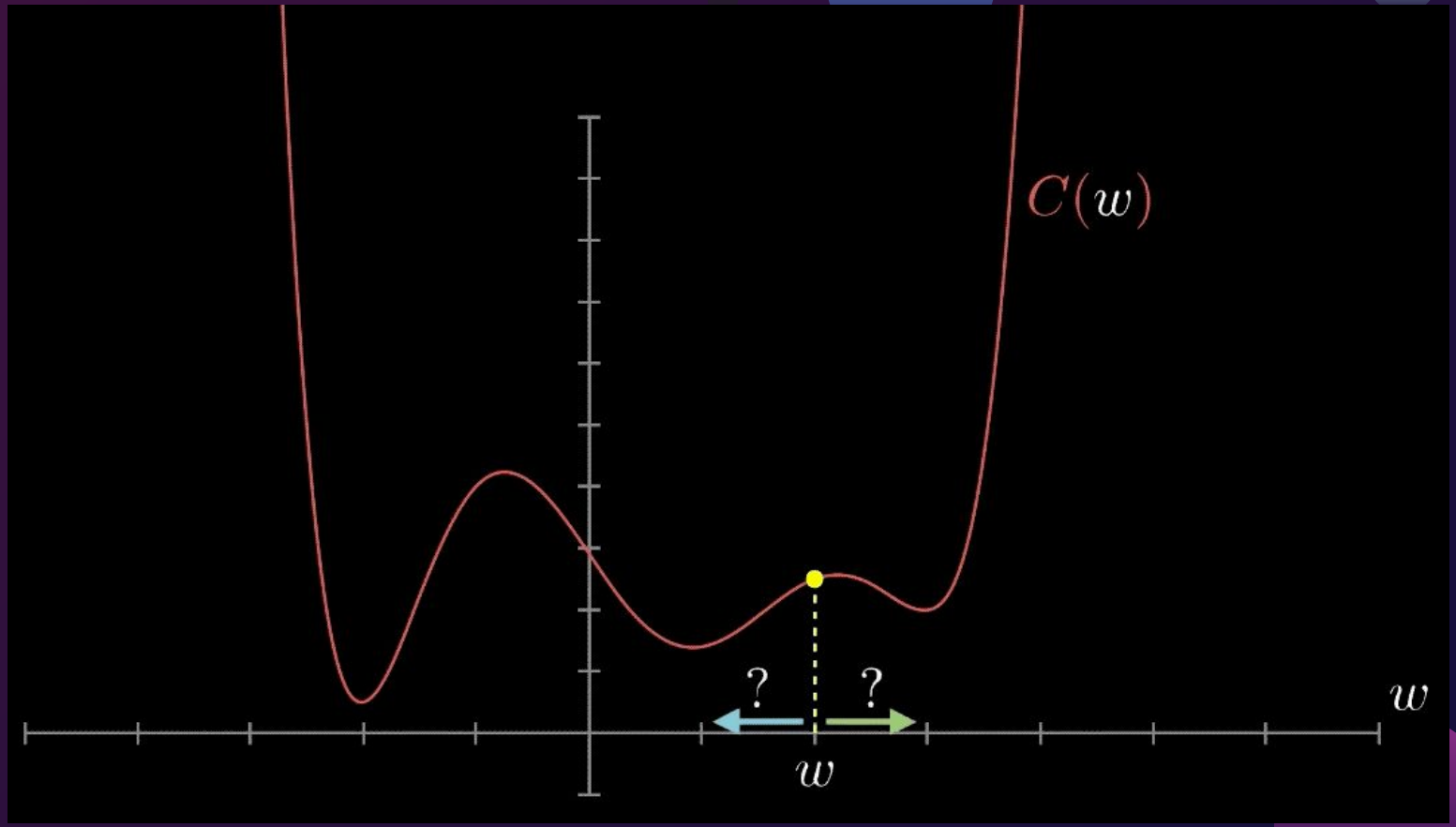
Smaller when the model makes more accurate predictions over the entire training set.

$$\text{COST} = \text{Avg}((\text{Output} - \text{Expected Output})^2)$$

A cost function is a measure for how good/bad the network is at making predictions!

Input: A neural network weights and biases + training data
Output: A score *(the smaller the better)*


$C(w)$
?
?
w
w


$C(w)$
?
?
w
w

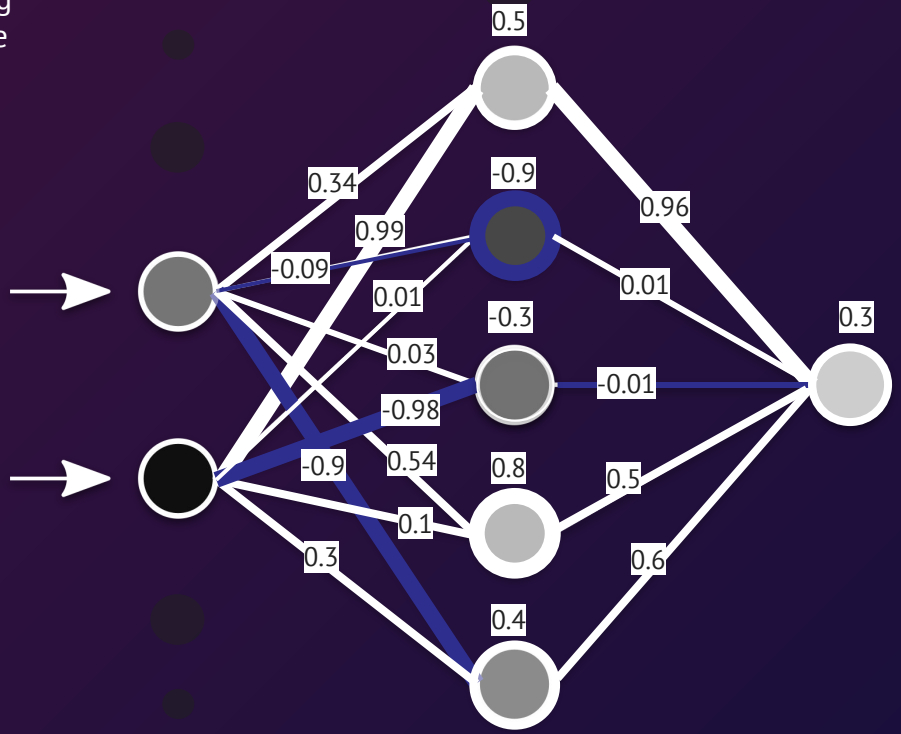
Backpropagation is an algorithm for computing the gradient of a cost function!

We can determine how we must adjust the weights and biases to reach a **local minimum**!

STEP 2&3

Test it on a **bunch** of training data and calculate the score (the difference)

17.99
10.38
…



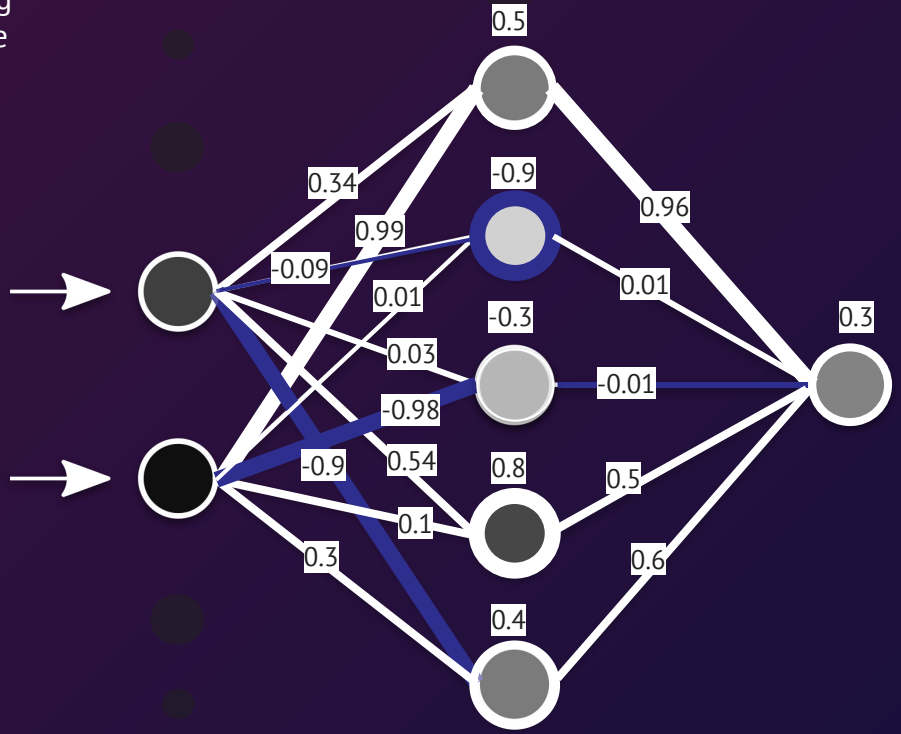
The **COST**:

$$(0.83 - 0.43)^2$$

STEP 2&3

Test it on a **bunch** of training data and calculate the score (the difference)

23.43
12.02
…



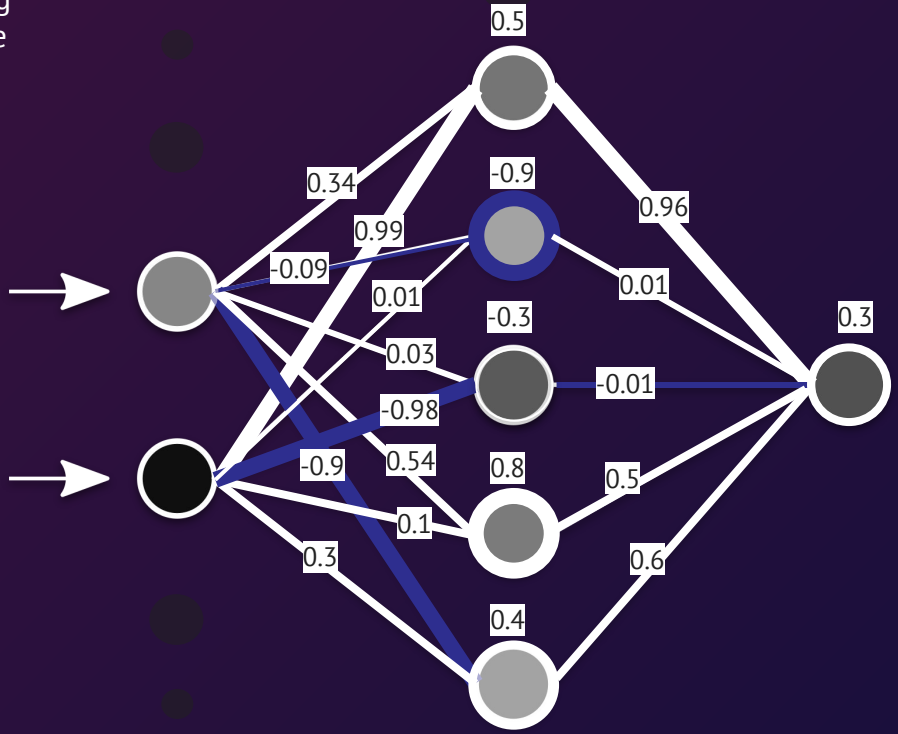
The **COST**:

$$(0.13 - 0.23)^2$$

STEP 2&3

Test it on a **bunch** of training data and calculate the score (the difference)

23.43
12.02
…



The **COST**:

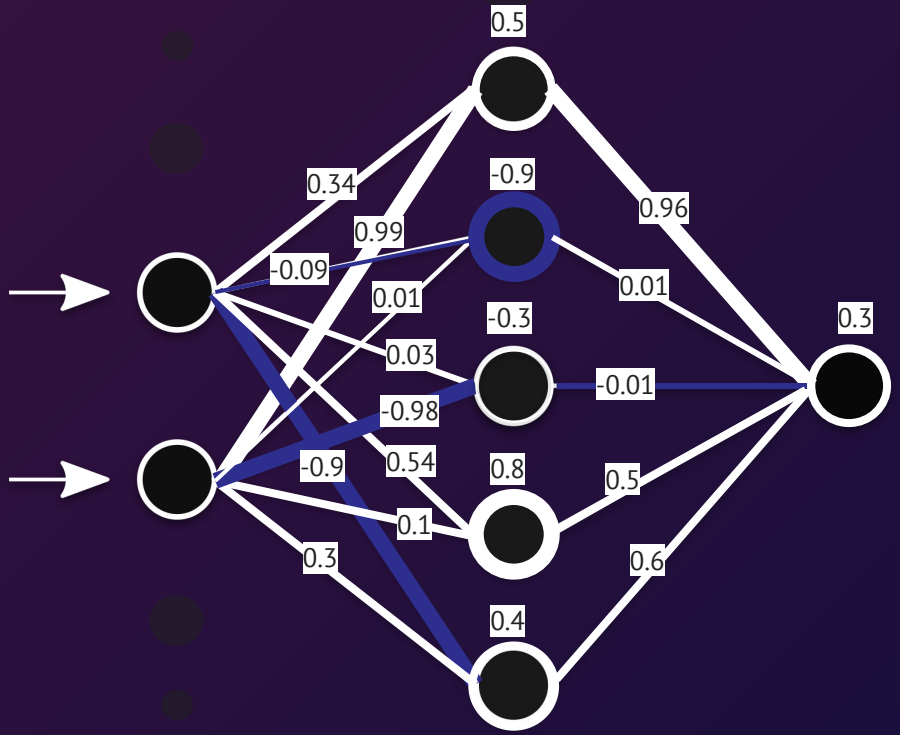
$$(0.27- 0.25)^2$$

$$\text{COST} = \text{Avg}((\text{Output} - \text{Expected Output})^2)$$

$$= 0.0568$$

STEP 4

Calculate the differences
for each weight and bias



The **COST**:

0.0568

$$\text{Output} = \sigma(\text{Input} \times W + B)$$

$$\text{COST} = (\text{Output} - \text{Expected Output})^2$$

a stands for activation

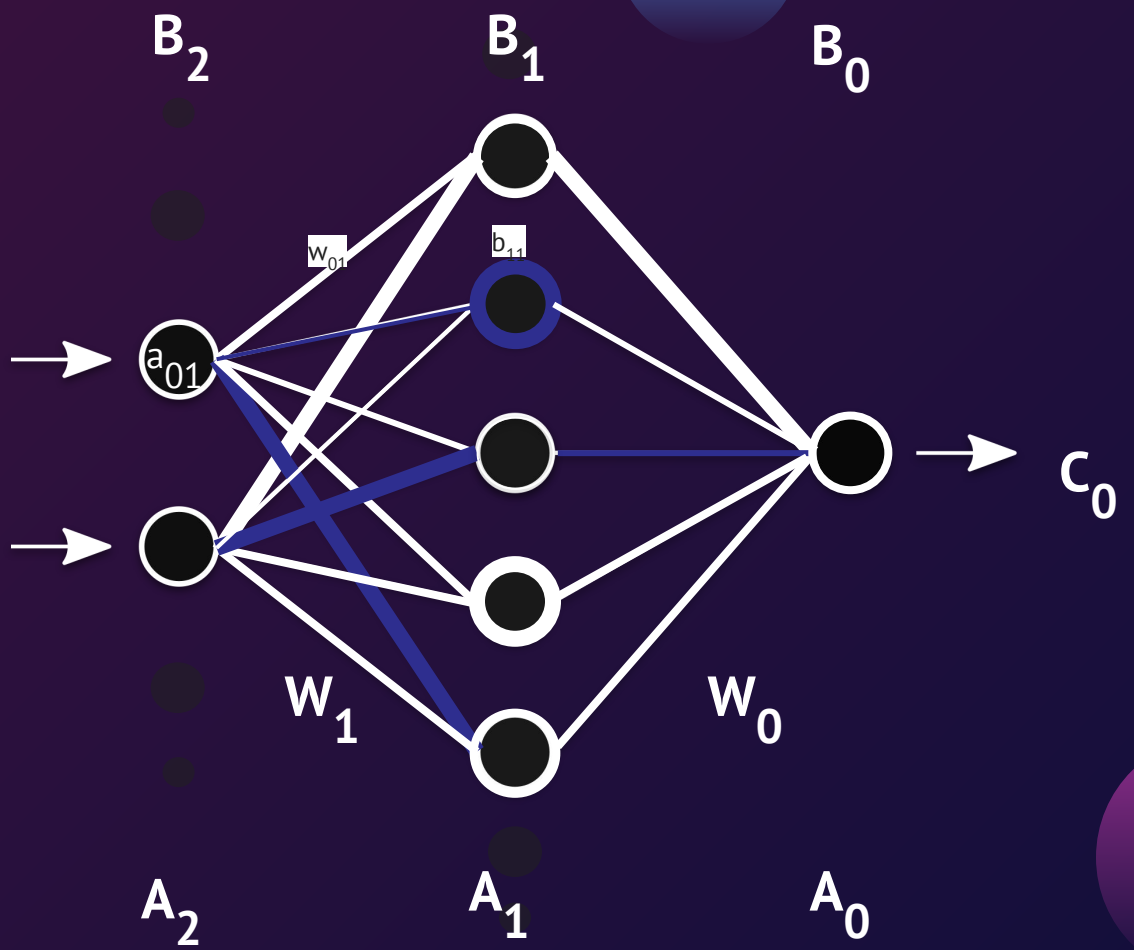
$$A_i = \sigma(A_{i-1} \times W_i + B_i)$$

$$C_0 = (A_0 - Y)^2$$

the cost is the difference between the activation on the neuron and the expected activation

STEP 4

Calculate the differences
for each weight and bias



$$A_i = \sigma(A_{i-1} \times W_i + B_i)$$

$$C_0 = (A_0 - Y)^2$$

$$Z_i = A_{i-1} \times W_i + B_i$$

$$A_i = \sigma(Z_i)$$

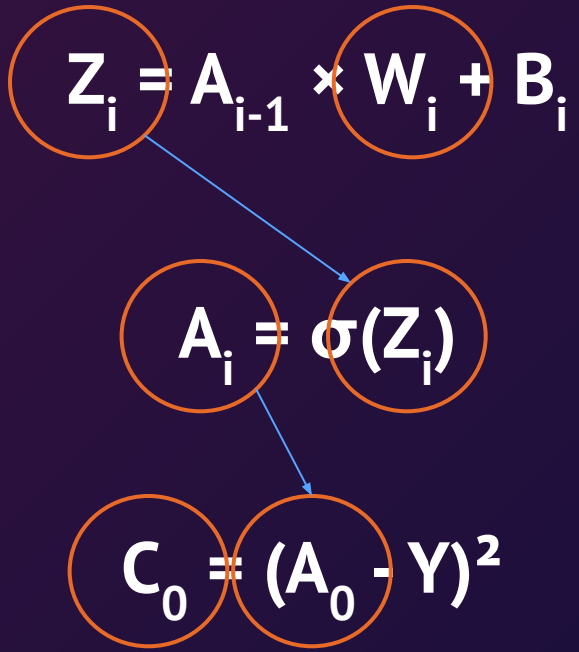
$$C_0 = (A_0 - Y)^2$$

Let’s calculate the difference in the weights!

the change in the cost function with respect to the change in weights

$$\Delta C_0 / \Delta W_0$$

This tells me "If I nudge the weights (W) by _, then I can expect the cost (C) to nudge by _."


$Z_i = A_{i-1} \times W_i + B_i$
$A_i = \sigma(Z_i)$
$C_0 = (A_0 - Y)^2$

$$Z_0(W_0) = A_1 \times W_0 + B_0$$

$$A_0(Z_0) = \sigma(Z_0)$$

$$C_0(A_0) = (A_0 - Y)^2$$


W_0
Z_0
A_0
C_0

chain rule

$$\frac{\delta C_0}{\delta W_0} = \frac{\delta Z_0}{\delta W_0} * \frac{\delta A_0}{\delta Z_0} * \frac{\delta C_0}{\delta A_0}$$

chain rule

$$\frac{\delta C_0}{\delta W_0} = \frac{\delta Z_0}{\delta W_0} * \frac{\delta A_0}{\delta Z_0} * \frac{\delta C_0}{\delta A_0}$$

$$C(A) = (A_0 - Y)^2$$

$$C'(A) = 2(A_0 - Y)$$

chain rule

$$\frac{\delta C_0}{\delta W_0} = \frac{\delta Z_0}{\delta W_0} * \frac{\delta A_0}{\delta Z_0} * \frac{\delta C_0}{\delta A_0}$$

$$A(Z) = \sigma(Z)$$

$$A'(Z) = \sigma'(Z)$$

chain rule

$$\frac{\delta C_0}{\delta W_0} = \frac{\delta Z_0}{\delta W_0} * \frac{\delta A_0}{\delta Z_0} * \frac{\delta C_0}{\delta A_0}$$

$$Z(W) = A \times W + B$$

$$Z'(W) = A$$

We can compute this!

$$\frac{\delta C_0}{\delta W_0} = \frac{\delta Z_0}{\delta W_0} * \frac{\delta A_0}{\delta Z_0} * \frac{\delta C_0}{\delta A_0}$$

$$A * \sigma'(Z) * 2(A_0 - Y)$$

Now, for the biases!

$$Z_0(B_0) = A_1 \times W_0 + B_0$$

$$A_0(Z_0) = \sigma(Z_0)$$

$$C_0(A_0) = (A_0 - Y)^2$$

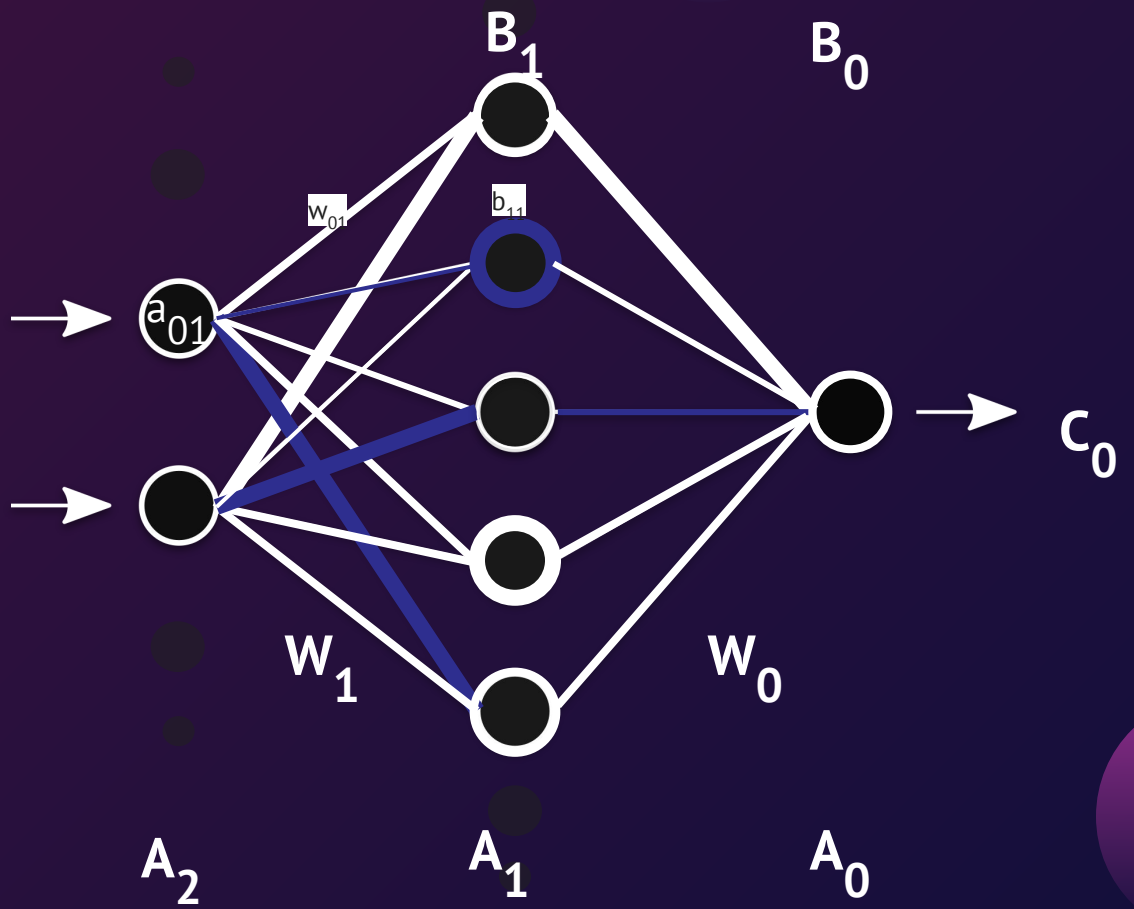
chain rule

$$\frac{\delta C_0}{\delta B_0} = \frac{\delta Z_0}{\delta B_0} * \frac{\delta A_0}{\delta Z_0} * \frac{\delta C_0}{\delta A_0}$$

$$1 * \sigma'(Z) * 2(A_0 - Y)$$

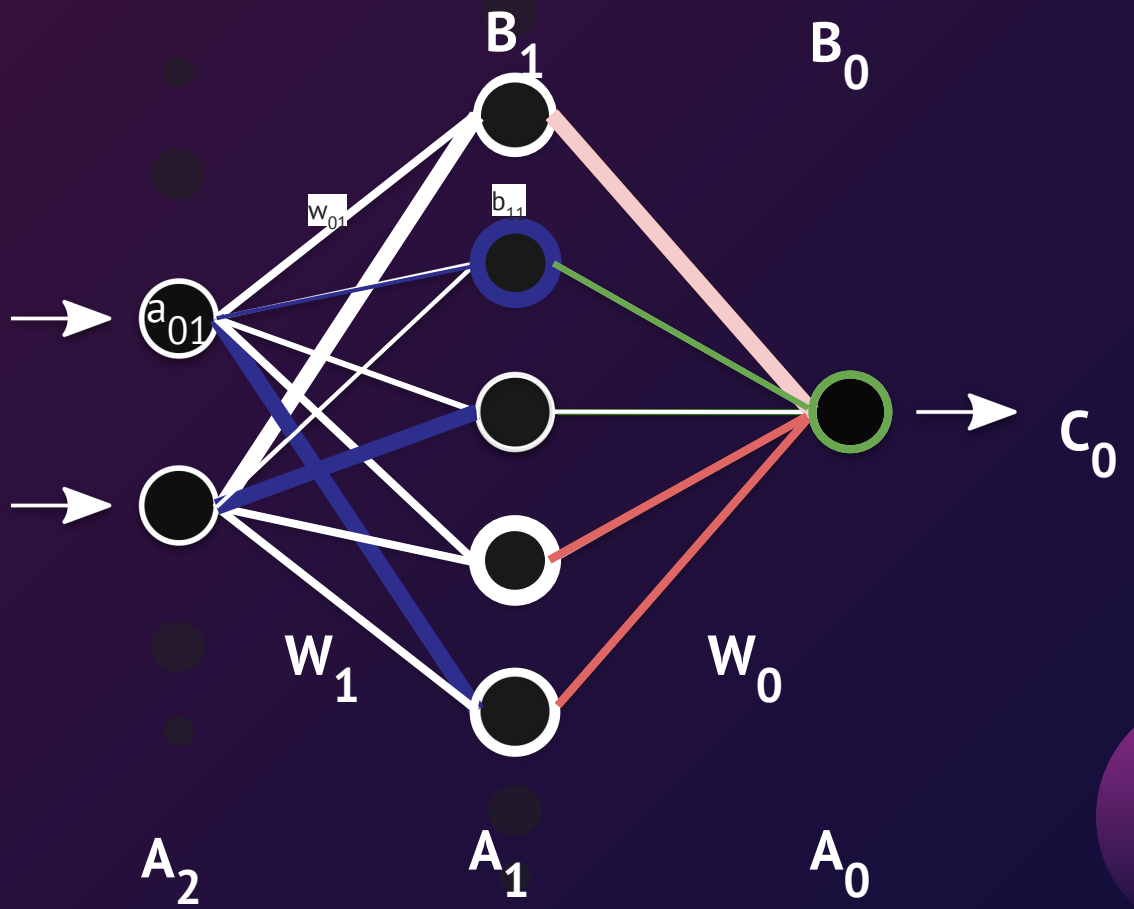
STEP 4

Calculate the differences
for each weight and bias



STEP 4

Calculate the differences
for each weight and bias



What about the rest?

$$Z_0(W_0) = A_1 \times W_0 + B_0$$

$$A_0(Z_0) = \sigma(Z_0)$$

$$C_0(A_0) = (A_0 - Y)^2$$

$$Z_1(W_1) = A_2 \times W_1 + B_1$$

$$\downarrow$$

$$A_1(Z_1) = \sigma(Z_1)$$

$$\downarrow$$

$$Z_0(W_0) = A_1 \times W_0 + B_0$$

$$\downarrow$$

$$A_0(Z_0) = \sigma(Z_0)$$

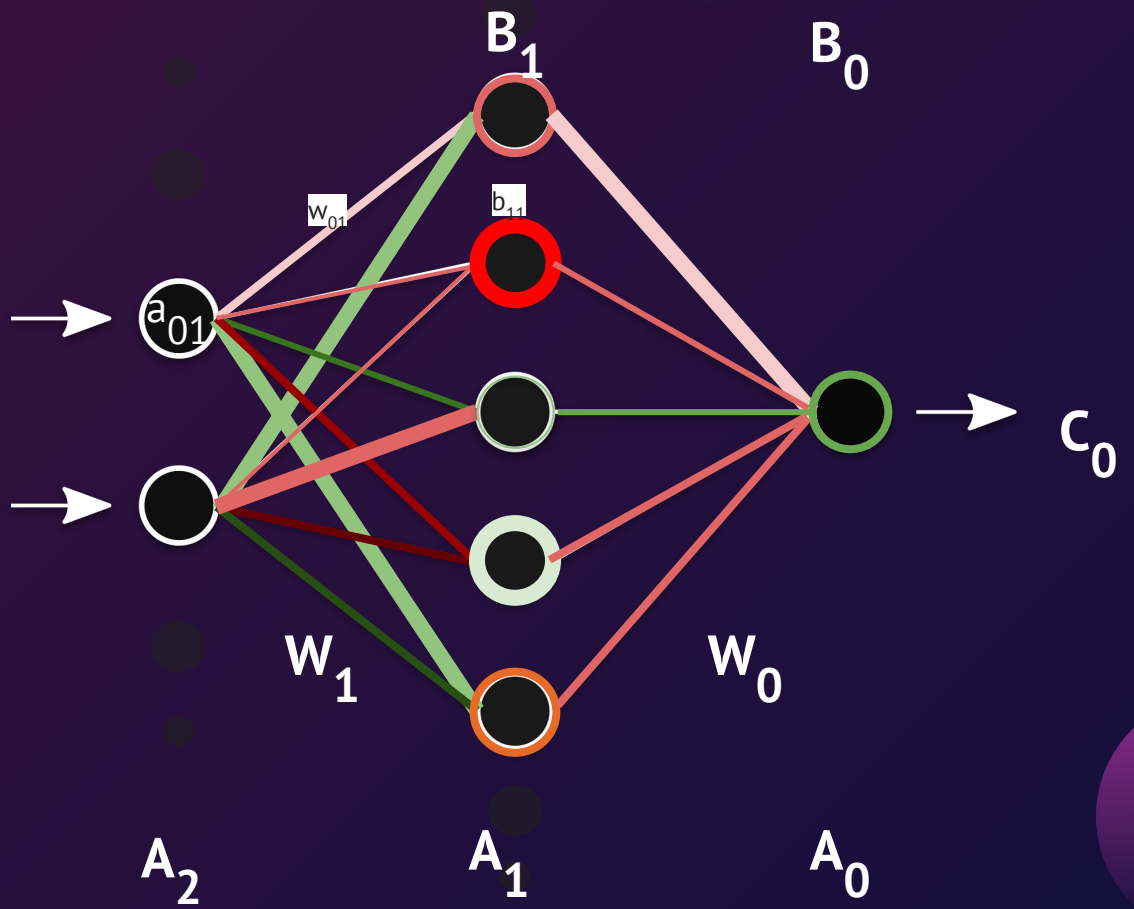
$$\downarrow$$

$$C_0(A_0) = (A_0 - Y)^2$$


W_1
Z_1
A_1
Z_0
A_0
C_0

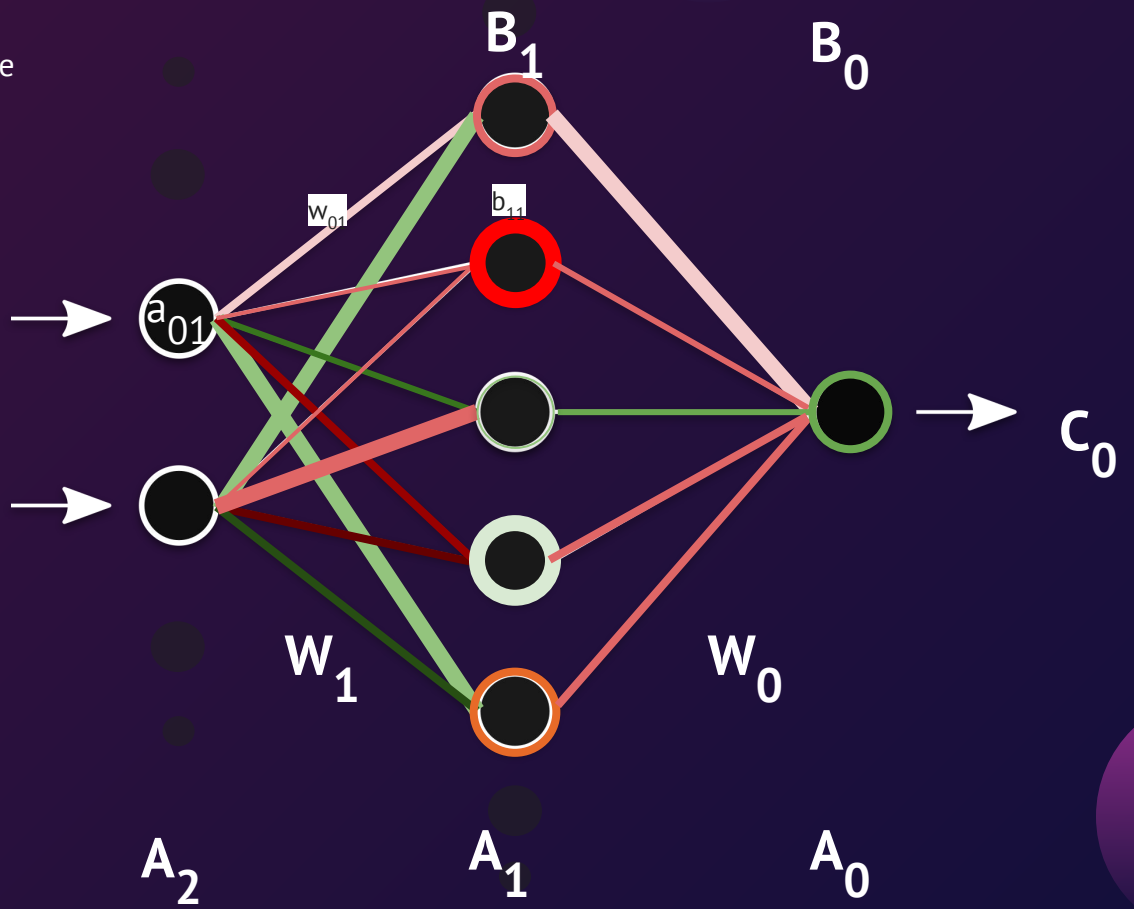
STEP 4

Calculate the differences
for each weight and bias



STEP 5

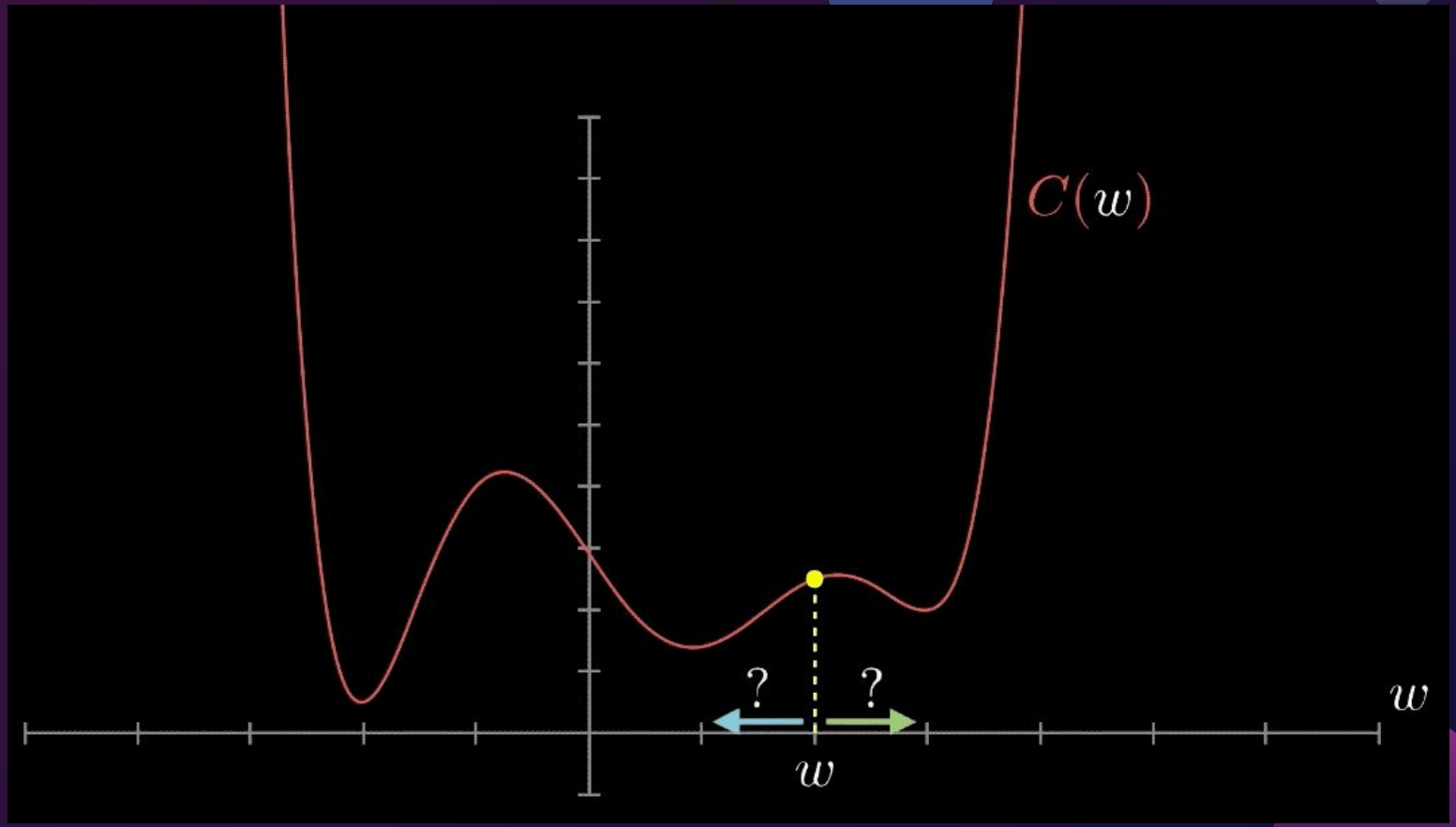
Updates the weights and biases towards a better score



update the weights!

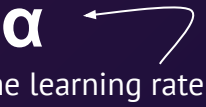
$$W \mathrel{+}= \Delta C/\Delta W * \alpha$$

the learning rate


$C(w)$
?
?
w
w

update the biases!

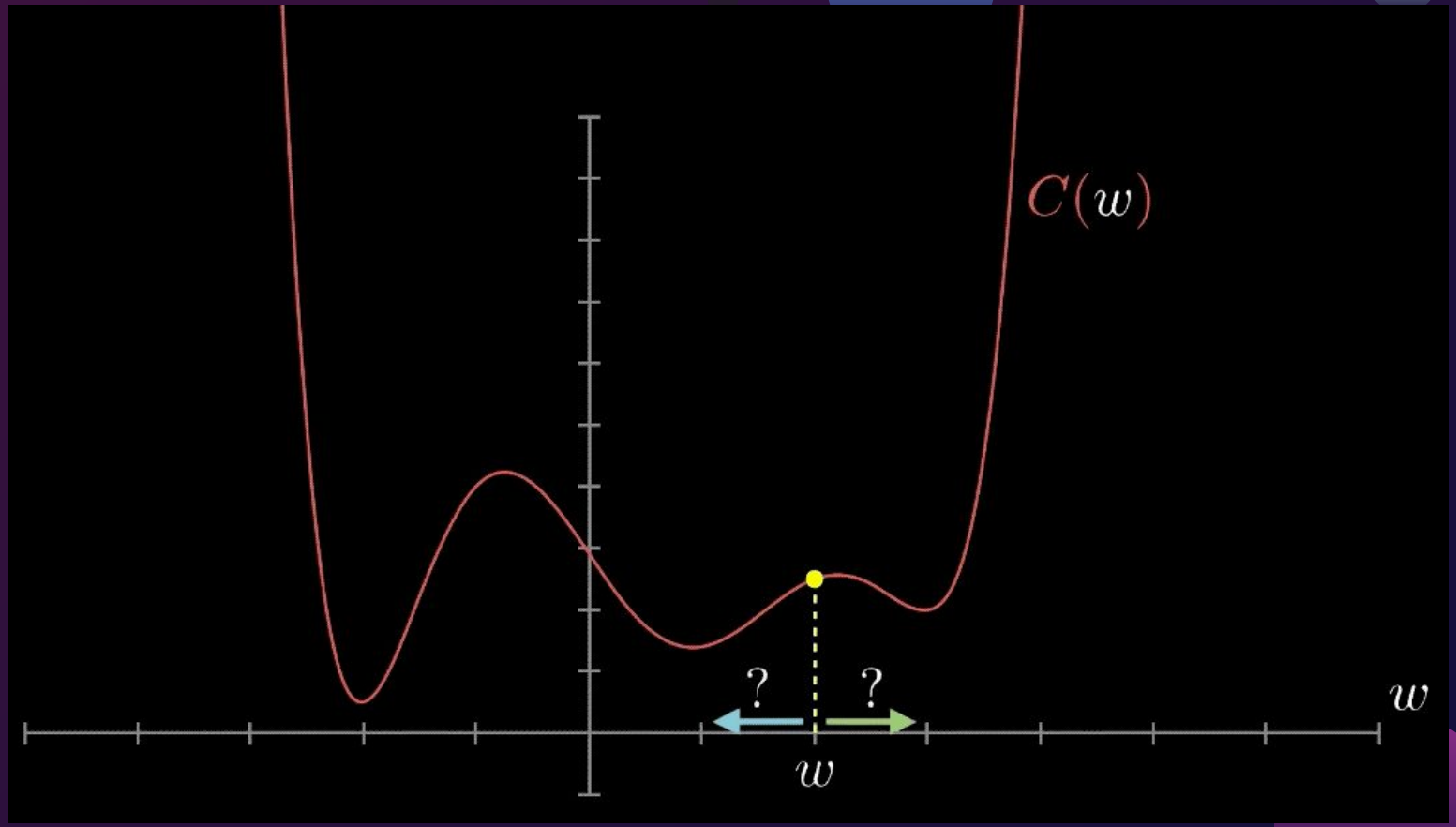
$$B \mathrel{+}= \Delta C/\Delta B * \alpha$$

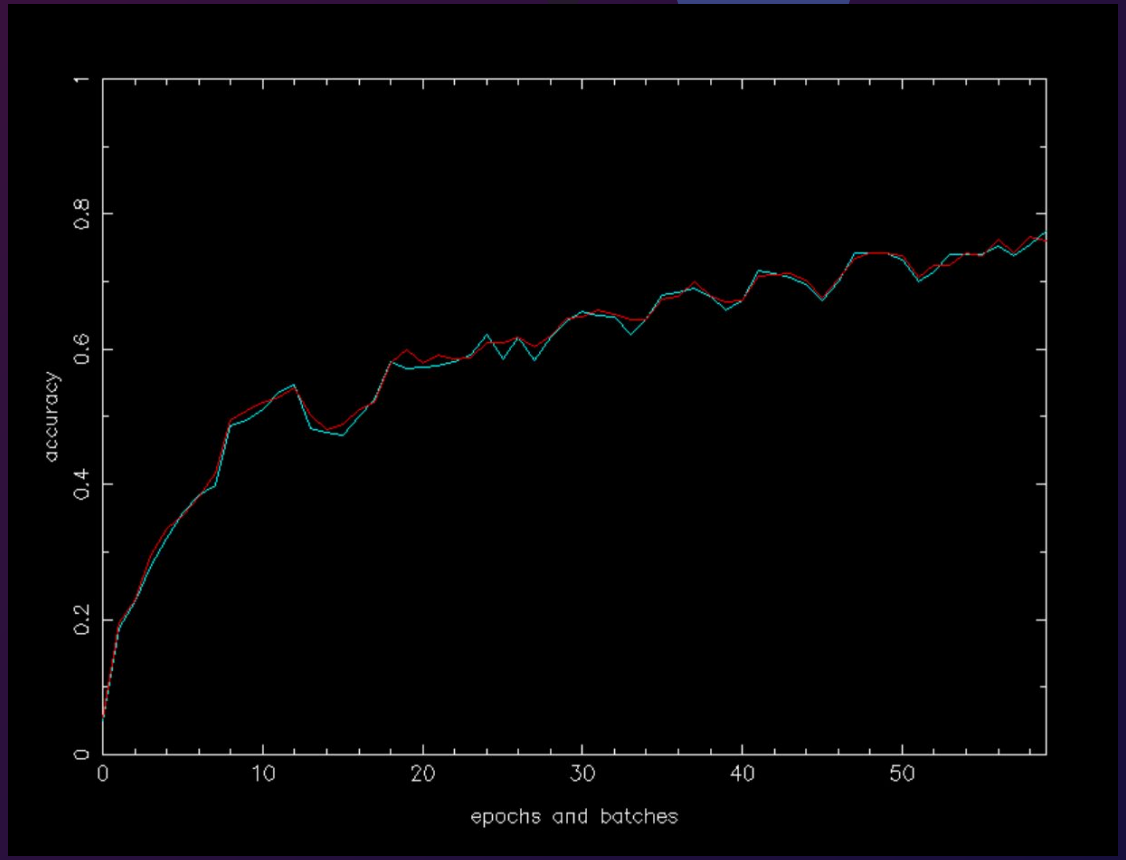


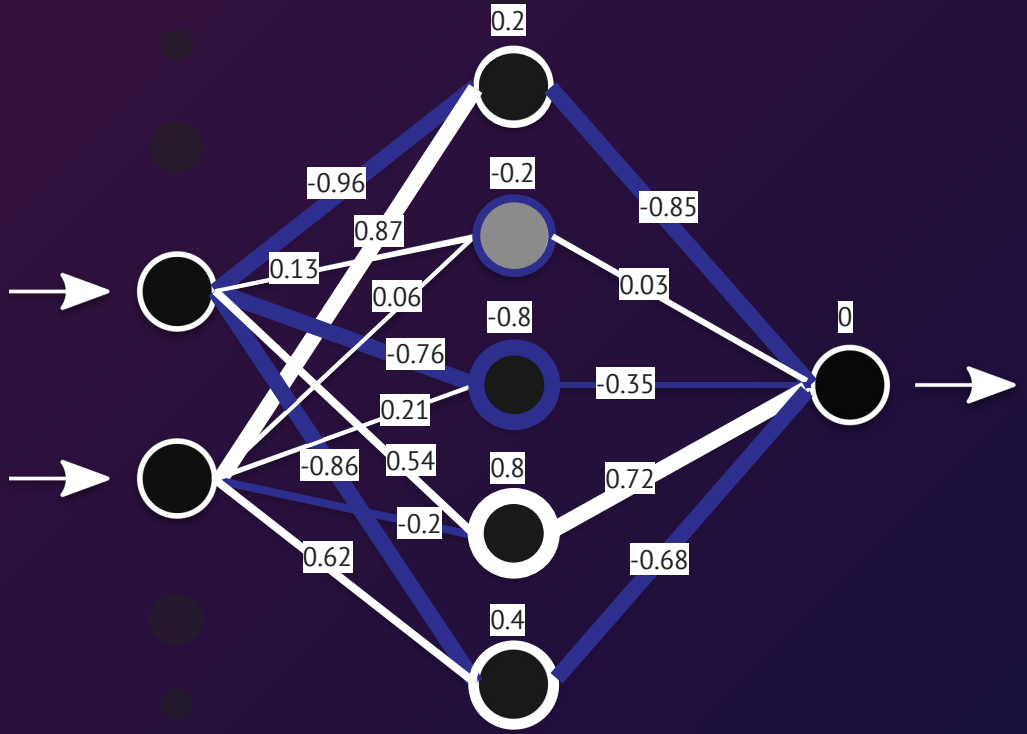
the learning rate

STEP 6

REPEAT!


$C(w)$
?
?
w
w


accuracy
0
0.2
0.4
0.6
0.8
1
0
10
20
30
40
50
epochs and batches


0.2
-0.96
-0.2
-0.85
0.87
0.13
0.06
0.03
-0.8
0
-0.76
-0.35
0.21
-0.86
0.54
0.8
0.72
-0.2
0.62
-0.68
0.4

Resources (GitHub)

https://github.com/ConnorUsaty/MacAIEducation2023

Today's Kahoot!

https://create.kahoot.it/details/9633268d-8ece-4cec-8561-e51f31f52e78