AN INTRODUCTION TO CLASSIFICATION IN ML

acm-w SRM

SPEAKERS

Ankita

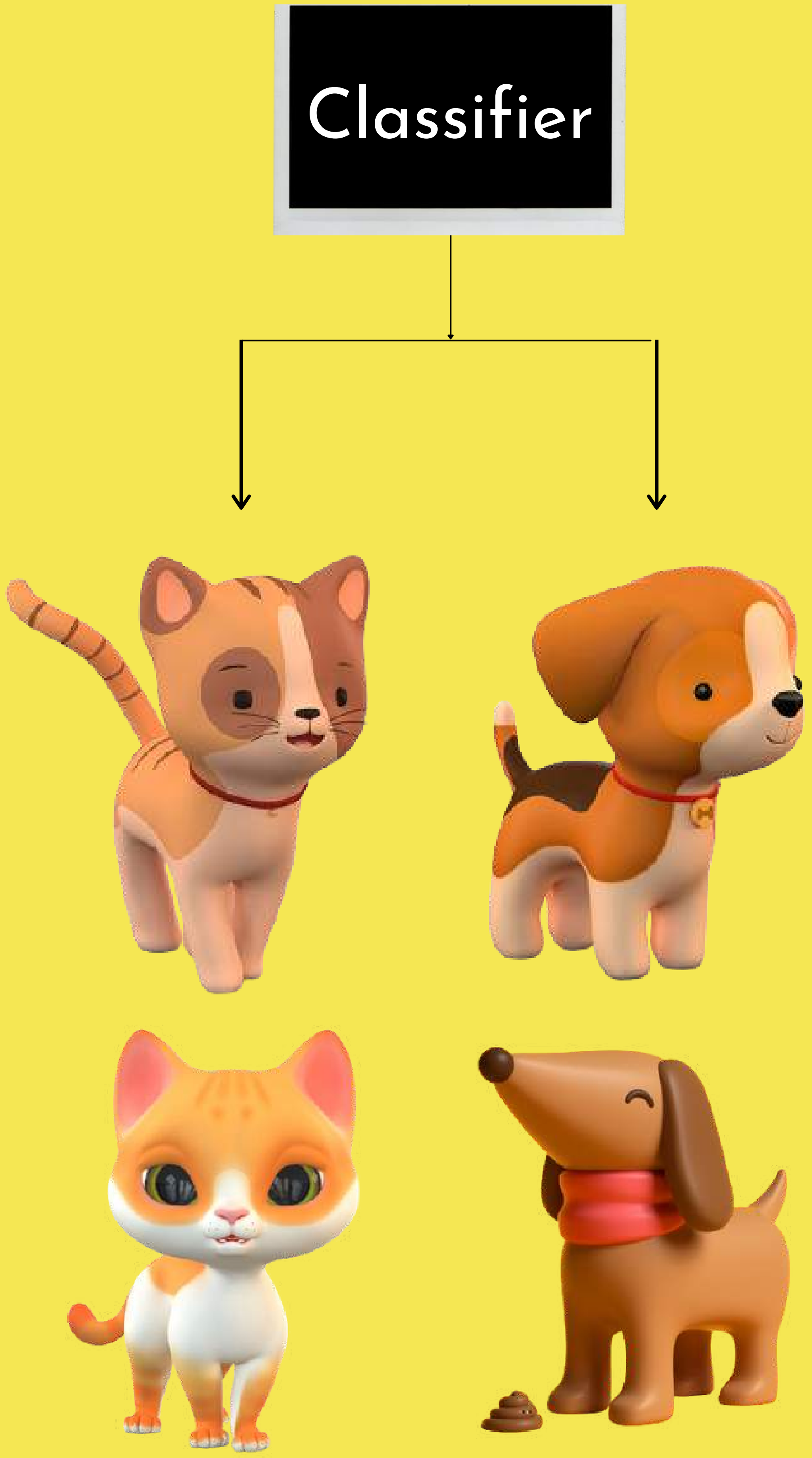
Kamalini

AGENDA

- What is Classification in ML?
- Terminologies
- Classification Algorithms
- CNN : First Step in Computer Vision

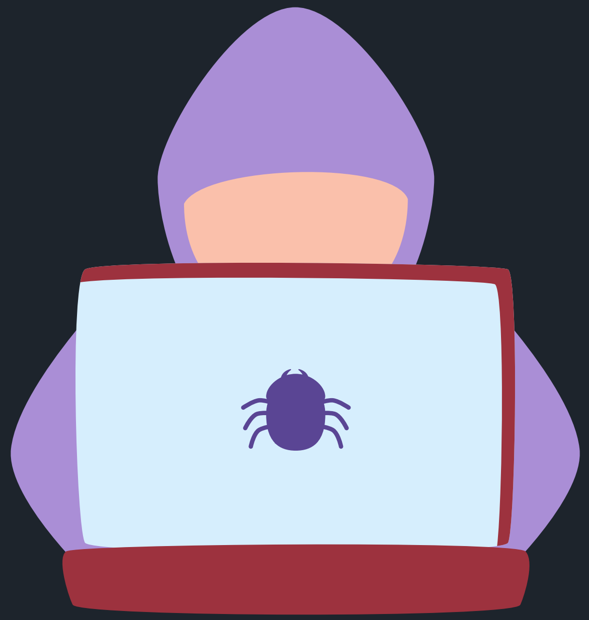
WHAT IS CLASSIFICATION?

Classification is a technique of categorizing a given set of statistics into different classes which can be applied to both structured and unstructured statistics.



REAL WORLD APPLICATIONS

EMAIL SPAM CLASSIFIER



FRAUD DETECTION

HANDWRITING DETECTION

SENTIMENTAL ANALYSIS

DISEASE PREDICTION

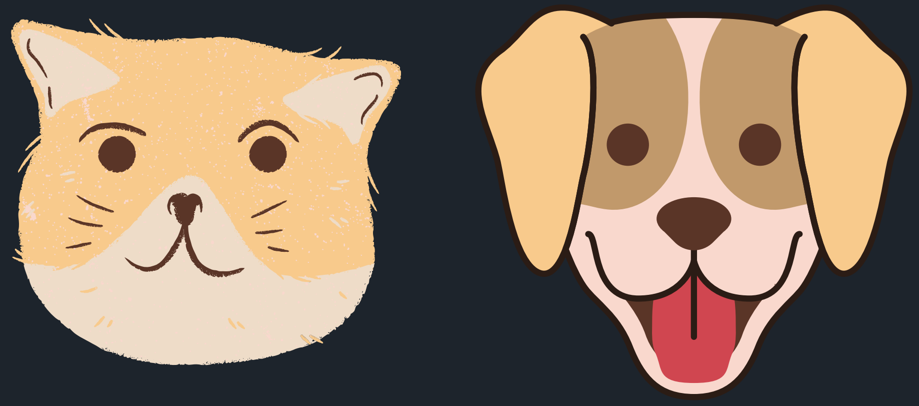


IMAGE CLASSIFICATION

UNDERSTANDING THE TERMINOLOGIES

Supervised Learning

The system is presented with example inputs and their desired outputs, and the goal is to learn a general rule that maps inputs to outputs.

Unsupervised Learning

No labels are given to the learning algorithm, leaving it on its own to find structure in its input.

Classifier

A classifier in ML is an algorithm that automatically orders or categorizes data into one or more of a set of “classes.”

Classification Model

A classification model tries to draw some conclusion from the input values given for training. It will predict the class labels/categories for the new data.

Training

The process of training an ML model involves providing an ML algorithm (that is, the learning algorithm) with training data to learn from.

Clustering

It is an unsupervised machine learning task.
It involves automatically discovering natural grouping in data.

Underfitting

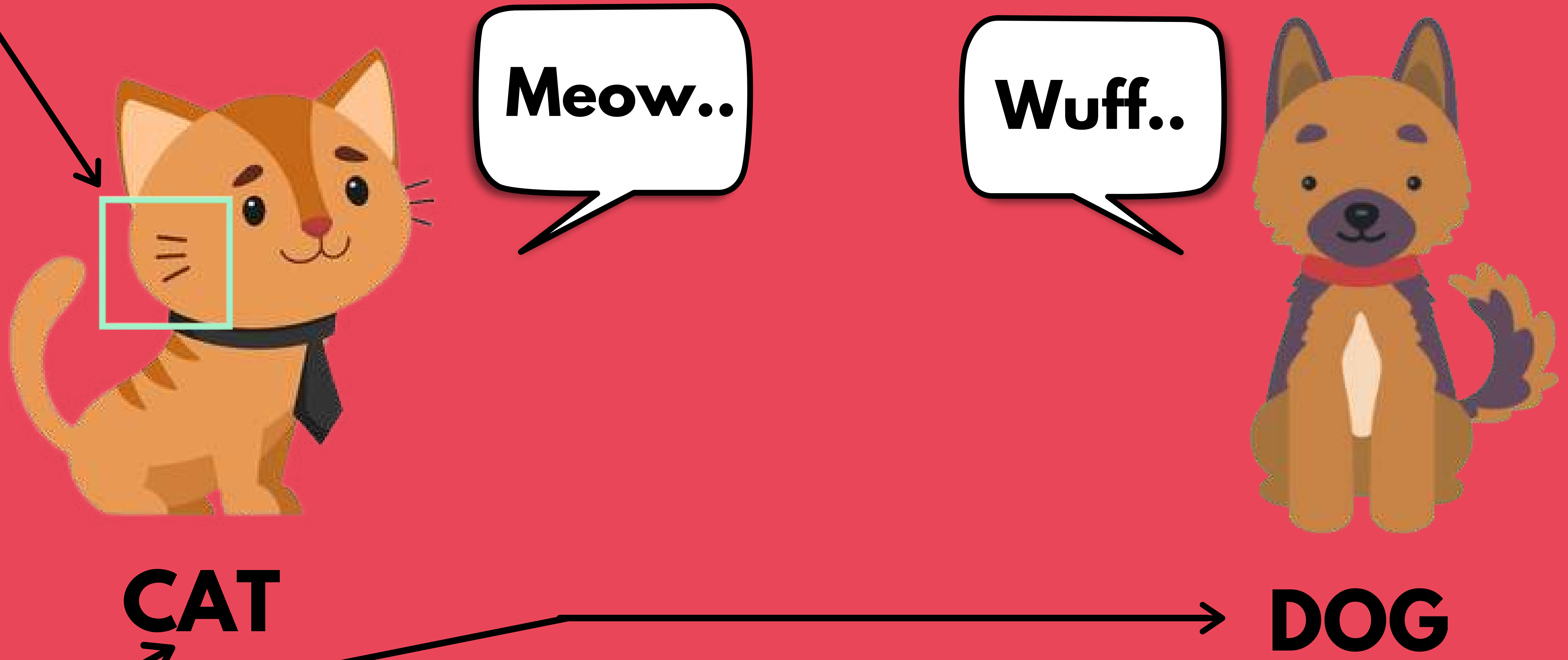
Underfitting occurs when our machine learning model is not able to capture the underlying trend of the data. The model is not able to learn enough from the training data, and hence it reduces the accuracy and produces unreliable predictions.

Overfitting

A model is said to be overfitted when we train it with a lot of data and it starts learning from the noise and inaccurate data entries in our data set.

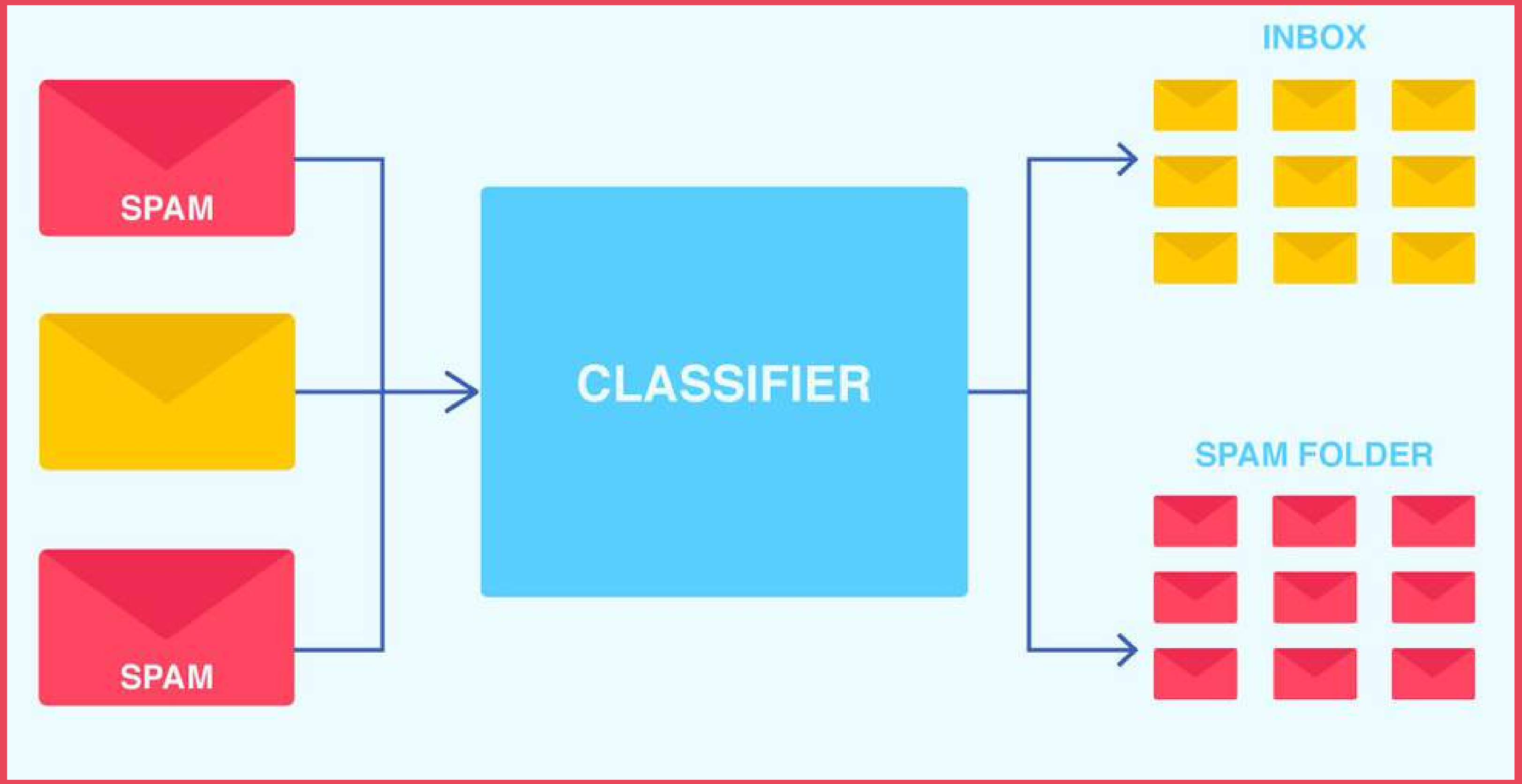
Feature

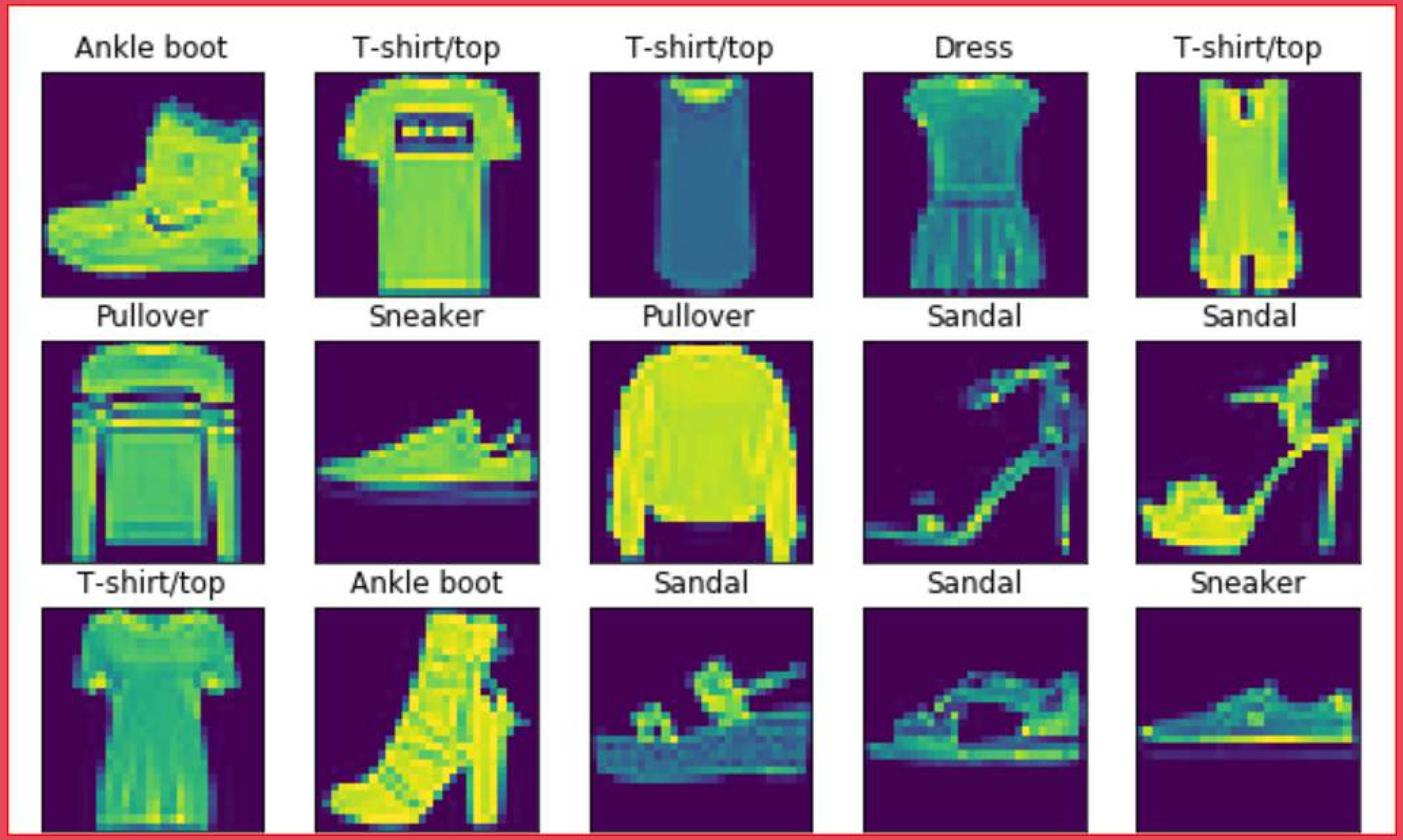
A feature is an individual measurable property of the phenomenon being observed.



Labels

They are the discrete attribute whose value you want to predict based on the values of other attributes.





BINARY CLASSIFICATION

It is a type of classification with two outcomes, for eg – either true or false.

MULTI- CLASS CLASSIFICATION

It is the classification with more than two classes, in this each sample is assigned to one and only one label or target.

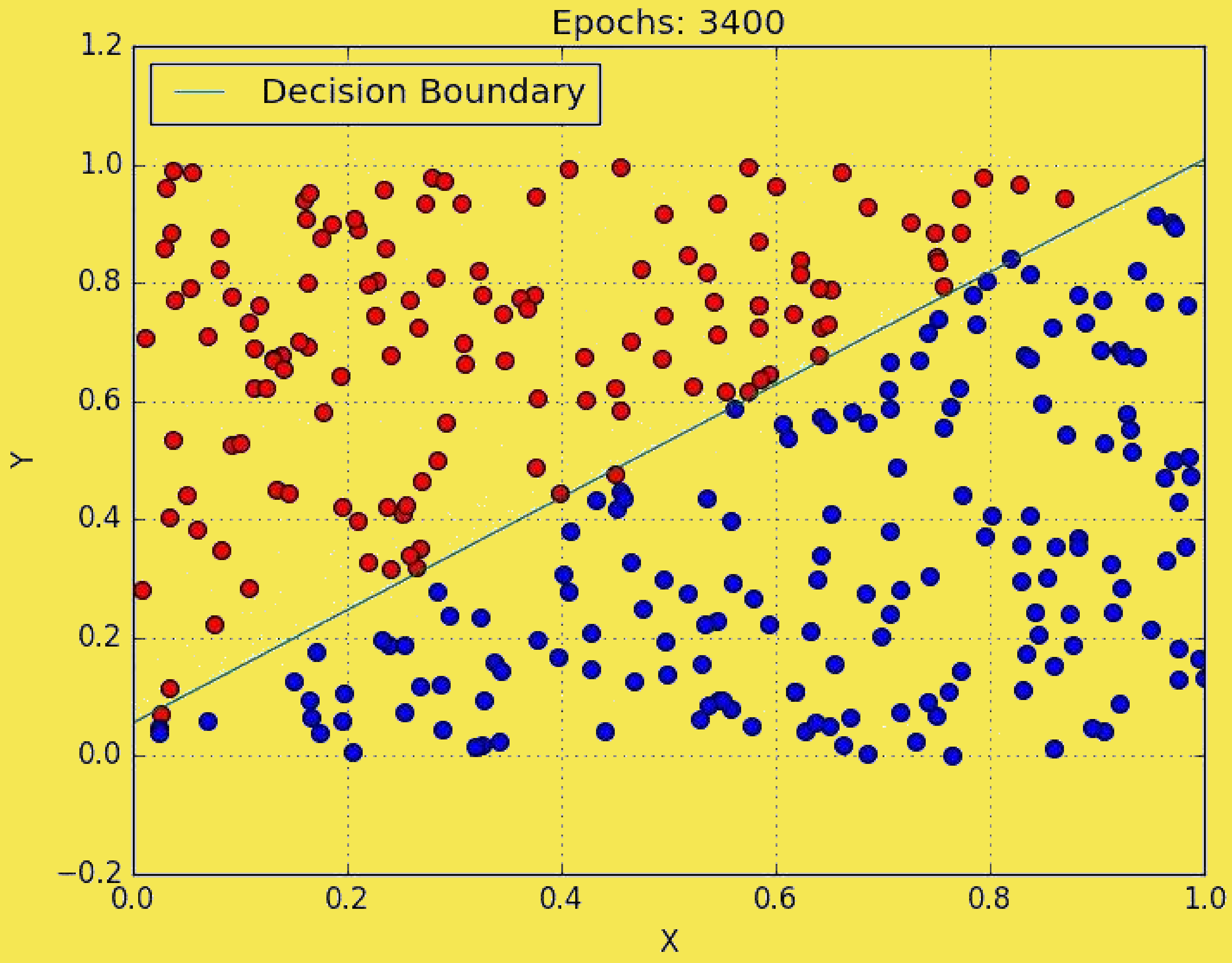
CLASSIFICATION ALGORITHMS

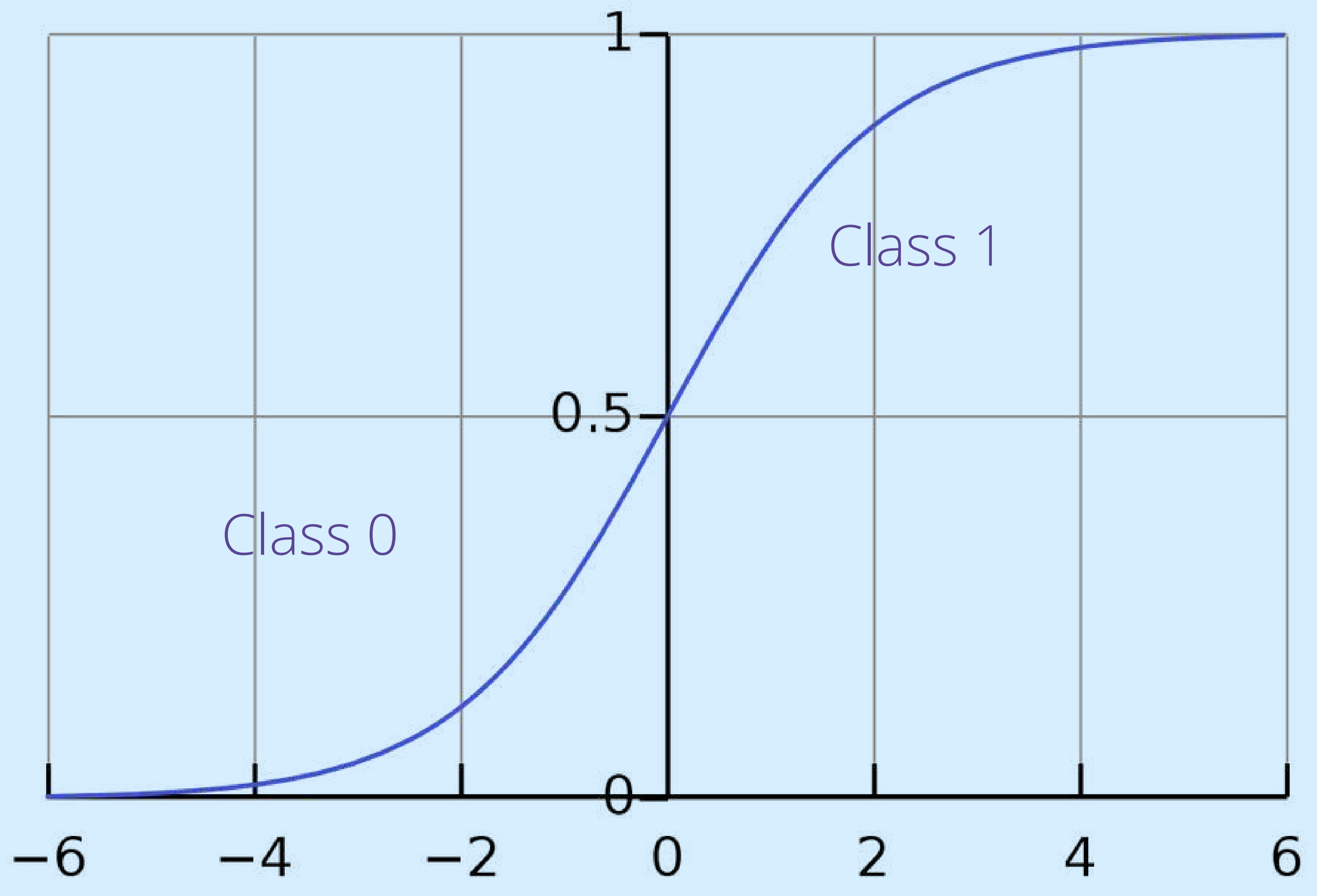
- Logistic Regression
- KNN
- Decision Tree
- Random Forest
- Naïve Bayes
- Stochastic Gradient Descent
- Support Vector
- ANN
- CNN

LOGISTIC REGRESSION

Logistic regression is a statistical analysis method used to predict a data value based on prior observations of a data set.

It provides the output using the sigmoid function to return a probability value that can be mapped to two or more discrete classes





Formula of a sigmoid function

$$f(x) = \frac{1}{1 + e^{-x}}$$

SIGMOID FUNCTION

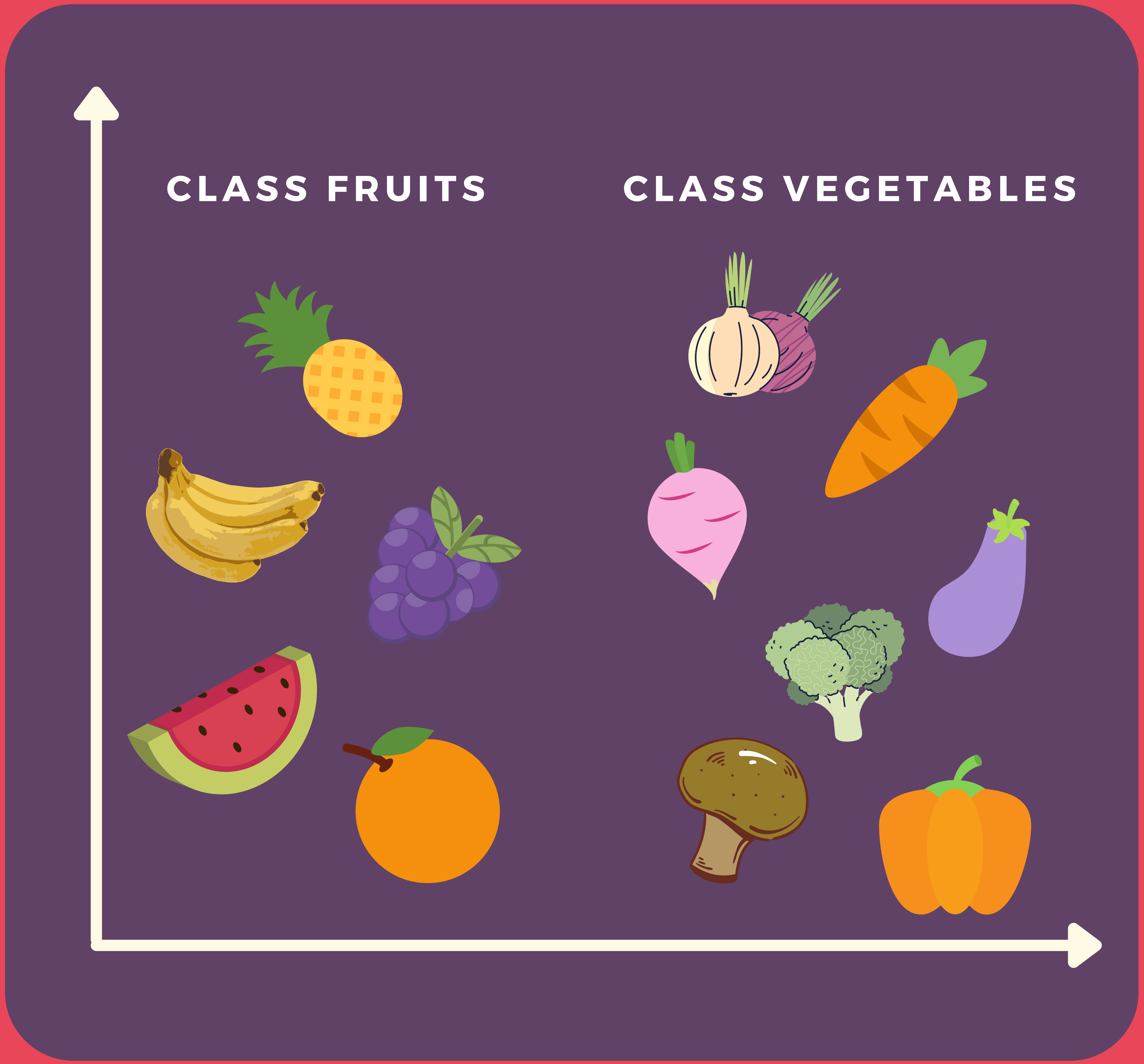
In machine learning, we use sigmoid to map predictions to probabilities. It's an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

It is used to add non-linearity in an ML model.

K-NEAREST NEIGHBOURS

A simple and easy-to-implement supervised ML algorithm that can be used to solve both classification and regression problems.

The KNN algorithm assumes that similar things exist in close proximity i.e. similar things are near to each other.

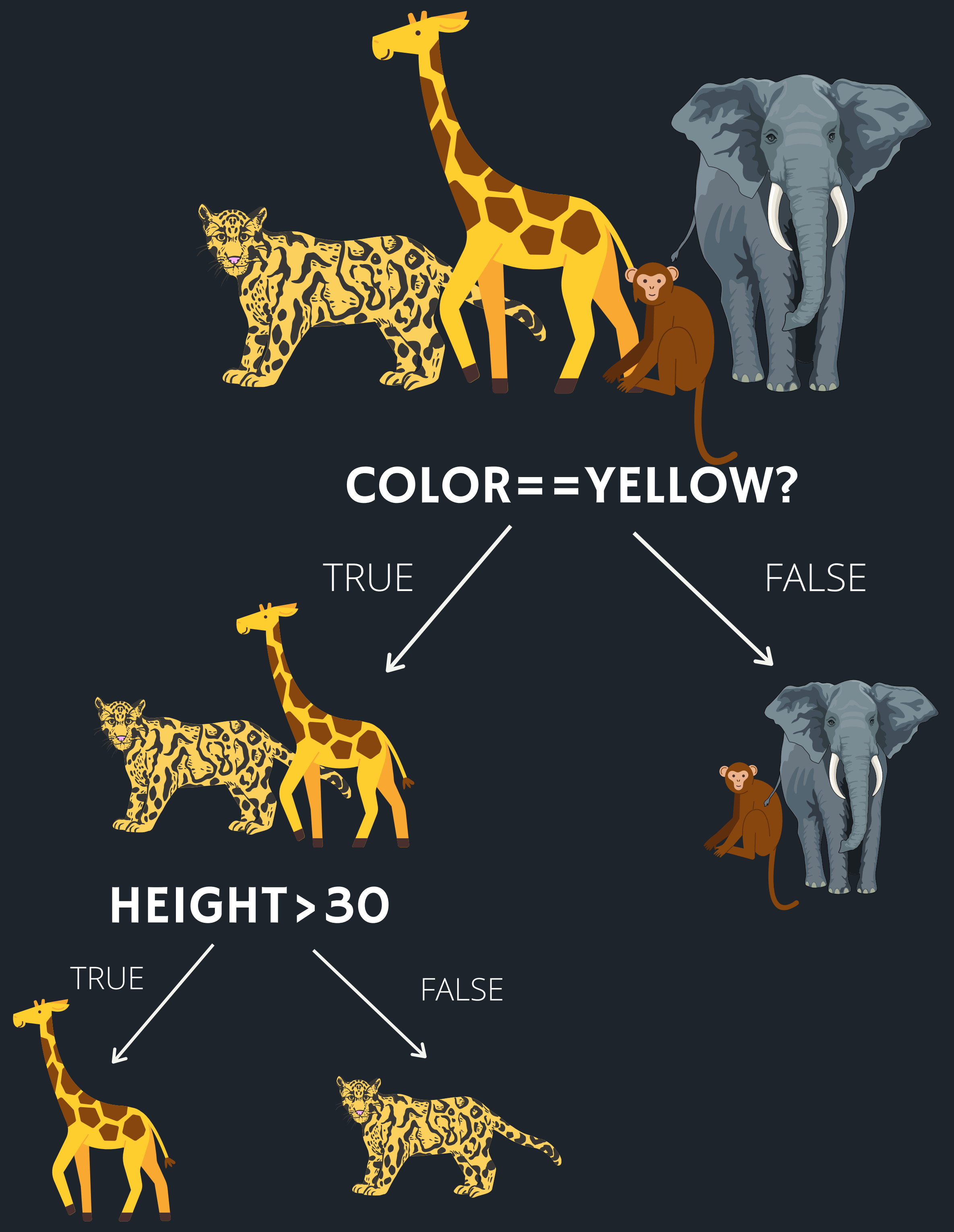


DECISION TREE

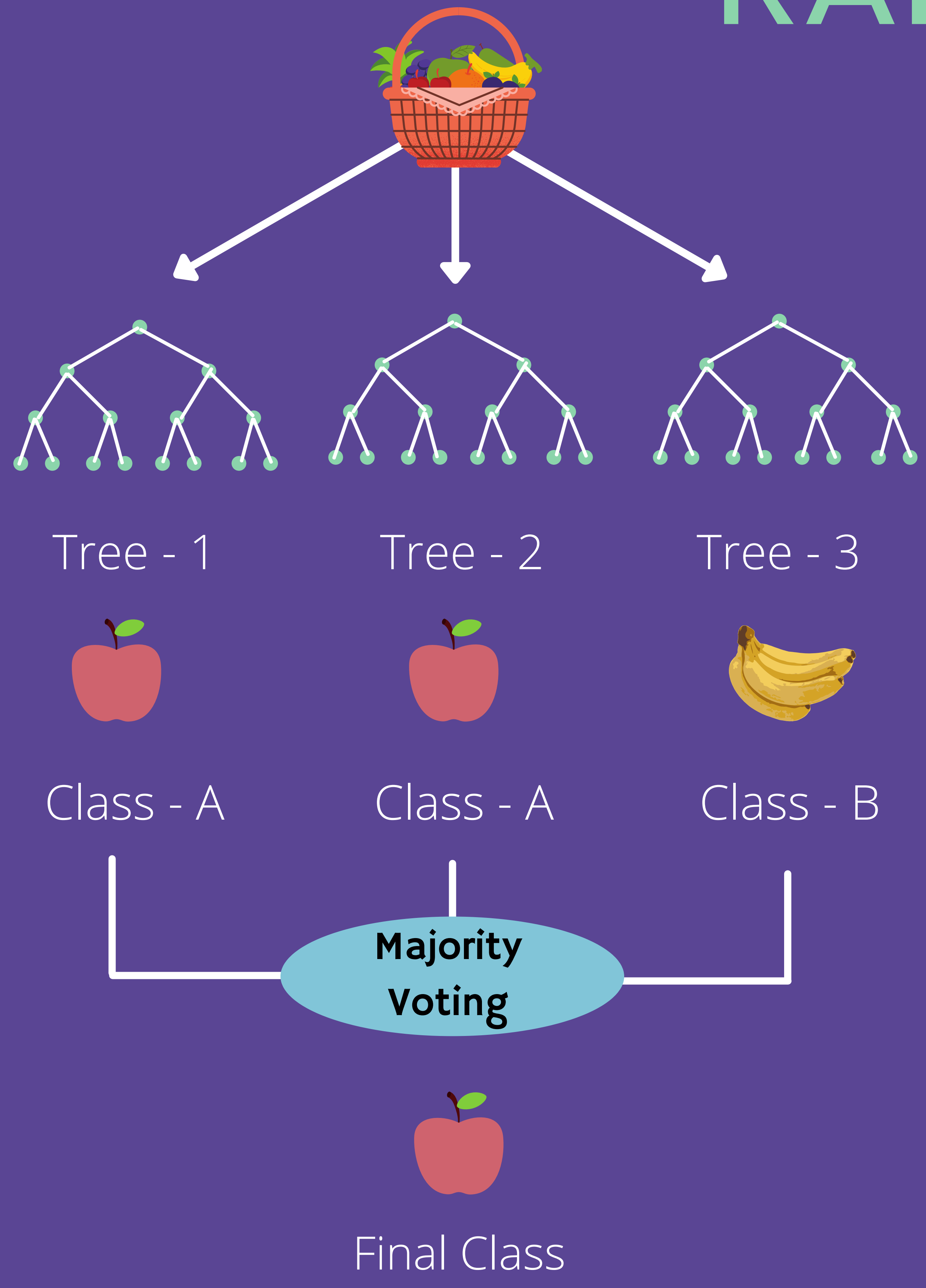
A decision tree builds classification or regression models in the form of a tree structure. It utilizes an if-then rule set that is mutually exclusive and exhaustive for classification.

The rules are learned sequentially using the training data one at a time. Each time a rule is learned, the tuples covered by the rules are removed. This process is continued until meeting a termination condition.

Is the animal a giraffe?



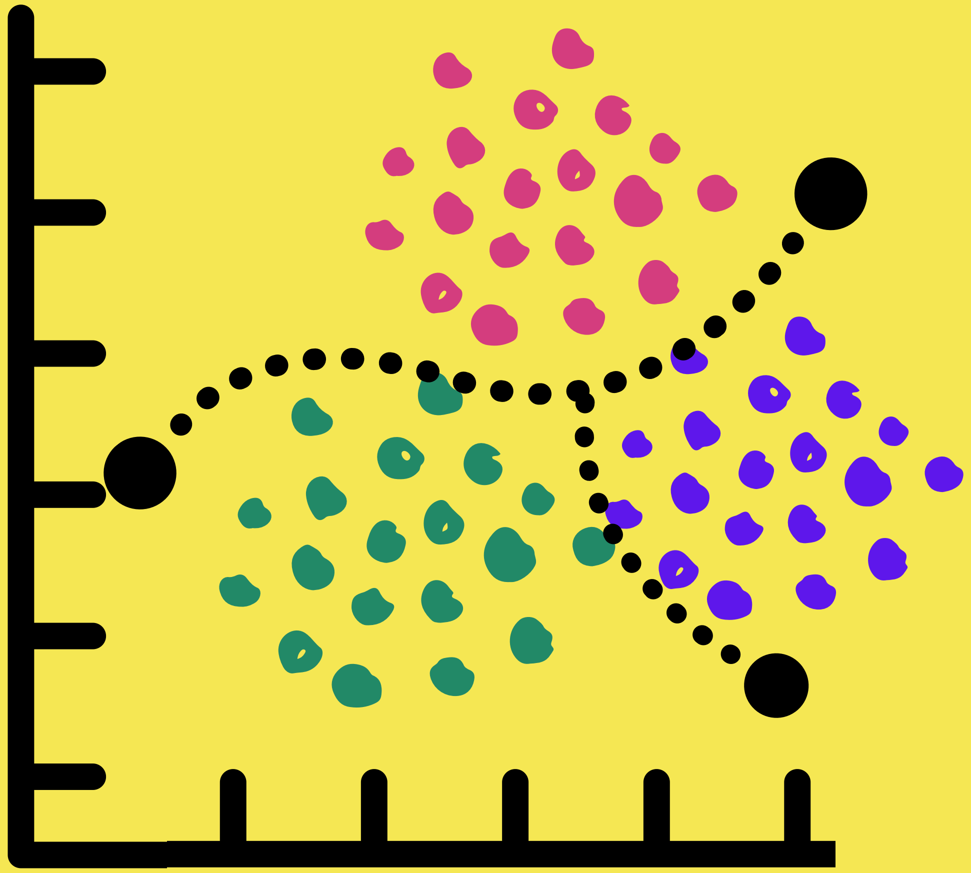
RANDOM FOREST



A classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.

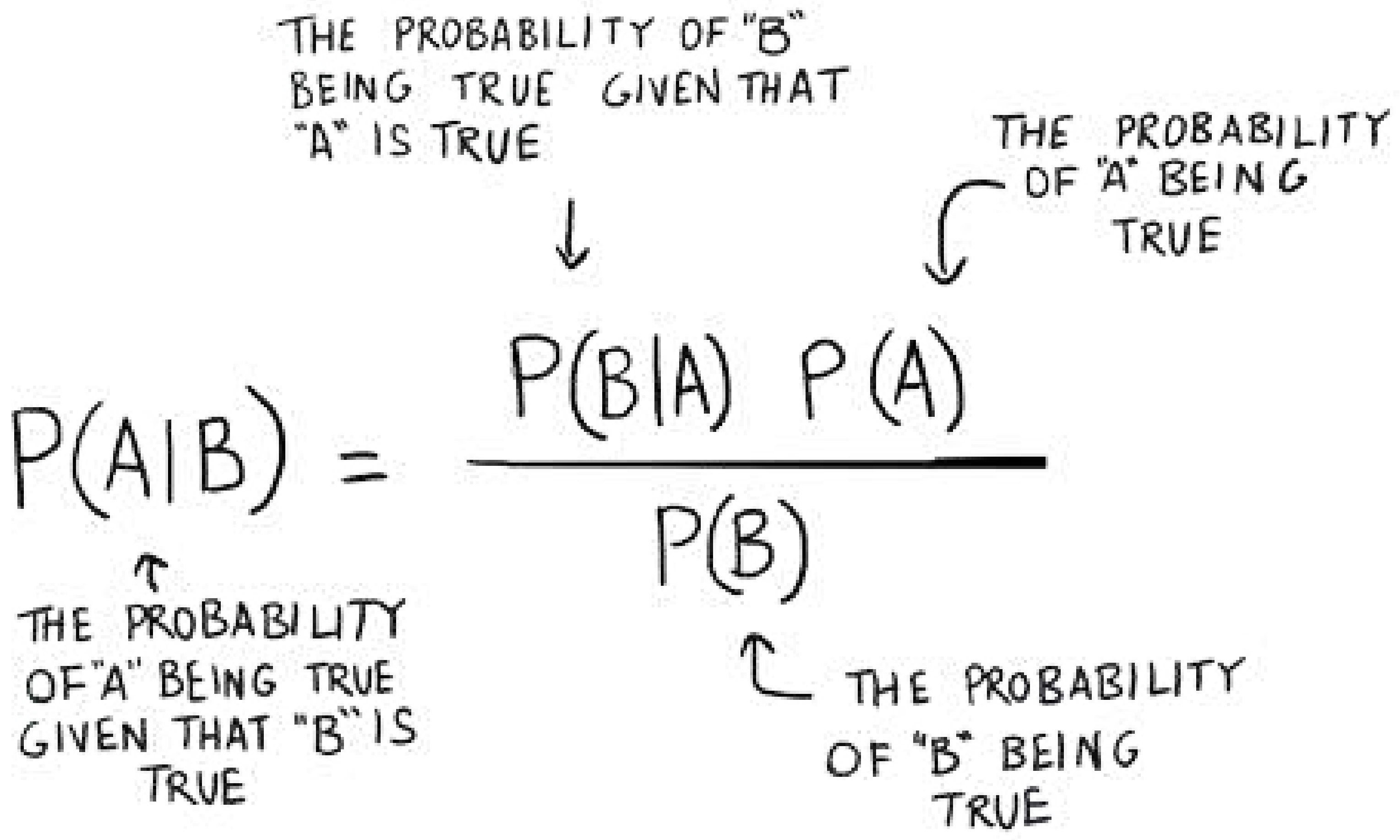
The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

NAÏVE BAYES CLASSIFIER



CLASSIFIER 1
CLASSIFIER 2
CLASSIFIER 3

Naive Bayes is a probabilistic classifier inspired by the Bayes theorem under a simple assumption which is the attributes are conditionally independent.



EXAMPLE

	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY GOLF
1	Rainy	Hot	High	False	No
2	Rainy	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	Sunny	Mild	High	False	Yes
5	Sunny	Cool	Normal	False	Yes
6	Sunny	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes
8	Rainy	Mild	High	False	No
9	Rainy	Cool	Normal	False	Yes
10	Sunny	MIld	Normal	False	Yes
11	Rainy	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Sunny	Mild	High	True	No

With relation to our dataset,

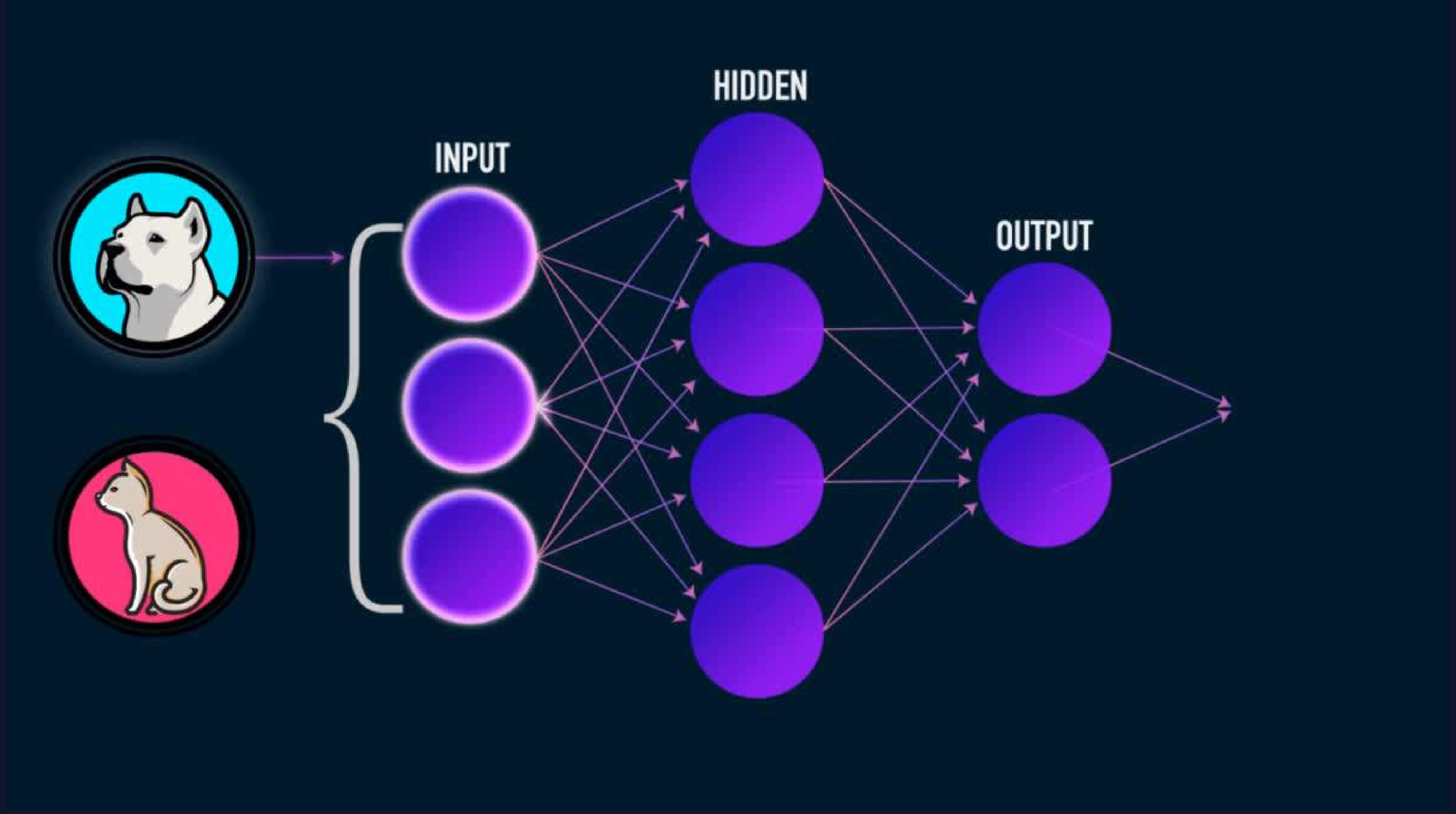
- We assume that no pair of features are dependent.
- Secondly, each feature is given the same weight(or importance).

Outlook			Temperature			Humidity			Windy			Play	
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	2	3	hot	2	2	high	3	4	false	6	2	9	5
overcast	4	0	mild	4	2	normal	6	1	true	3	3		
rainy	3	2	cool	3	1								
sunny	2/9	3/5	hot	2/9	2/5	high	3/9	4/5	false	6/9	2/5	9/14	5/14
overcast	4/9	0/5	mild	4/9	2/5	normal	6/9	1/5	true	3/9	3/5		
rainy	3/9	2/5	cool	3/9	1/5								

today= {Sunny, Hot, Normal Humidity, Not Windy}

$$P(Yes|today) = \frac{P(SunnyOutlook|Yes)P(HotTemperature|Yes)P(NormalHumidity|Yes)P(NoWind|Yes)P(Yes)}{P(today)}$$

ARTIFICIAL NEURAL NETWORKS (ANN)



$$z = x_1 * w_1 + x_2 * w_2 + \ldots\ldots + x_n * w_n + b * 1$$

$$\hat{y} = a_{out} = sigmoid(z)$$

$$sigmoid(z) = \frac{1}{1 + e^{-z}}$$

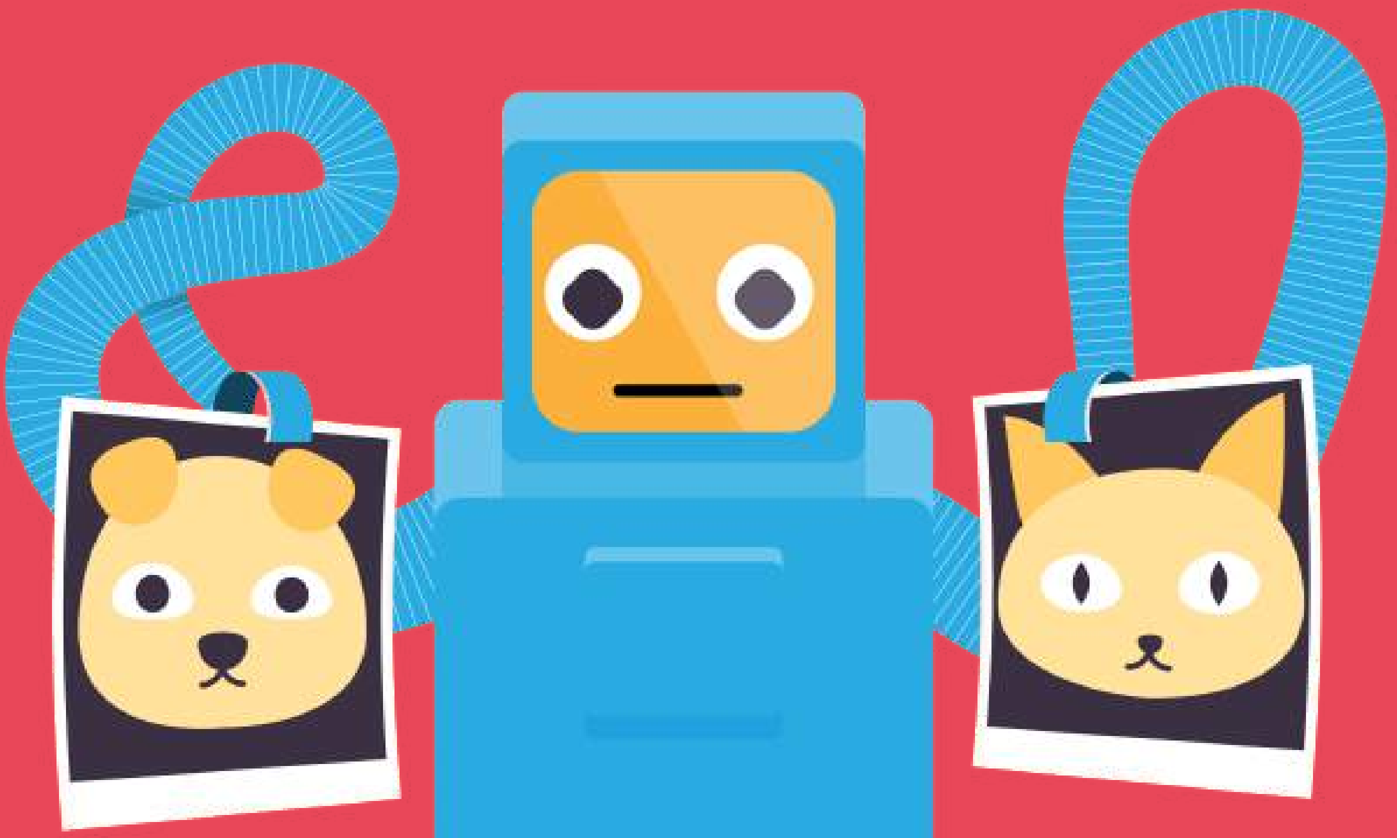
An ANN is based on a collection of nodes called artificial neurons, which loosely represent the neurons in a biological brain.

Your brain fires off groups of neurons that communicate with each other. In an ANN, a neuron is grouped with a bunch of other neurons to form a layer, and these layers stack on top of each other through connections. The more layers a neural network has, the “deeper” it is. That’s where the idea of “deep learning” comes from.

LET'S TAKE FIVE!

CNN

CONVOLUTIONAL NEURAL NETWORK

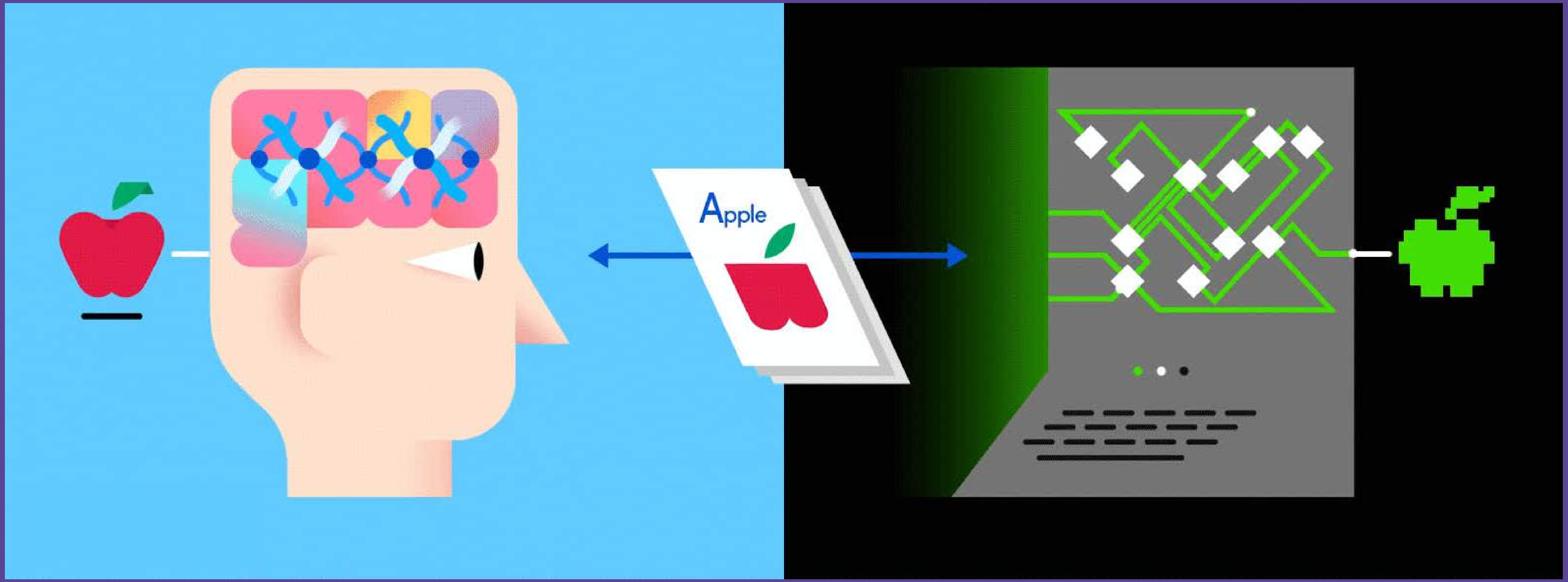


INTRODUCTION TO CLASSIFICATION

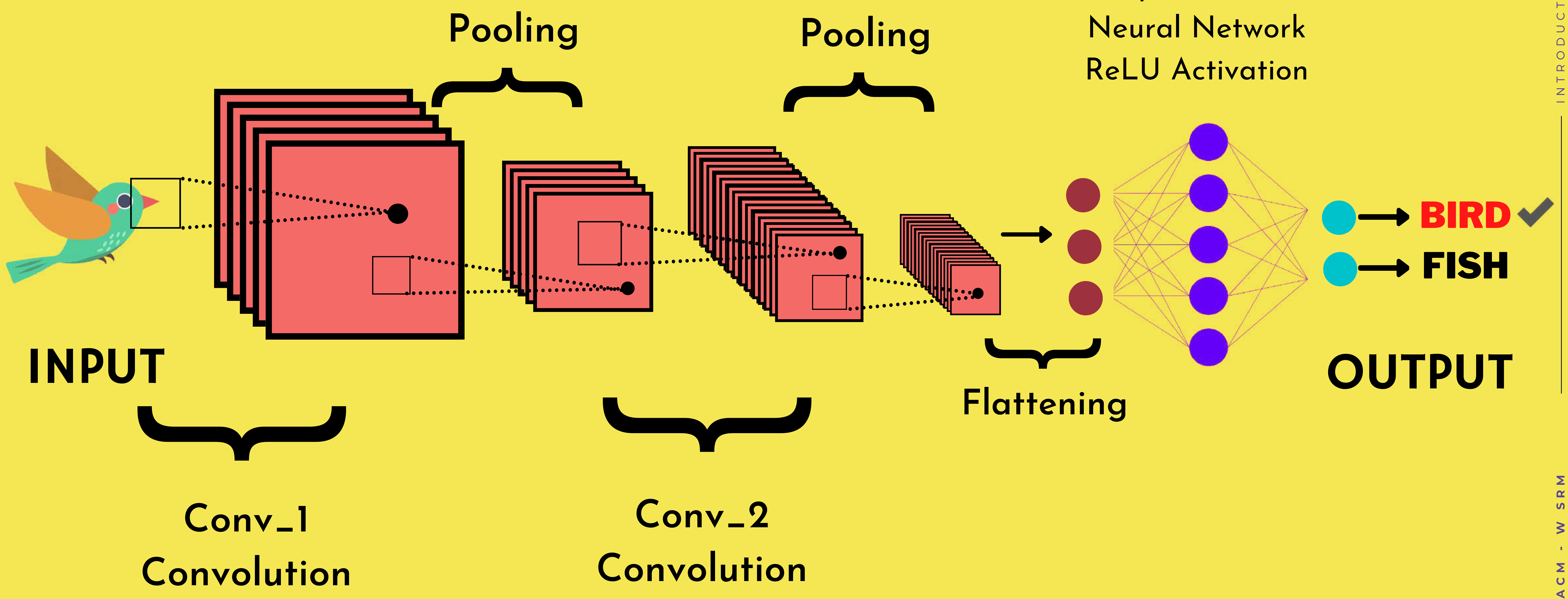
ACM - W SRM

WHAT IS CNN?

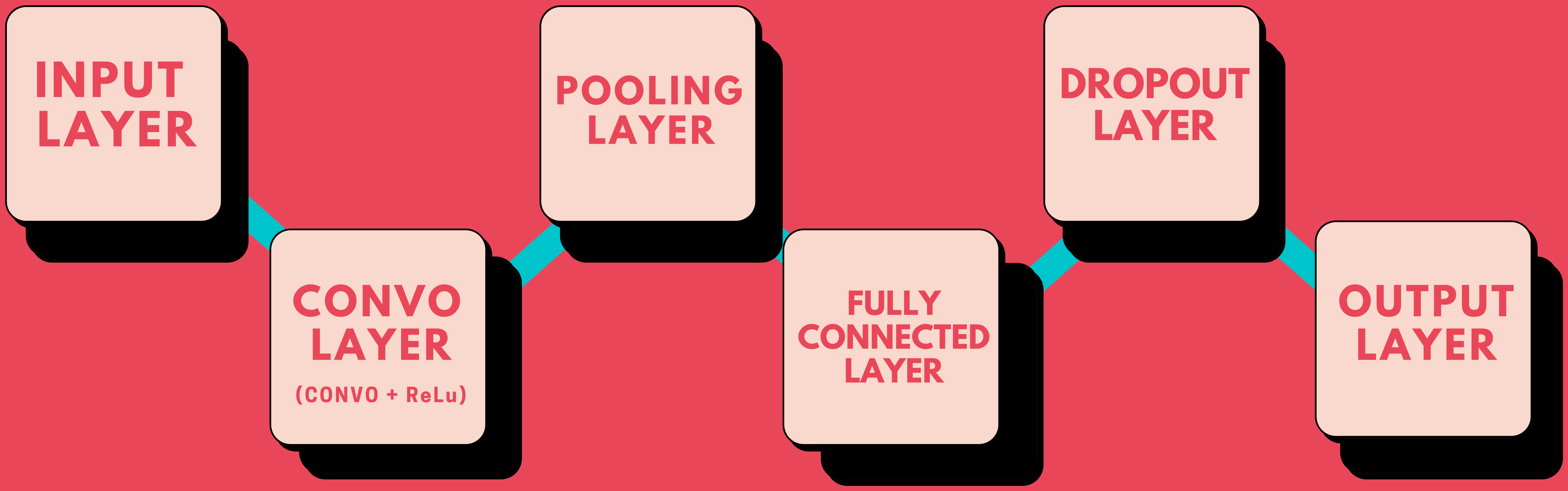
It is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.



CNN ARCHITECTURE



DIFFERENT LAYERS OF CNN



INPUT LAYER

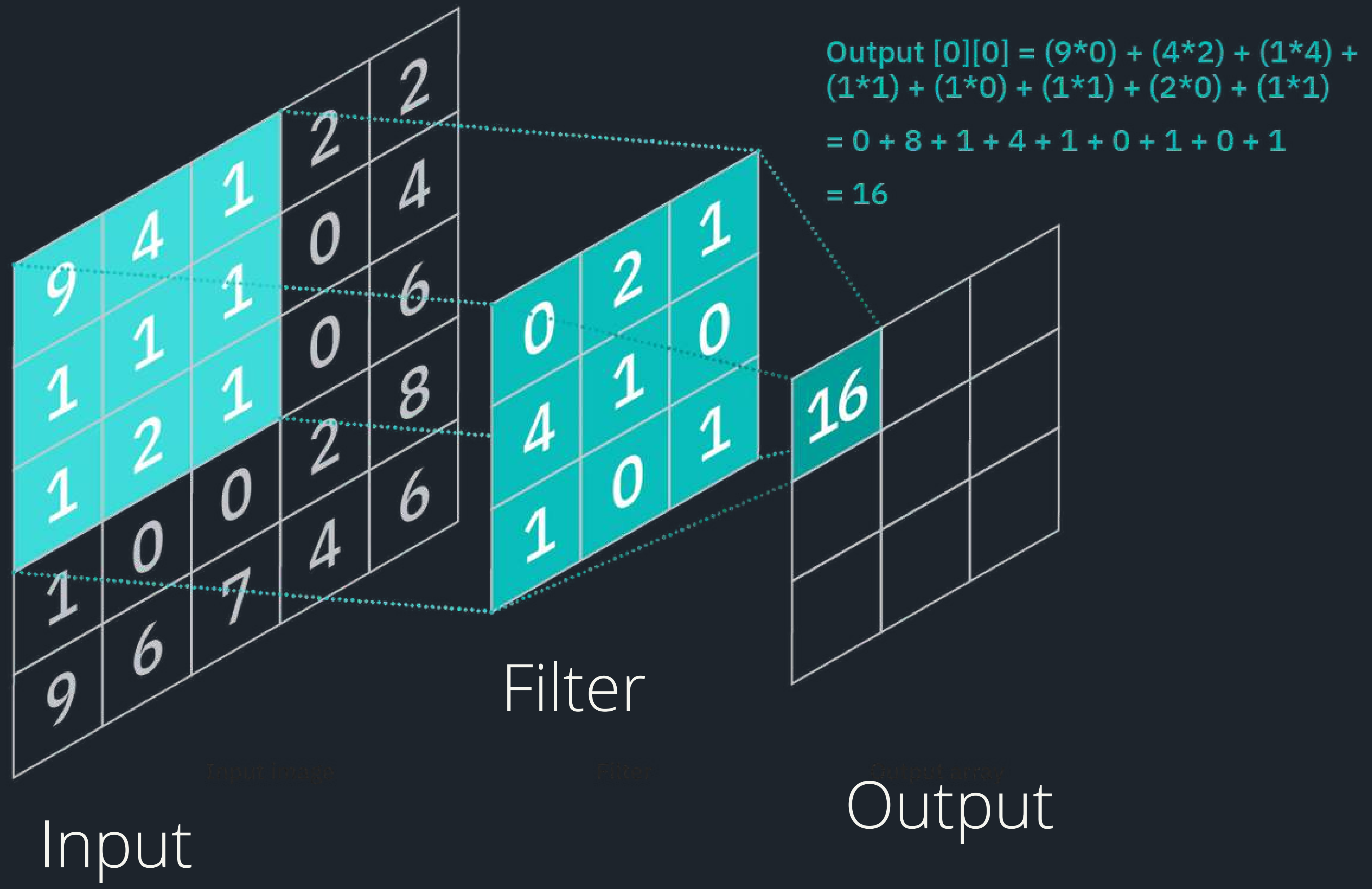
The input layer of a neural network is composed of artificial input neurons, and brings the initial data into the system for further processing by subsequent layers of artificial neurons. The input layer is the very beginning of the workflow for the artificial neural network. The pixels from the input image help recognize the features in the image

CONVO LAYER

A CNN is a neural network with some convolutional layers. It's the first layer that is used to extract the various features from the input images.
This layer has a number of filters that do the convolutional operation. Convo layer also contains ReLU activation to make all negative values to zero.

CONVOLUTION

Computation :



The mathematical operation of convolution is performed between the input image and a filter of a particular size FxF. By sliding the filter over the input image, the dot product is taken between the filter and the parts of the input image with respect to the size of the filter (FxF).

VERTICAL EDGE DETECTION

Vertical edges can be detected by using a horizontal gradient operator followed by a threshold operation to detect the extreme values of the gradient. The gradient produces a doublet of extremes, positive-negative or negative-positive, depending on the direction of the transition

6 x 6 Grayscale Image, Stride=1

10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0

(10 x 1) + (10 x 1) +(10 x 1) +
(10 x 0) + (10 x 0) + (10 x 0) +
(10 x -1) + (10 x -1) + (10 x -1)=0

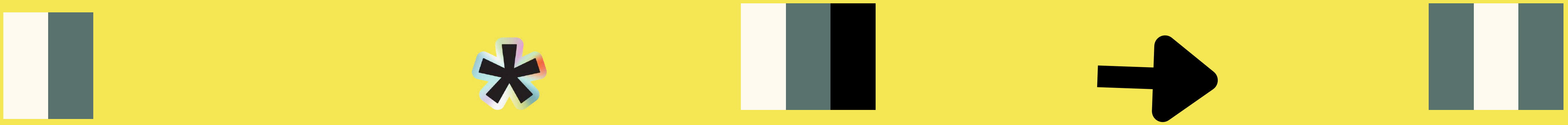
*

3 x 3 Filter

1	0	-1
1	0	-1
1	0	-1

4 x 4 Feature Map

0	30	30	0
0	30	30	0
0	30	30	0
0	30	30	0



SOME OTHER FILTERS

Horizontal Edge Detection

Just as in Vertical Edge Detection, this matrix is used to detect Horizontal Edges

1	1	1
0	0	0
-1	-1	-1

Sobel Filter

The Sobel filter is used for edge detection. It works by calculating the gradient of image intensity at each pixel within the image. It finds the direction of the largest increase from light to dark and the rate of change in that direction.

1	0	-1
2	0	-2
1	0	-1

Scharr Filter

The Scharr operator is similar to the method used by the sobel operator. The difference is that the Scharr operator has a relatively large kernel value so that the surrounding pixels will have a larger influence on the edge, and the edge will be more.

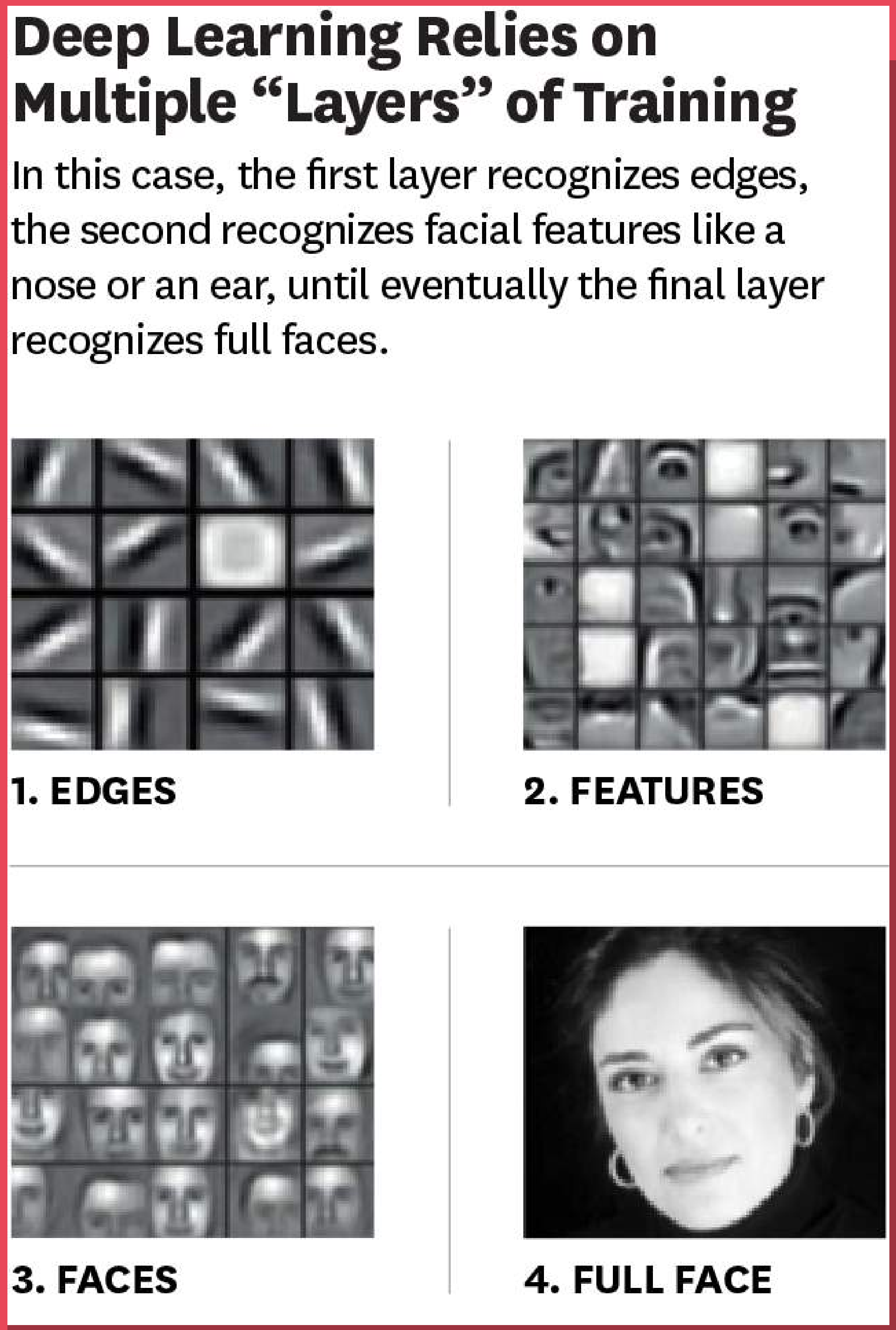
3	0	-3
10	0	-10
3	0	-3

DIFFERENT LEVELS OF ABSTRACTION

VERTICAL EDGE DETECTION

INPUT

HORIZONTAL EDGE DETECTION



PADDING

We have observed that a convolution of a 6 x 6 Input Image and a 3 x 3 Filter results is an output image of 4 x 4.

Hence we can say that If the input Image is n x n and filter is f x f, the output will be of the dimension (n-f+1) x (n-f+1).

We notice that our output dimensions are shrinking.
On further computation, the corners will not be used. Therefore there is a loss of information.

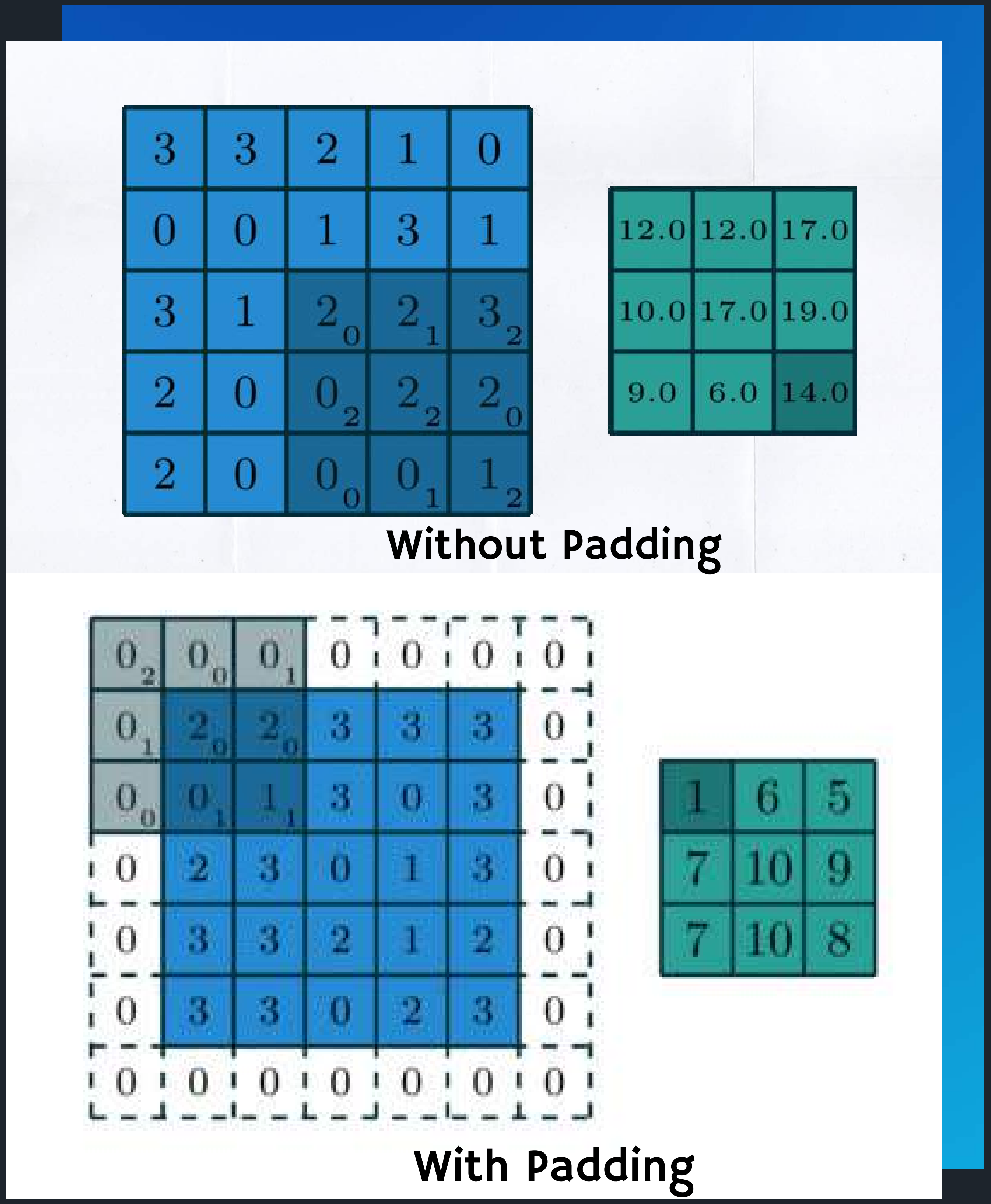
Padding is simply a process of adding layers of zeros to our input images so as to avoid shrinking.

There are two types of convolutions:

1. Valid- No Padding
2. Same- Same padding means the size of output feature-maps is the same as the input feature-maps

If p = number of layers of zeros added to the border of the image, then our (n x n) image becomes (n+2p) x (n+2p) image after padding, i.e., the output dimensions are changed to (n+2p-f+1) x (n+2p-f+1) from the original (n-f+1) x (n-f+1).

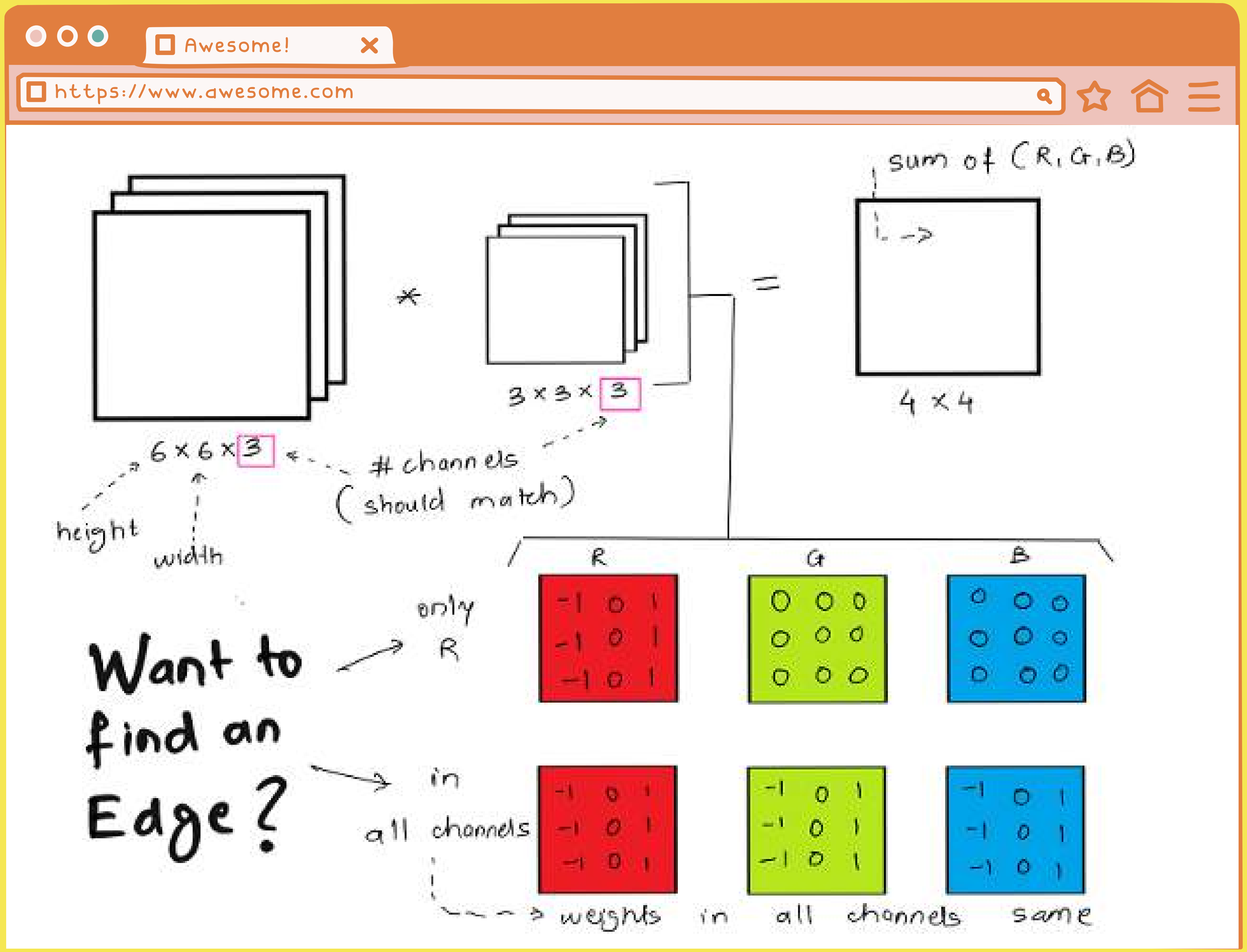
If we go back to our eg with input 6 x 6 and filter 3 x 3, say we use a padding with p=1 here. Then our output dimensions change to (6+ 2x1 -3 +1=6) i.e., 6 x 6 in contrast with the initial 4 x 4 where there was loss of information.



CONVOLUTION OVER VOLUME

Suppose, instead of a 2-D image, we have a 3-D input image of shape 6 X 6 X 3 (n x n x nc , where nc=no. of filters). We will use a 3 X 3 X 3 filter instead of a 3 X 3 filter.

The dimensions above represent the height, width, and 3 channels(R,G,B) in the input and filter. Keep in mind that the number of channels in the input and filter should be the same. This will result in an output of 4 X 4.



POOLING LAYER

Input

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

Max Pool with 2x2 filters and Stride 2

8	6
9	9

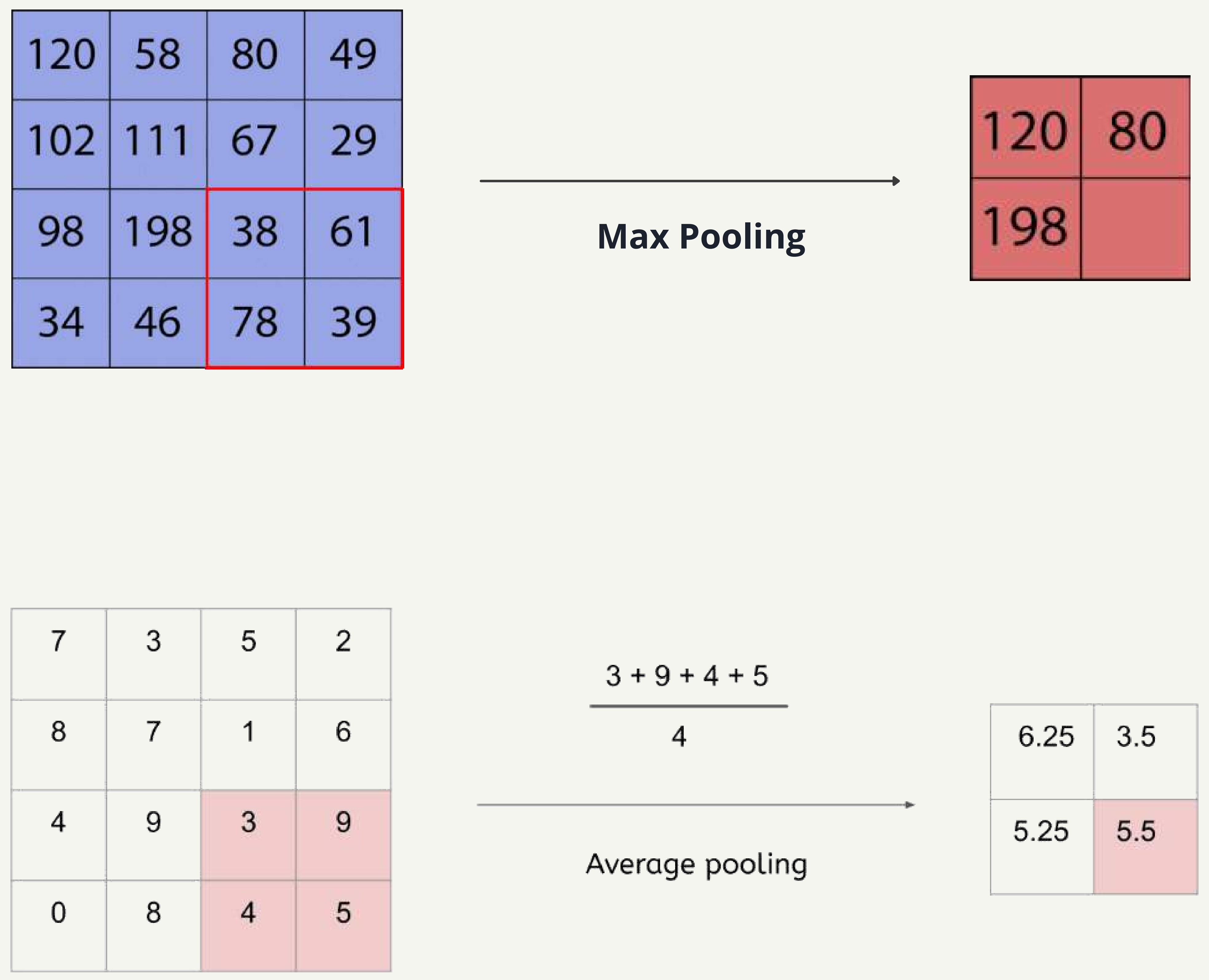
Output

The pooling layer is responsible for reducing the spatial size of the input image after convolution. This is to decrease the computational power required to process the data through dimensionality reduction. There are two types of Pooling:

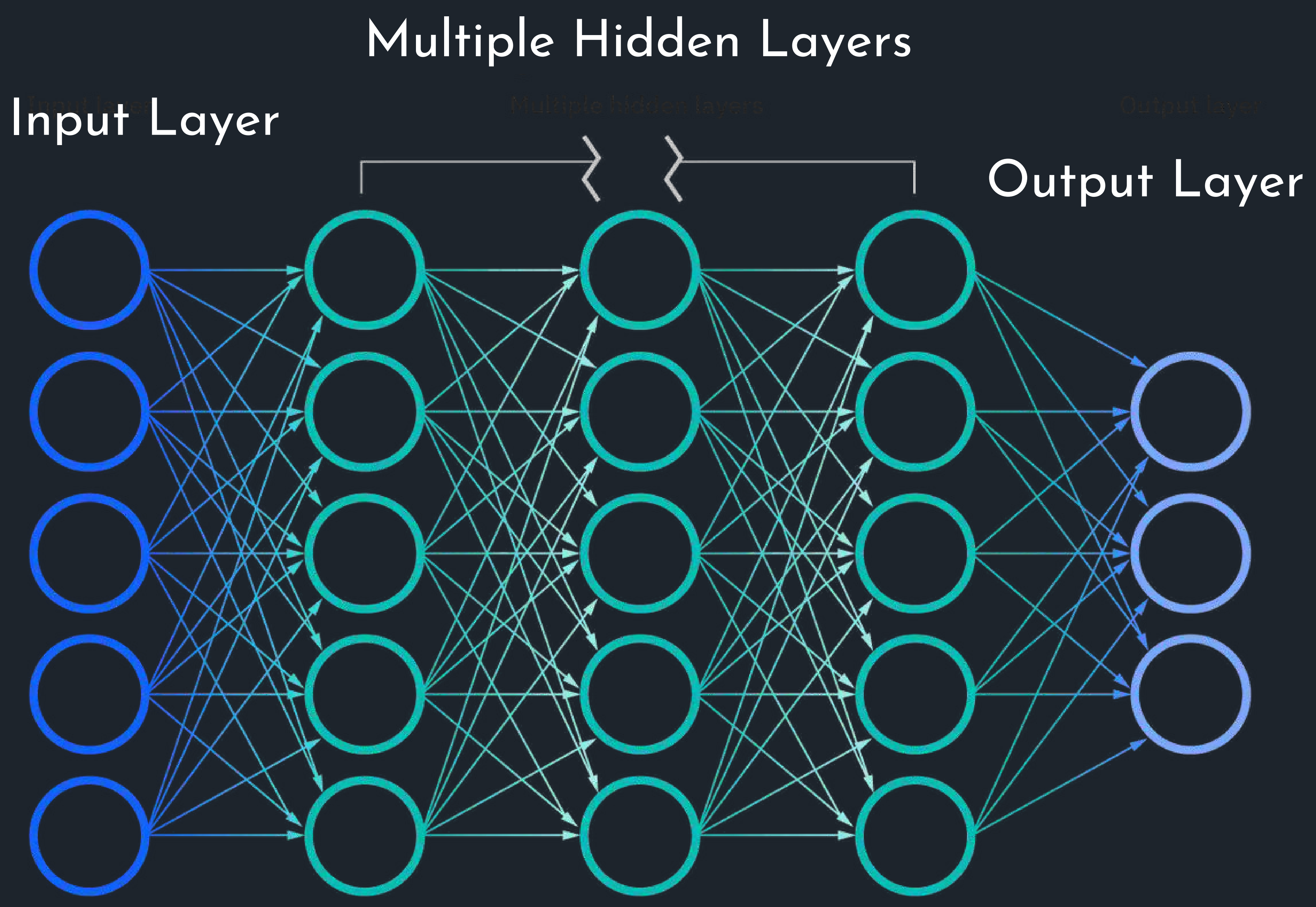
1. Max Pooling
2. Average Pooling

Max Pooling returns the maximum value from the portion of the image covered by the Filter.

Average Pooling returns the average of all the values from the portion of the image covered by the Filter.



FULLY CONNECTED LAYER (FC LAYER)



A fully connected layer involves weights, biases, and neurons and is used to connect the neurons between two different layers. It is used to classify images between different categories by training.

DROPOUT LAYER

Usually, when all the features are connected to the FC layer, it can cause overfitting in the training dataset. To overcome this problem, a dropout layer is utilized wherein a few neurons are dropped from the neural network during the training process resulting in reduced size of the model.

OUTPUT LAYER

Output layer contains the label which is in the form of one-hot encoded. It is responsible for producing the final result. There must always be one output layer in a neural network.

ACTIVATION FUNCTIONS

The activation function is an important part of an artificial neural network. They decide whether a neuron should be activated or not and it is a non-linear transformation that can be done on the input before sending it to the next layer of neurons or finalizing the output.

There are several commonly used activation functions such as the ReLU, Softmax, tanH, and the Sigmoid functions. The choice of activation function in the output layer will define the type of predictions the model can make.

For a binary classification CNN model, sigmoid and ReLU are preferred and for multi-class classification, softmax is used.

TYPES OF ACTIVATION FUNCTIONS

Sigmoid

$$f(x) = \frac{1}{1 + e^{-x}}$$

Tanh

$$f(x) = \frac{2}{1 + e^{-2x}} - 1$$

ReLU

$$f(x) = max(0, x)$$

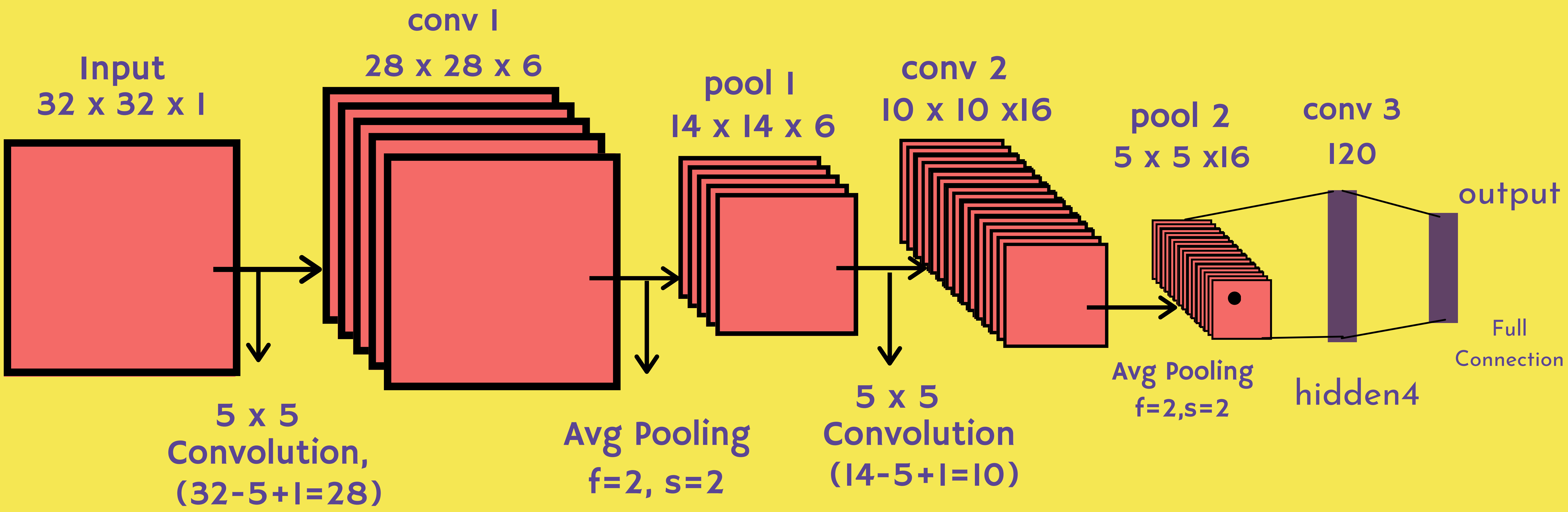
Softmax

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

TYPES OF CNN

- LeNet-5
- GoogLeNet
- AlexNet
- VGGNet
- ZFNet
- ResNet

LENET-5 (1998)



LeNet-5 CNN architecture is made up of 7 layers. The layer composition consists of 3 convolutional layers, 2 Pooling layers and 2 fully connected layers. It classifies digits and was applied by several banks to recognise hand-written numbers on checks (cheques) digitized in 32x32 pixel greyscale inputimages.

Why are CNNs so powerful?

In simple terms a large enough CNN can solve any solvable problem.
Notable CNN architecture's that perform exceptionally well across many different image processing tasks are the VGG models (K. Simonyan and A. Zisserman), the ResNet models (Kaiming He et al) and the Google Inception models (Christian Szegedy et al). These models have millions of trainable parameters.

The Universal approximation theorem essentially states if a problem can be solved it can be solved by deep neural networks, given enough layers of functions layered with non-linear functions.

WHAT'S FOR TOMORROW?

Hands - On Session!

IT'S KAHOOT TIME!

GAME PIN

6165531

Let's Connect!

https://twitter.com/acmwsrm?s=09

https://linktr.ee/acmwsrm

https://www.facebook.com/acmwsrm/

acmwsrm@gmail.com

https://www.linkedin.com/company/acm-women-srmist-student-chapter

https://instagram.com/acmwsrm?utm_medium=copy_link

THANK YOU!