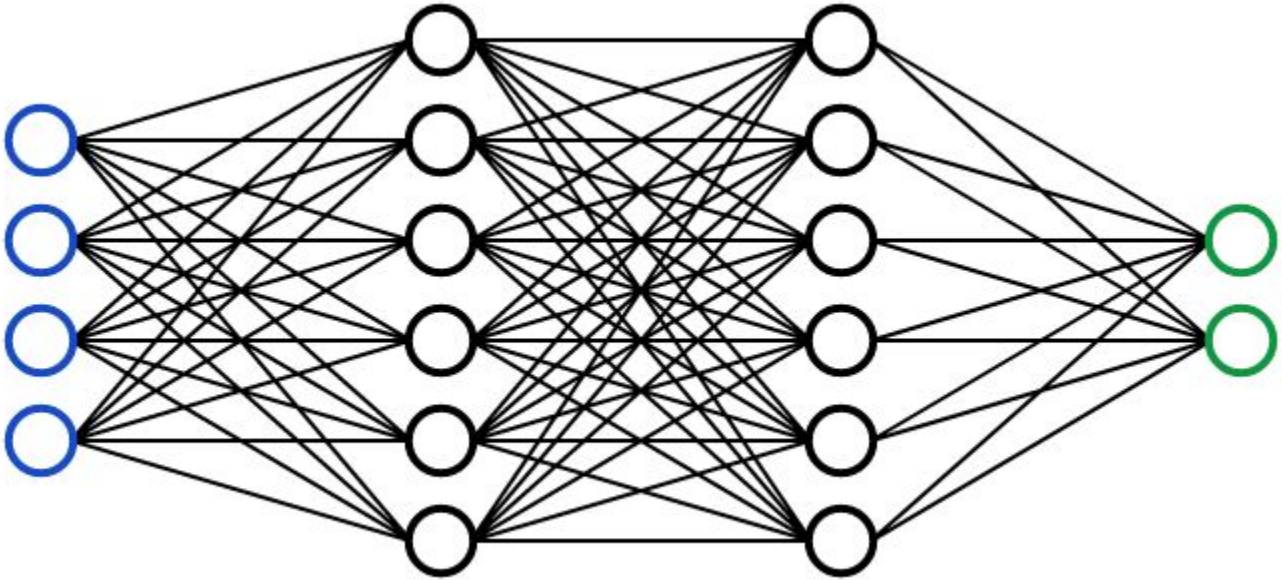
# Neural Networks

The Jubilee Institute, January 2023

# MNIST Dataset

# Neural Network

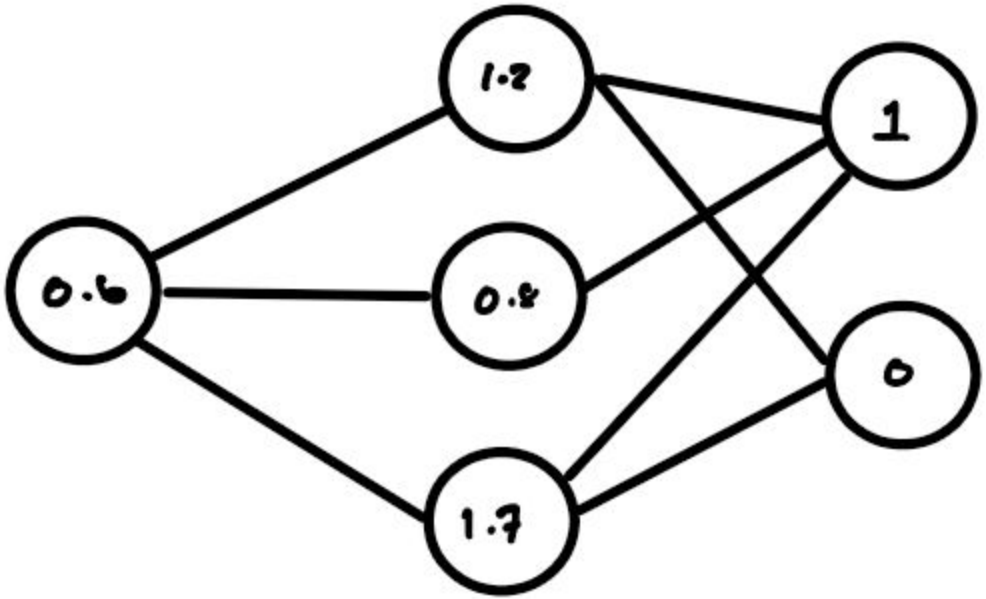
an interconnected group of nodes (neurons) connected with weights and biases that can be trained to learn complex tasks

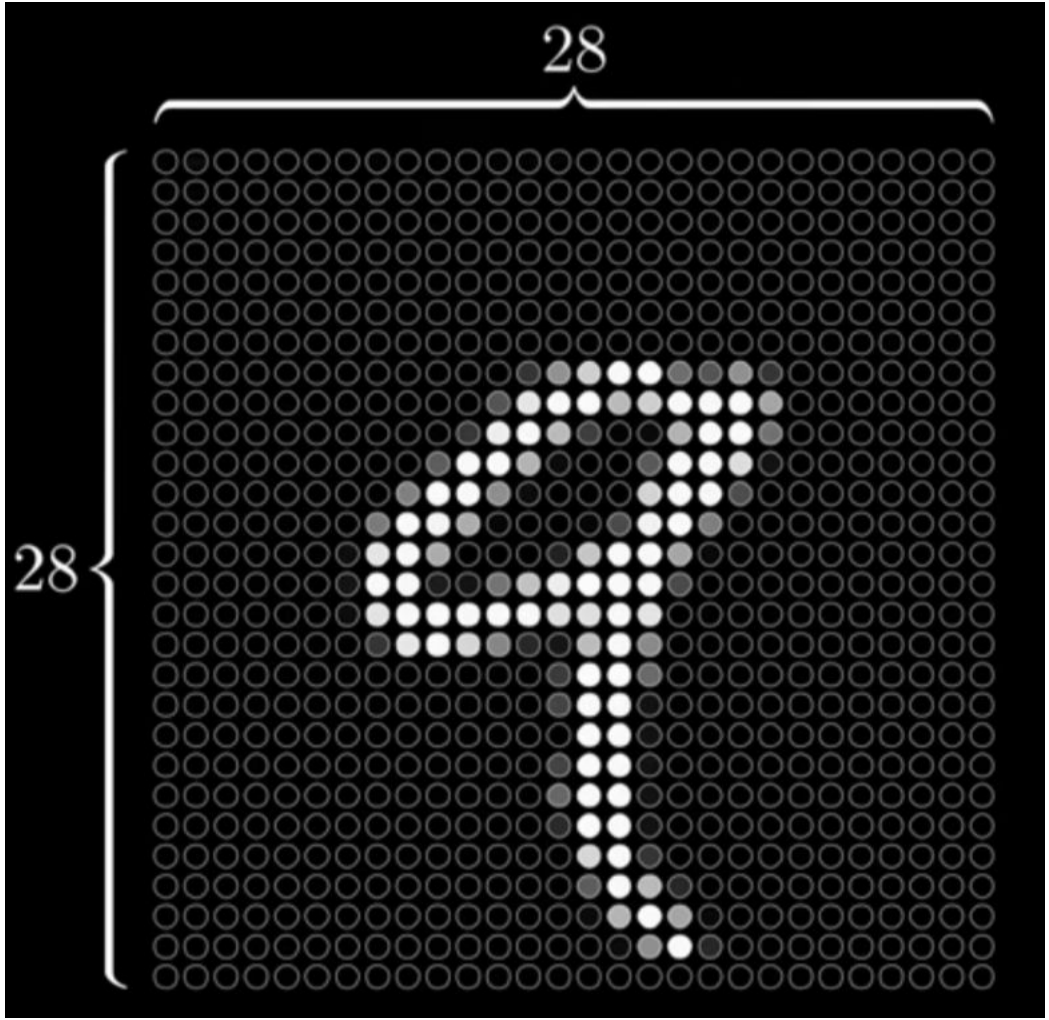


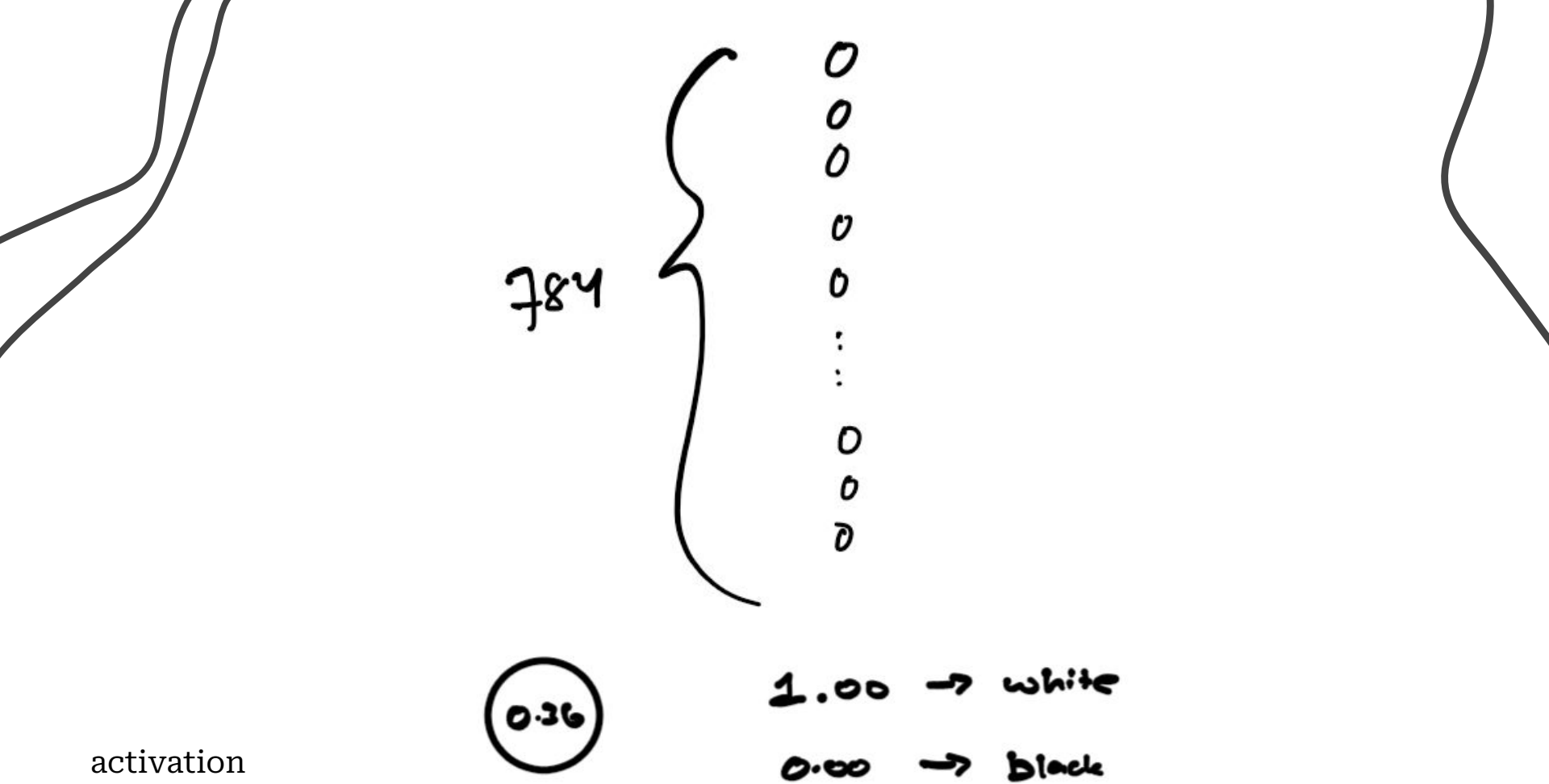
# What is a neuron?

# Neuron/Node

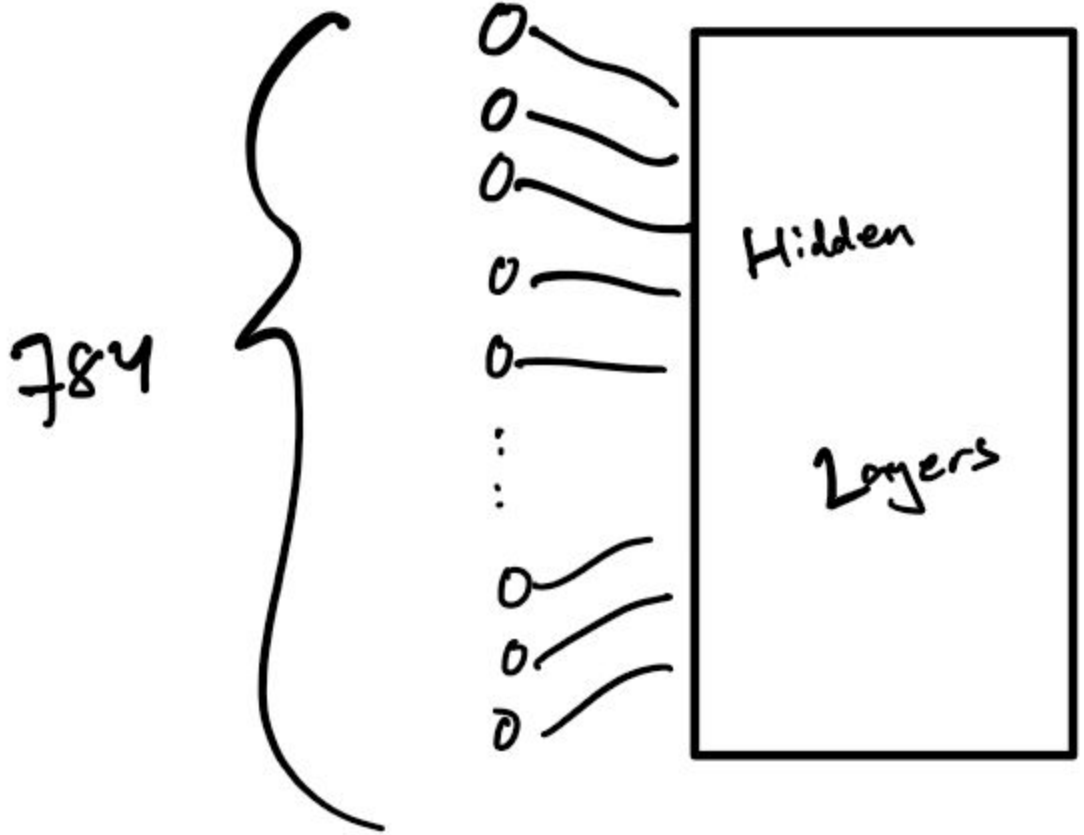
a thing that holds a number, and is connected to other nodes/neurons

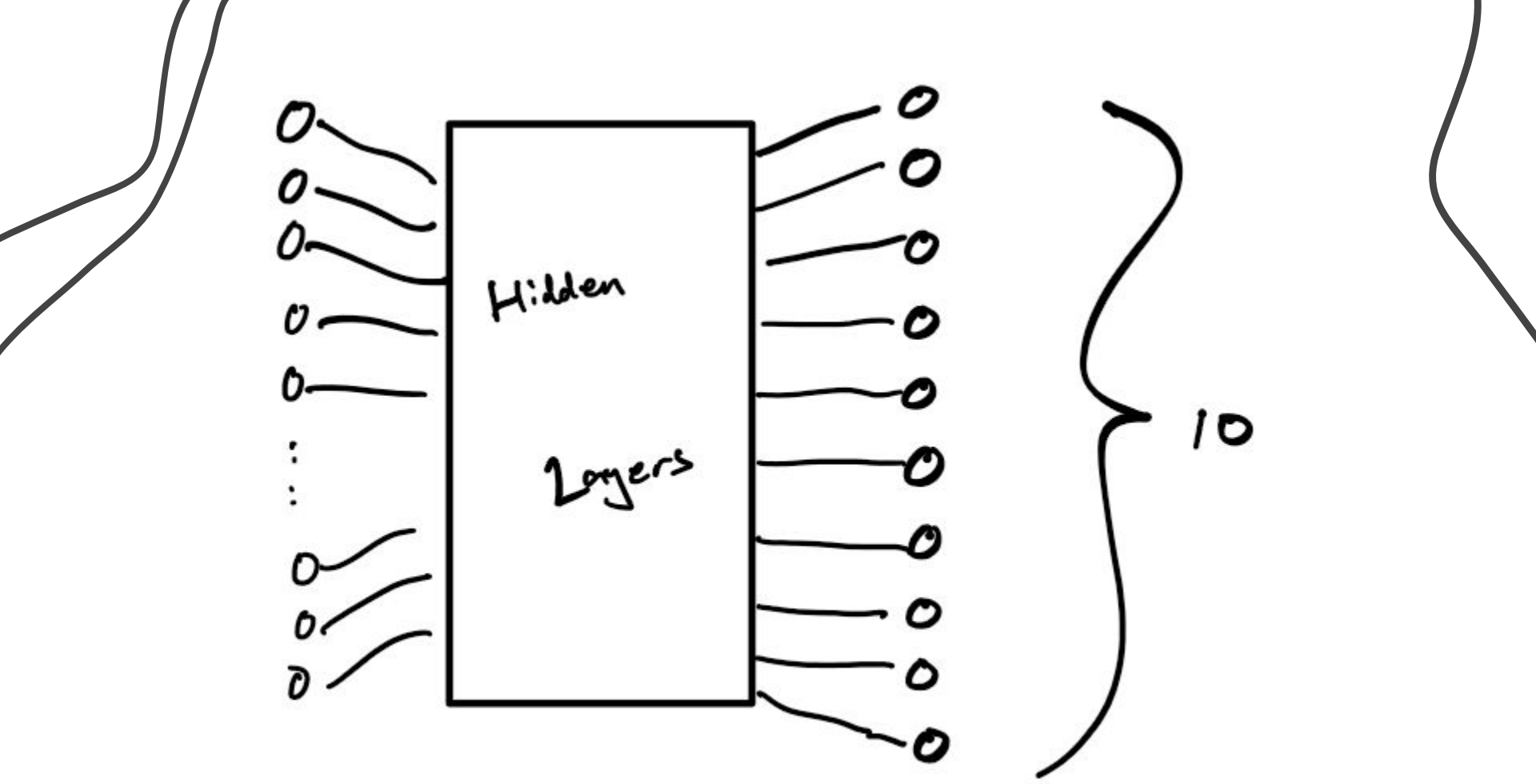

1.2
1
0.6
0.8
0
1.7

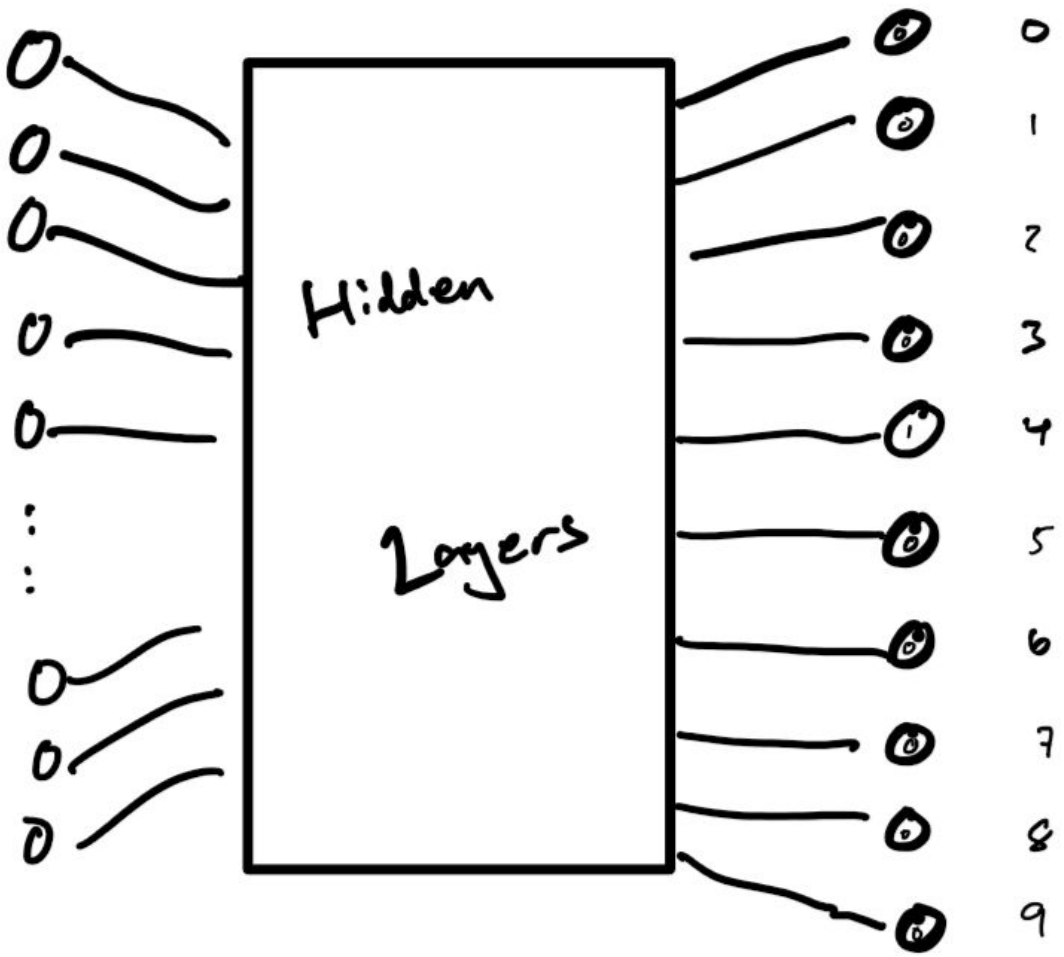

28
28

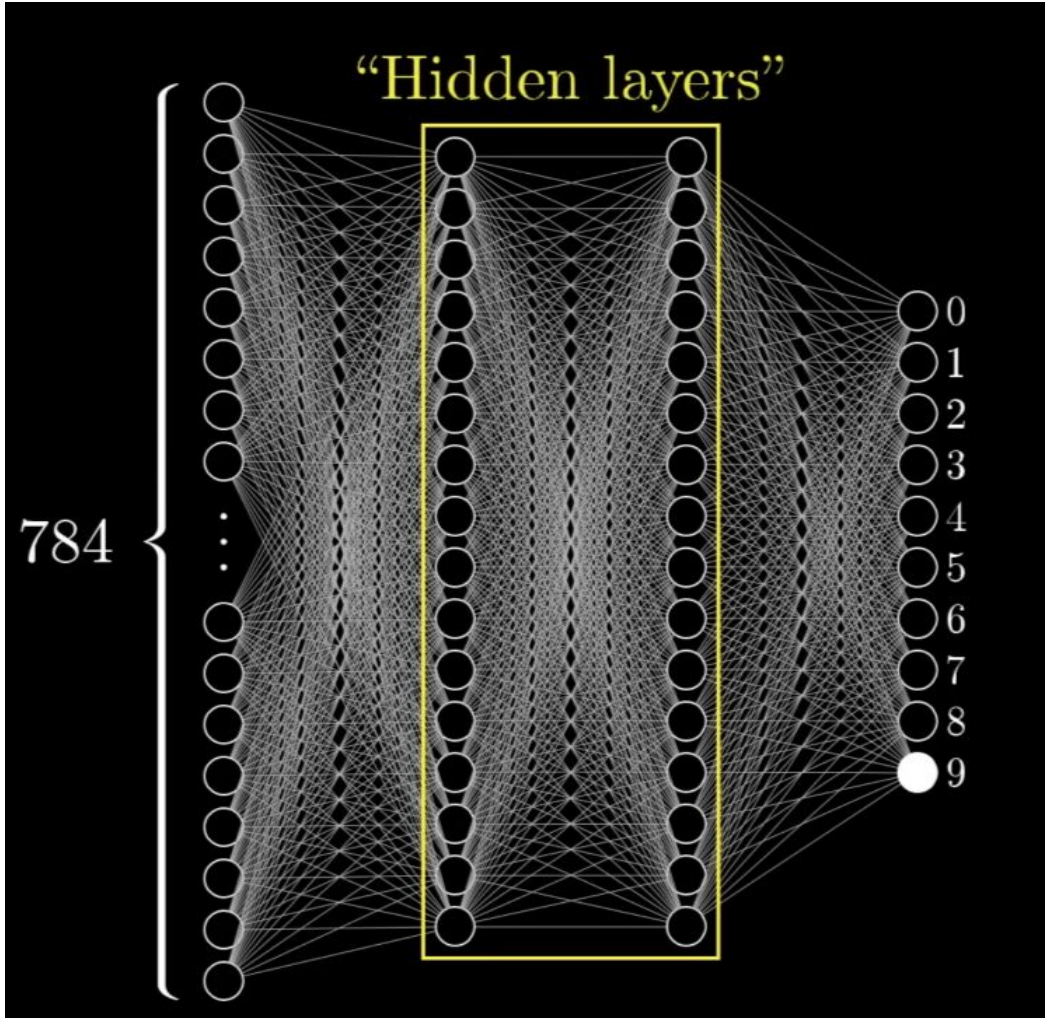



activation


784
Hidden
Layers

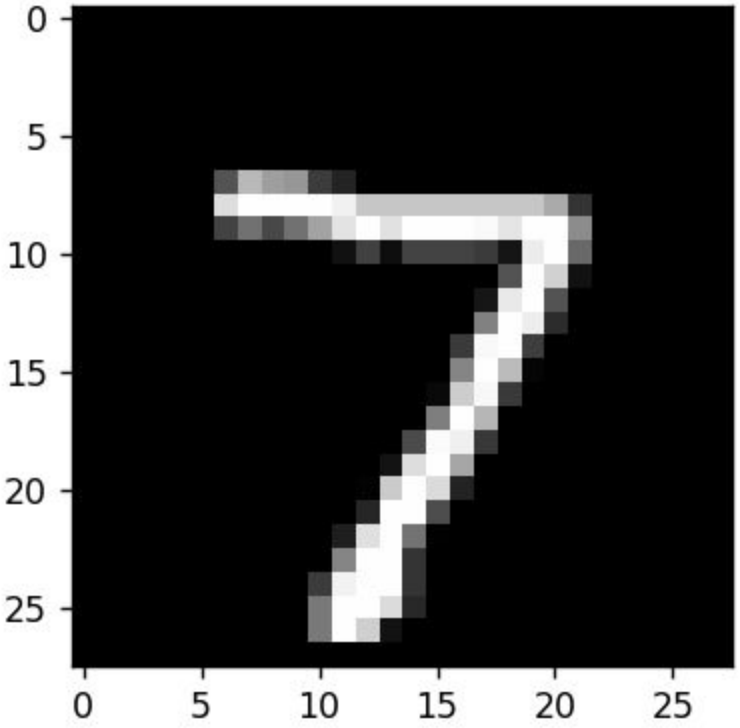

Hidden
Layers
10

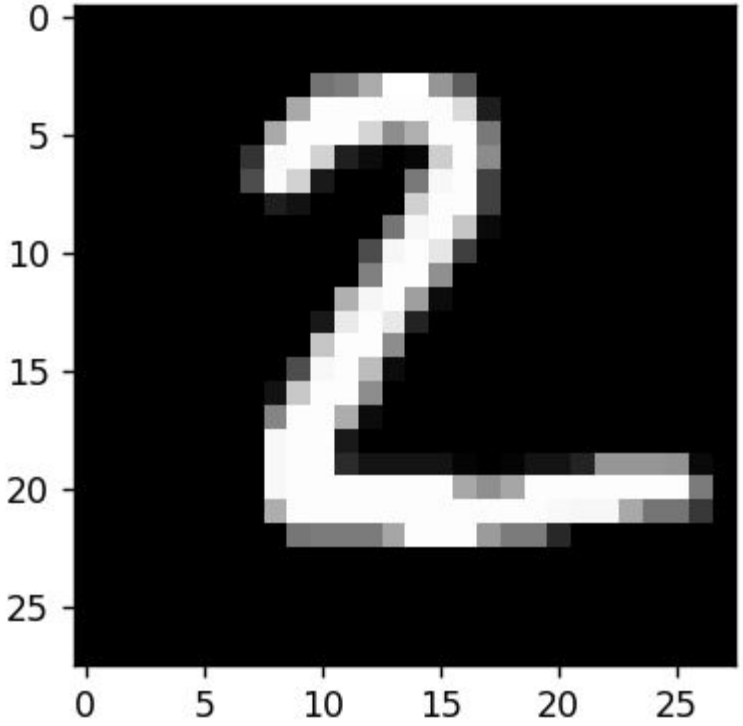

Hidden
Layers
0
1
2
3
4
5
6
7
8
9

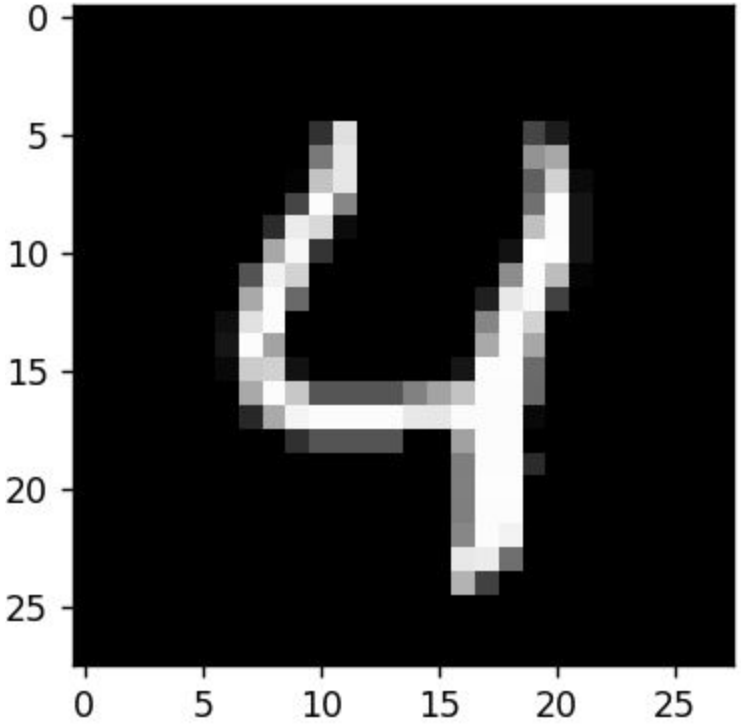

"Hidden layers"
784
0
1
2
3
4
5
6
7
8
9

# Why do we think a layered structure like this may work?

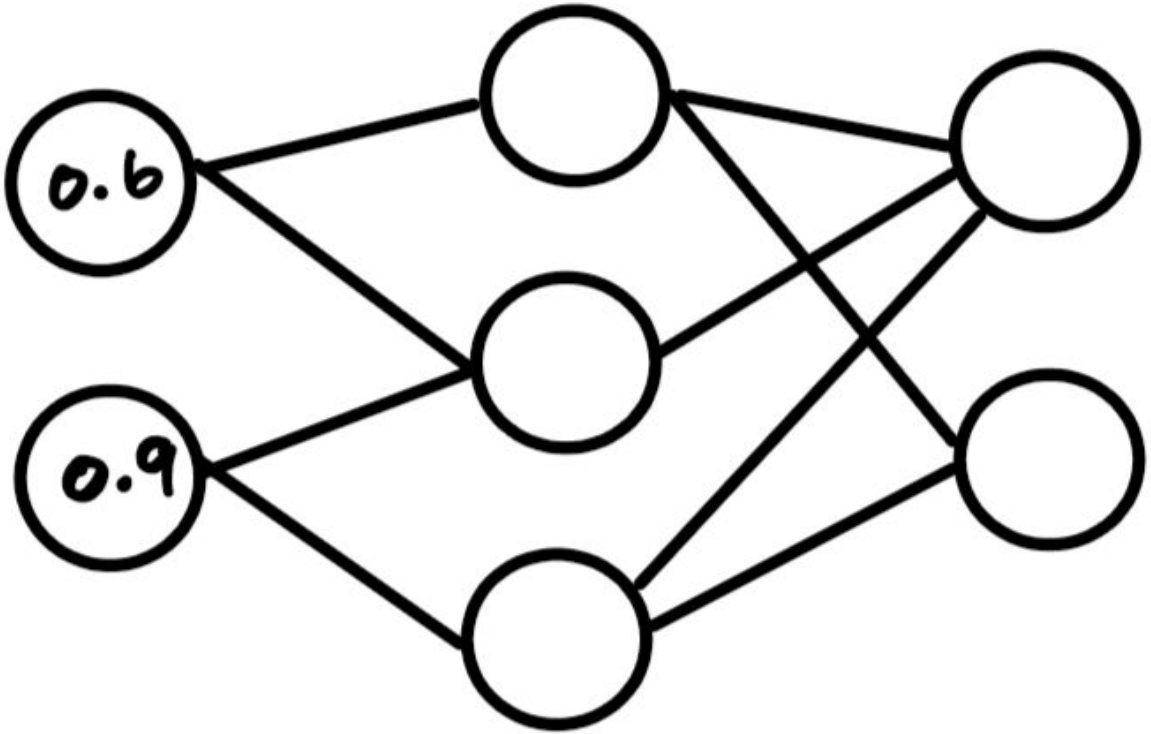
# What could each layer try to identify?

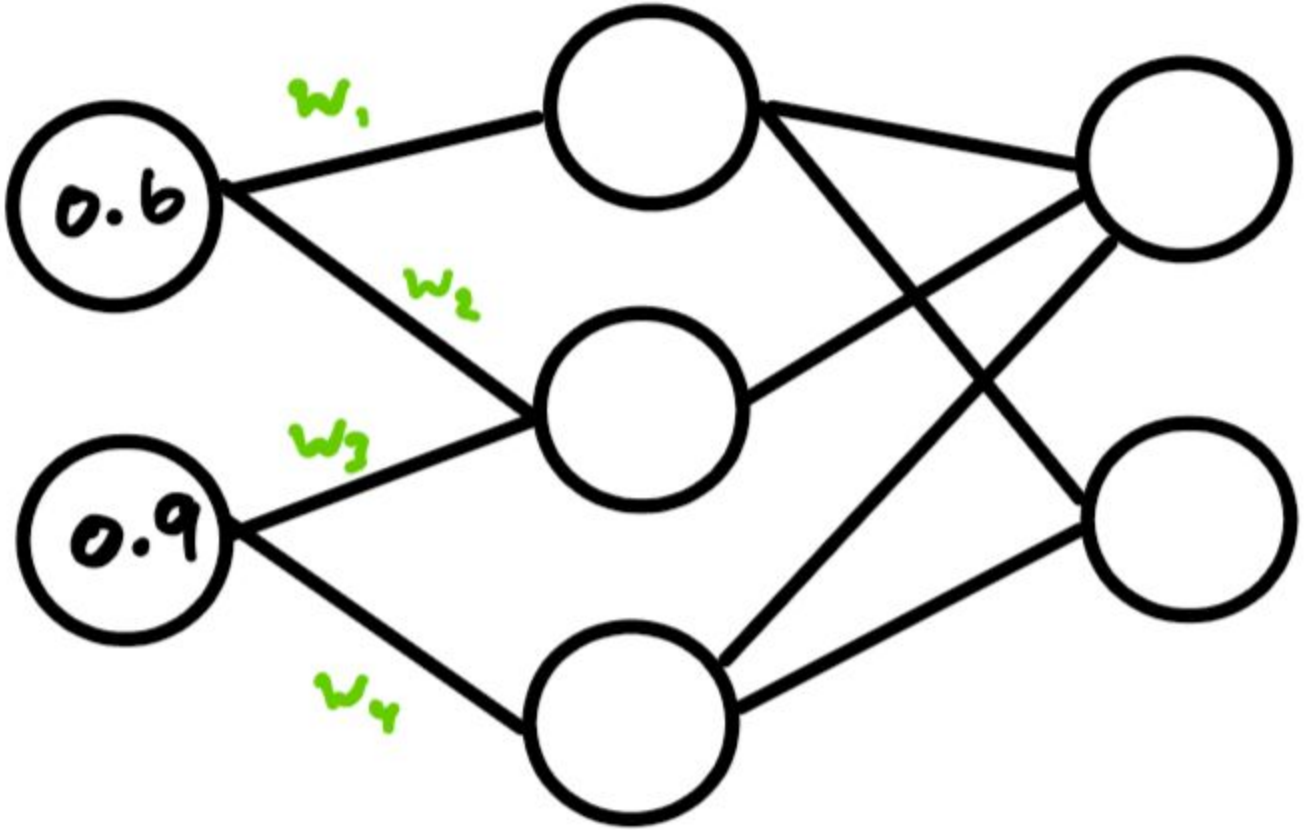

0
5
10
15
20
25
0
5
10
15
20
25


0
5
10
15
20
25
0
5
10
15
20
25

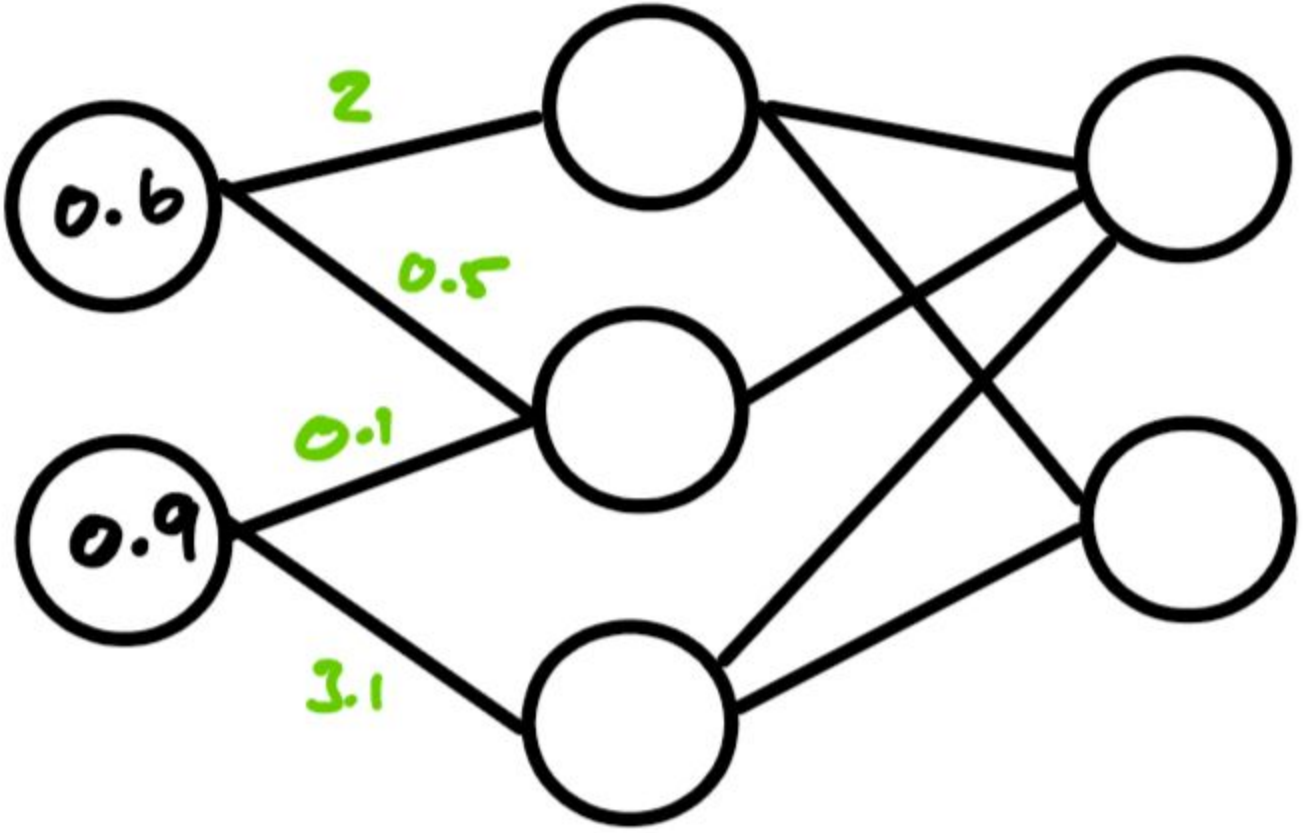

0
5
10
15
20
25
0
5
10
15
20
25

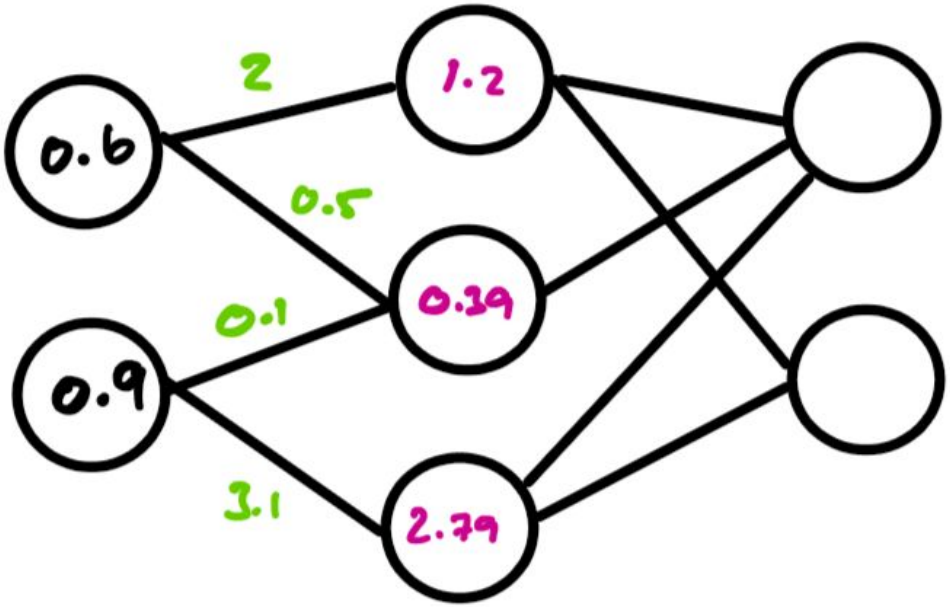
# How does one layer affect the next?


0.6
0.9




**weights**

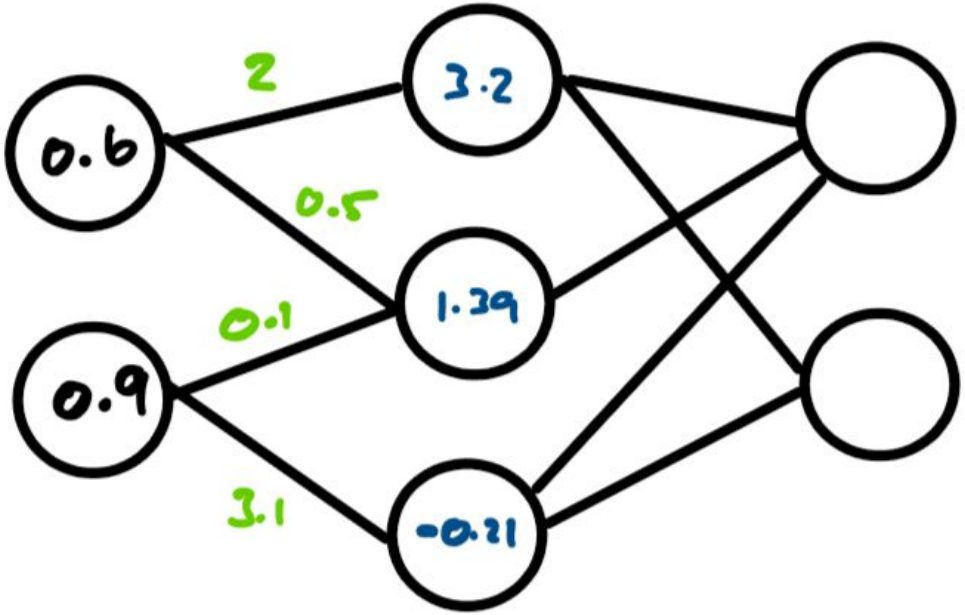

0.6
0.9
2
0.5
0.1
3.1




$0.6 \times 2 = 1.2$

$0.6 \times 0.5 + 0.9 \times 0.1 = 2.79$

**step 1:**
multiply by weights




1.2 + 2 = 3.2

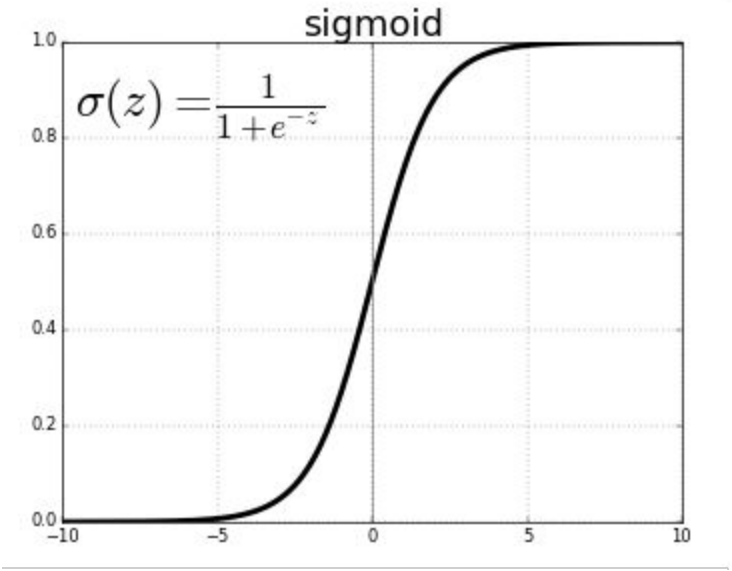
0.39 + 1 = 1.39

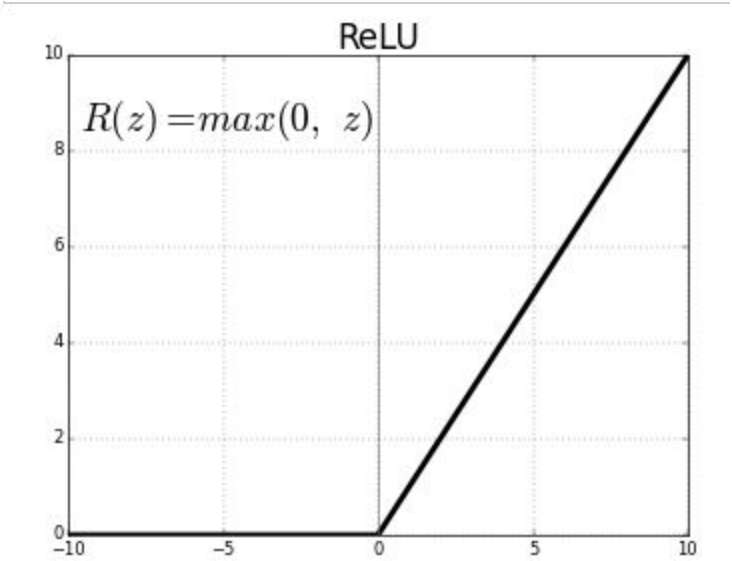
2.79 + -3 = -0.21

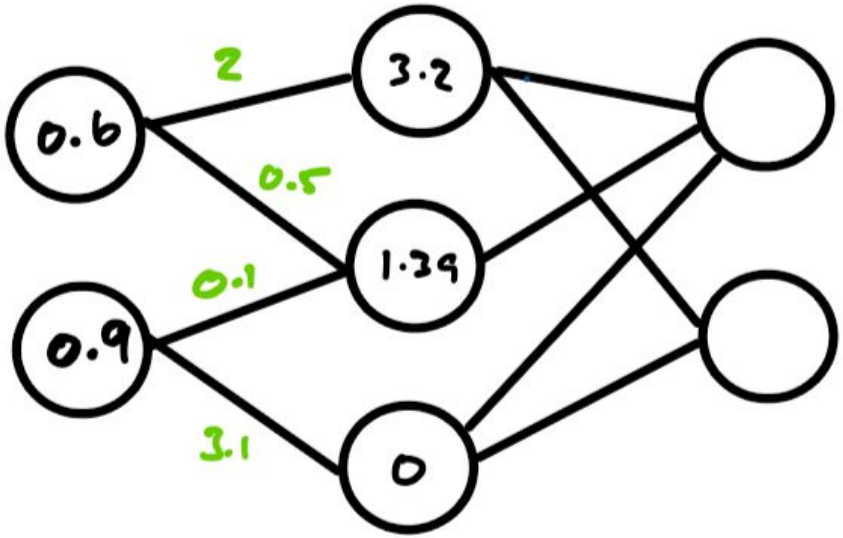
**step 2:**
add biases

# Activation Function

determines whether a neuron should
be activated or not


sigmoid
$\sigma(z) = \frac{1}{1+e^{-z}}$
1.0
0.8
0.6
0.4
0.2
0.0
−10
−5
0
5
10


ReLU
$R(z) = max(0,\ z)$
10
8
6
4
2
0
−10
−5
0
5
10

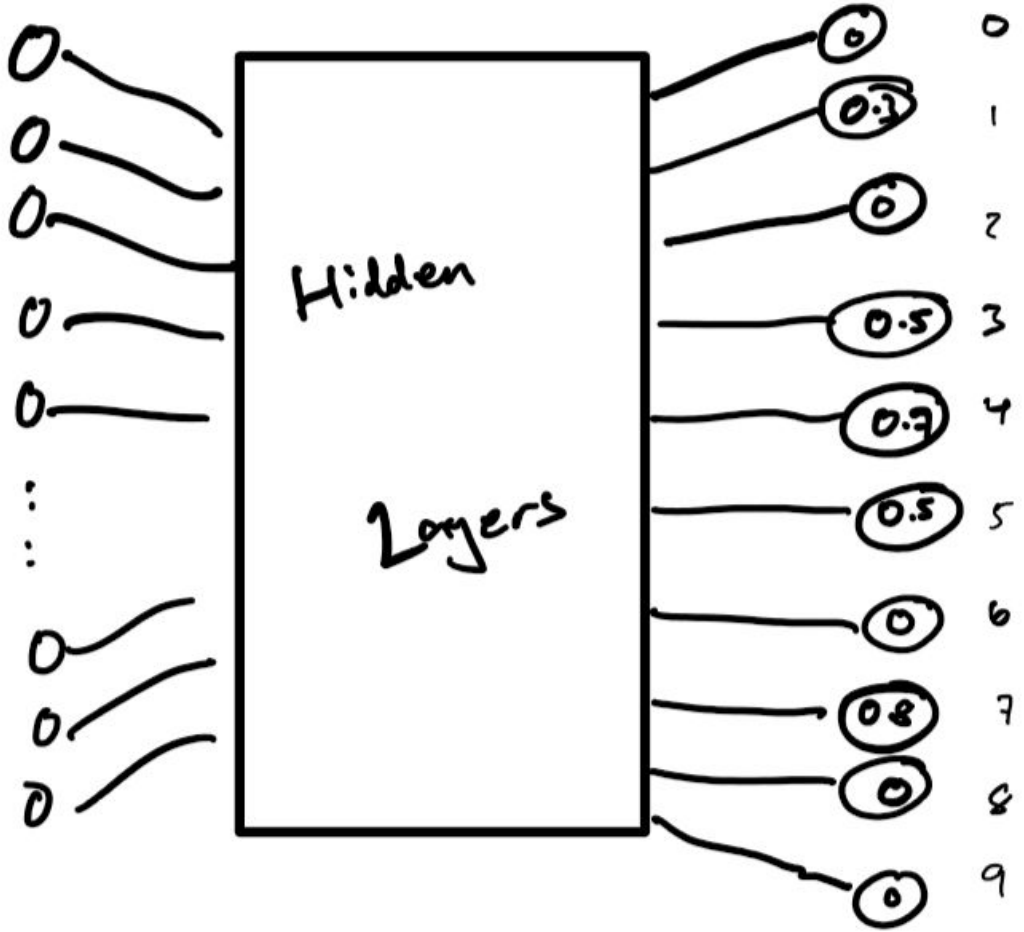



$$ReLU(3.2) = \max(0, 3.2) = 3.2$$

$$ReLU(1.39) = 1.39$$

$$ReLU(-0.21) = 0$$

**step 3:**
apply activation


Hidden
Layers
0
0
0
0
0
0
0
0
0.3
0
0.5
0.3
0.5
0
0.8
0
0
0
1
2
3
4
5
6
7
8
9

# Use Gradient Descent to minimize the cost function

# Backpropagation

algorithm to compute the gradients of
a neural network