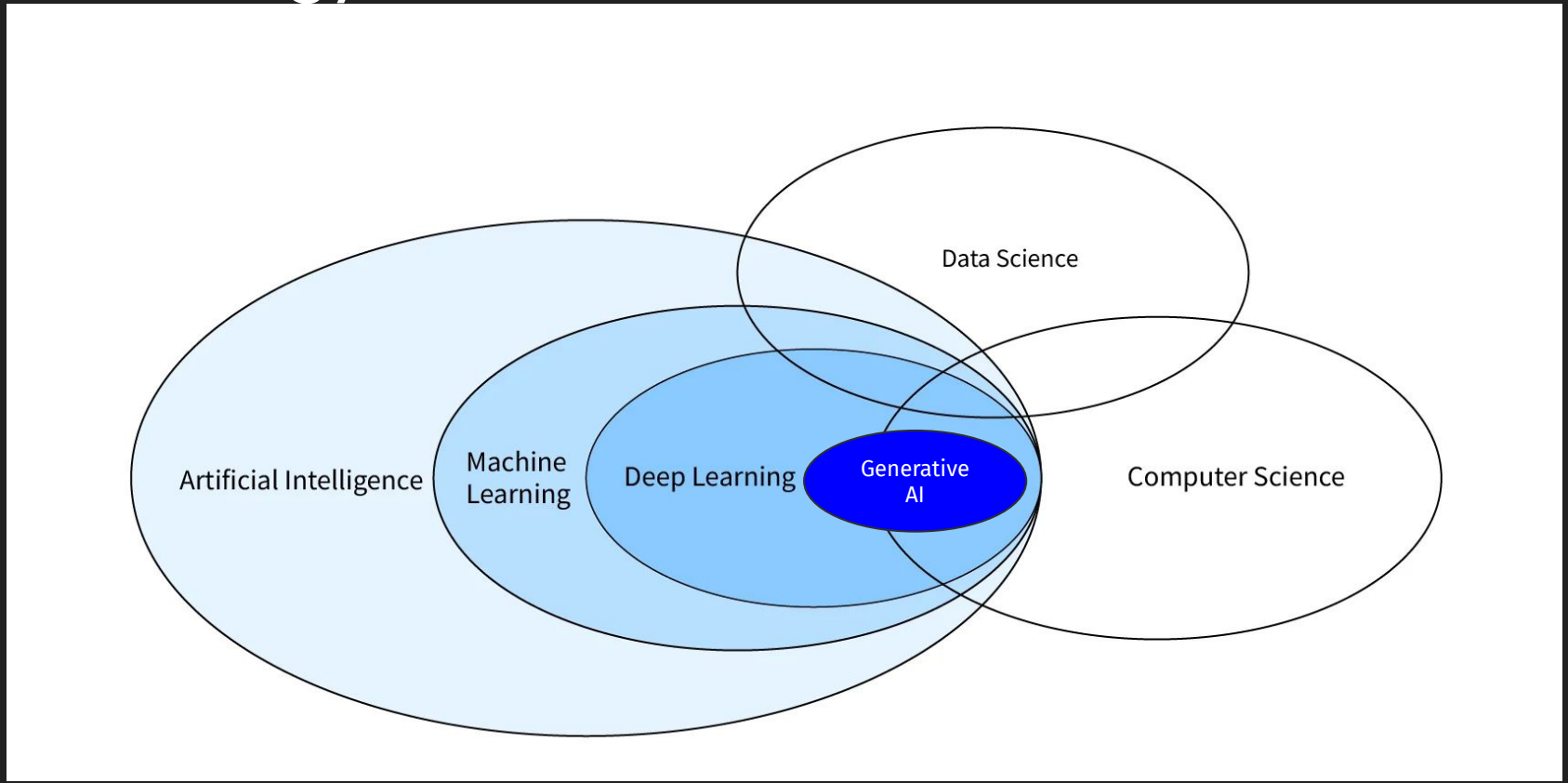
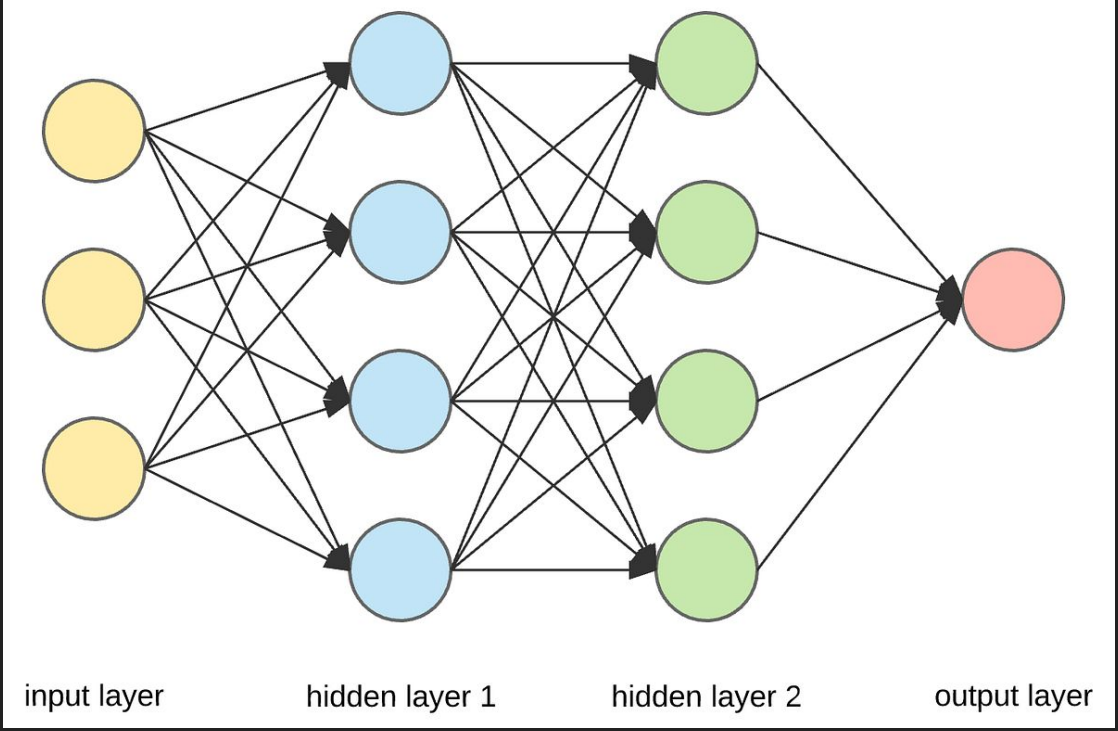
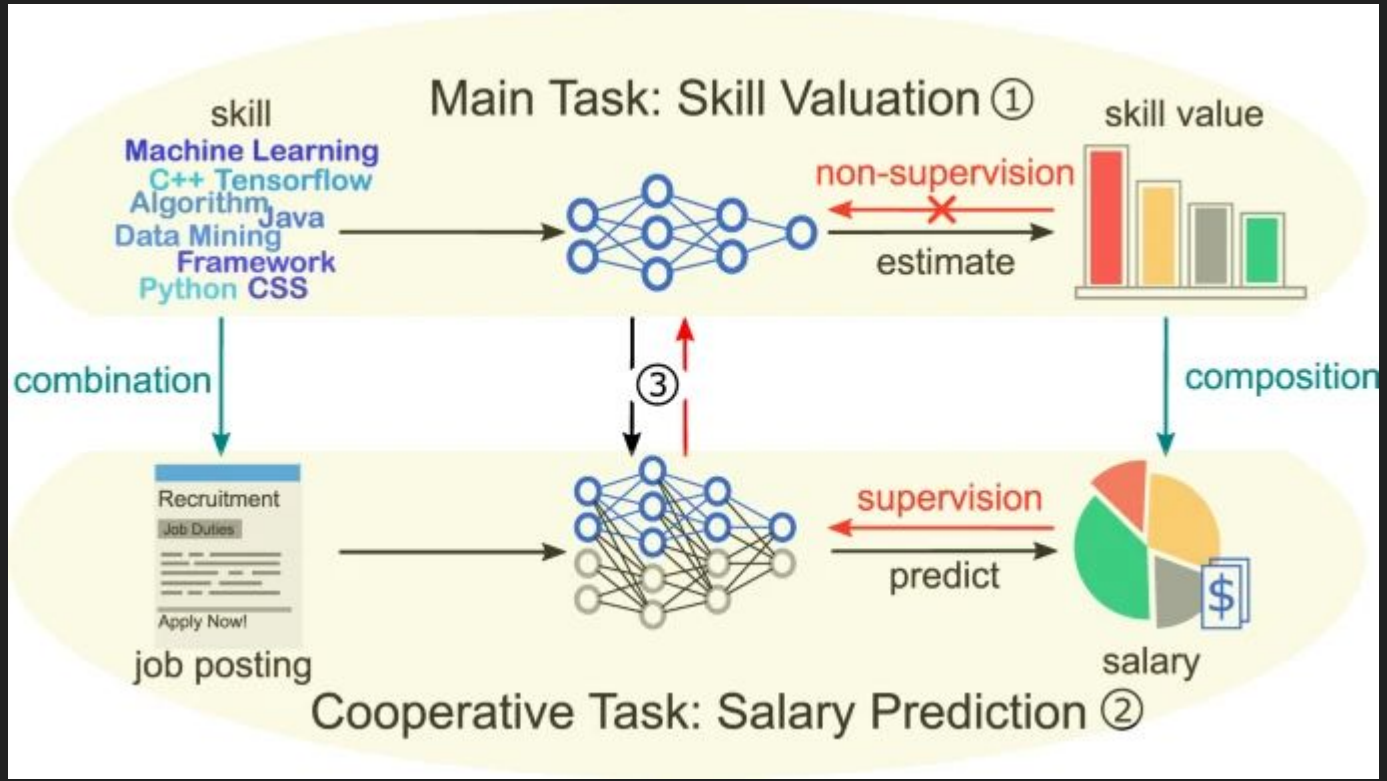
An insufficient introduction to neural networks

Professor Paul Ohm
Georgetown Law
Spring 2025

Terminology

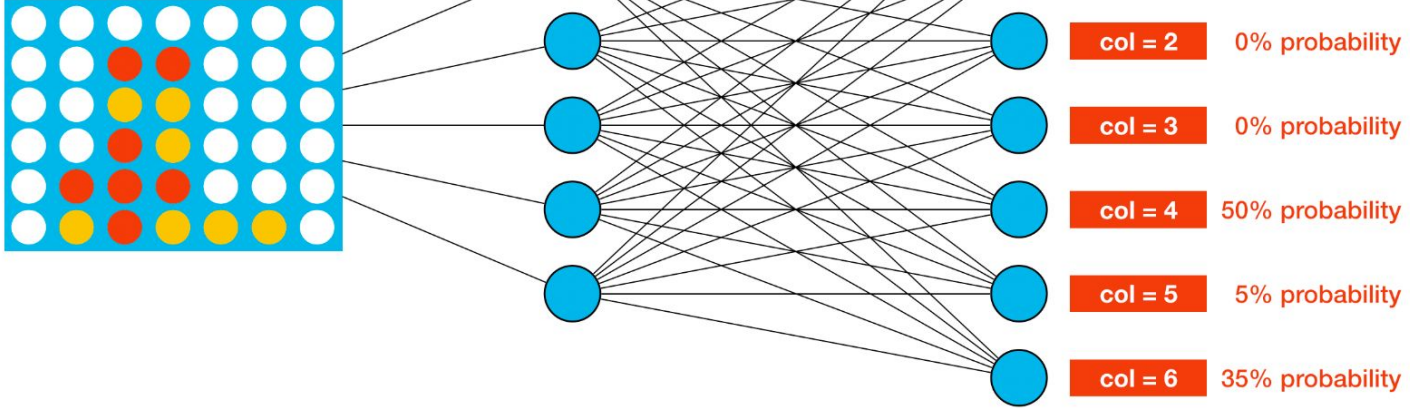



input layer
hidden layer 1
hidden layer 2
output layer



https://www.nature.com/articles/s41467-021-22215-y

Reinforcement Learning



Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

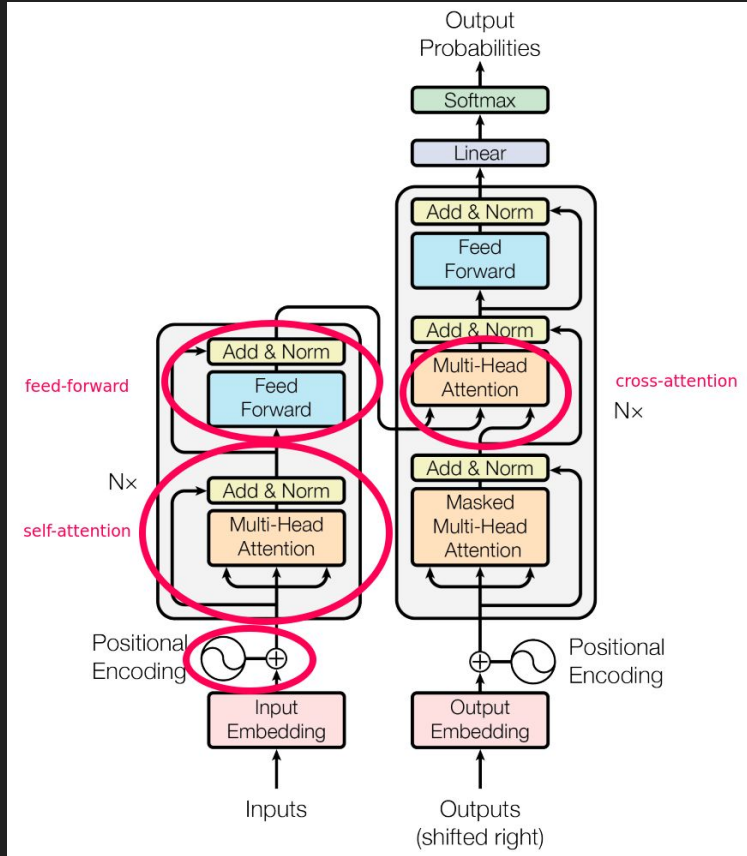


Figure 1: The Transformer - model architecture.

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

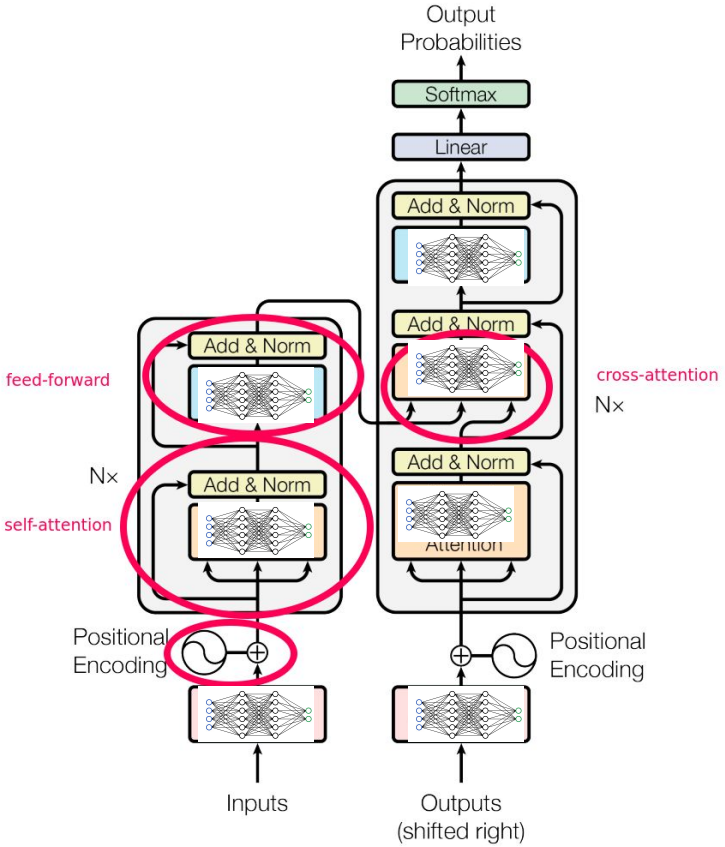
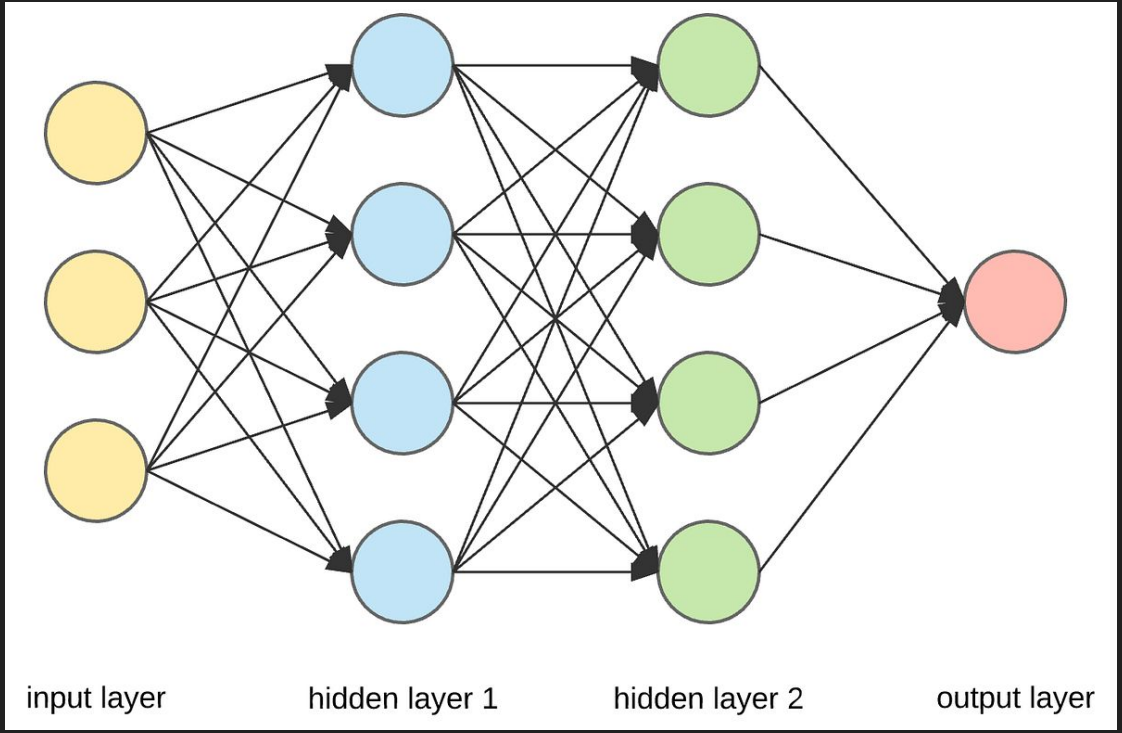


Figure 1: The Transformer - model architecture.

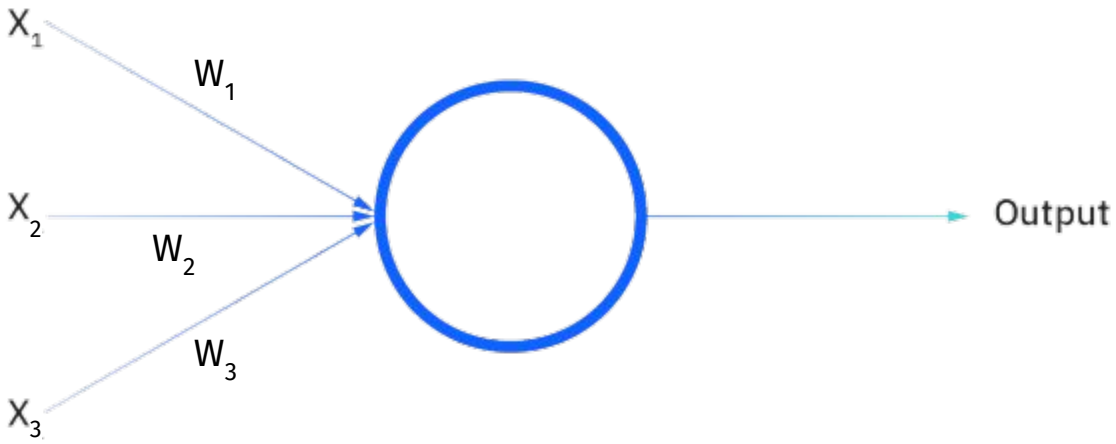
What is a Neural Network?

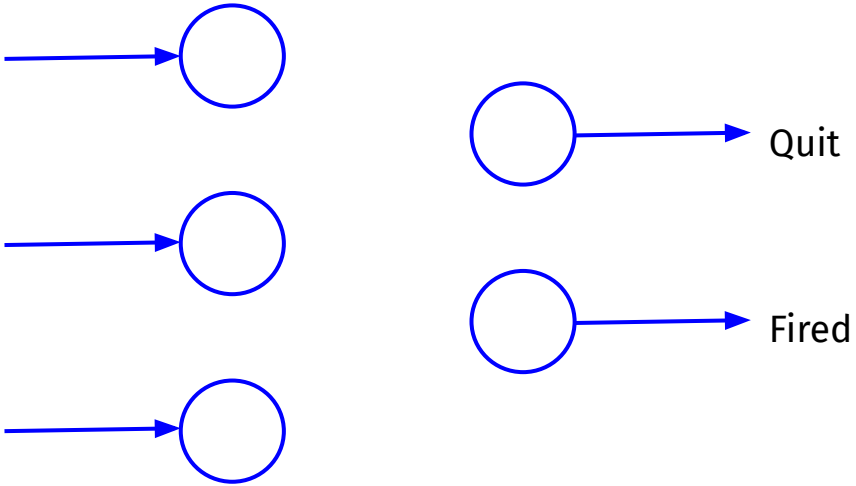
A machine to find relationships between inputs and outputs modeled on data


input layer
hidden layer 1
hidden layer 2
output layer

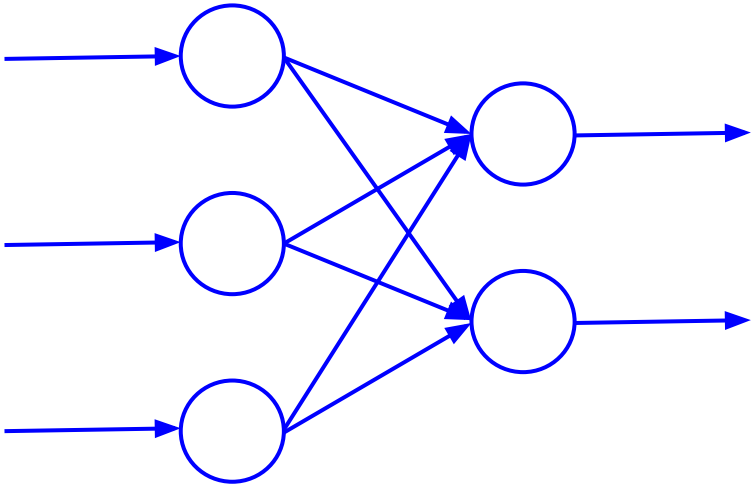
Neurons / Nodes

$$x_1w_1 + x_2w_2 + x_3w_3 = \text{prediction}$$

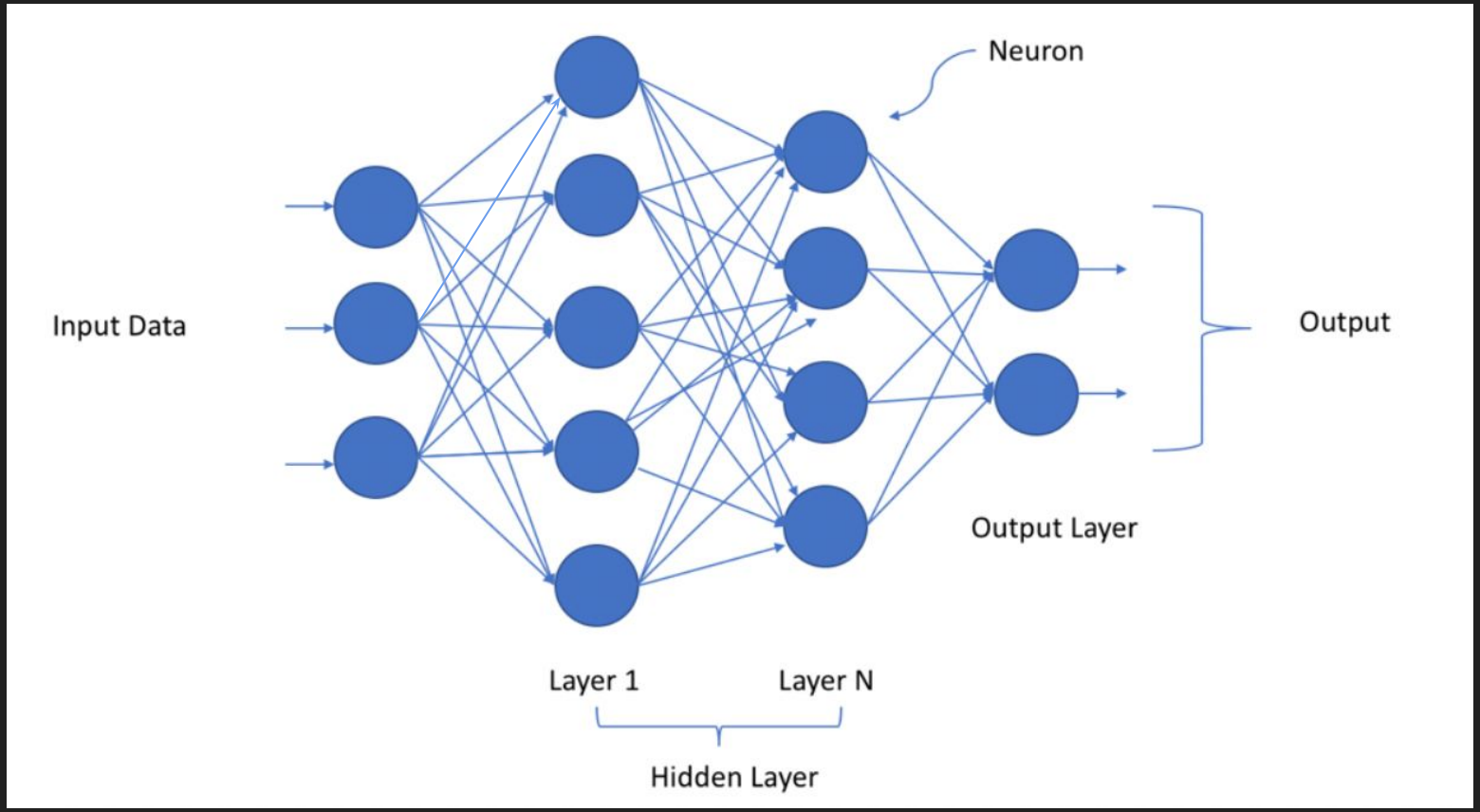



Quit
Fired

$$x_1 w_{1q} + x_2 w_{2q} + x_3 w_{3q} = \text{quit}$$
$$x_1 w_{1f} + x_2 w_{2f} + x_3 w_{3f} = \text{fired}$$



Going “deep”


Neuron
Input Data
Output
Output Layer
Layer 1
Layer N
Hidden Layer

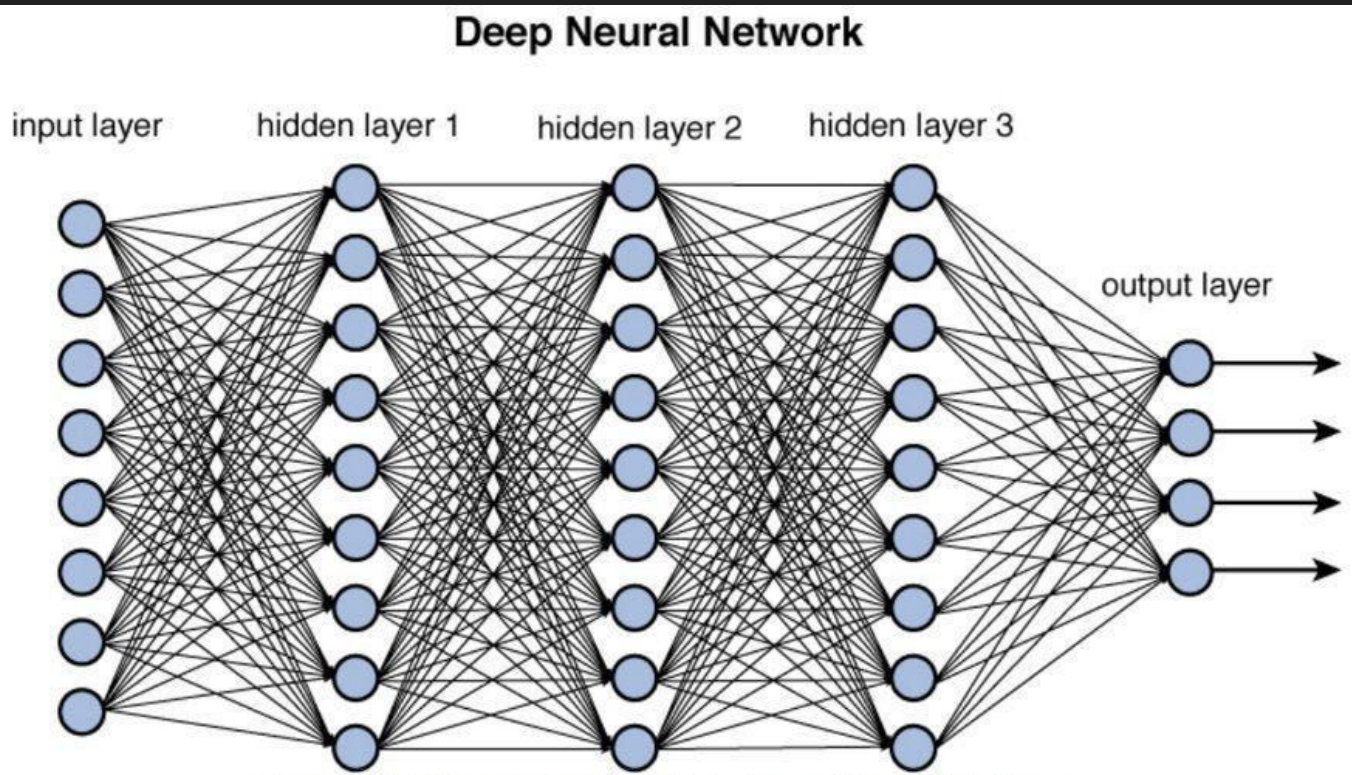
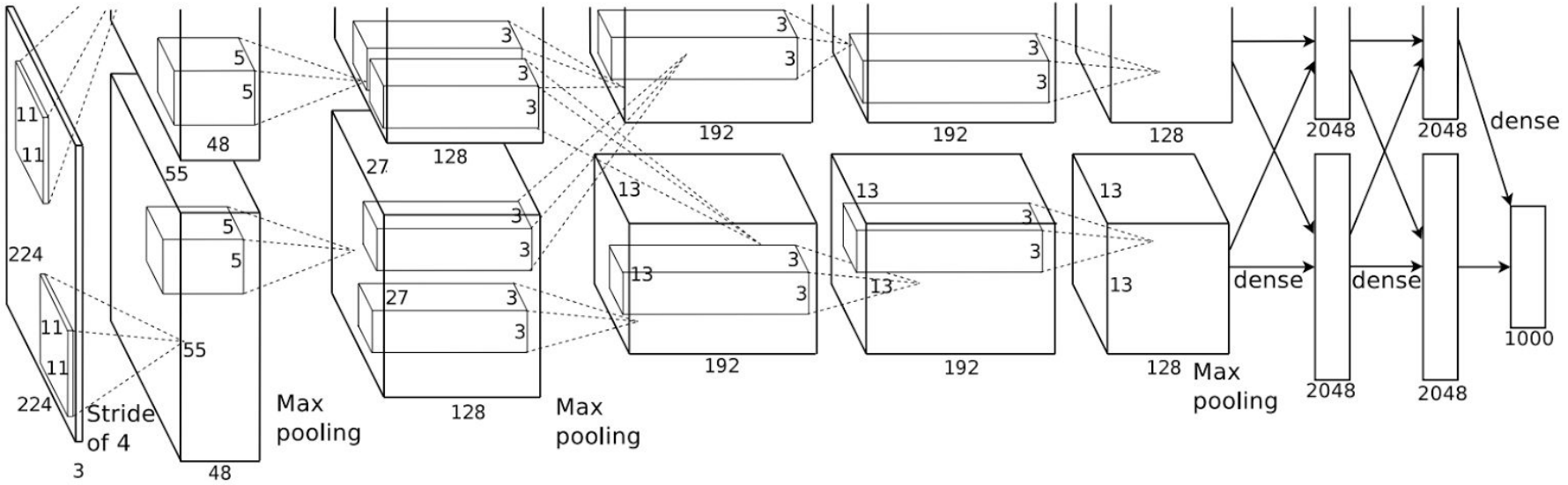
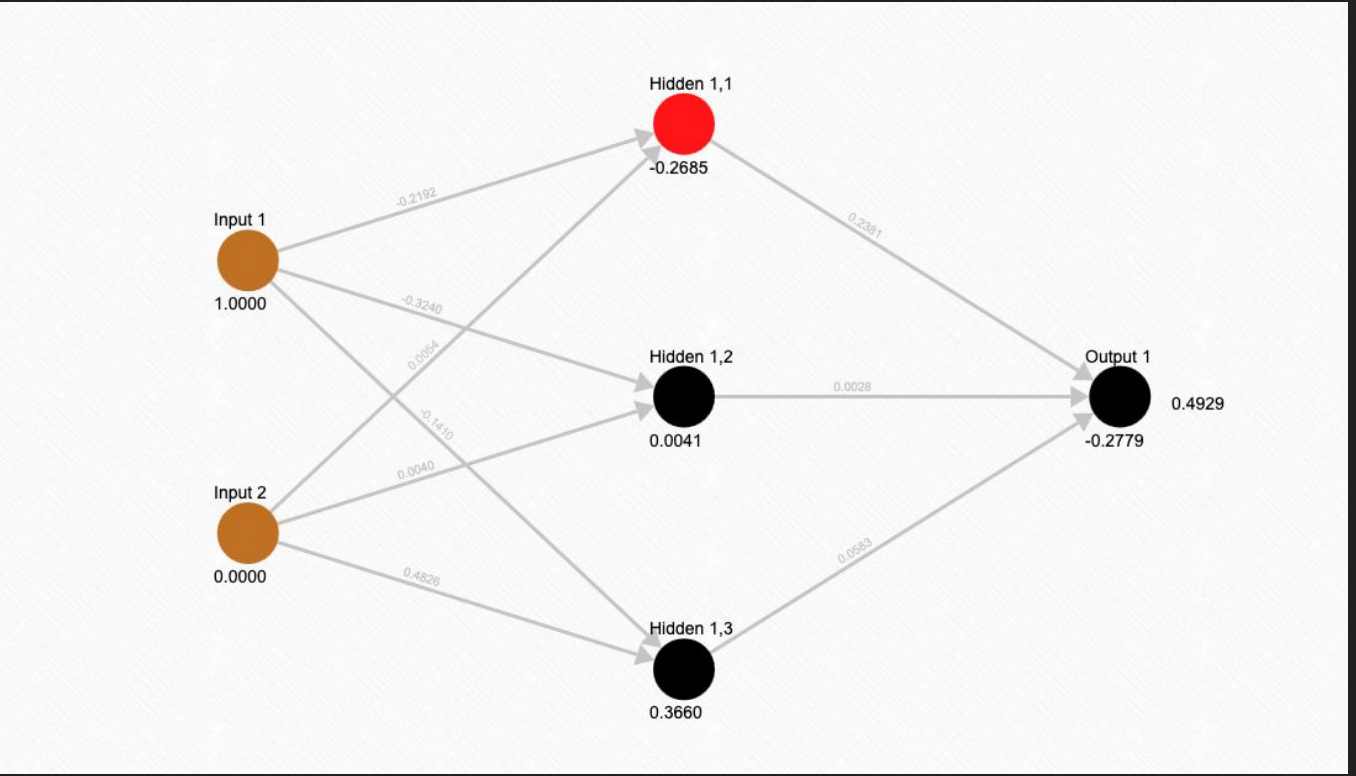


Figure 12.2 Deep network architecture with multiple layers.


11
11
224
224
3
Stride
of 4
5
5
48
55
5
5
48
55
Max
pooling
3
3
3
128
27
3
3
3
128
27
Max
pooling
3
3
192
3
3
192
128
13
13
3
3
192
13
13
3
3
192
13
13
128
Max
pooling
dense
dense
2048
2048
2048
2048
dense
1000

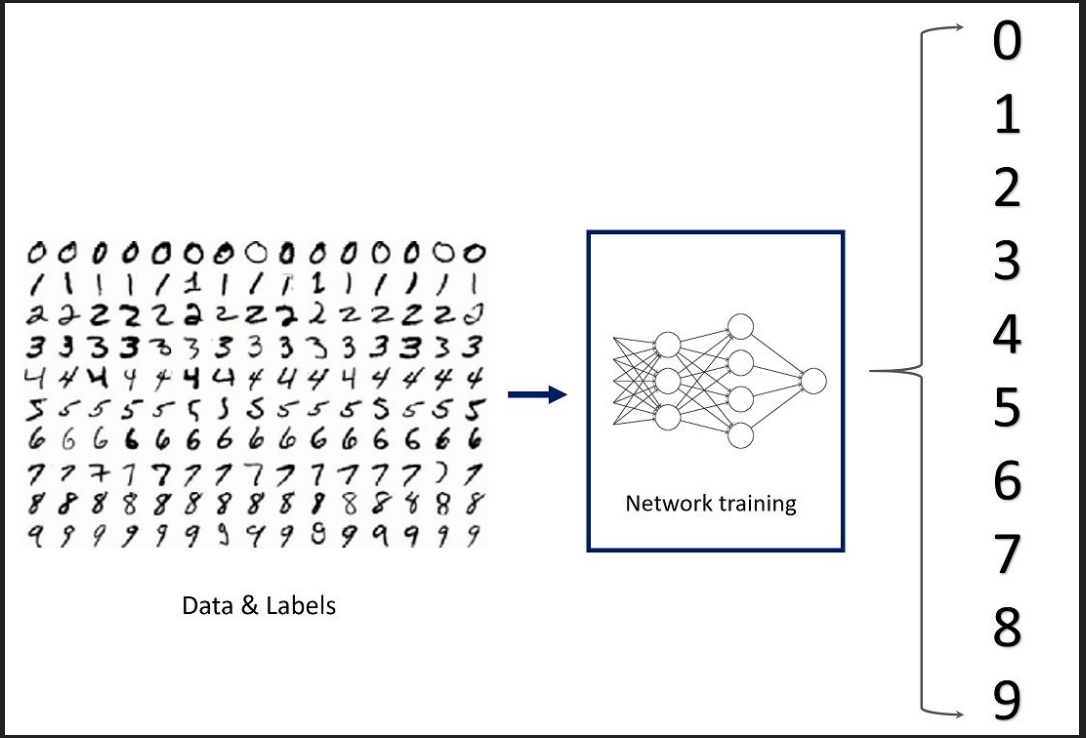
Training (intuition not deep understanding)

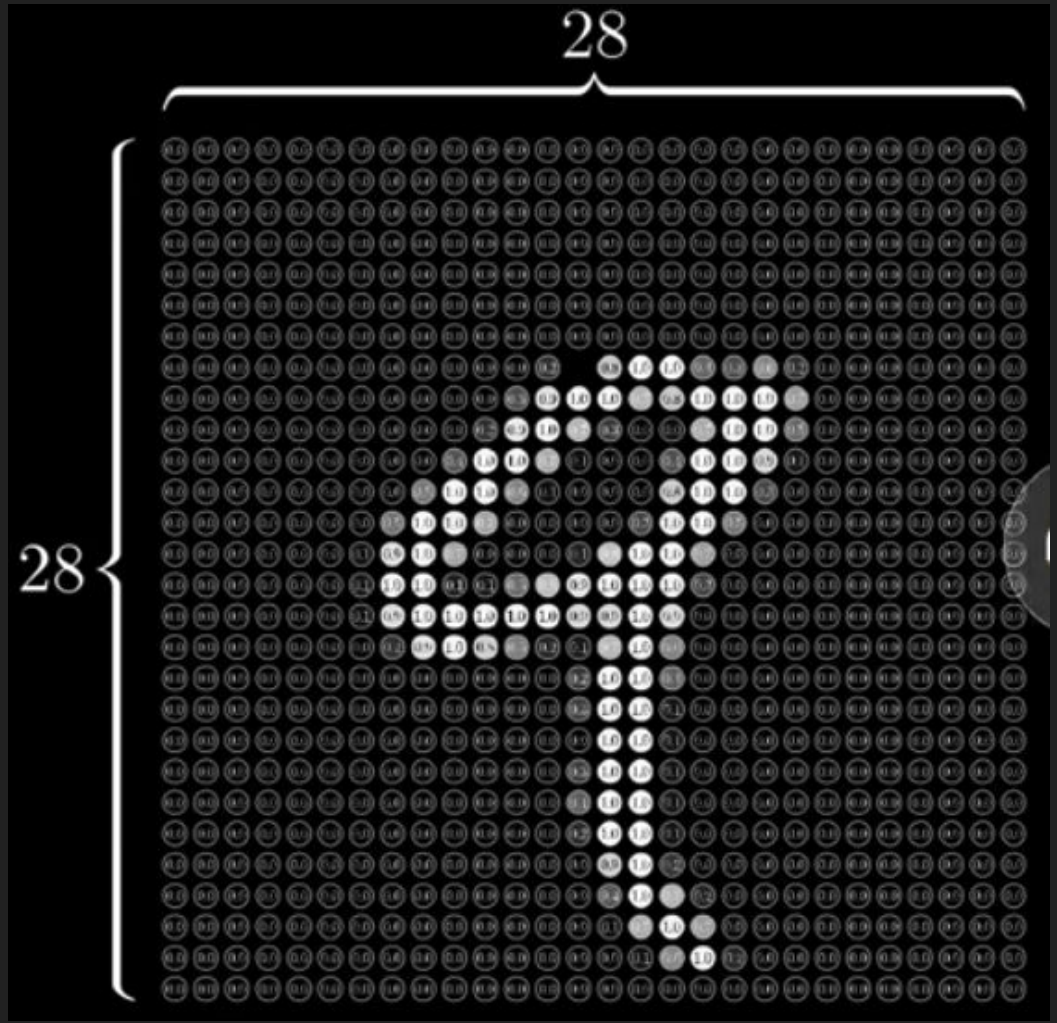


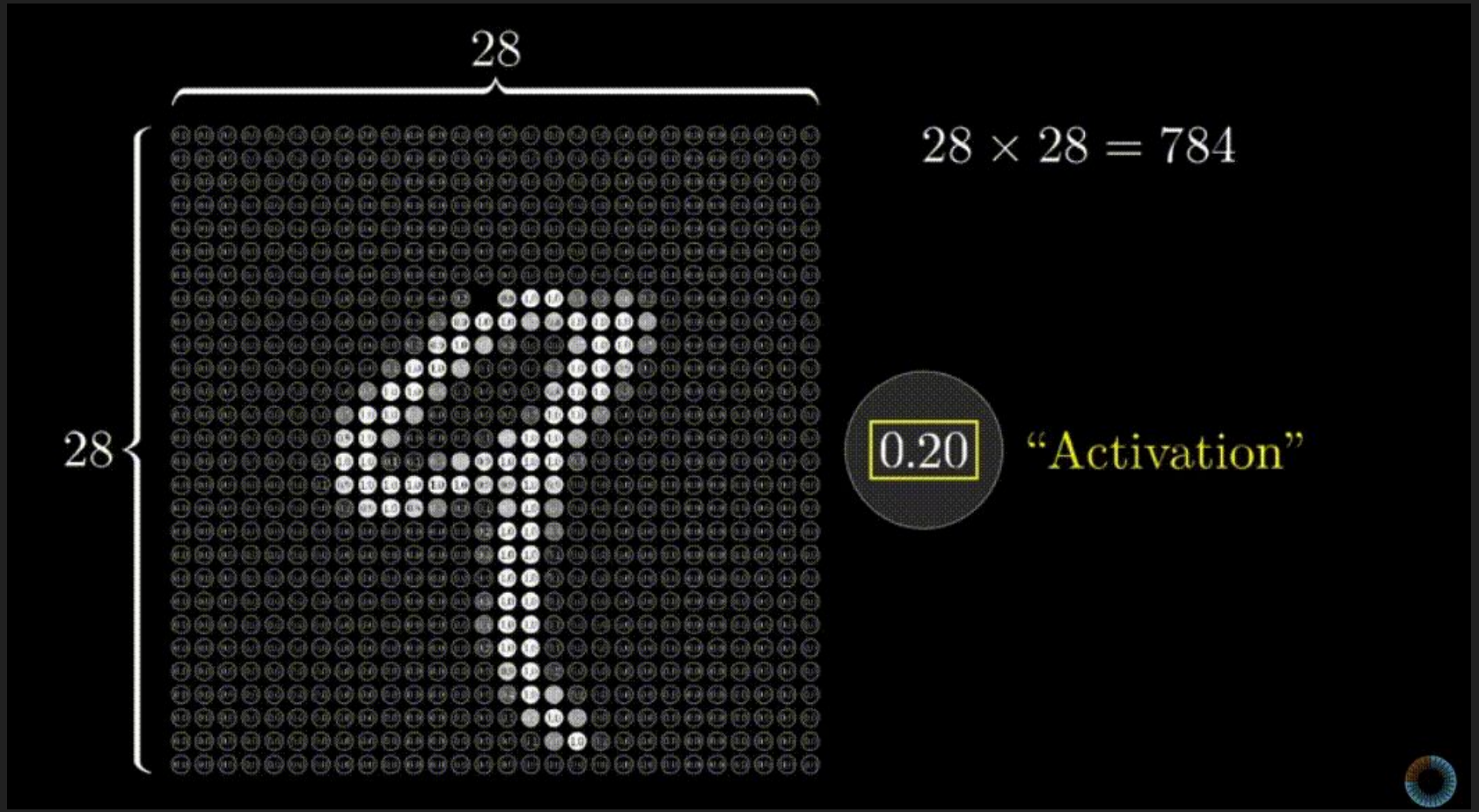
https://experiments.mostafa.io/ffbpann/

Case Study:
Image Identification

Supervised Learning - MNIST




28
28


28
28
$28 \times 28 = 784$
0.20
“Activation”

Convolutional Neural Networks