Exploring The Power Of **Conv'Nets**

01 Introduction to Computer Vision

Yahia Chammami

ESSAI MACHINE LEARNING CLUB

What is Machine Learning?

Machine learning is a **subdomain of computer science** that focuses on algorithms which help a **computer learn from data without explicit programming**

Types of Machine Learning

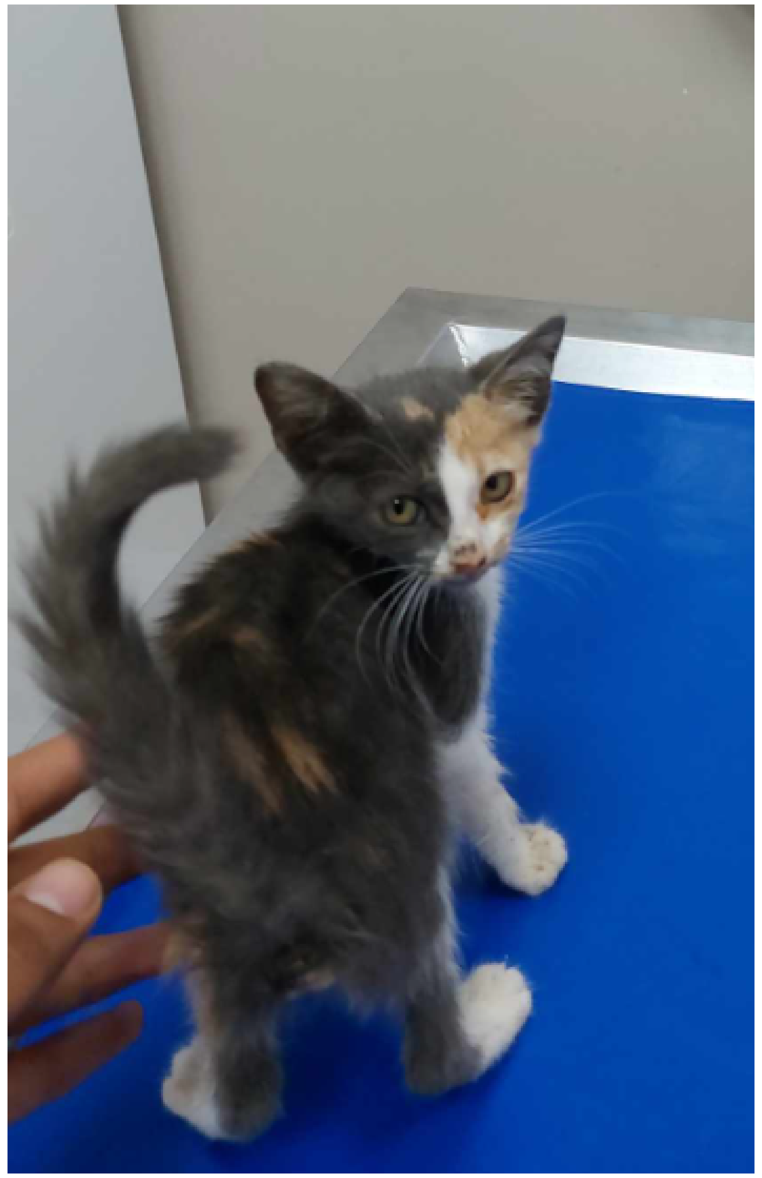
01 **Supervised learning** - uses labeled inputs (meaning the input has a corresponding output label) to train models and learn outputs

02 **Unsupervised learning** - uses unlabeled data to learn about patterns in data

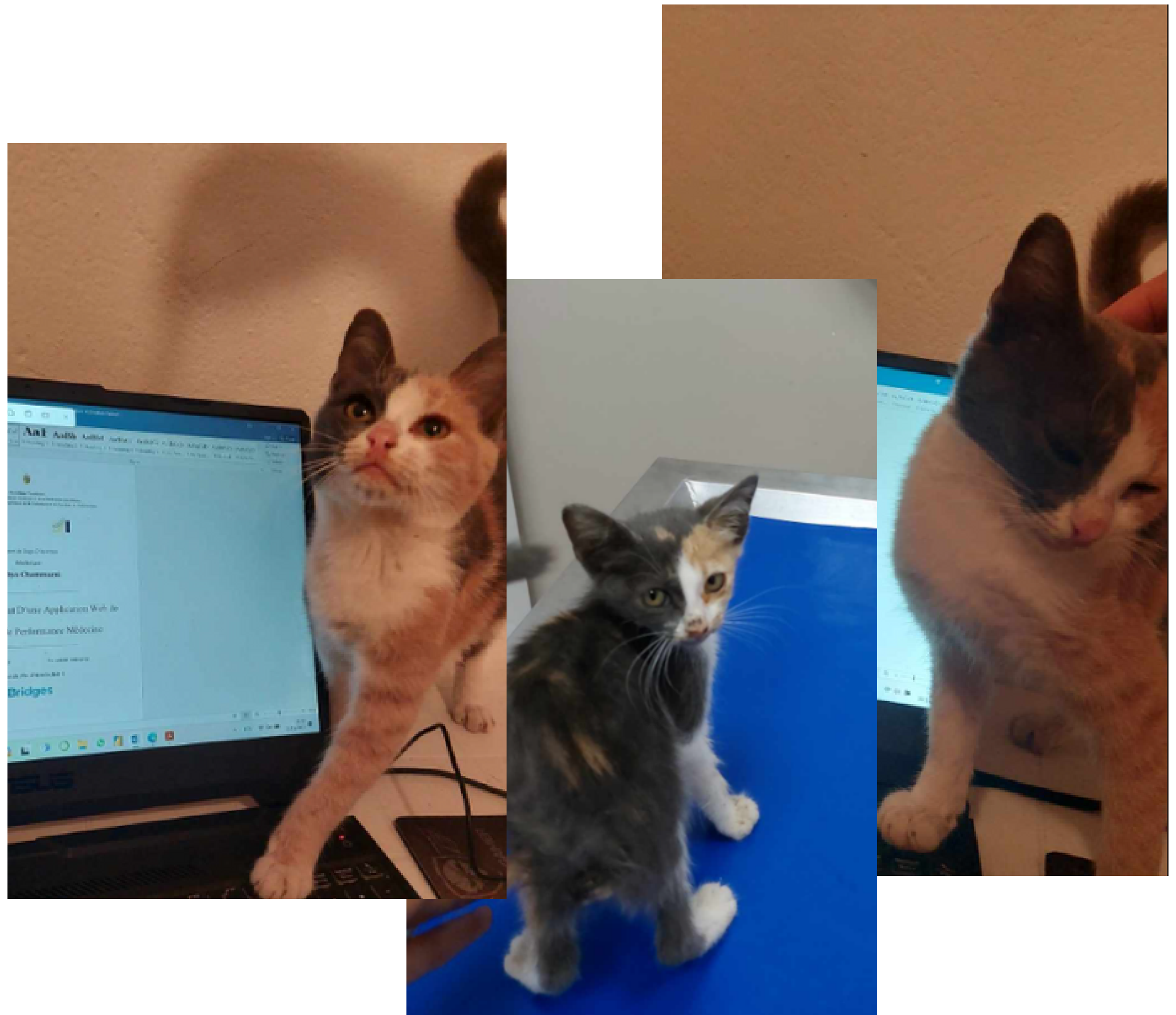
03 **Reinforcement learning** - agent learning in interactive environment based on rewards and penalties

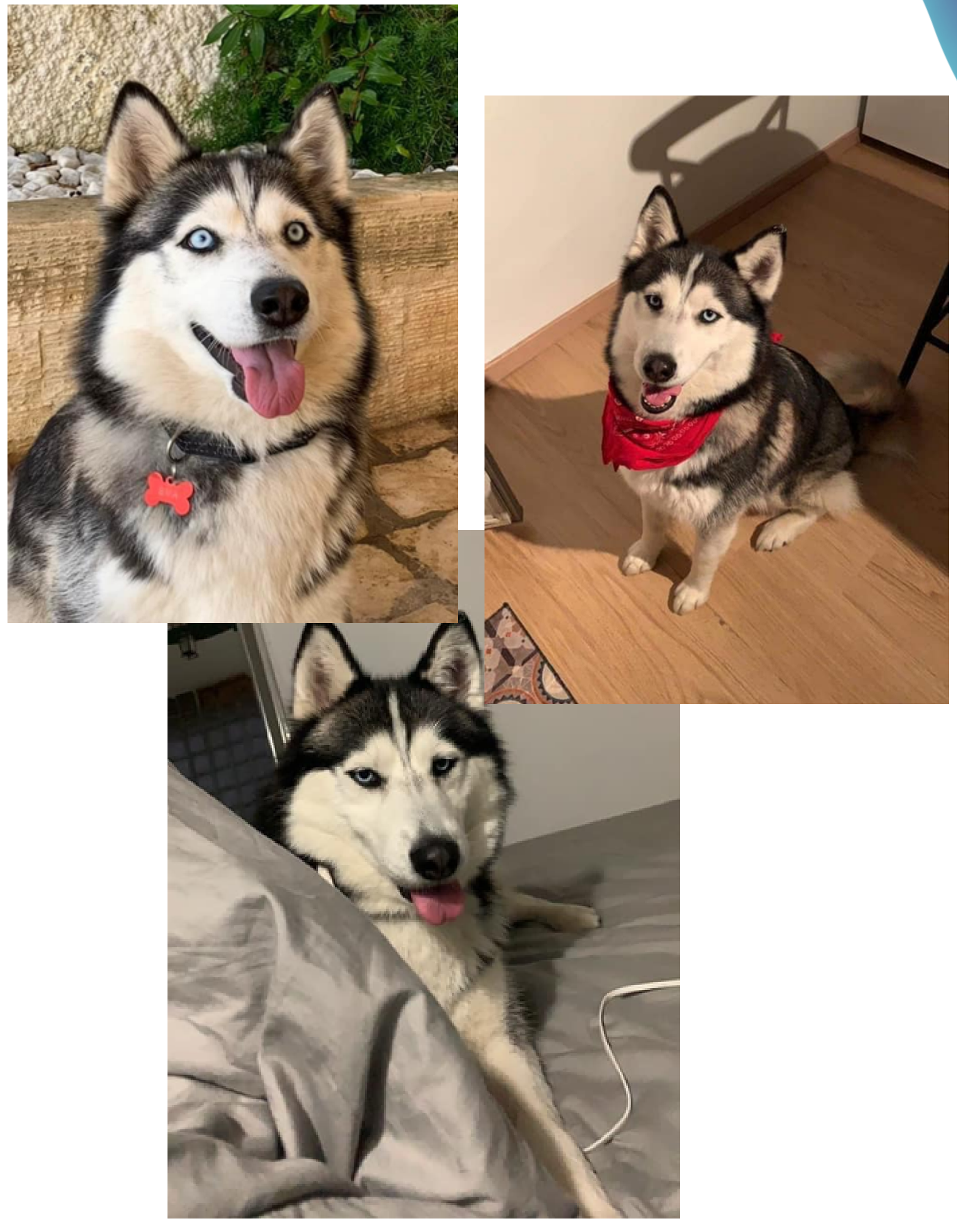
ESSAI MACHINE LEARNING CLUB

Supervised Learning

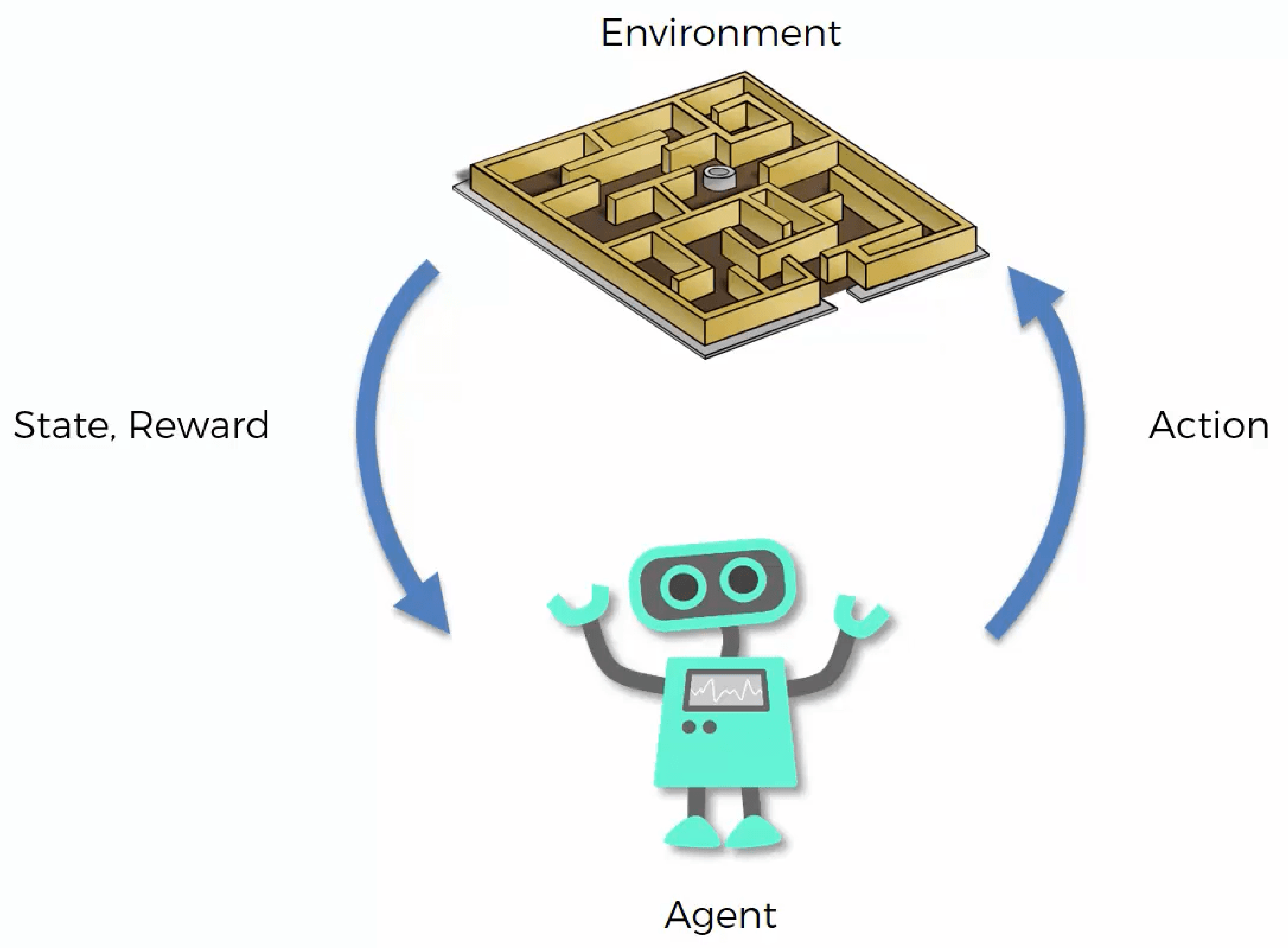


UnSupervised Learning





Reinforcement learning



Supervised Learning Tasks

01 Regression - predict continuous values

02 Classification - predict discrete classes

BINARY CLASSIFICATION

Positive/negative

Cat/dog

Spam/not spam

MULTICLASS CLASSIFICATION

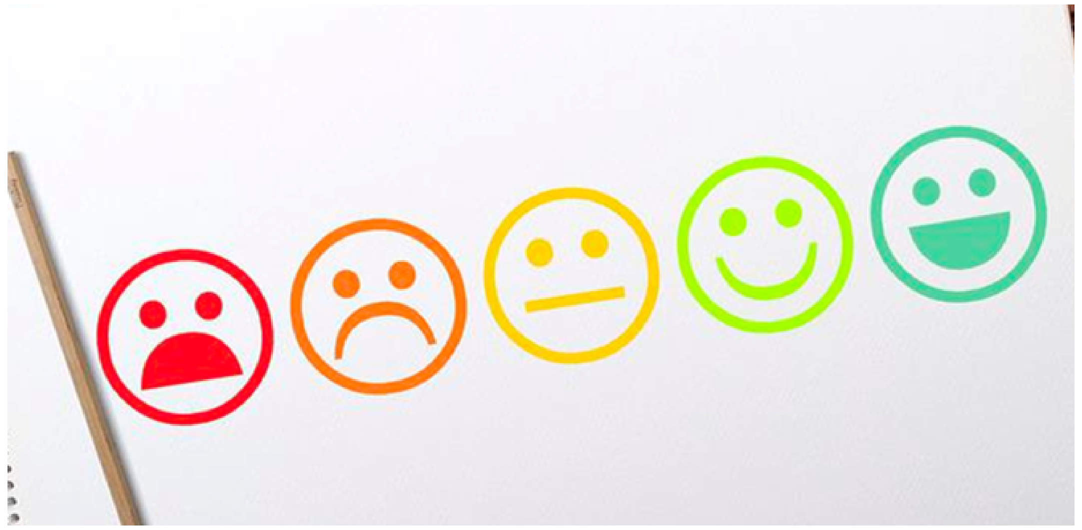
Cat/dog/lizard/dolphin

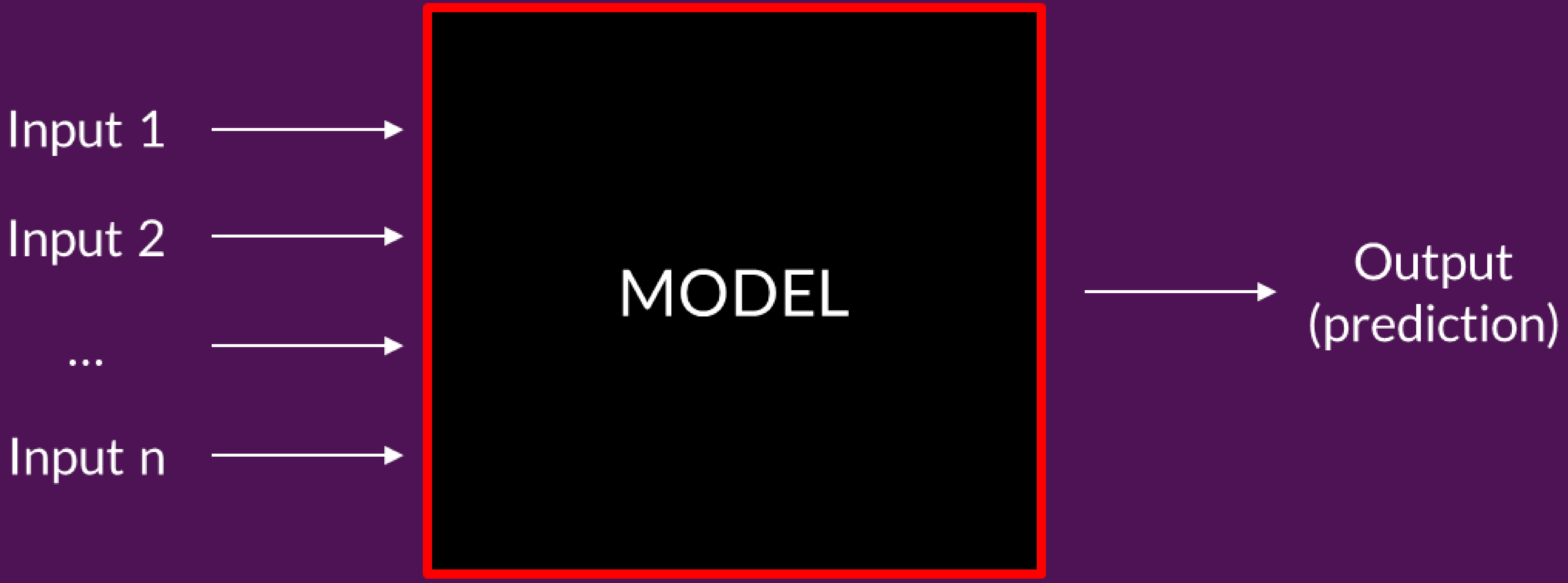
Orange/apple/pear

Plant species

Quantitative - numerical valued data (could be discrete or continuous)

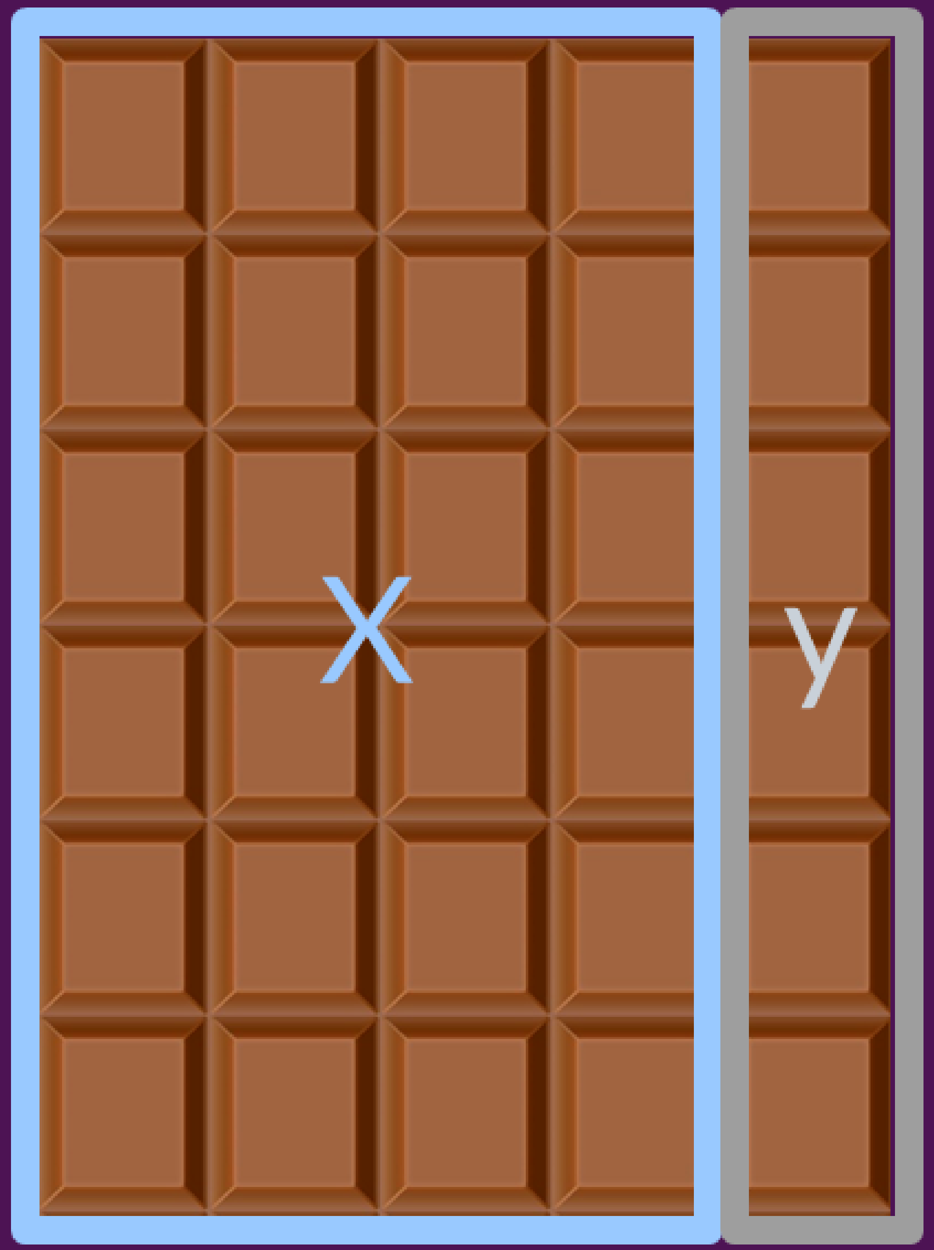
Qualitative - categorical data (finite number of categories or groups)





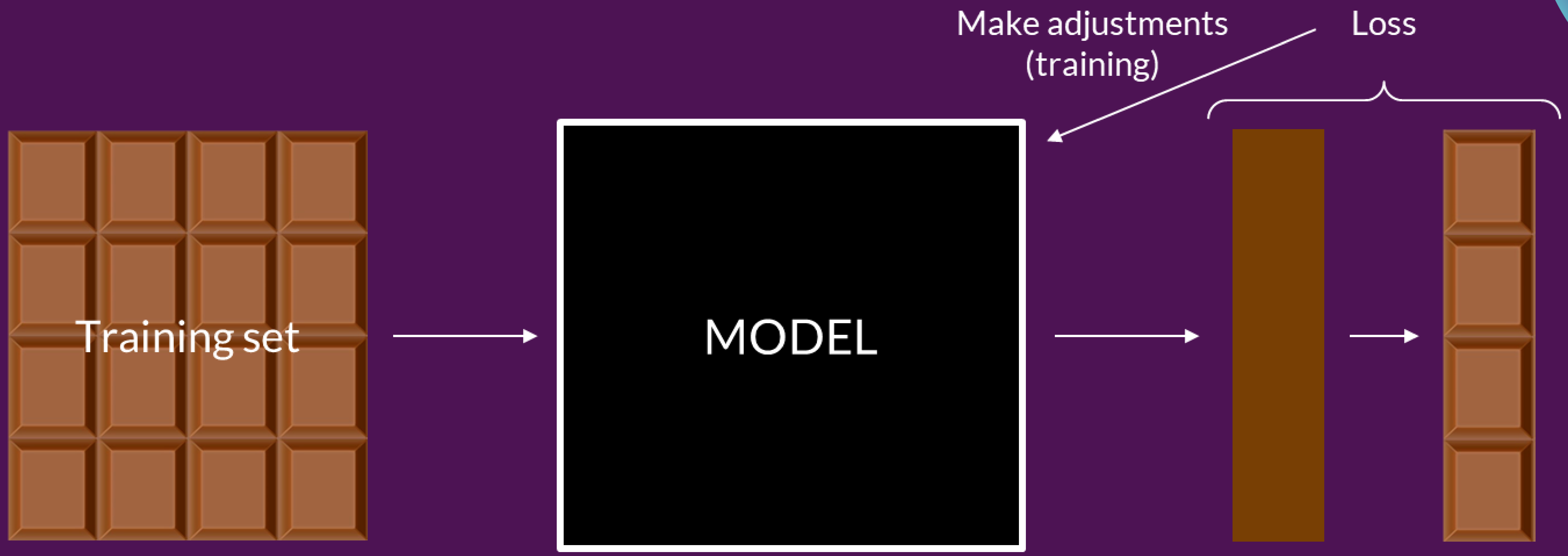
How do we make the model learn?
How can we tell whether or not it's learning?

Supervised Learning Dataset

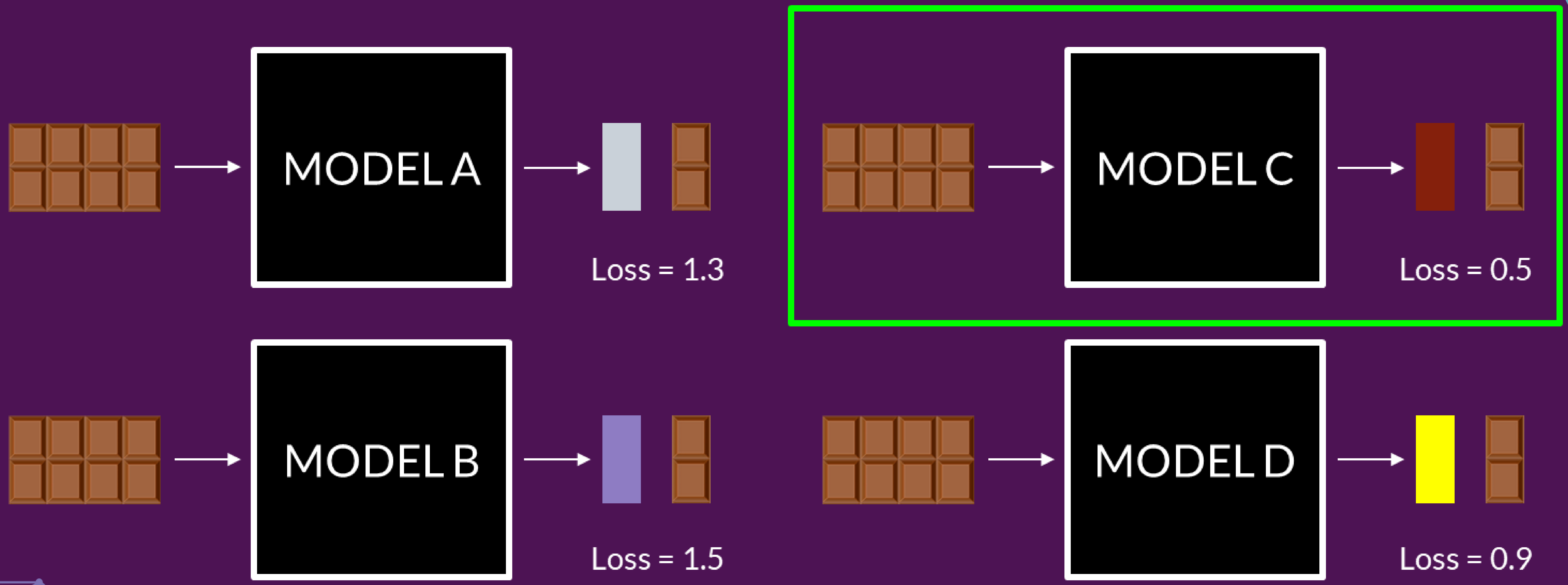


Supervised Learning Dataset

Supervised Learning Dataset



Supervised Learning Dataset



Metrics of Performance

Classification Tasks

True Positives (TP): The model predicted the positive class correctly, and the actual class is indeed positive.

True Negatives (TN): The model predicted the negative class correctly, and the actual class is indeed negative.

False Positives (FP): The model predicted the positive class, but the actual class is negative.

False Negatives (FN): The model predicted the negative class, but the actual class is positive.

Metrics of Performance

Classification Tasks

Accuracy : Is the metric to evaluate how correcct the model is to predict the values.
Accuracy = (TP + TN / (TP + FP + FN + TN)

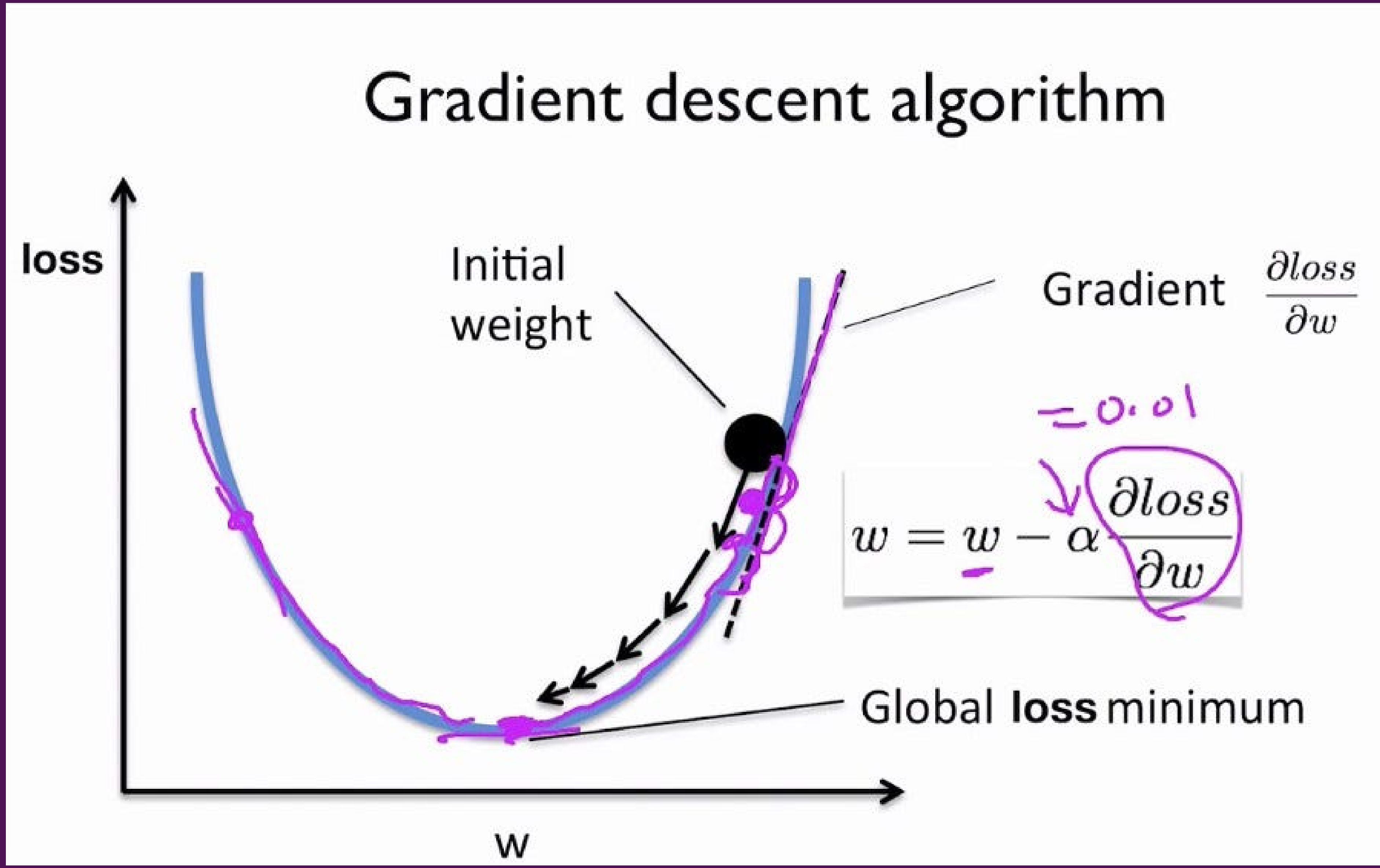
Precision : Out of all actual values, how many are correctly predicted
Precision = TP /(TP +FP)

Recall : From all the predicted values how many are correctly predicted.
Recall = TP/(TP+FN)

F1 score : Is the best choice if your distrubution is uneven.
F1_score = 2(Recall Precision)/(Recall + Precision)

Optimization Techniques

Optimization refers to the process of adjusting the <u>parameters of a model</u> to minimize or maximize some objective function.



Gradient Descent

$$\nu_t = \beta_1 * \nu_{t-1} - (1 - \beta_1) * g_t$$
$$s_t = \beta_2 * s_{t-1} - (1 - \beta_2) * g_t^2$$
$$\Delta\omega_t = -\eta \frac{\nu_t}{\sqrt{s_t + \epsilon}} * g_t$$
$$\omega_{t+1} = \omega_t + \Delta\omega_t$$

η : *Initial Learning rate*
g_t : *Gradient at time t along* ω^j
ν_t : *Exponential Average of gradients along* ω_j
s_t : *Exponential Average of squares of gradients along* ω_j
β_1, β_2 : *Hyperparameters*

ADAM = momentum + Root Mean Square Propagation

RegularizationTechniques

Regularization is a set of techniques used to prevent overfitting and improve the generalization ability of a model.

L2 Regularization (Weight Decay) : adds a penalty term to the loss function, which penalizes large weights in the network.

Dropout : technique where randomly selected neurons are ignored during training.

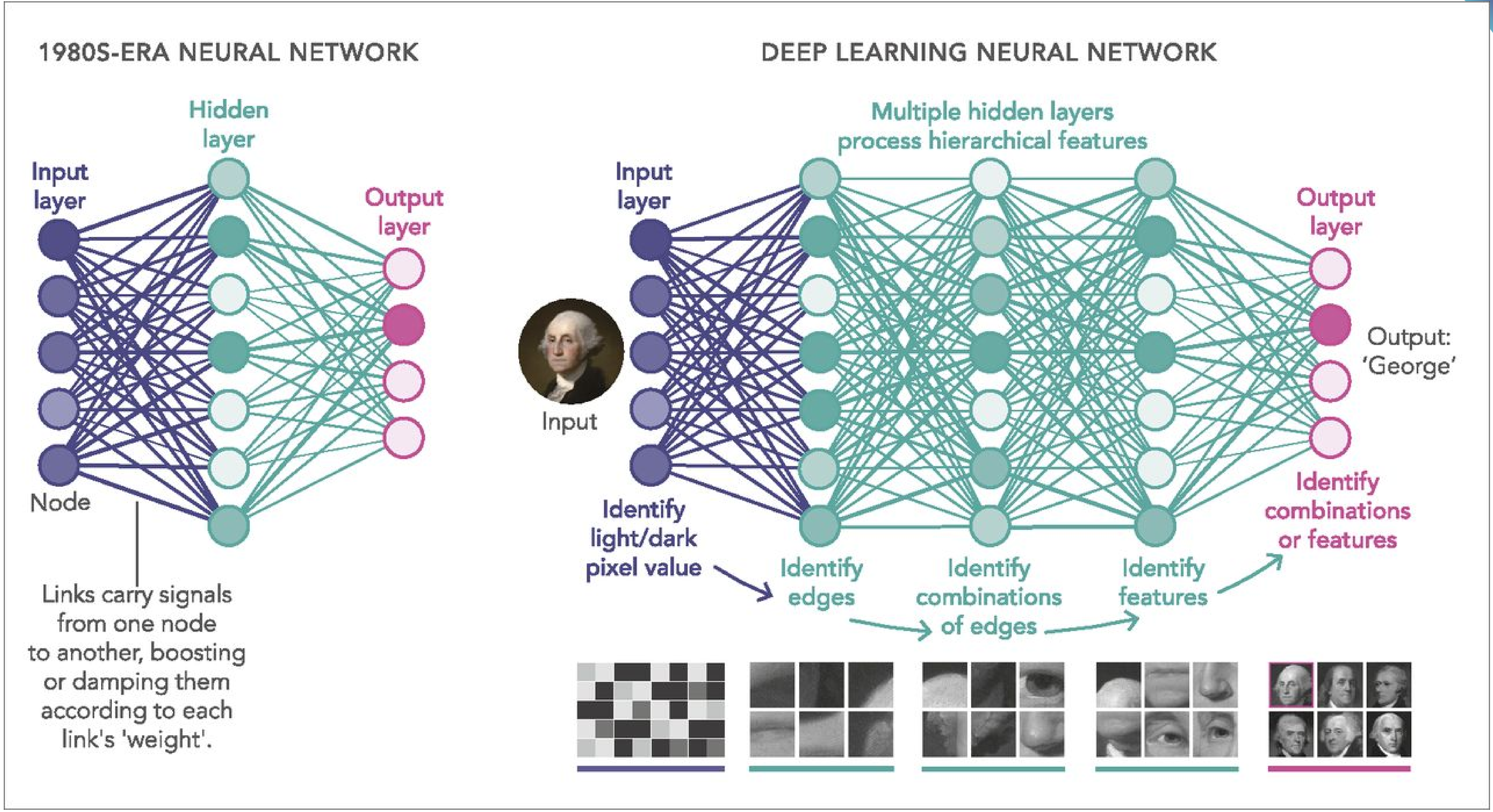
Data Augmentation : involves artificially increasing the size of the training dataset by applying transformations such as rotation, translation, scaling, flipping, and cropping to the input images.

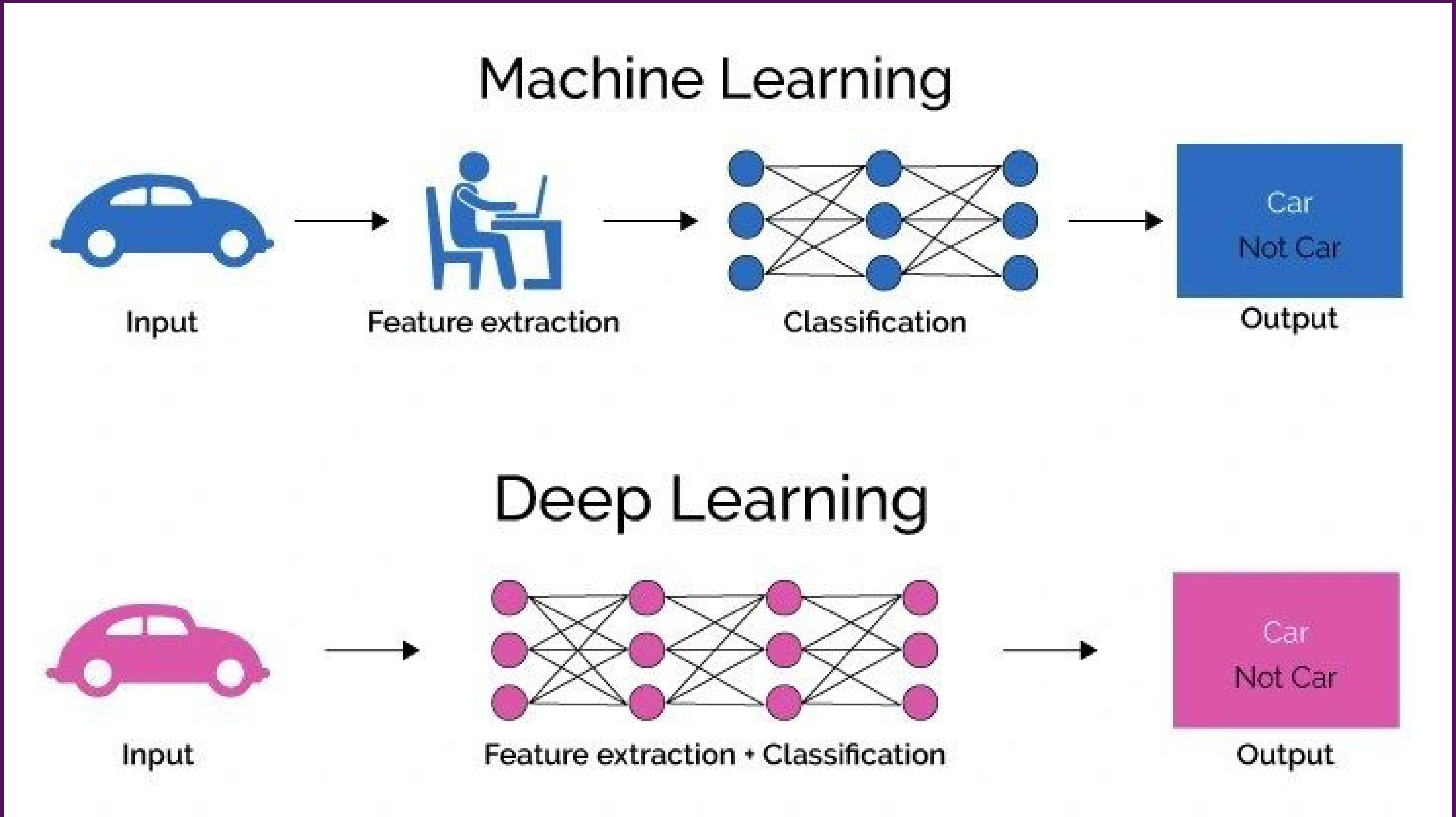
Batch Normalization : stabilizing training by reducing internal covariate shift and acting as a regularizer by adding noise.

Early Stopping : stopping the training process when the performance starts to degrade.

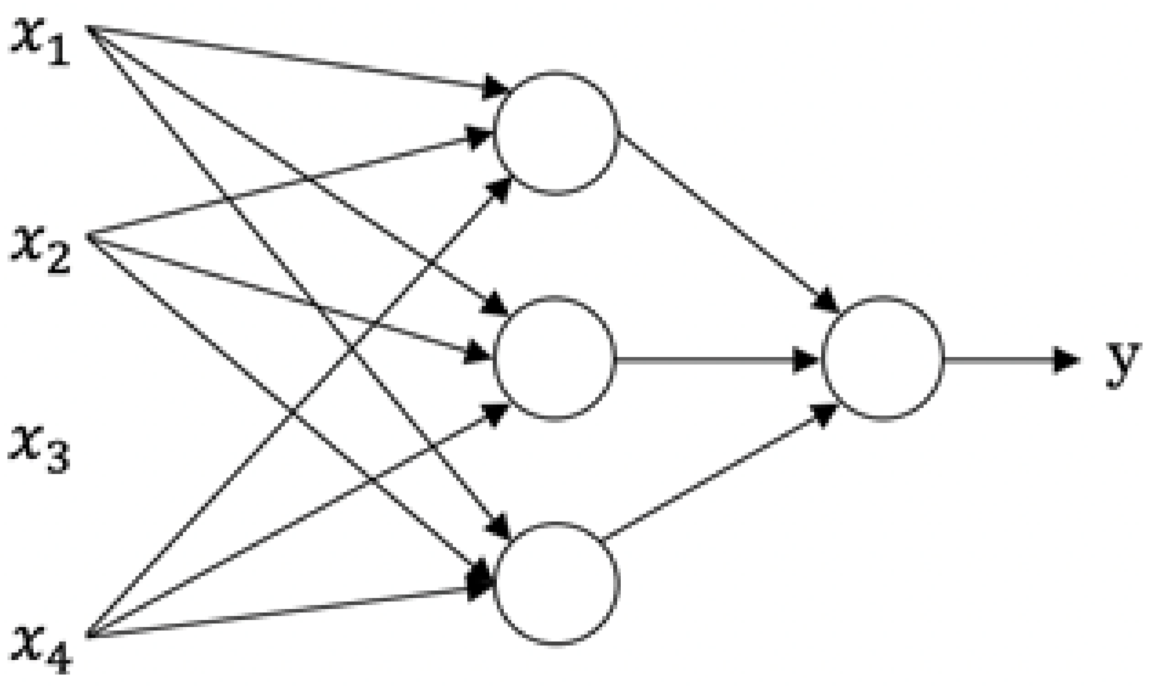
NEURAL NETS

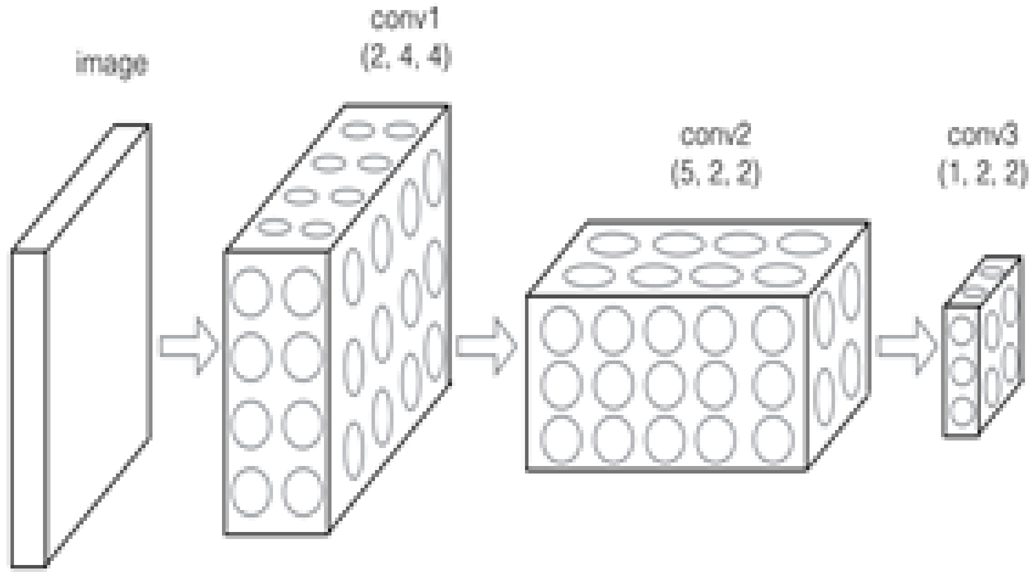
ESSAI MACHINE LEARNING CLUB

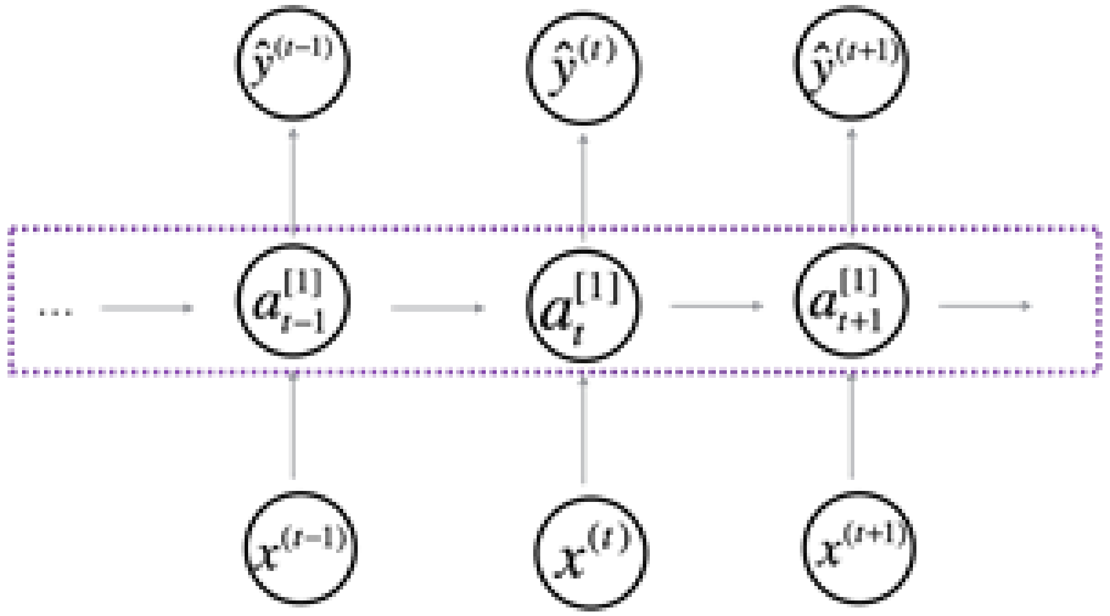

1980S-ERA NEURAL NETWORK
Hidden layer
Input layer
Output layer
Node
Links carry signals from one node to another, boosting or damping them according to each link's 'weight'.
DEEP LEARNING NEURAL NETWORK
Multiple hidden layers process hierarchical features
Input layer
Output layer
Input
Output: 'George'
Identify light/dark pixel value
Identify edges
Identify combinations of edges
Identify features
Identify combinations or features


Machine Learning
Car
Not Car
Input
Feature extraction
Classification
Output
Deep Learning
Car
Not Car
Input
Feature extraction + Classification
Output

Neural Network examples



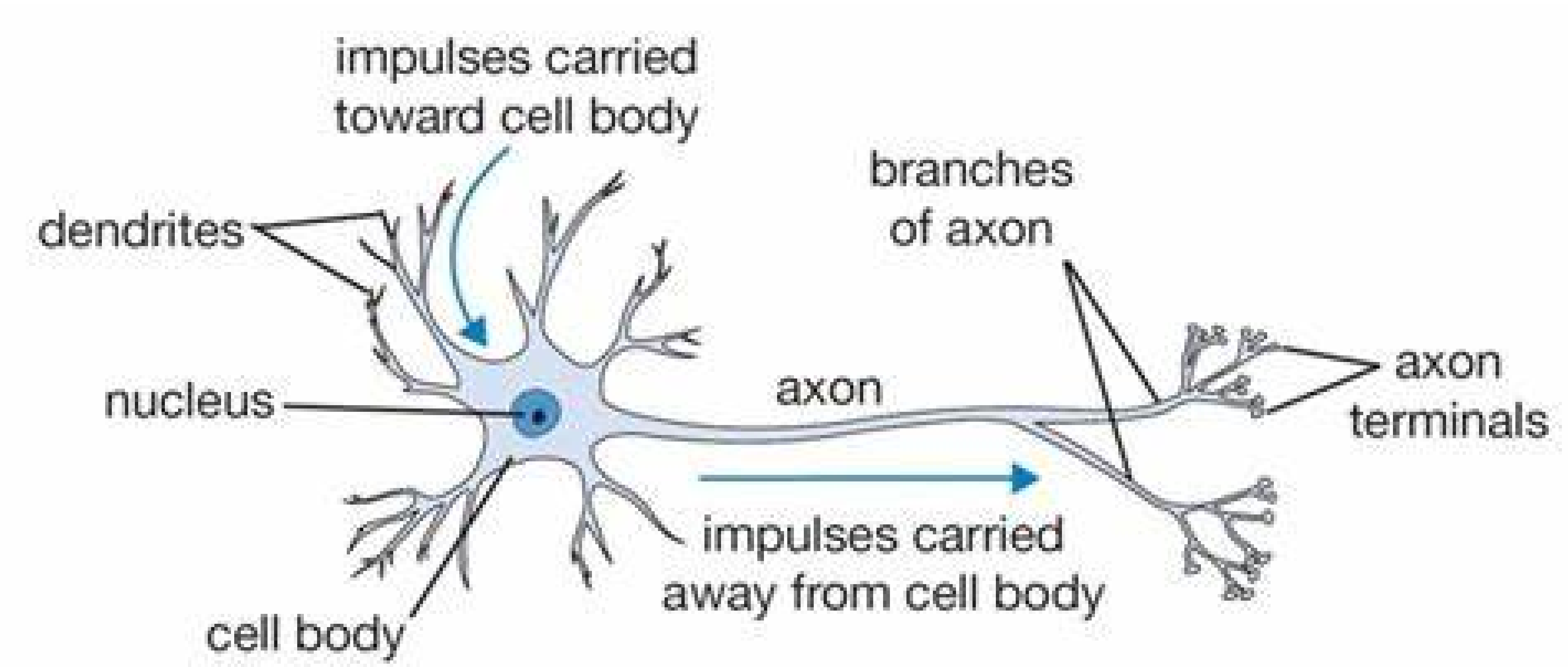


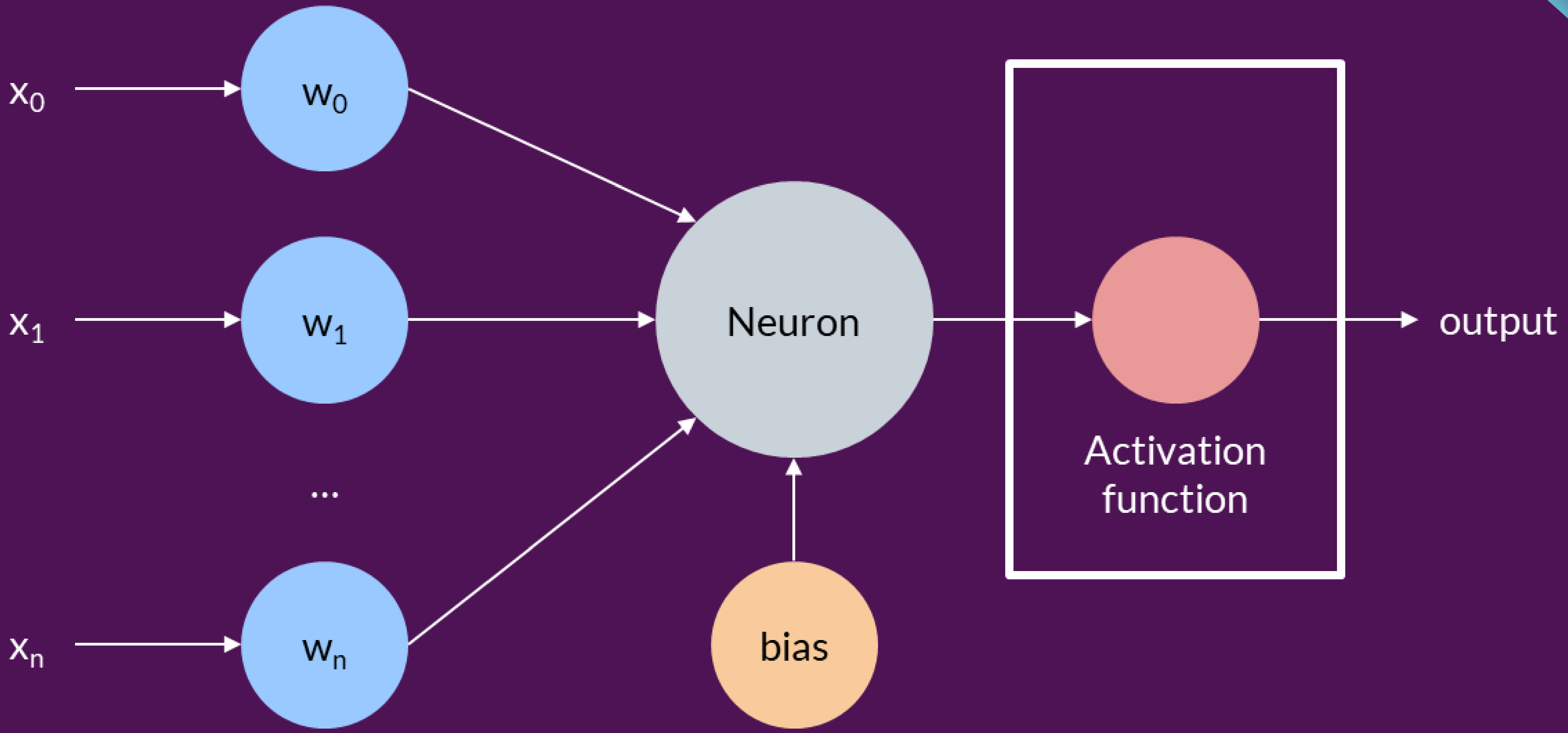


Standard NN

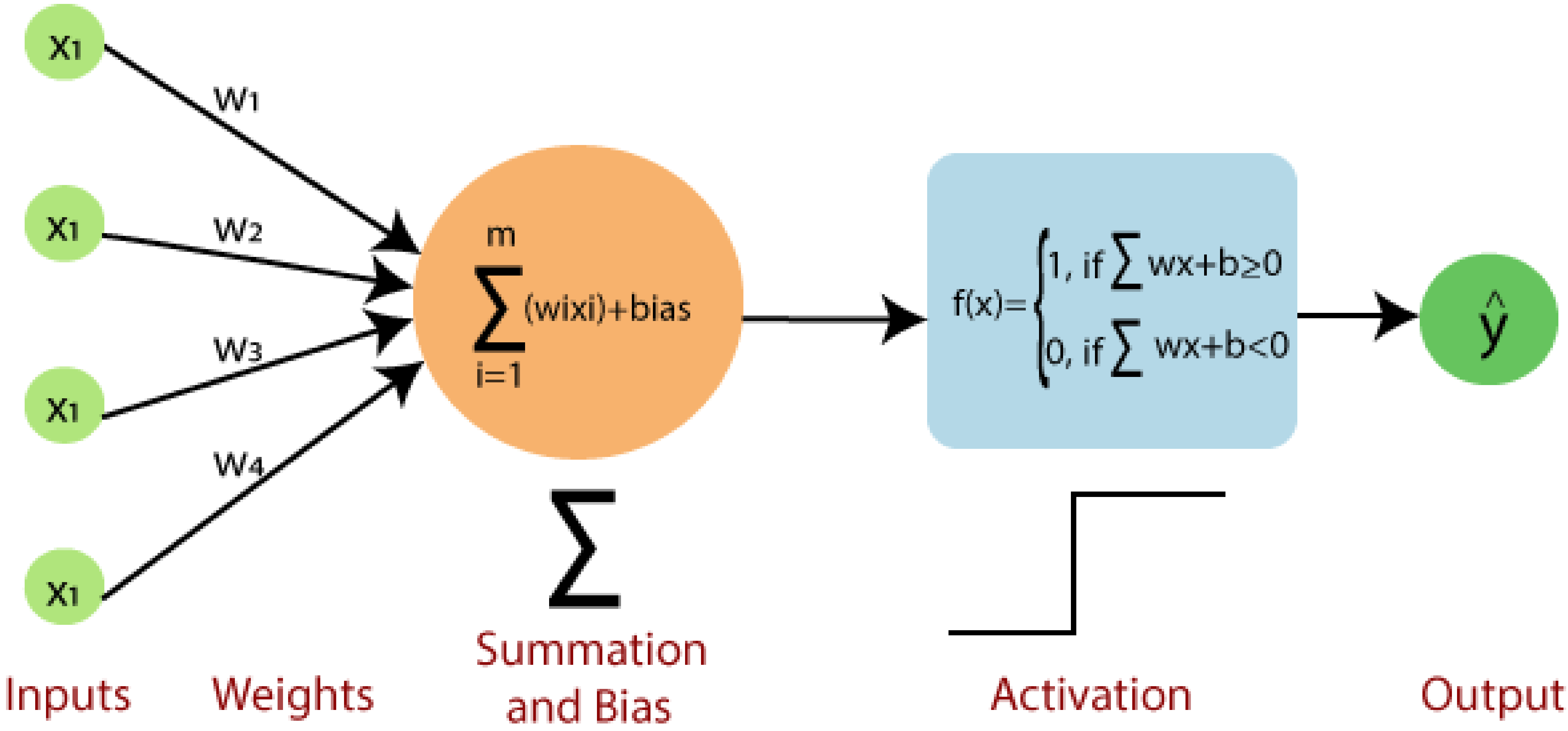
Convolutional NN

Recurrent NN

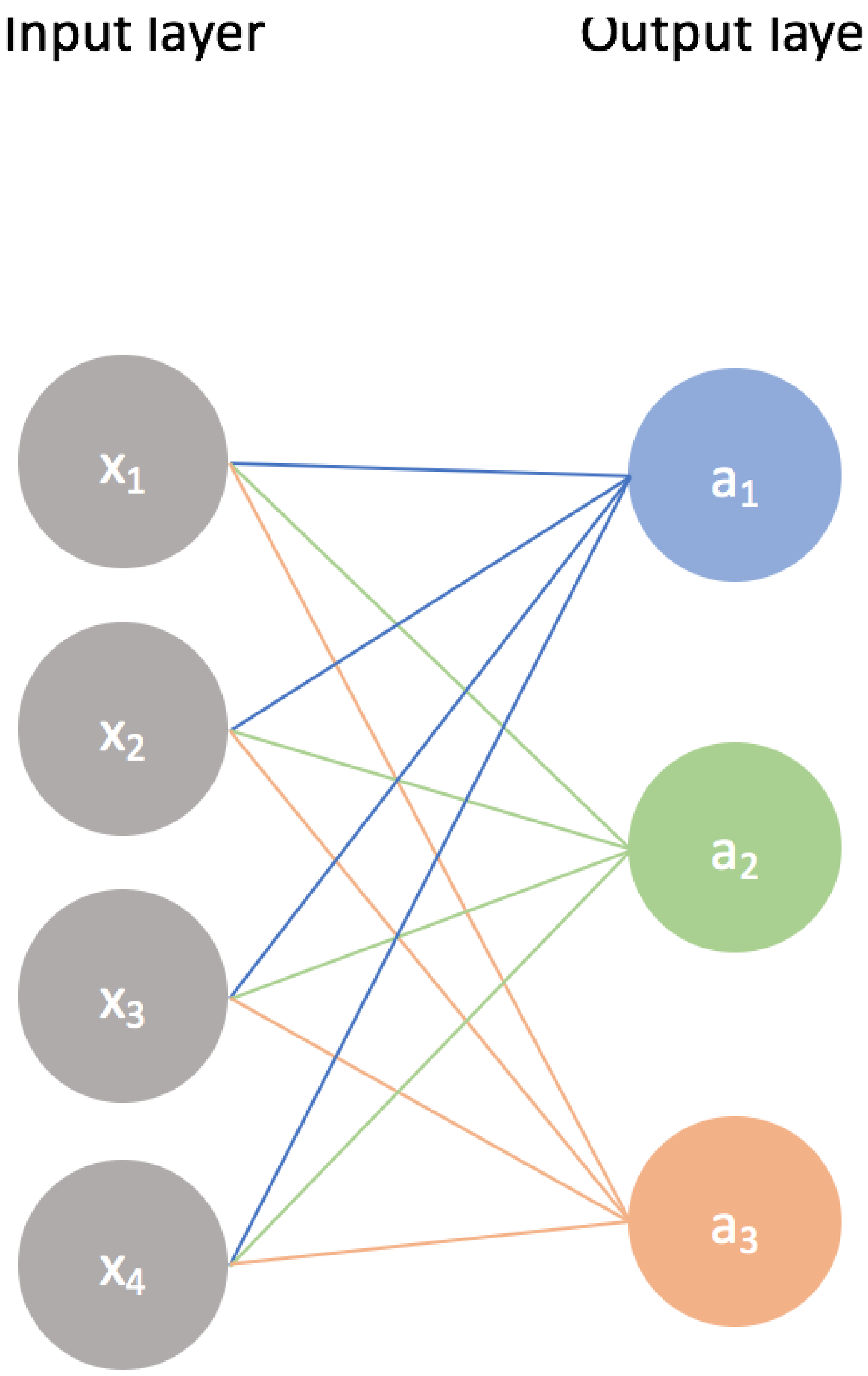

impulses carried
toward cell body
dendrites
branches
of axon
axon
axon
terminals
nucleus
impulses carried
away from cell body
cell body



A node, also called a neuron, is a unit that contains a value calculated through a mathematical operation.



x1
x1
x1
x1
w1
w2
w3
w4
m
∑ (wixi)+bias
i=1
f(x)= 1, if ∑ wx+b≥0
0, if ∑ wx+b<0
ŷ
Inputs
Weights
Summation and Bias
Activation
Output



A simple neural network

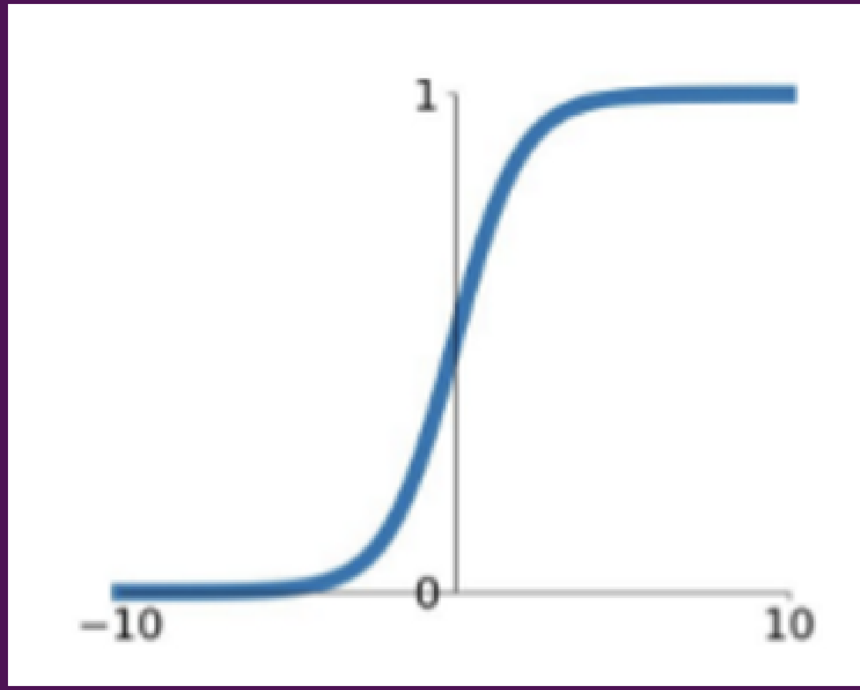
$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b \\ b \\ b \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \end{bmatrix} \xrightarrow[\textit{activation}]{} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Every connection is associated a weight (can be random at first)

Every neuron is associated a bias (can be random at first)

Activation Fonctions

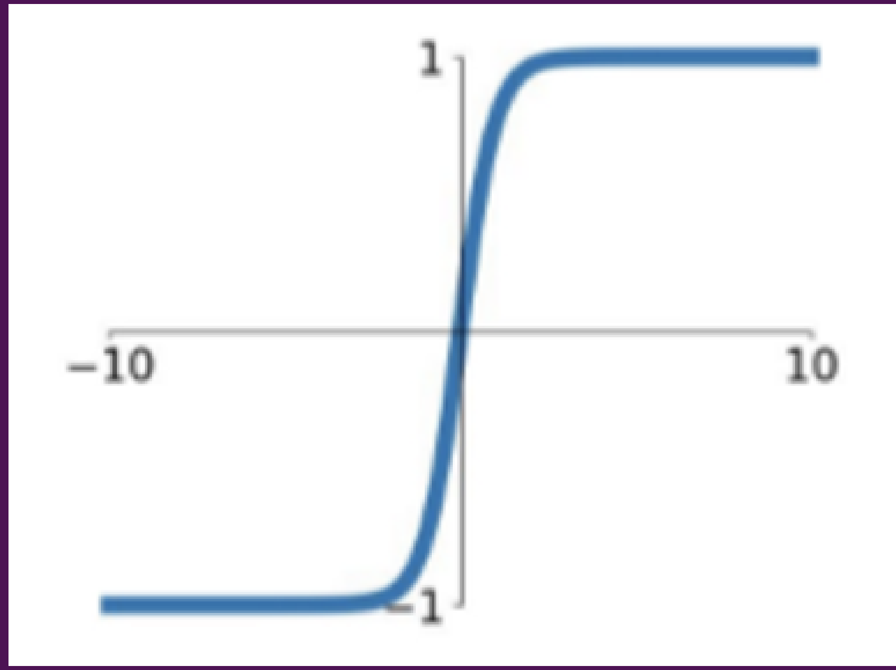
Sigmoid



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

binary classification problem

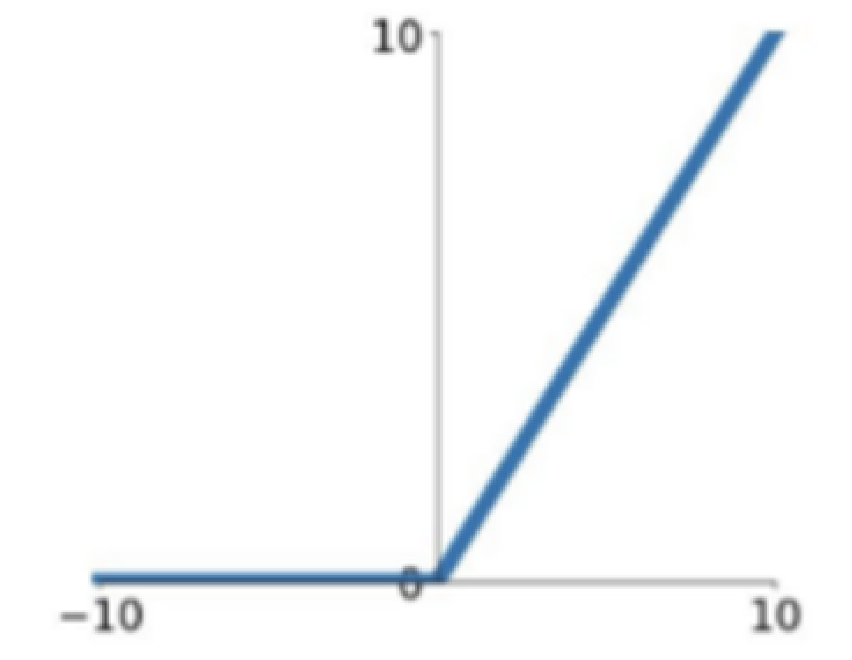
Tanh



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

Descent is faster

RELU

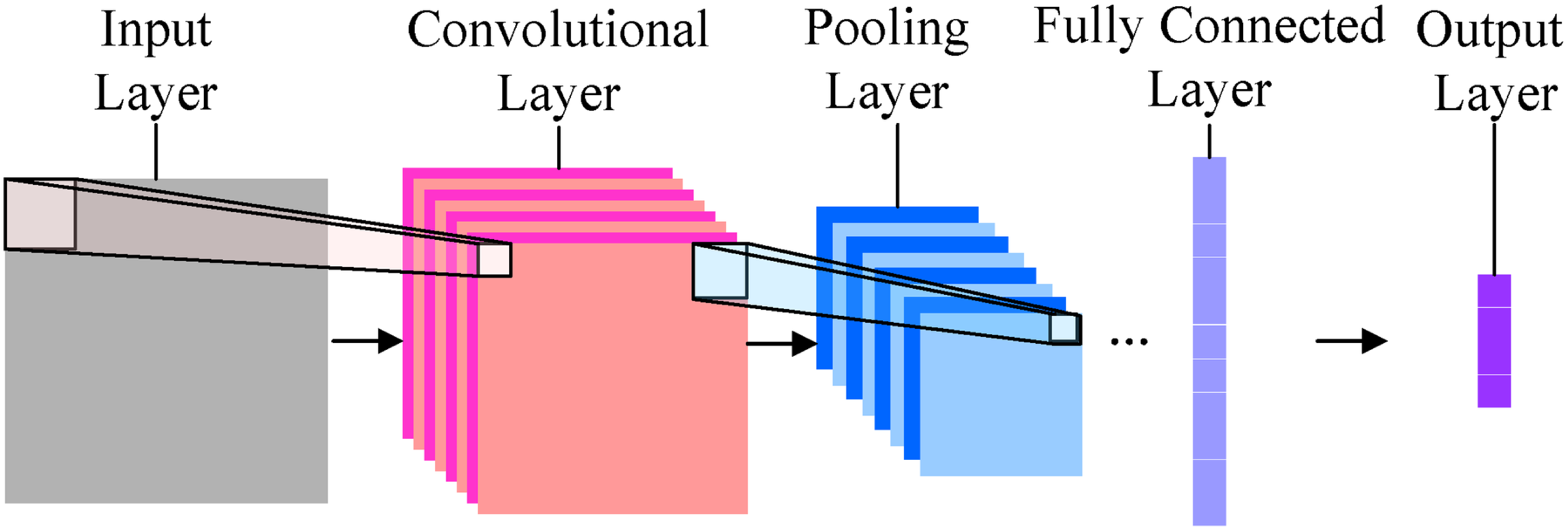


$$relu(x) = \max(0, x)$$

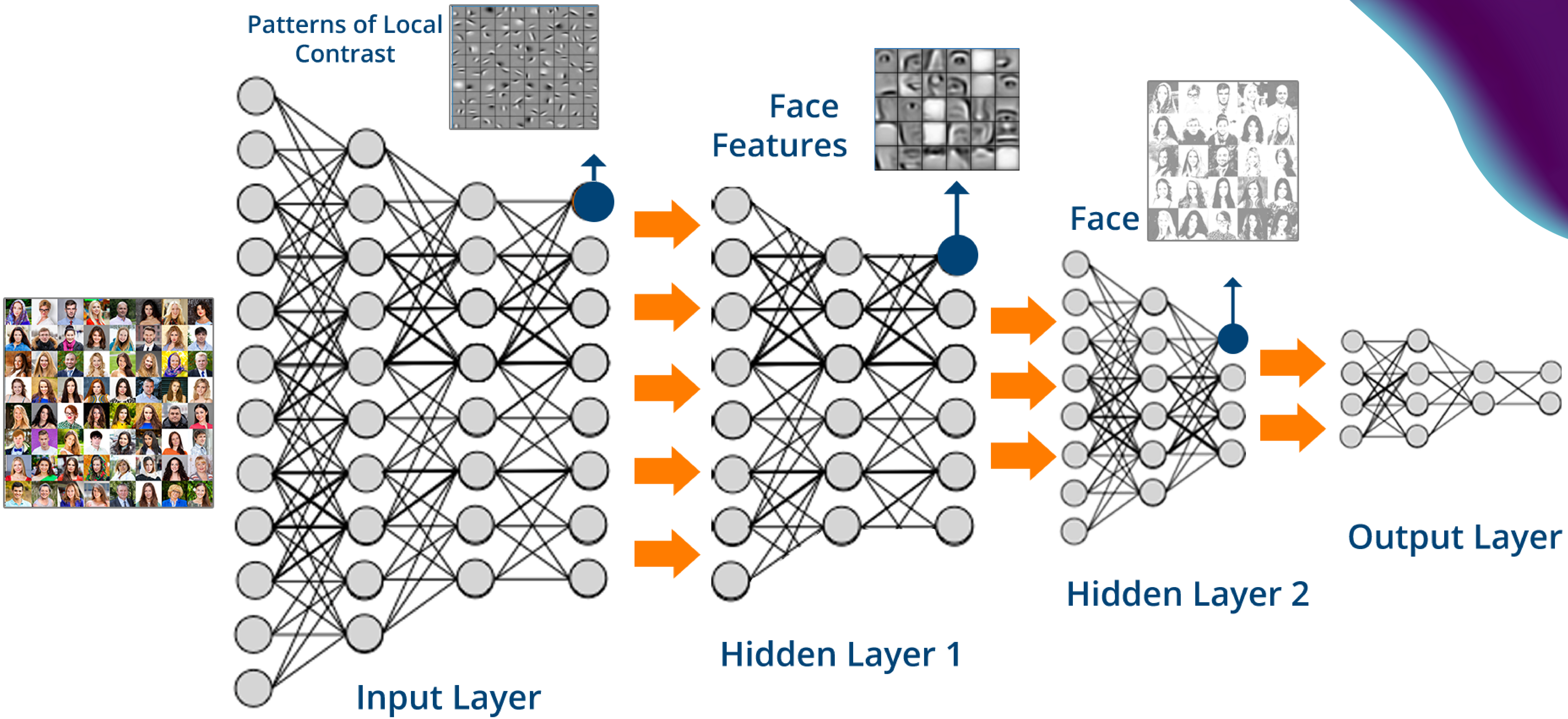
Default choice for activation

Without activation functions, this becomes a linear model!

Convolutional Neural Network

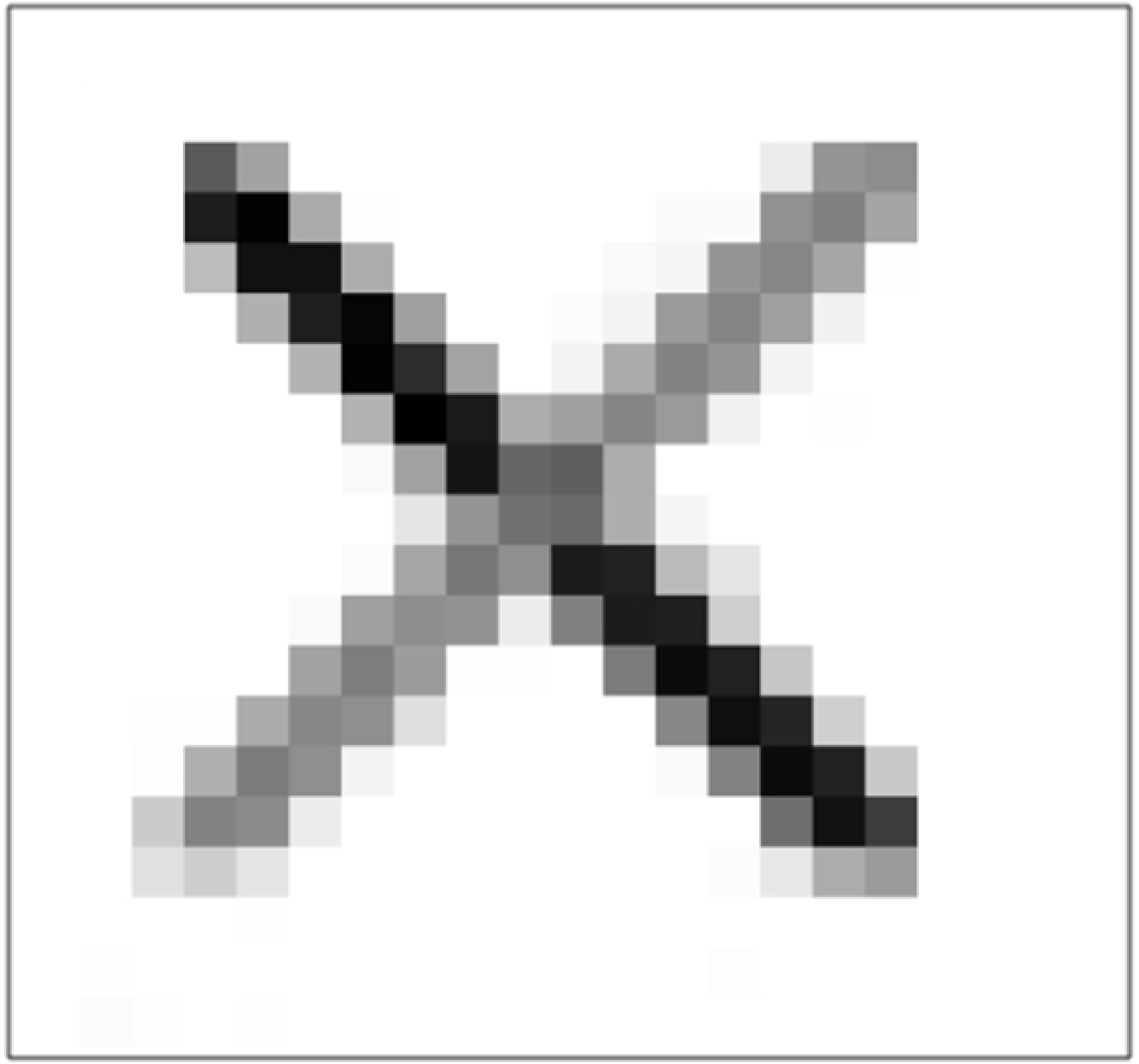


Convolutional Neural Networks (CNNs) are a category of Neural Network that have proven very effective in areas such as image recognition and classification.



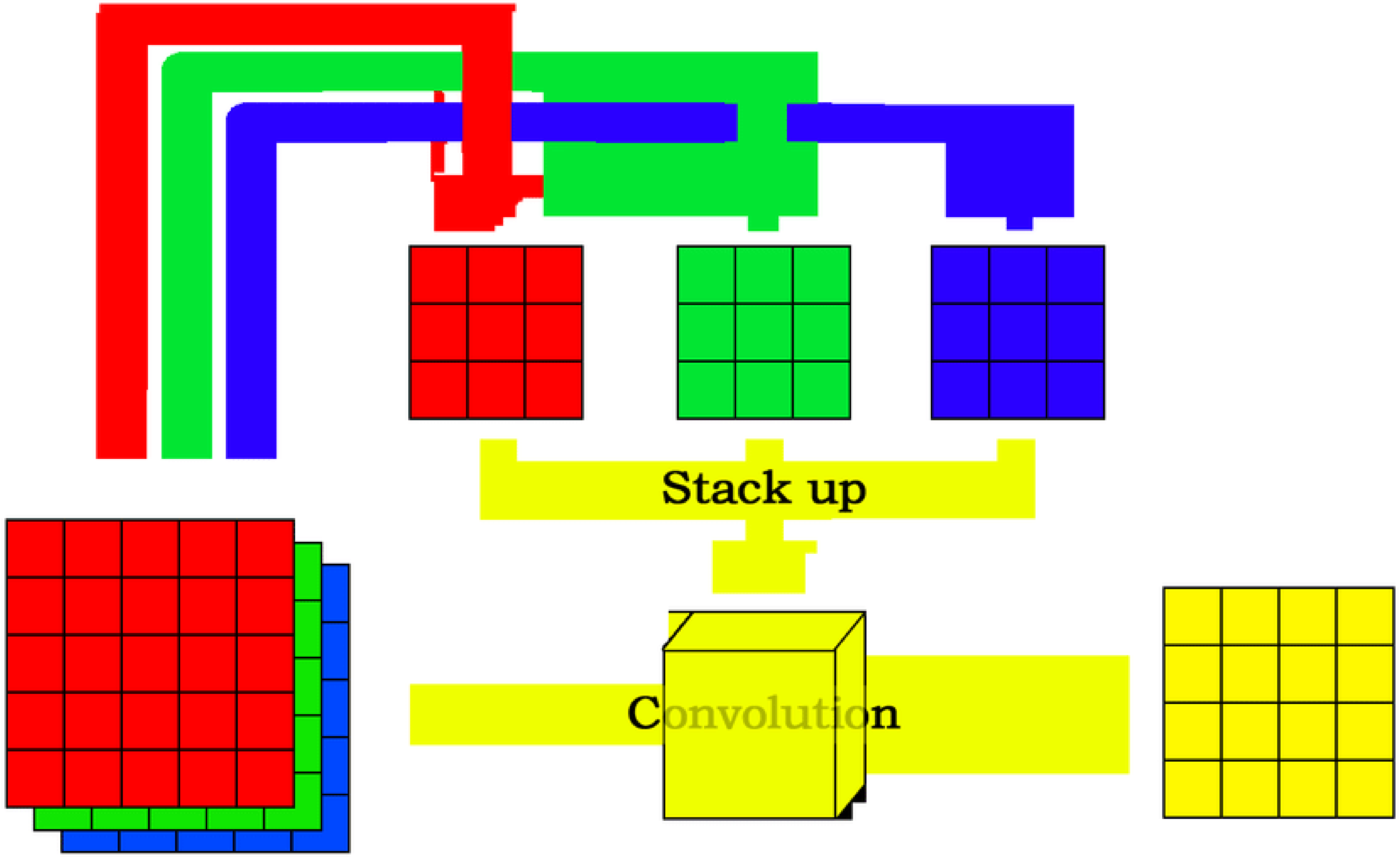
Patterns of Local Contrast
Face Features
Face
Input Layer
Hidden Layer 1
Hidden Layer 2
Output Layer

Input image



254	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255
255	255	254	255	255	255	255	255	255	254	255	255	255	255	255	255	255	255	255	255
255	255	89	163	255	255	255	255	255	255	255	255	254	236	150	142	255	255	255	255
255	255	24	1	171	253	255	255	255	255	255	250	250	146	129	166	255	255	255	255
255	255	189	10	11	173	255	255	255	255	248	245	150	135	168	252	255	255	255	255
255	255	255	177	26	2	160	255	255	251	242	156	133	160	241	255	255	255	255	255
255	255	255	255	179	1	41	164	255	243	172	132	150	242	255	255	255	255	255	255
255	255	255	255	255	179	0	22	174	161	134	155	240	255	253	254	255	255	255	255
255	255	255	255	255	249	162	15	102	95	173	255	255	255	255	254	255	255	255	255
255	255	255	255	255	255	230	150	114	106	175	244	255	255	255	255	255	255	255	255
255	255	255	255	255	251	167	120	144	23	28	187	230	255	255	255	255	255	255	255
255	255	255	255	249	161	143	147	237	129	23	27	207	255	255	255	255	255	255	255
255	255	255	254	163	126	157	252	252	255	124	6	29	199	255	255	255	255	255	255
254	253	253	172	136	144	224	255	255	255	255	136	7	32	208	255	255	255	255	255
255	252	176	124	144	243	255	255	255	255	255	248	131	5	28	201	255	255	255	255
255	204	131	141	237	255	255	255	255	255	255	255	255	110	11	59	255	255	255	255
255	225	206	230	255	255	255	255	255	255	255	255	251	232	173	156	255	255	255	255
254	255	255	253	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255
253	255	255	255	255	255	255	255	255	255	255	255	252	255	255	255	255	255	255	255
251	252	255	253	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255

ESSAI MACHINE LEARNING CLUB

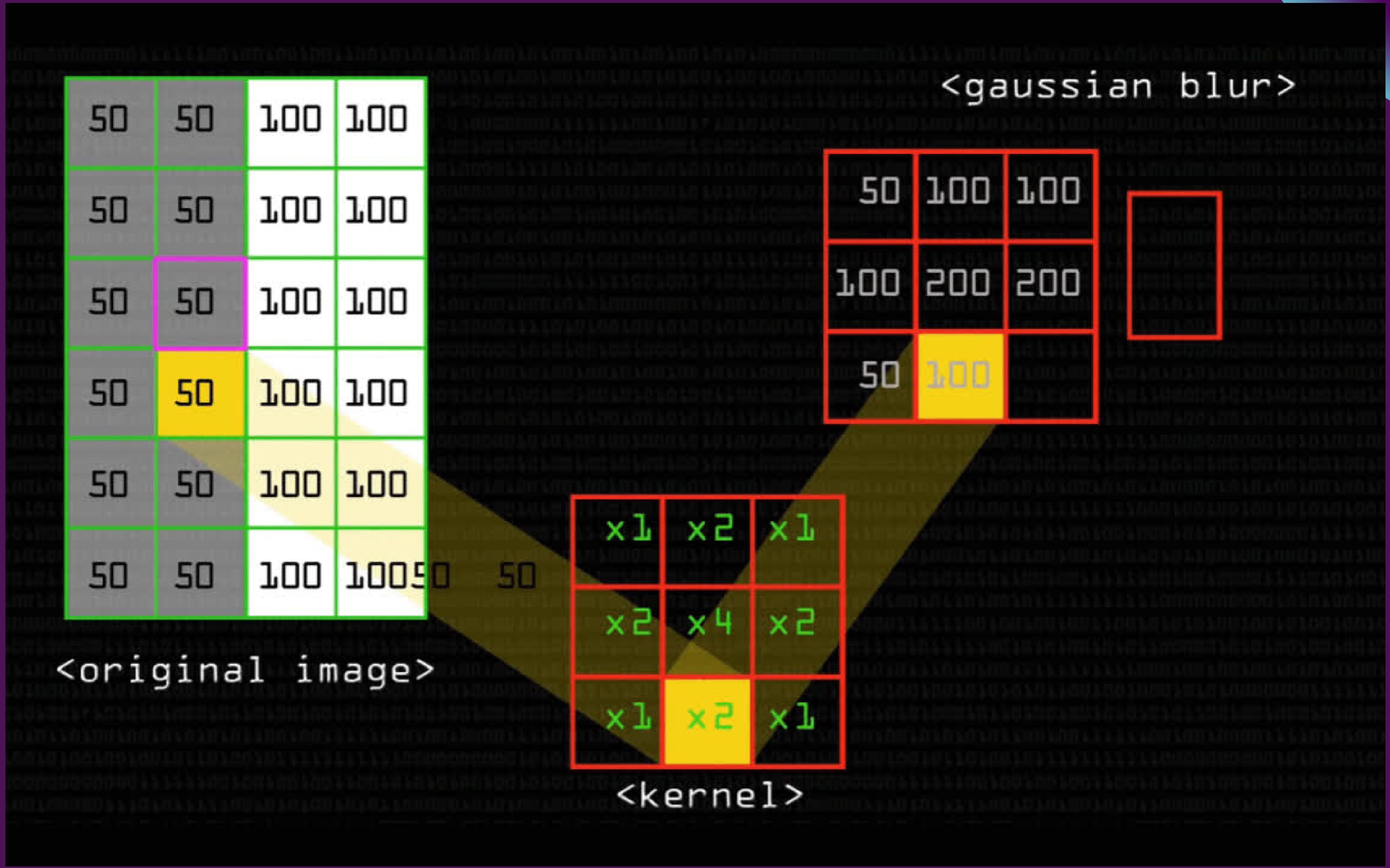


[redacted] [redacted] [redacted]

ESSAI MACHINE LEARNING CLUB

Convolutional layers

Convolutional layers are a type of neural network layer primarily used in deep learning models forprocessing data that has a grid-like topology, such as images. These layers perform a mathematicaloperation called convolution



Convolutional layers

- **Convolutional Layer**: The primary purpose of Convolution is to extract features from the input image. Convolution preserves the spatial relationship between input by learning input features.

  Convolution layer output y = $\sum_{i}^{n} x \; * f$

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & -1 & 1 & 1 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 2 \\ 0 & 2 & 1 \\ -2 & 1 & 0 \end{bmatrix}$$

Input Image(x) (nxn)

Filter (fxf)

Output Image (y) (n-f+1) x (n-f+1)

Convolutional layers

- Kernel is nothing but a filter used to extract the features from the images

 - The kernel is a matrix that moves over the input data, performs the dot product with the sub-region of input data, and etst e output as the matrix of dot products.
- Kernel moves on the input data by the stride value

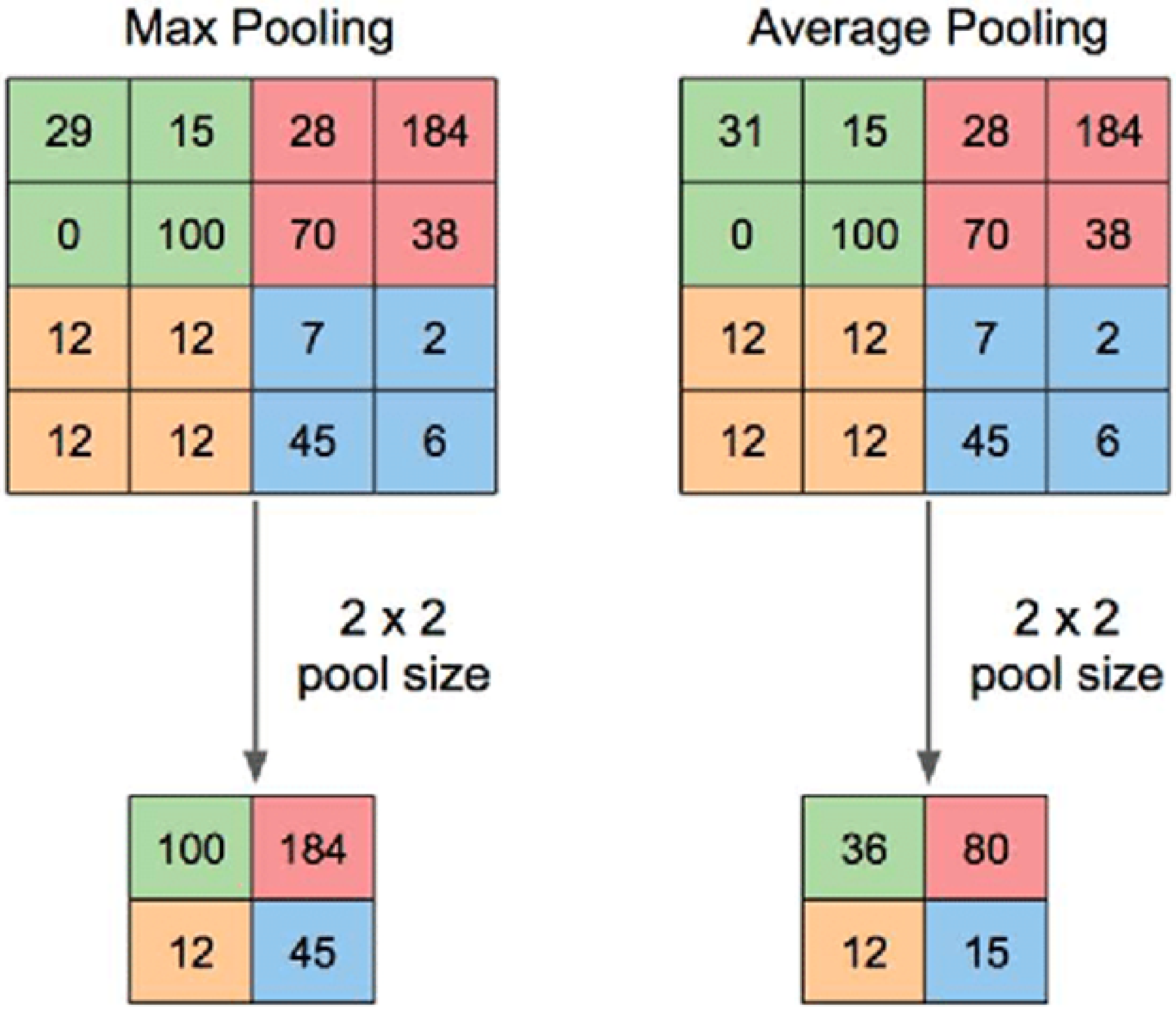
- Strides: The number of pixels shifts over the input matrix by the filter. When the stride is 1 then we move the filters to 1 pixel at a time.

Reference: https://www.quora.com/How-are-convolutional-filters-kernels-initialized-and-learned-in-a-convolutional-neural-network-CNN

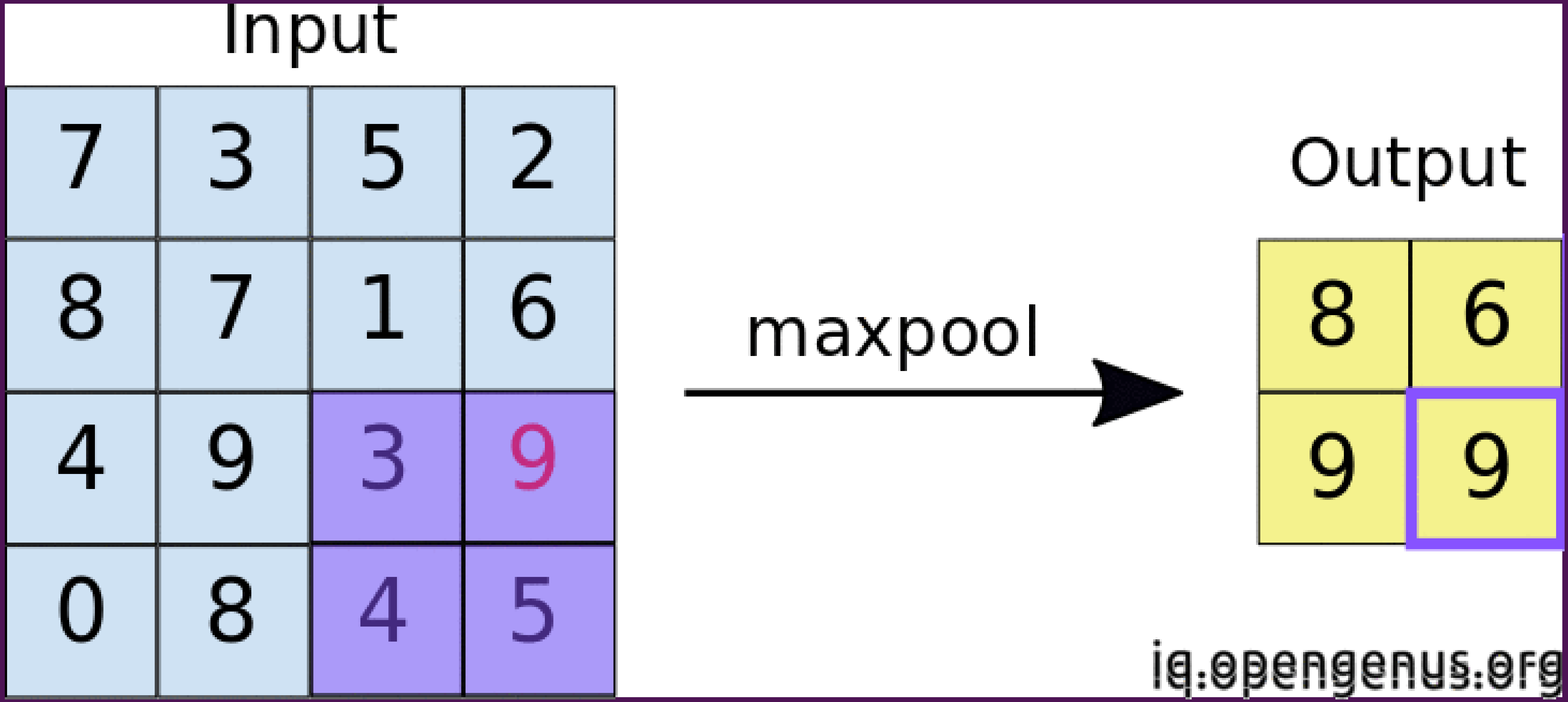
ESSAI MACHINE LEARNING CLUB

Pooling

Pooling layers would reduce the number of parameters when the inputs are too large.
Pooling also called down sampling which reduces the dimensionality of each map but retains important information.
There are three types of pooling namely, Max Pooling, Average Pooling, Sum Pooling.



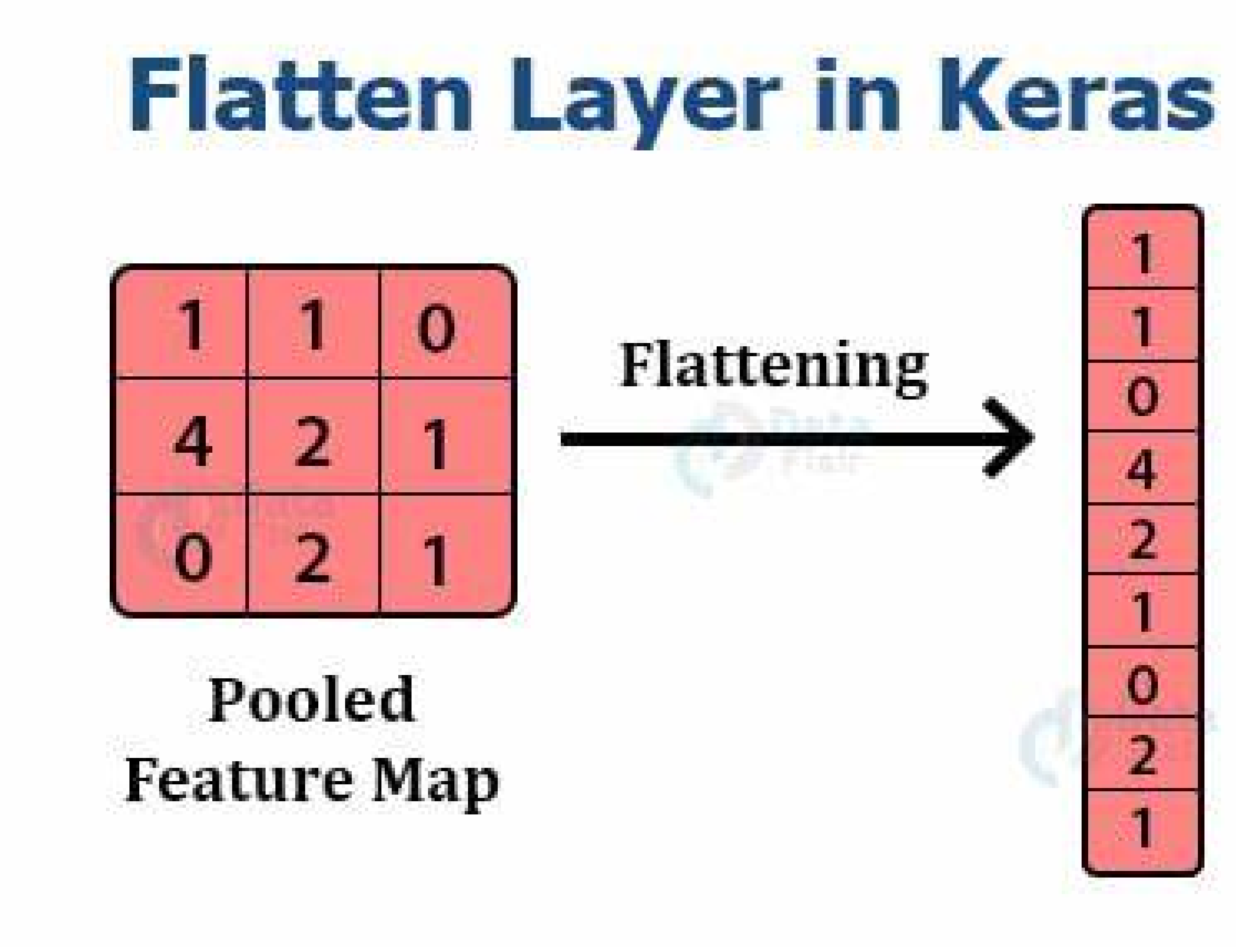
Pooling



Input
7
3
5
2
8
7
1
6
4
9
3
9
0
8
4
5
maxpool
Output
8
6
9
9
iq.opengenus.org

Flatten Layer

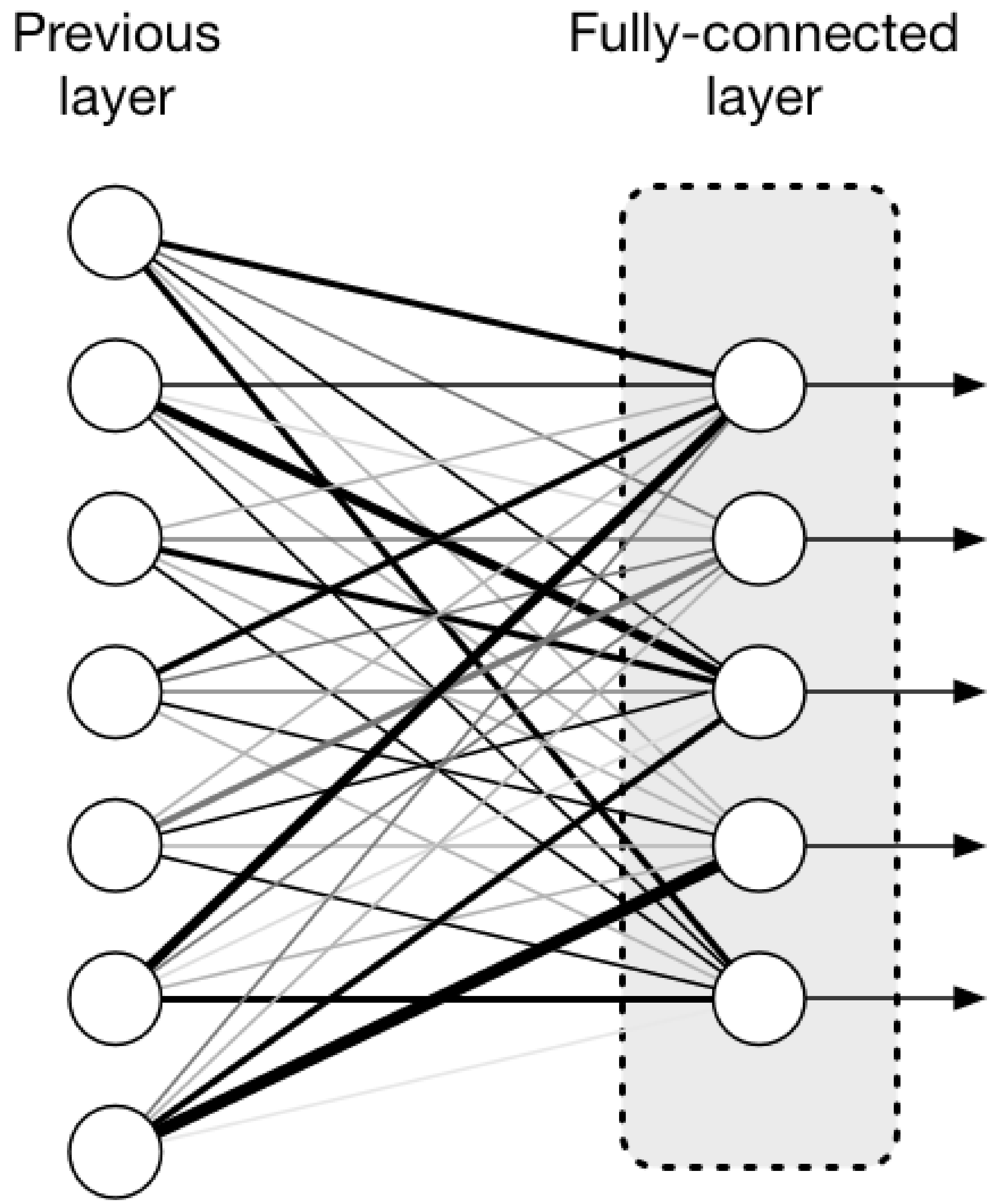
The Flatten layer is introduced to convert the multi-dimensional tensor into a one-dimensional array. It essentially unravels or flattens the spatial structure of the data while retaining the learned features.



Fully connected layer

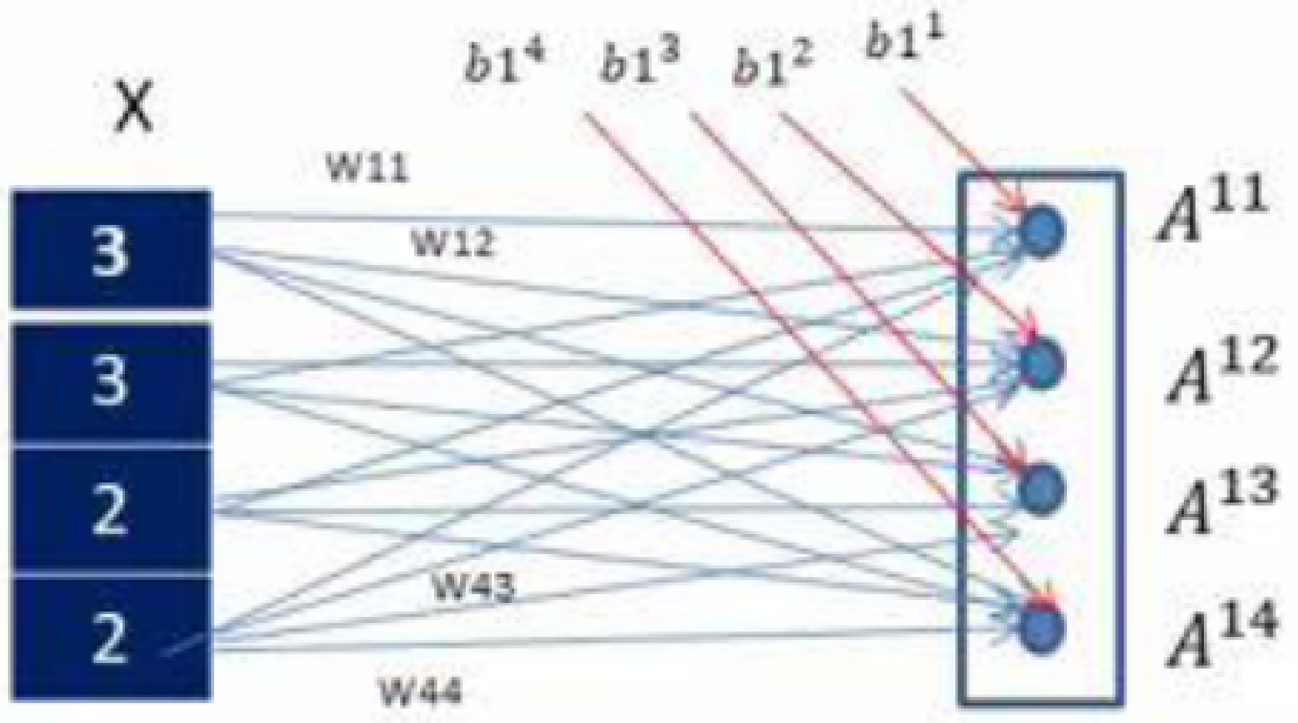
Dense (Fully Connected) Layers are typical in the later stages of a CNN. These layers can learn complex relationships across different parts of the image and produce the final output of the network

These layers are responsible for making predictions based on the high-level features learned by the convolutional layers



Fully connected layer

Fully Connected layer with ReLU

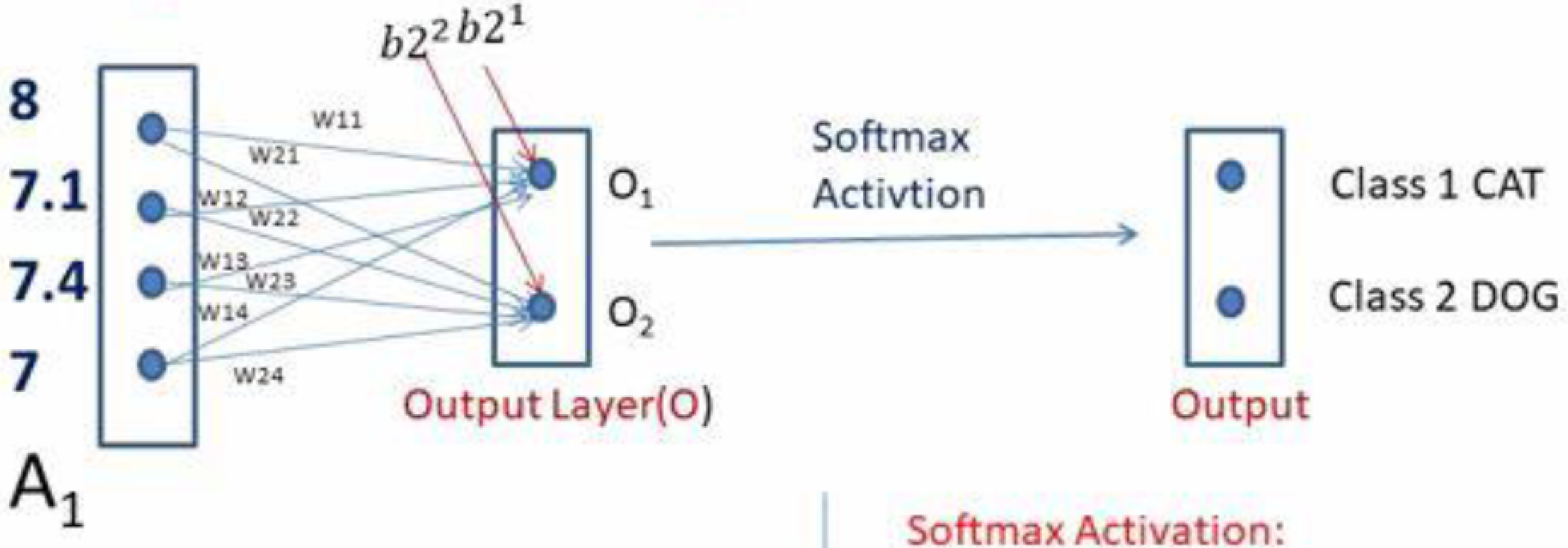


$$A_1 = W_1 X + b_1$$

$$\begin{bmatrix} A11 \\ A12 \\ A13 \\ A14 \end{bmatrix} = \begin{bmatrix} 1.0 & 1.0 & 0.2 & 0.8 \\ 1.0 & 0.5 & 0.5 & 0.8 \\ 0.8 & 1.0 & 0.2 & 0.8 \\ 0.5 & 0.5 & 1.0 & 1.0 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 8 \\ 7.1 \\ 7.4 \\ 7 \end{bmatrix}$$

$$\text{ReLU}\begin{bmatrix} 8 \\ 7.1 \\ 7.4 \\ 7 \end{bmatrix} = \begin{bmatrix} 8 \\ 7.1 \\ 7.4 \\ 7 \end{bmatrix}$$

Fully Connected layer & Softmax Activation



$$O = W_2 A_1 + b_2$$

$$\begin{bmatrix} O_1 \\ O_2 \end{bmatrix} = \begin{bmatrix} 0.2 & 0.1 & 0.5 & 0.5 \\ 0.1 & 0.5 & 0.1 & 0.2 \end{bmatrix} \begin{bmatrix} 8 \\ 7.1 \\ 7.4 \\ 7 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}$$

$$= \begin{bmatrix} 9.61 \\ 6.49 \end{bmatrix}$$

Softmax Activation:

$$\sigma(O_i) = \frac{e^{O_i}}{\sum e^{O_i}}$$

$$\sigma(O1) = \frac{e^{O1}}{e^{O1}+e^{O2}} = \frac{14913.17}{14913.17+658.52} = 0.95$$

$$\sigma(O2) = \frac{e^{O2}}{e^{O1}+e^{O2}} = \frac{658.52}{14913.17+658.52} = 0.04$$

Transfer Learning : Transfer learning involves taking a pre-trained model (usually trained on a large dataset) and using it as a starting point for a new task or dataset.

examples

VGG (Visual Geometry Group) ResNet (Residual Network) MobileNet DenseNet EfficientNet

Fine-tuning : Fine-tuning is a specific form of transfer learning where instead of keeping all the pre-trained model parameters frozen, you unfreeze some or all of the layers and allow them to be further trained on the new dataset.

ESSAI MACHINE LEARNING CLUB

CS231n Convolutional Neural Networks for Visual Recognition

Course materials and notes for Stanford class CS231n: Convolutional Neural Networks for Visual Recognition.

github.io

A Comprehensive Survey on Transfer Learning

Transfer learning aims at improving the performance of target learners on target domains by transferring the knowledge contained in different but related source domains. In this way, the...

arXiv.org

What is the difference between Transfer Learning vs Fine Tuning...

I finished my undergraduate thesis in transfer learning, keeping the most of the CNN layer freeze...

ResearchGate

TensorFlow

Keras

PyTorch

ESSAI MACHINE LEARNING CLUB

Deep Learning Aproach For Safty Roads

01 Introduction to Computer Vision

Yahia Chammami

ESSAI MACHINE LEARNING CLUB