Training a Convolutional Neural Network to Classify Images

Pratheerth Padman
Freelance Data Scientist

Module Overview

What are convolutional neural networks (CNNs)?

Convolutions, activation, and pooling

Classification using a CNN

Demo: Creating the CNN architecture

Demo: Training the model

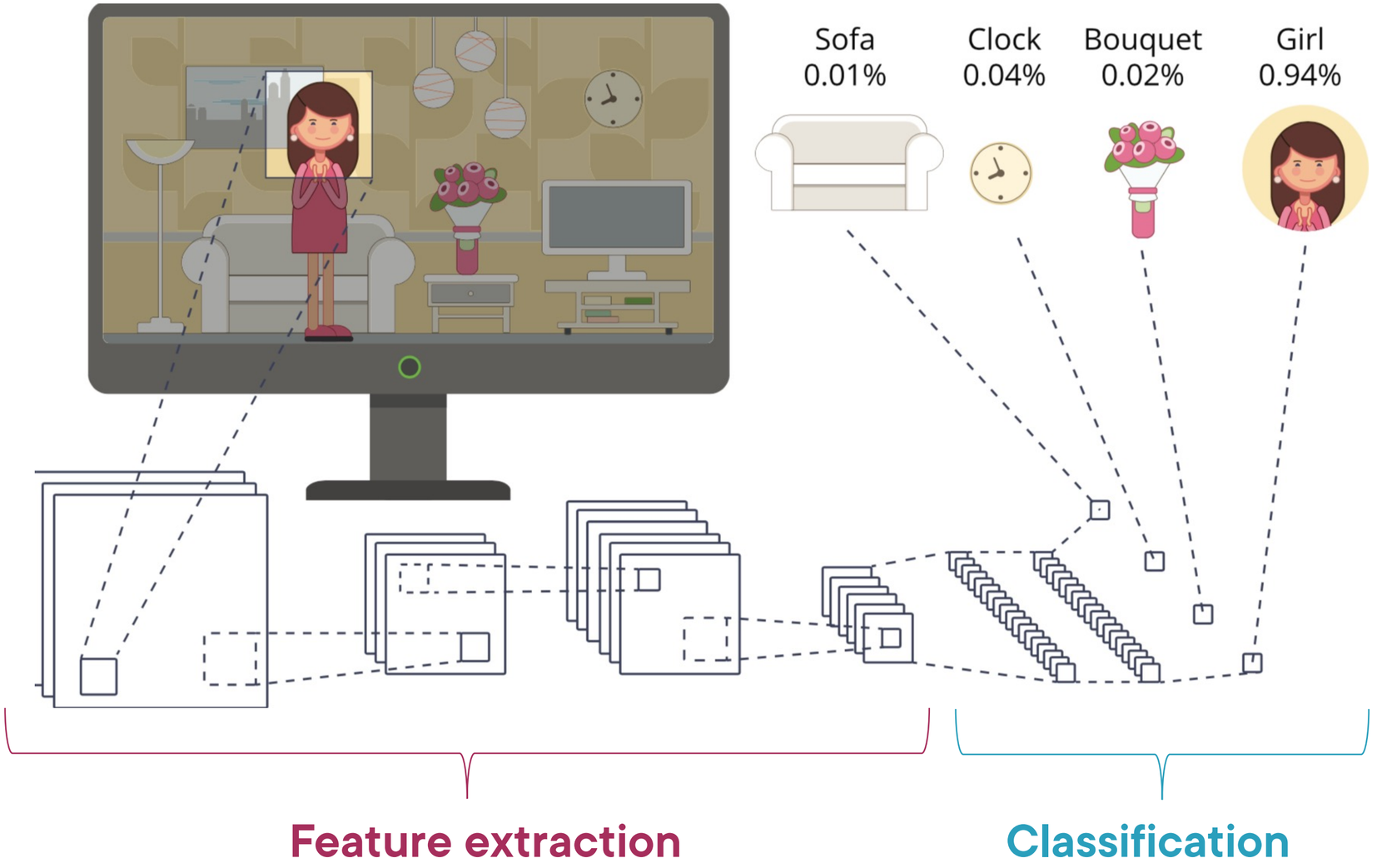
Demo: Performance metrics – how well did our model do?

What Are Convolutional Neural Networks?

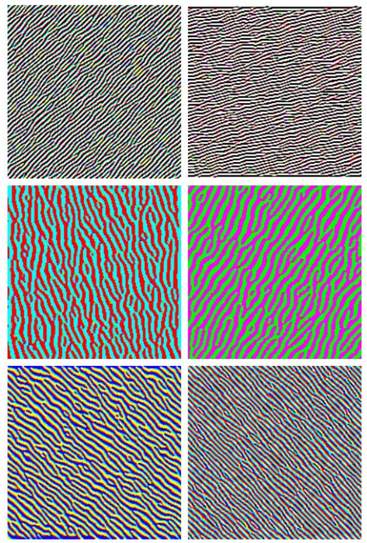
Convolutional Neural Network (CNN)

A Convolutional Neural Network is a type of deep learning algorithm that can take images as inputs, assign importance to various parts or aspects of these, and then differentiate one image from the others.

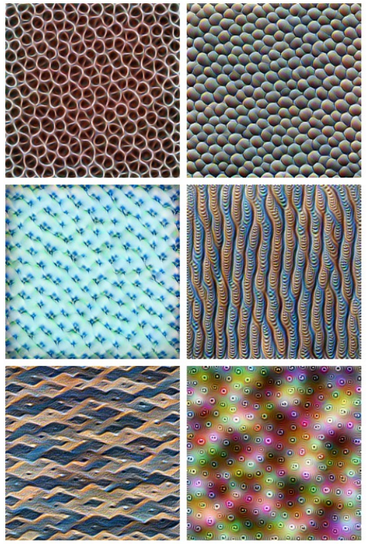
Convolutional Neural Network - Layout



What Does a CNN "See"?

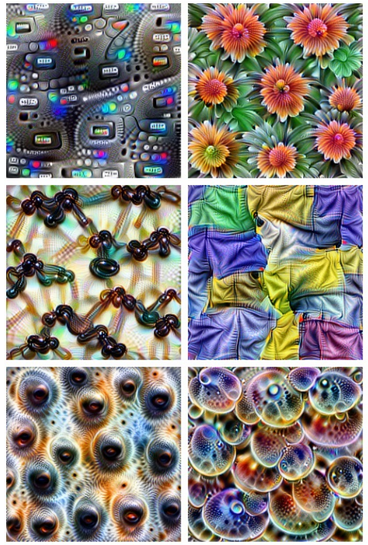


Edges



Textures

Patterns



Parts

Objects

Citation: Olah, et al., "Feature Visualization", Distill, 2017 - https://distill.pub/2017/feature-visualization/

Feature Extraction

Convolutions

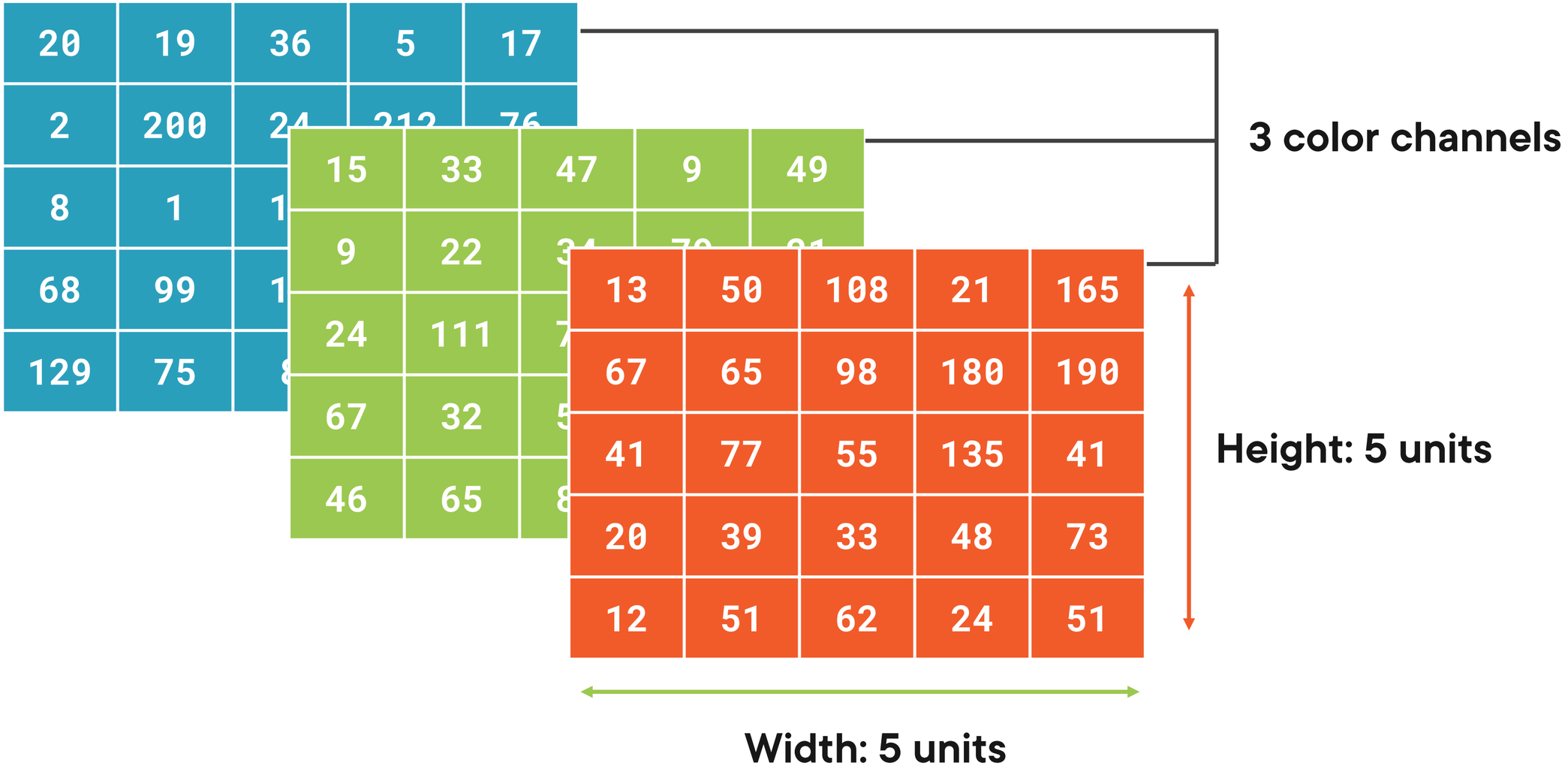
Activation

Pooling

CNN: Convolutions

Input image
Filter
Feature maps

Input Image


20 19 36 5 17
2 200
8 1
68 99
129 75
15 33 47 9 49
9 22
24 111
67 32
46 65
13 50 108 21 165
67 65 98 180 190
41 77 55 135 41
20 39 33 48 73
12 51 62 24 51
3 color channels
Height: 5 units
Width: 5 units

Convolution

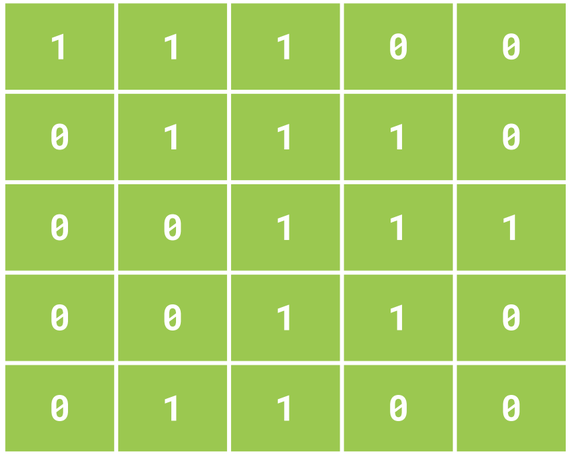
1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Input image

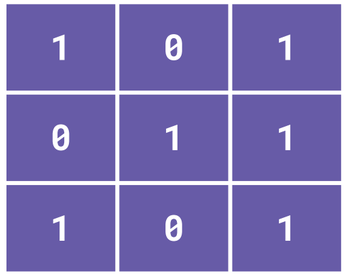
1	0	1
0	1	1
1	0	1

Filter

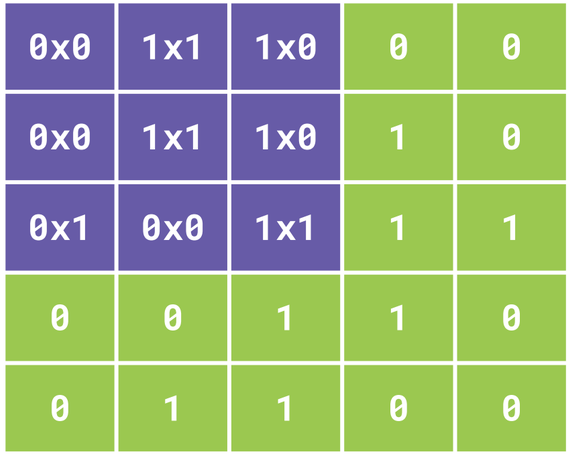
Convolution



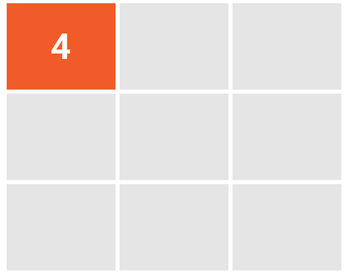
Input



Filter



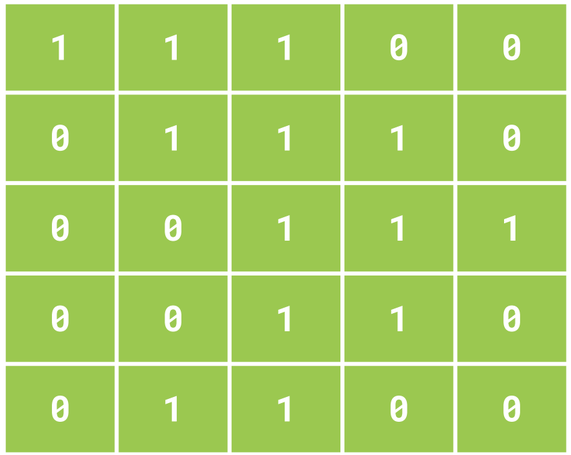
Sliding
(Stride = 1)



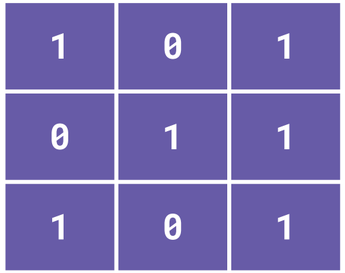
Feature map

Calculations - (1x1 + 0x1 + 1x1) + (0x0 + 1x1 + 1x0) + (0x1 + 0x0 + 1x1) = 4

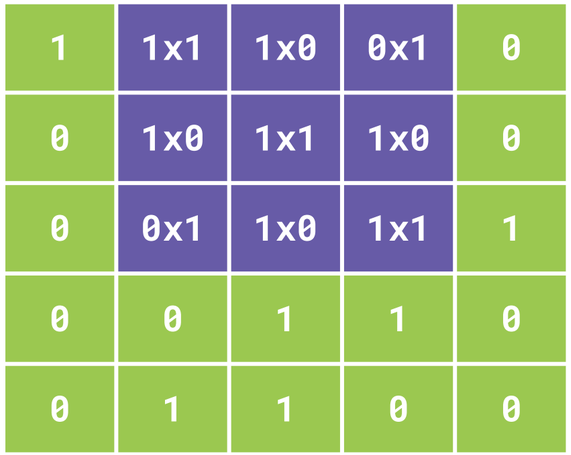
Convolution



Input



Filter



Sliding 2
(Stride = 1)

Stride

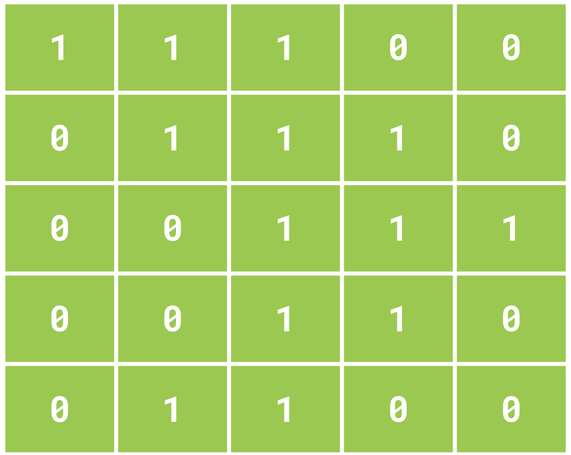
1x1	1x0	1x1	0	0
0x0	1x1	1x0	1	0
0x1	0x0	1x1	1	1
0	0	1	1	0
0	1	1	0	0

Sliding
(Stride = 2)

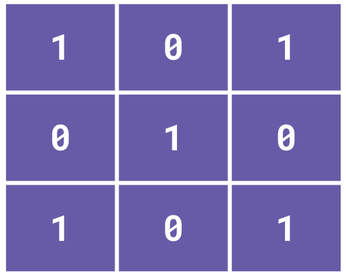
1	1	1x1	0x0	0x1
0	1	1x0	1x1	0x0
0	0	1x1	1x0	1x1
0	0	1	1	0
0	1	1	0	0

Sliding
(Stride = 2)

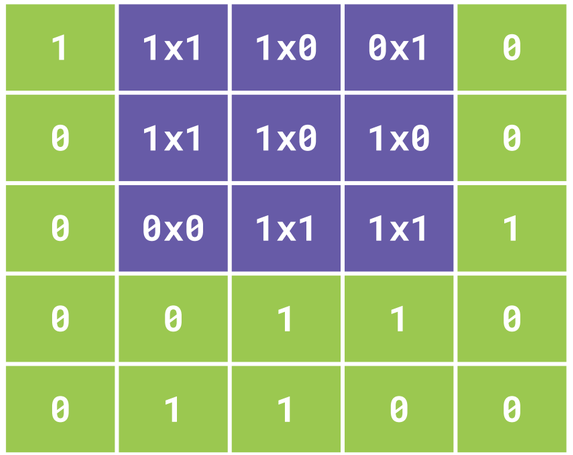
Convolution



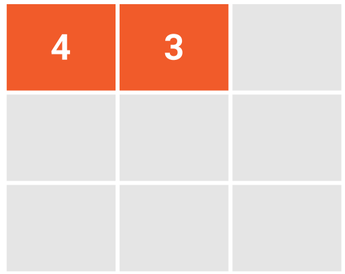
Input



Filter



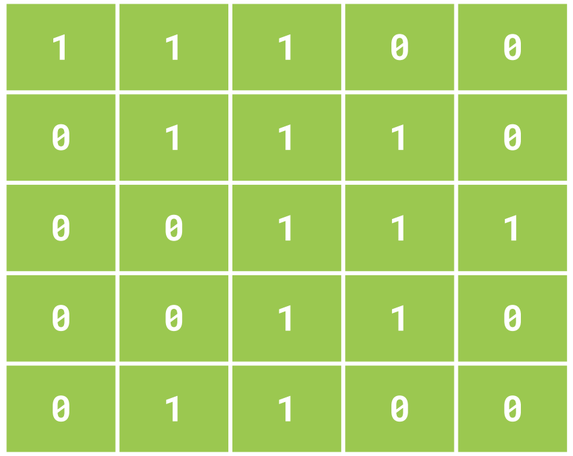
Sliding 2
(Stride = 1)



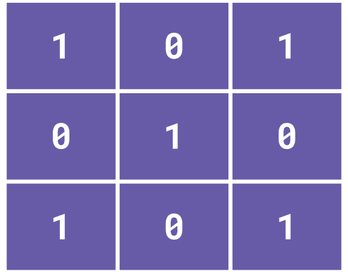
Feature map

Calculations - (1x1 + 1x0 + 0x1) + (1x0 + 1x1 + 1x0) + (0x1 + 1x0 + 1x1) = 3

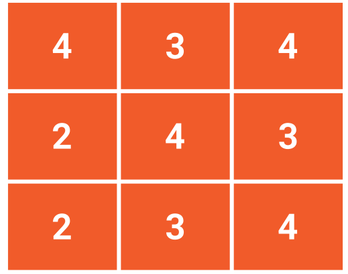
Convolution



Input



Filter



Final feature map

Convolutions

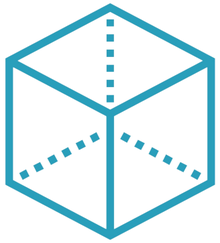
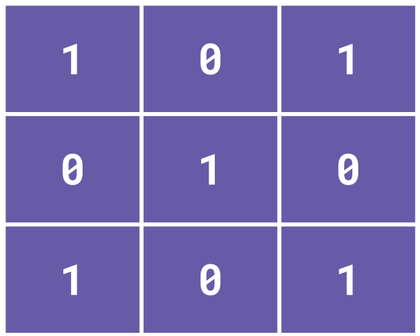


Image in 3D (height, weight, depth), then filter also in 3D

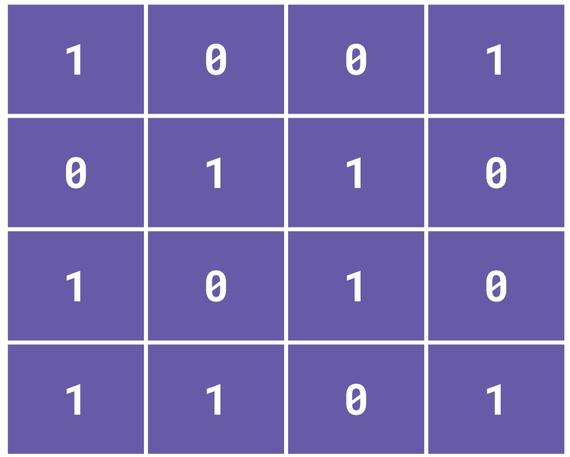
Multiple convolutions = multiple filters = multiple feature maps

All feature maps stacked together forms the output of the convolution layer

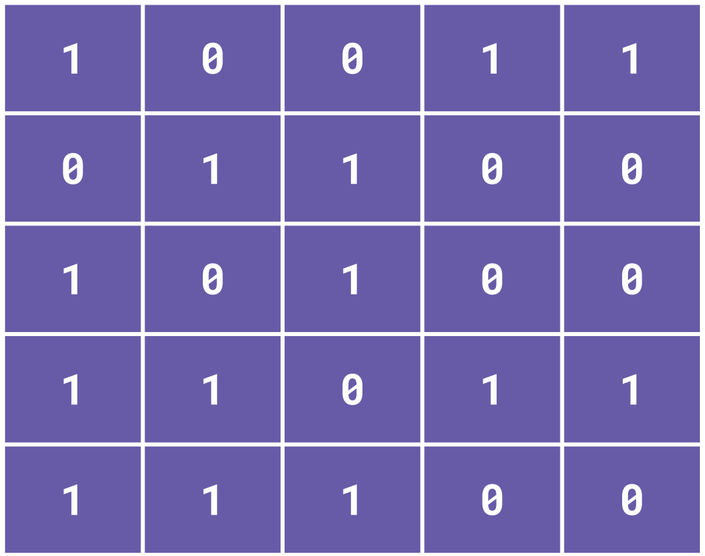
Filter



3x3 Filter

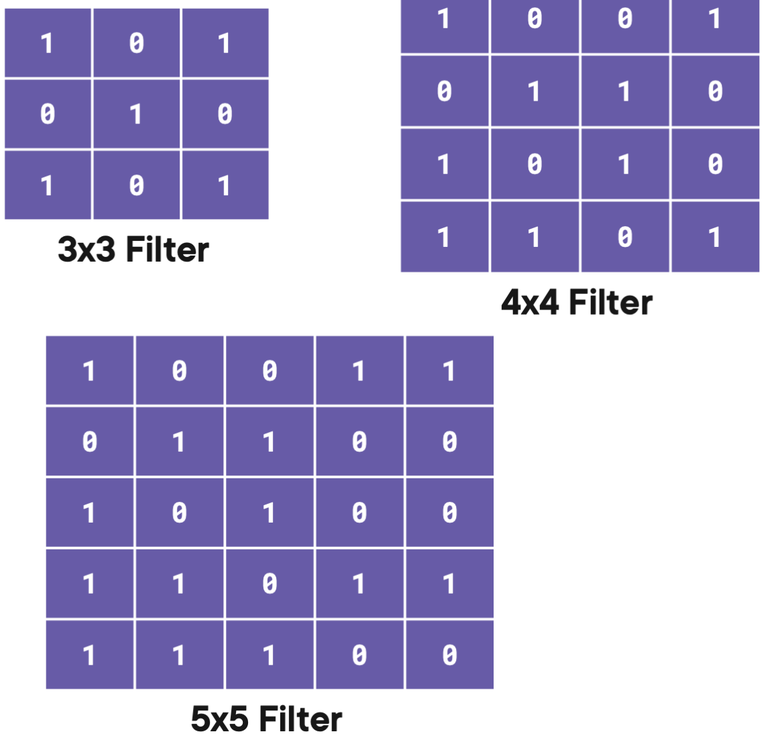


4x4 Filter



5x5 Filter

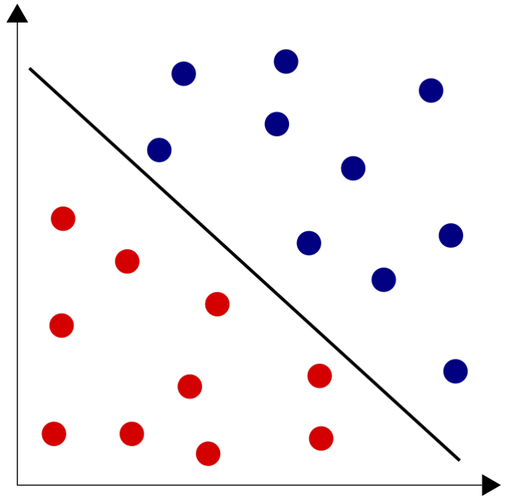
Filter



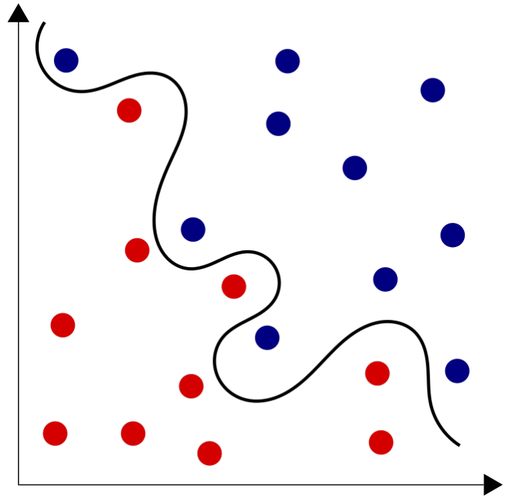
- Occurs in various sizes
- They are feature identifiers
- Feature maps produce high value when portion of image containing feature is multiplied by filter
- Start with random initialization
- Values change on training through backpropagation

CNN: Activation

Data Separability

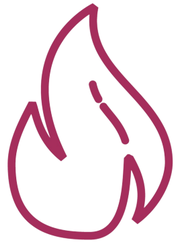


Linearly separable

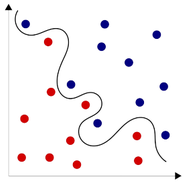


Non-linearly separable

Activation Functions



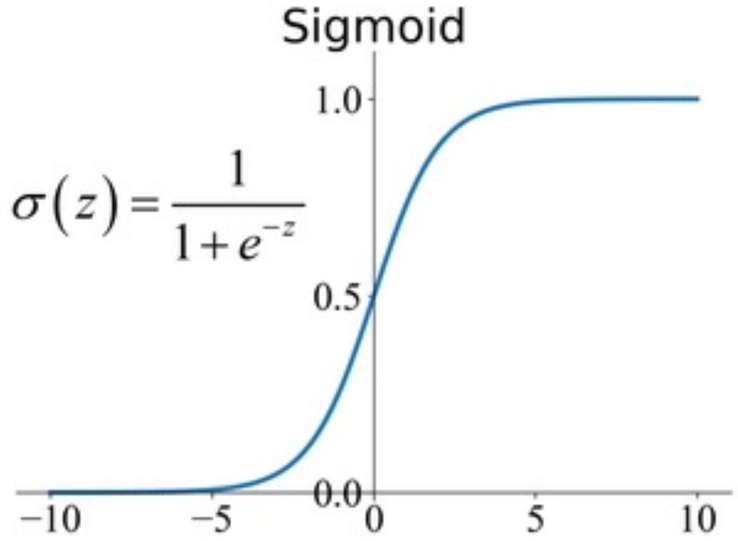
Activation function decides whether a neuron or node in a neural network should fire or not



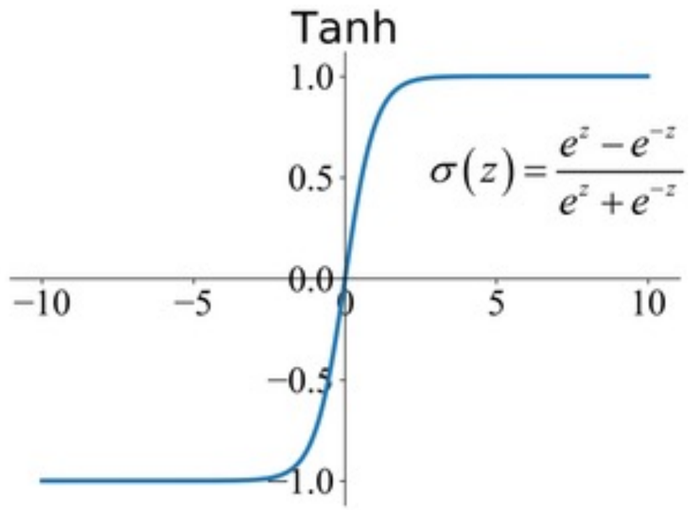
Activation functions help introduce non-linearity to the network

Many activation functions - Sigmoid, Tanh, ReLU etc.

Non-linear Activation Function



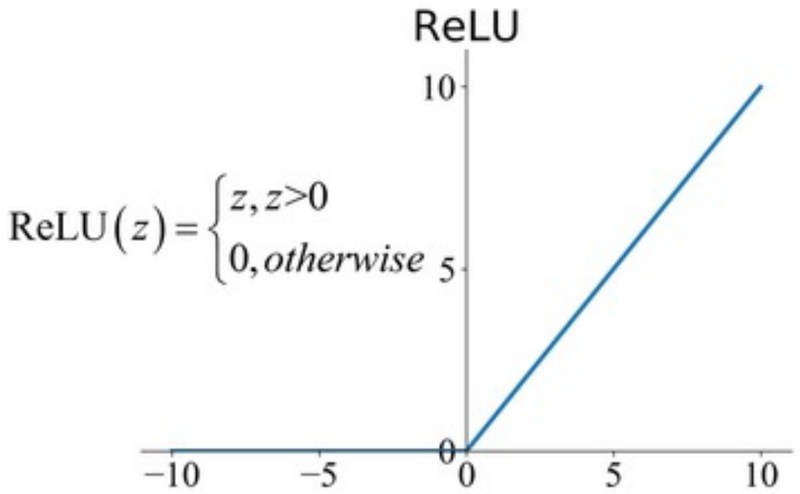
Sigmoid Activation Function



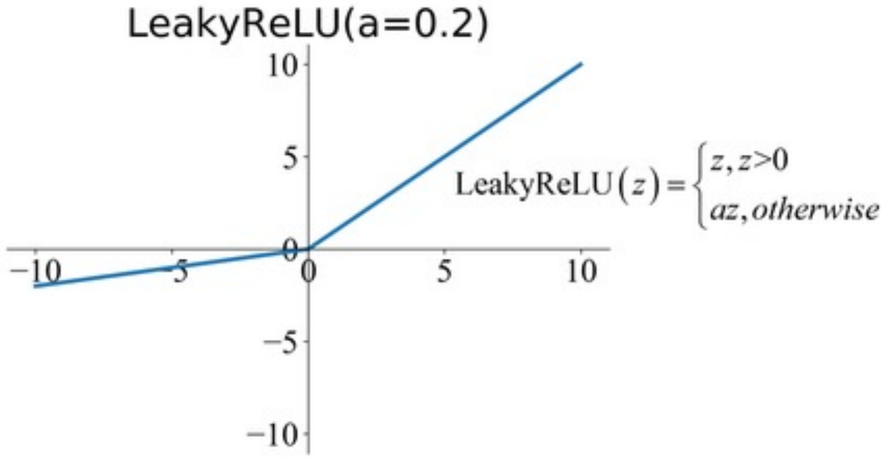
Tanh Activation Function

Image Citation - Feng, Junxi & He, Xiaohai & Teng, Qizhi & Ren, Chao & Chen, Honggang & Li, Yang. (2019). Reconstruction of porous media from extremely limited information using conditional generative adversarial networks.

Non-linear Activation Function



ReLU Activation Function

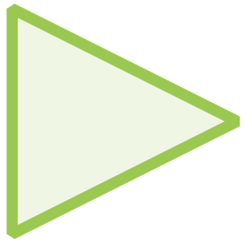


Leaky ReLU Activation Function

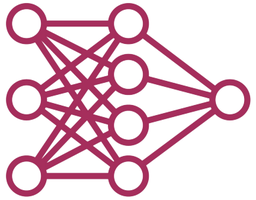
Image Citation - Feng, Junxi & He, Xiaohai & Teng, Qizhi & Ren, Chao & Chen, Honggang & Li, Yang. (2019). Reconstruction of porous media from extremely limited information using conditional generative adversarial networks.

Factors for Choosing an Activation Function

Type of prediction we want



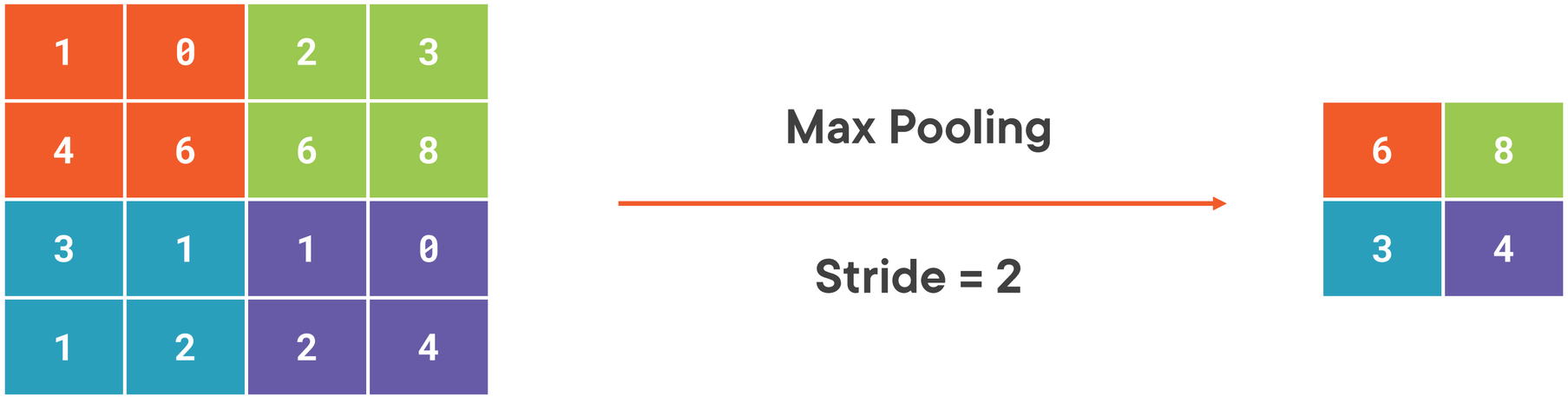
Current layer in the neural network



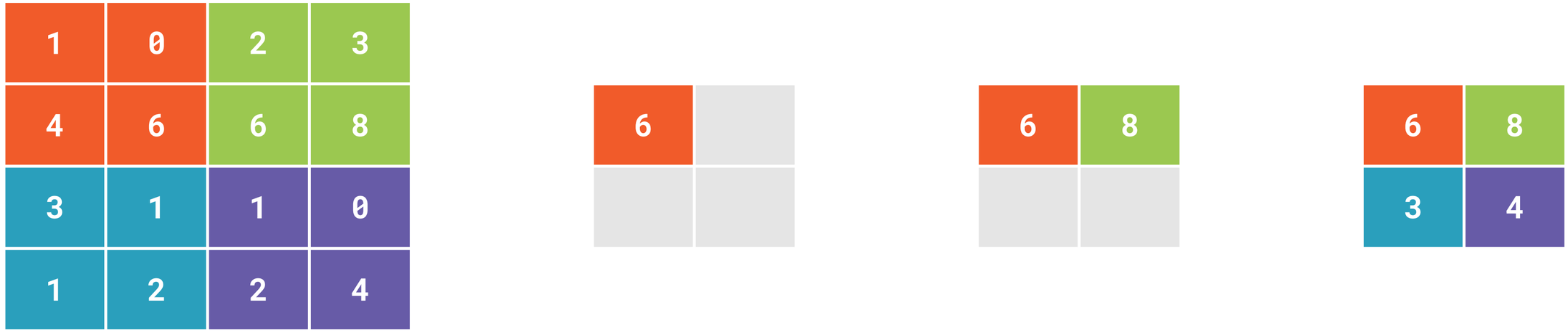
Type and architecture of the neural network

CNN: Pooling

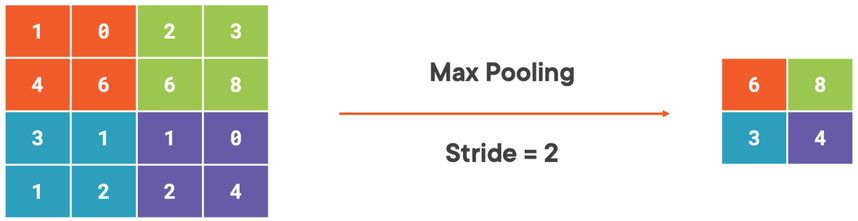
Pooling



Max Pooling (Stride = 2)



Pooling

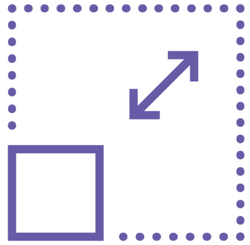


Performed after convolution and activation

There are different types of pooling

Max pooling is most popular

Why Is Pooling Done?



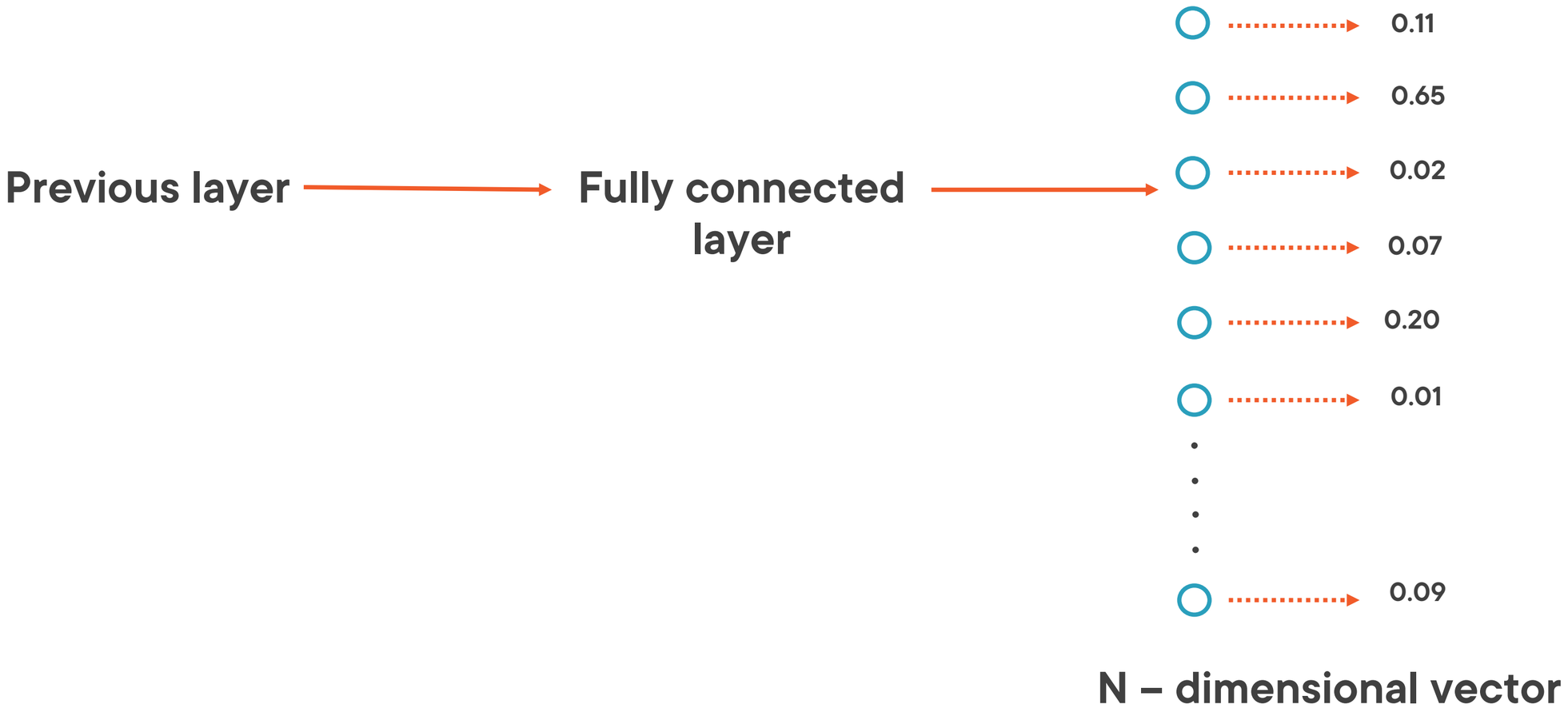
Reduces dimensionality – keeps depth, reduces height and width

Preserves important information

Reduces network training time

CNN: Classification

Classification

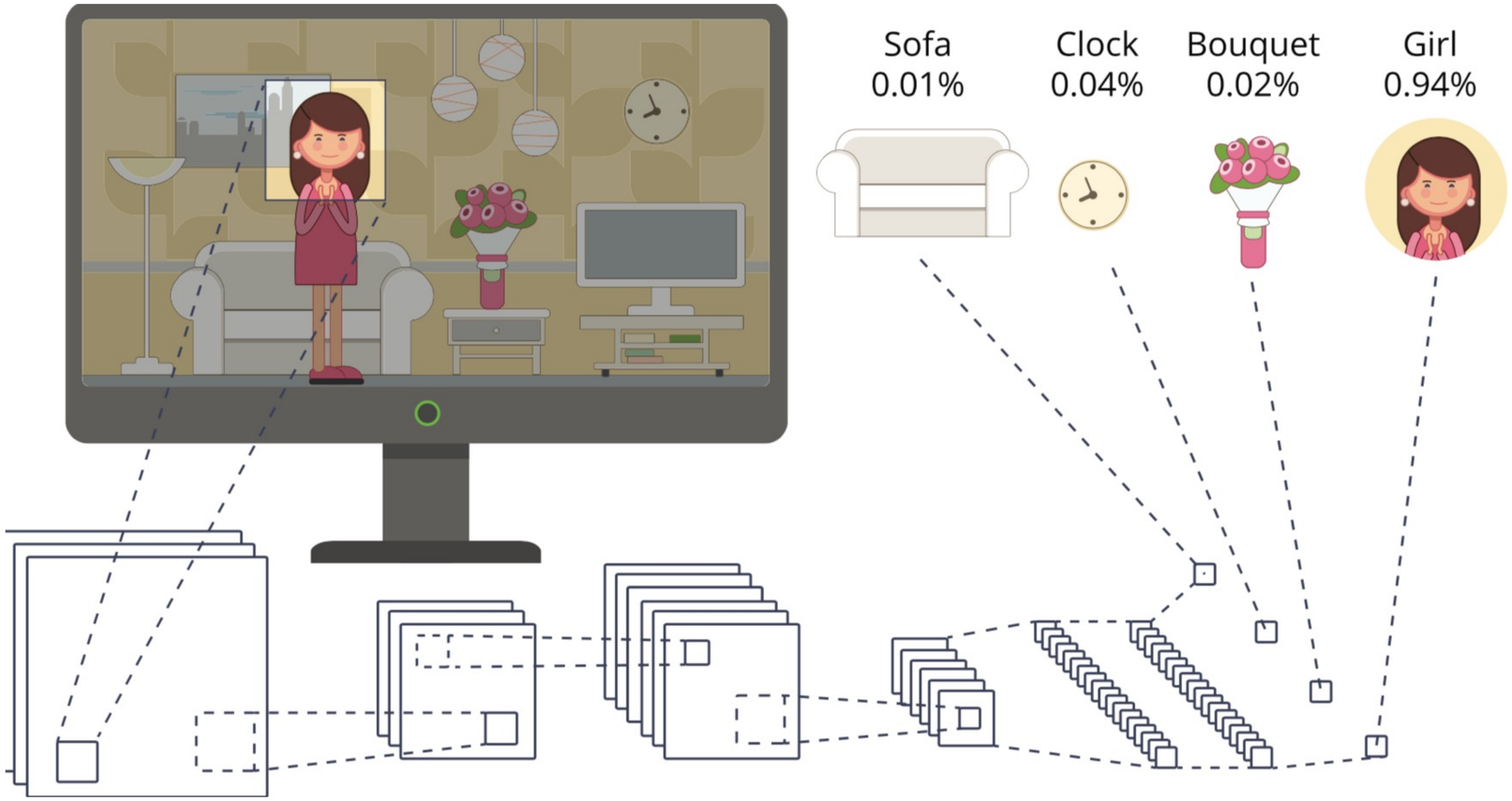


Classification

N = 2

N = 10

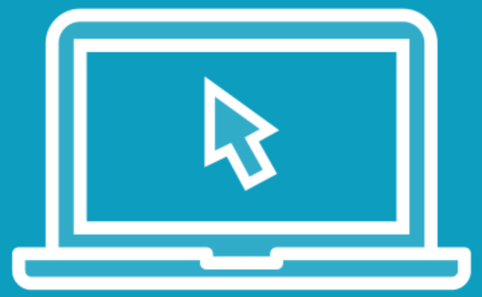
Convolutional Neural Network: Layout



Demo

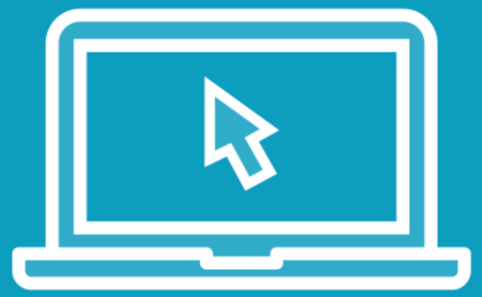
Creating the CNN architecture

Demo



Training the model

Demo



Performance metrics – how well did our model do?

Summary

- CNN is a deep learning algorithm that can differentiate one image from another
- A CNN gradually builds up its understanding of input images
- Convolutional layer performs a dot product of two matrices
- Activation function introduces non-linearity to the network
- Pooling layer helps in dimensionality reduction preserving important information
- Learned how to create, train and test a CNN

Up Next:

Improving Performance of the Convolutional Neural Network