Machine Learning

Unit 2: Supervised Learning

Prof. Priti Srinivas Sajja

P G Department of Computer Science,

Sardar Patel University, Vallabh Vidyanagar, Gujarat, India

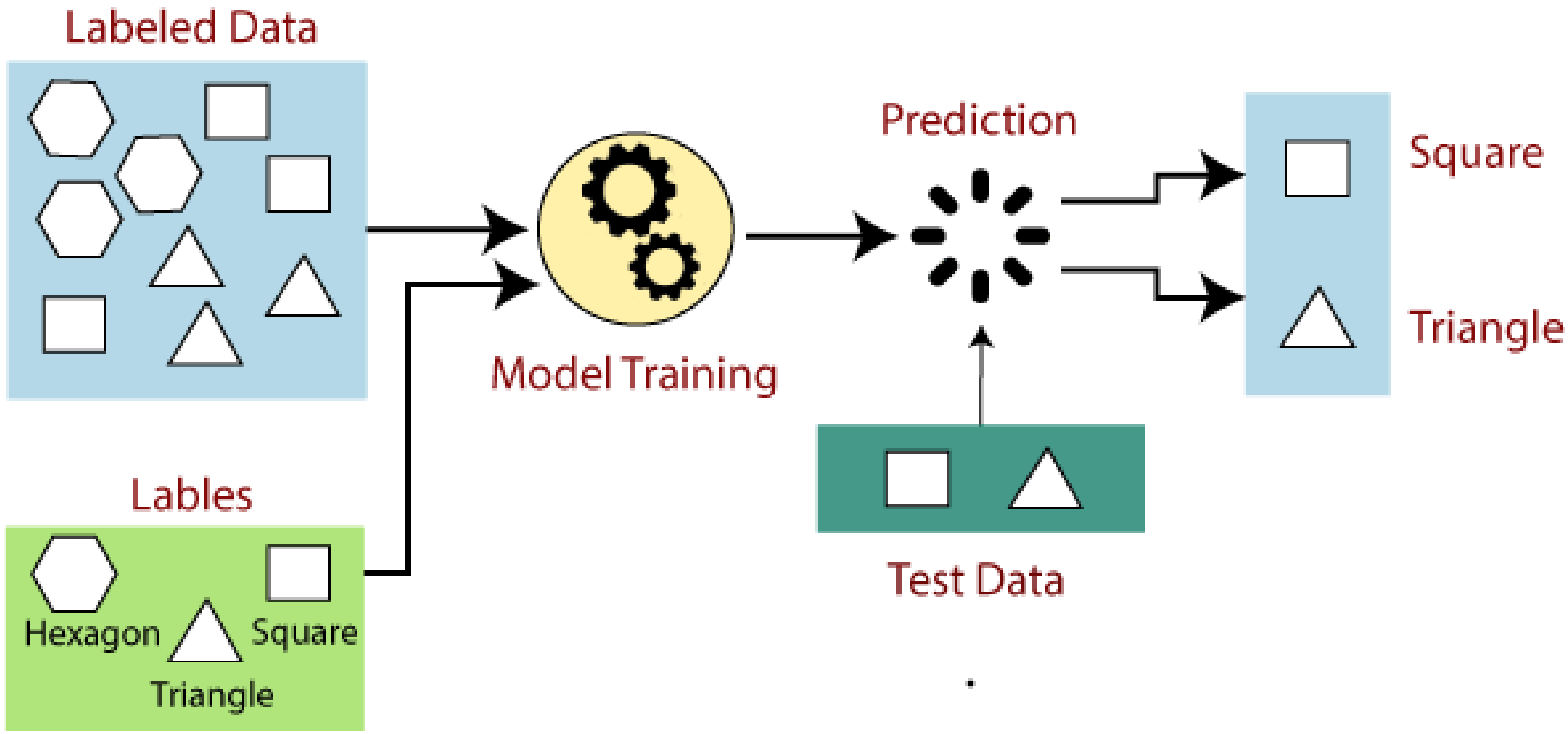
About the Speaker

- **Name: Dr. Priti Srinivas Sajja**
- **Email : priti@pritisajja.info**
- **Mobile : +91 98249 26020**
- **URL: http://pritisajja.info**
- **Academic qualifications : Ph. D in Computer Science**
- **Thesis Title: Knowledge-Based Systems for Socio-Economic Rural Development**
- **Subject area of specialization : Artificial Intelligence**
- **Publications : 219 in Books, Book Chapters, Journals, Proceedings, etc.**

Supervised Learning

- Supervised learning is the types of machine learning in which **machines are trained using well "labelled" training data**, and on basis of that data, machines predict the output.
- The labelled data means some input data is already **tagged with the correct output**.
- In supervised learning, **the training data** provided to the machines **work as the supervisor** that **teaches the machines to predict the output correctly**. It applies the same concept as a student learns in the supervision of the teacher.
- Supervised learning is a process of providing input data as well as correct output data to the machine learning model.
- The aim of a supervised learning algorithm is to **find a mapping function to map the input variable(x) with the output variable(y)**.

Supervised Learning: Example 1



Supervised Learning – Major Steps

- First Determine **the type/domain of training dataset and Collect/Gather the labelled** training data.
- Split the training dataset into training **dataset, training dataset, and test/validation dataset.**
- Determine the **input features** of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
- Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
- **Execute** the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
- **Evaluate the accuracy** of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

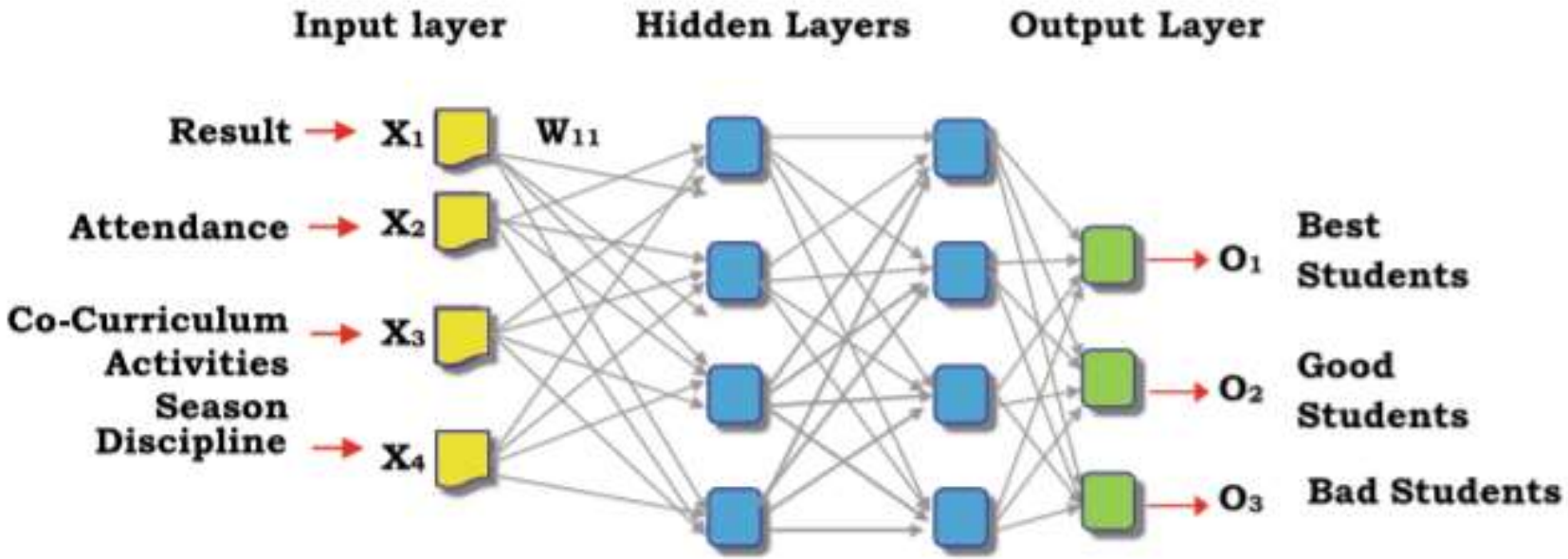
Supervised Learning – Advantages

- With the help of supervised learning, the model can **accurately predict** the output on the basis of prior experiences.
- In supervised learning, we can have an **exact idea about the classes of objects**.
- Supervised learning model helps us to solve various real-world problems such as **fraud detection, spam filtering**, etc.

Supervised Learning – Disadvantages

- Supervised learning models **are not suitable for handling the complex tasks**.
- Supervised learning cannot **predict the correct output if the test data is different from the training dataset**.
- Training required **lots of computation times**.
- In supervised learning, we need **enough knowledge about the classes** of object.

Supervised Learning: Example 2



Inputs				Outputs		
X_1	X_2	X_3	X_4	O_1	O_2	O_3
0.5	1.0	0.5	0.5	0	1	0
1.0	0.5	1.0	0.6	0	1	0
0.4	0.3	1.0	0.3	0	0	1
…	…	…	…	…	…	…

Supervised Learning: Example 3

Inputs									Output
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	Price($)
2800	4.0	0.5	2.0	1.0	0.5	0.5	1.0	1.0	935K
1380	3.0	0.0	1.0	1.0	1.0	1.0	1.0	0.0	560K
1790	3.0	1.0	3.0	0.0	0.0	1.0	0.5	0.5	600K
..	..	..	..	..	..	..	..	..	..

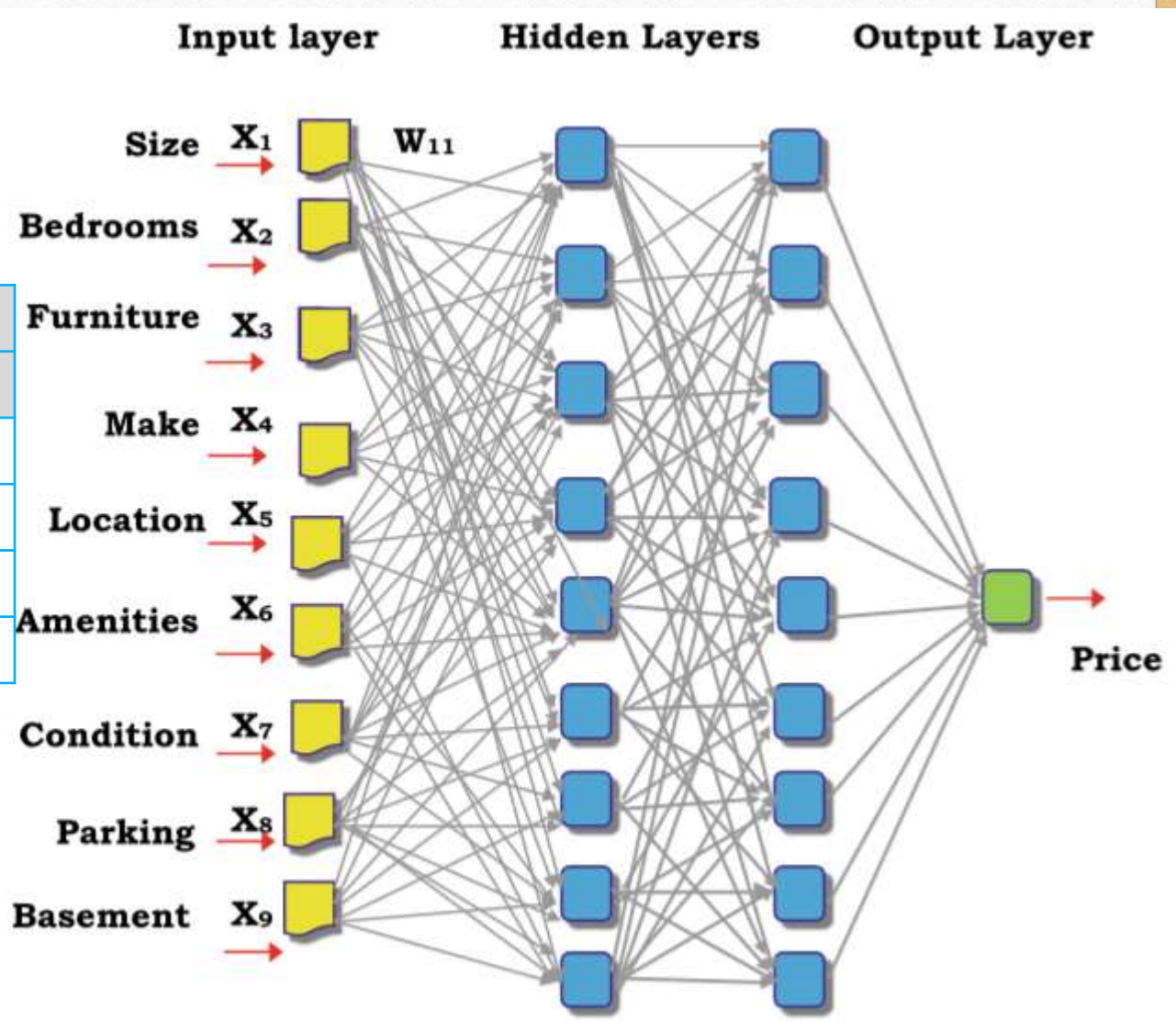


Fig. 4.13 Neural network to learn the price of a house

Supervised Learning: Example 4

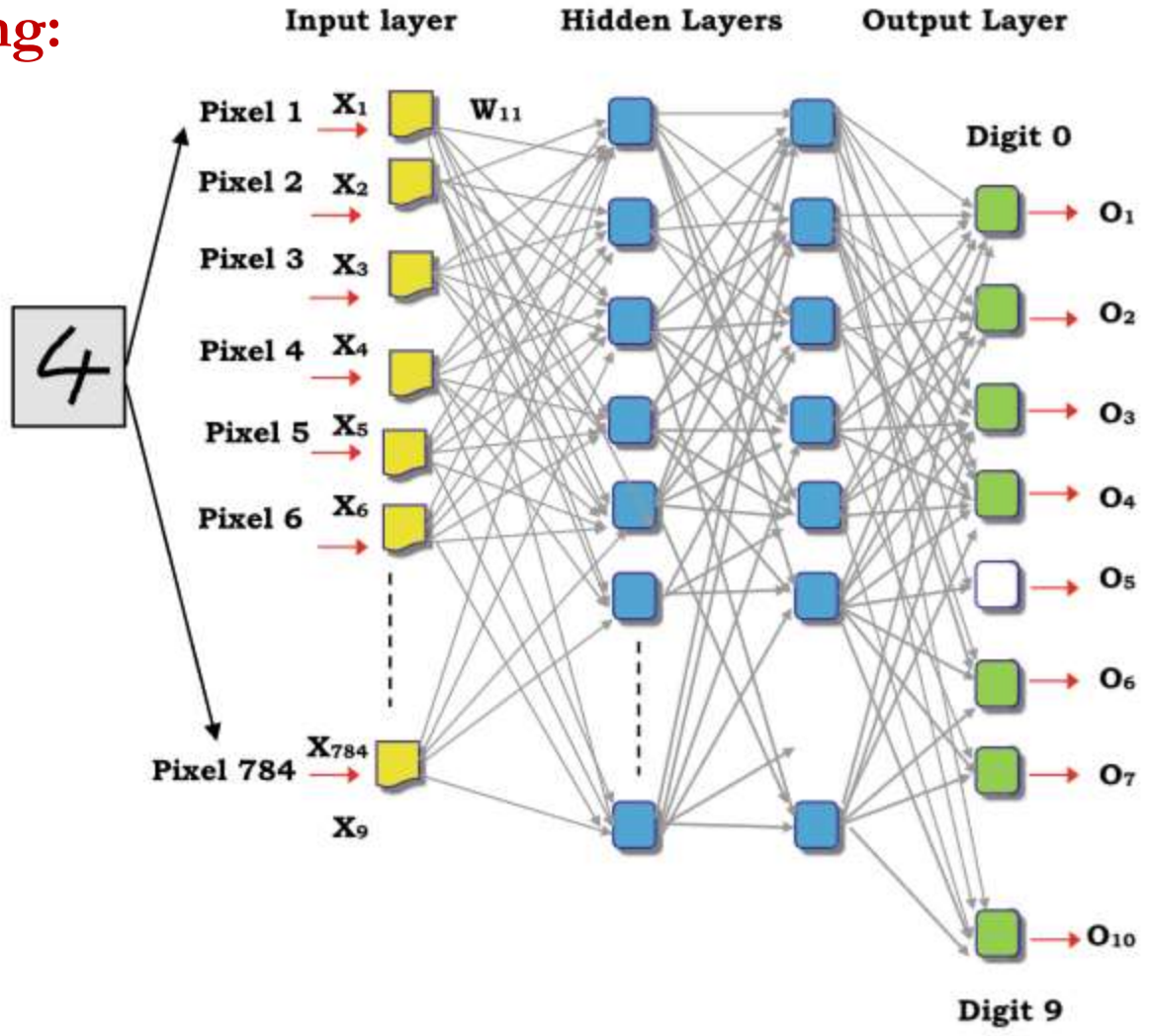
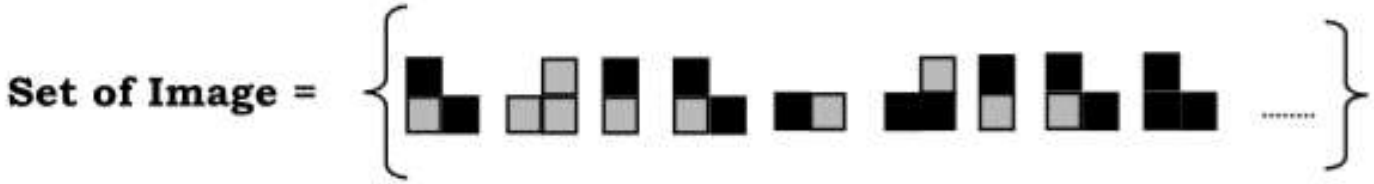
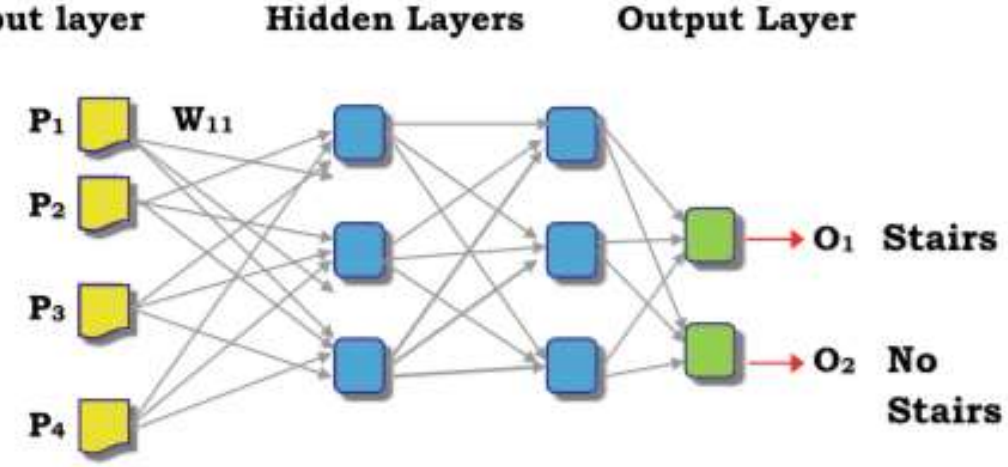


Fig. 4.14 Neural network for handwritten digits identification

Supervised Learning: Example 5





Inputs				Outputs	
P1	P2	P3	P4	Stairs	No Stairs
255	145	255	0	1	0
0	145	145	145	1	0
255	145	0	0	0	1
..	..	..	..	..	..

Classification

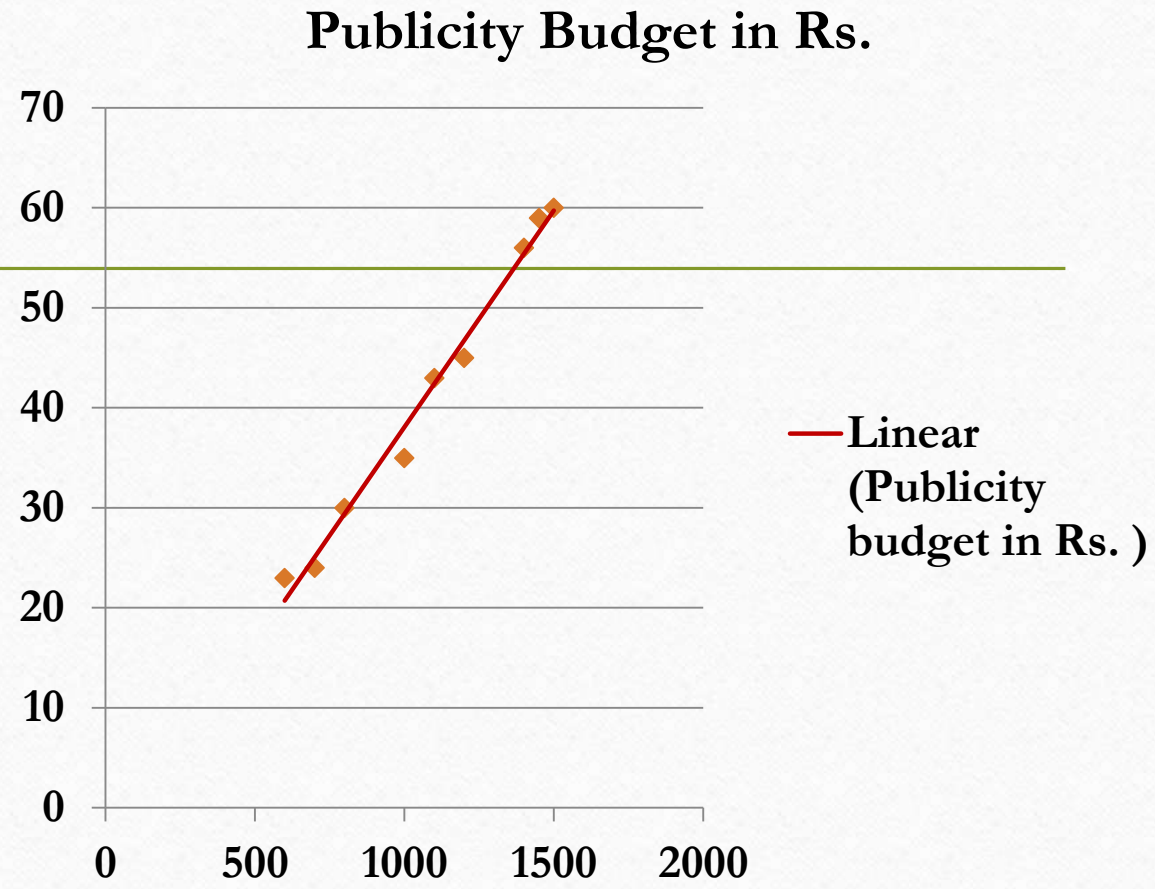
- Class labels are necessary to classify things easily
- Binary Classification and Multi Class Classification
- Examples of classification
  - Web page classification;
  - Mail Classification;
  - Image classification;
  - Students classification;
  - Etc.

Linear Regression

- Linear Regression is a kind of **supervised learning** algorithm popularly used to **estimate numerical values based on some independent** continuous variables.
- Examples can be **prediction of total annual sales** in a company, **cost of house, price of gold**, etc. Such functions are linear in nature.
- Linear regression having a form given as follows.
- $\mathbf{Y = \alpha_0 + \alpha_1 X_1}$
- The above equation shows **relation between two variables X and Y**, where **X is the independent** variable and **Y is the dependent** or output variable. Here α_0 and α_1 are known as regression coefficients.
- **Positive Relationship (Age, Experience), (Salary, Responsibility), (Power, Responsibility) etc.**
- **Negative Relationship (Cost, Affordability), (High Absence in a Class, Grades) , etc**.

Linear Regression

Year	Sales	Publicity Budget in Rs.
1	600	23
2	700	24
3	800	30
4	1000	35
5	1100	43
6	1200	45
7	1400	56
8	1450	59
9	1500	60

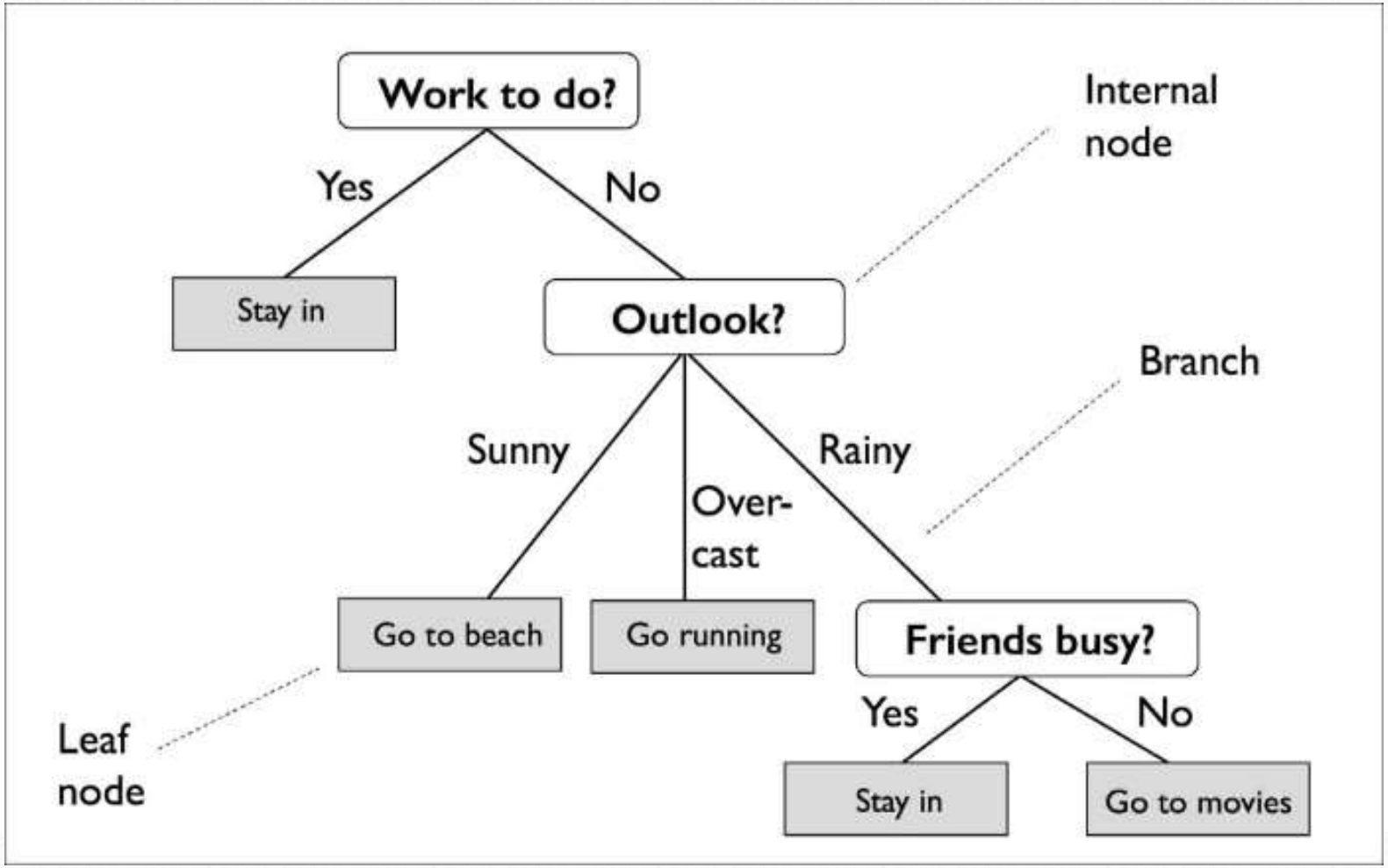


Decision Tree

- **Decision Tree consists of :**
- **Nodes** : Test for the value of a certain attribute.
- **Edges/ Branch** : Correspond to the outcome of a test and connect to the next node or leaf.
- **Leaf nodes** : Terminal nodes that predict the outcome (represent class labels or class distribution).

https://medium.com/swlh/decision-tree-classification-de64fc4d5aac

Decision Tree



https://towardsdatascience.com/https-medium-com-lorrli-classification-and-regression-analysis-with-decision-trees-c43cdbc58054

Decision Tree : Steps

- **Step-1:** Begin the tree with the root node, says S, which contains the complete dataset.
- **Step-2:** Find the best attribute in the dataset using **Attribute Selection Measure (ASM). Use Gini index or Information Gain**
- **Step-3:** Divide the S into subsets that contains possible values for the best attributes.
- **Step-4:** Generate the decision tree node, which contains the best attribute.
- **Step-5:** Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_decision_tree.htm

Decision Tree : Information Gain

- Information gain is the **measurement of changes in entropy** after the segmentation of a dataset based on an attribute.
- It calculates **how much information a feature provides us about a class**.
- **According to the value of information gain, we split the node and build the decision tree.**
- A decision tree algorithm always tries to **maximize the value of information gain**, and a **node/attribute having the highest information gain is split first**. It can be calculated using the below formula:

Decision Tree : Information Gain

- Information gain can be calculated using the below formula:
  - **Information Gain= Entropy(S)- [(Weighted Avg) *Entropy(each feature)**
- **Entropy:** Entropy is a metric to **measure the impurity in a given attribute**. It specifies randomness in data. Entropy can be calculated as:
  - **Entropy(s)= -P(yes)log2 P(yes)- P(no) log2 P(no)**

  Where,
  - **S= Total number of samples**
  - **P(yes)= probability of yes**
  - **P(no)= probability of no**

Decision Tree : Steps

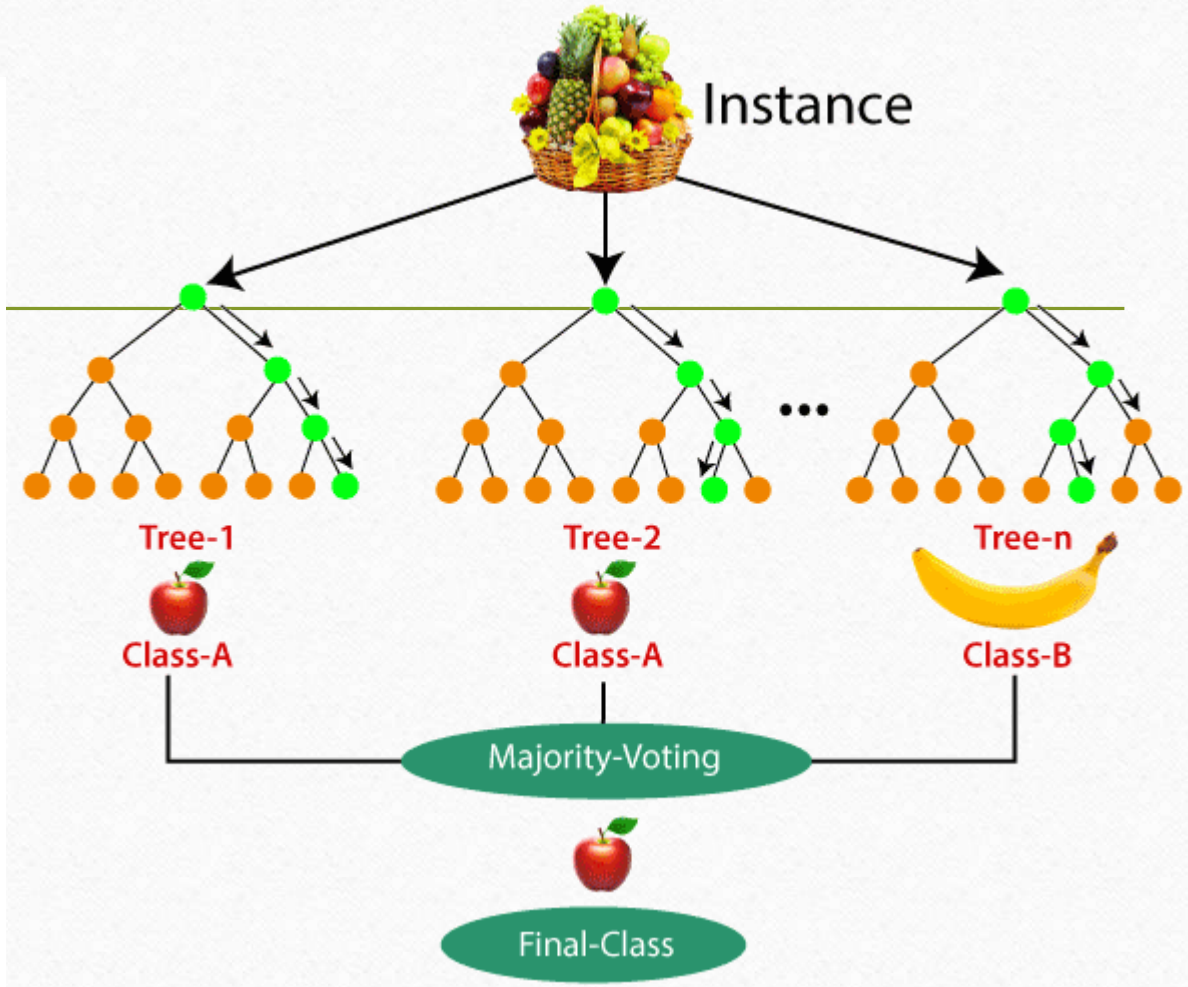
- **Refer**

- **https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_decision_tree.htm**

- For automatic generation of decision tree in Python

Random Forest

- A random forest algorithm consists of **many decision trees**.
- A random forest is a machine learning technique that's used to solve regression and classification problems.
- It utilizes **ensemble learning**, which is a technique that combines many classifiers to provide solutions to complex problems.

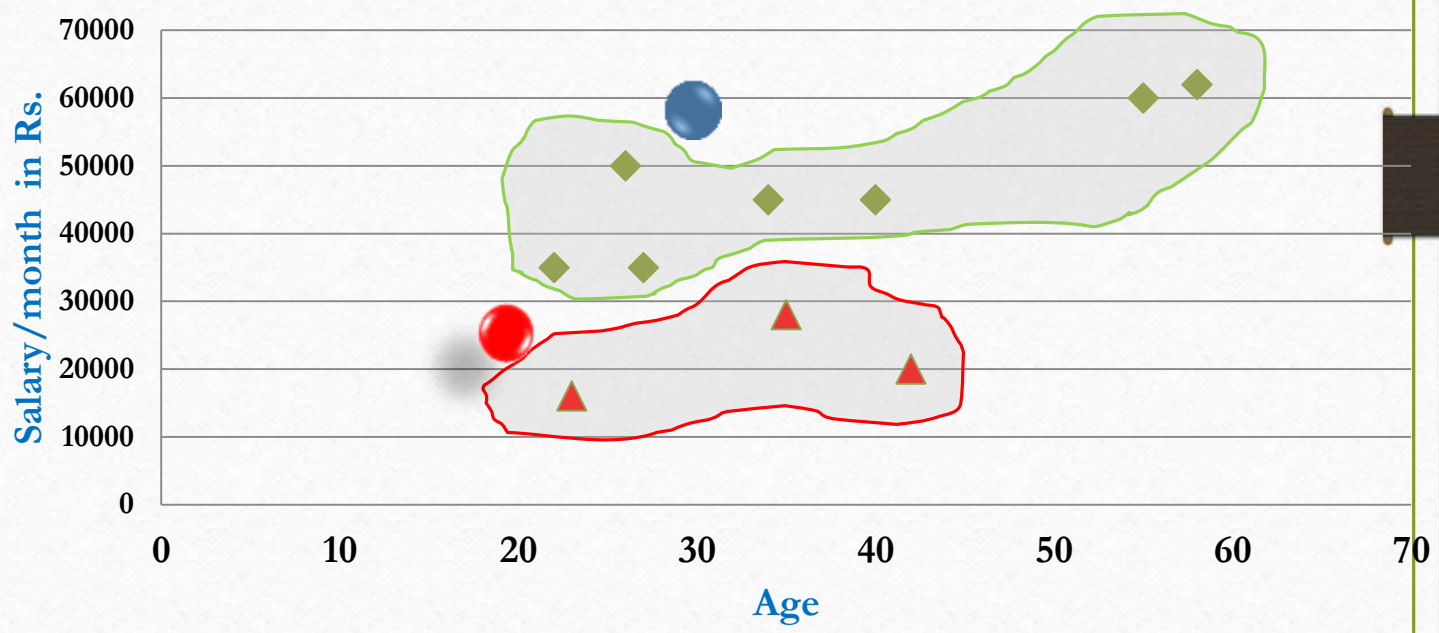


https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/

K Nearest Neighborhood: Choice of Transport Example

Sr. No.	Age	Salary / Month in Rs.	Label A: Private Car, B:Public Transport
1	22	35000	A
2	23	16000	B
3	35	28000	B
4	34	45000	A
5	40	45000	A
6	42	20000	B
7	27	35000	A
8	26	50000	A
9	55	60000	A
10	58	62000	A
11	30	60000	???
12	20	20000	???

Test Data



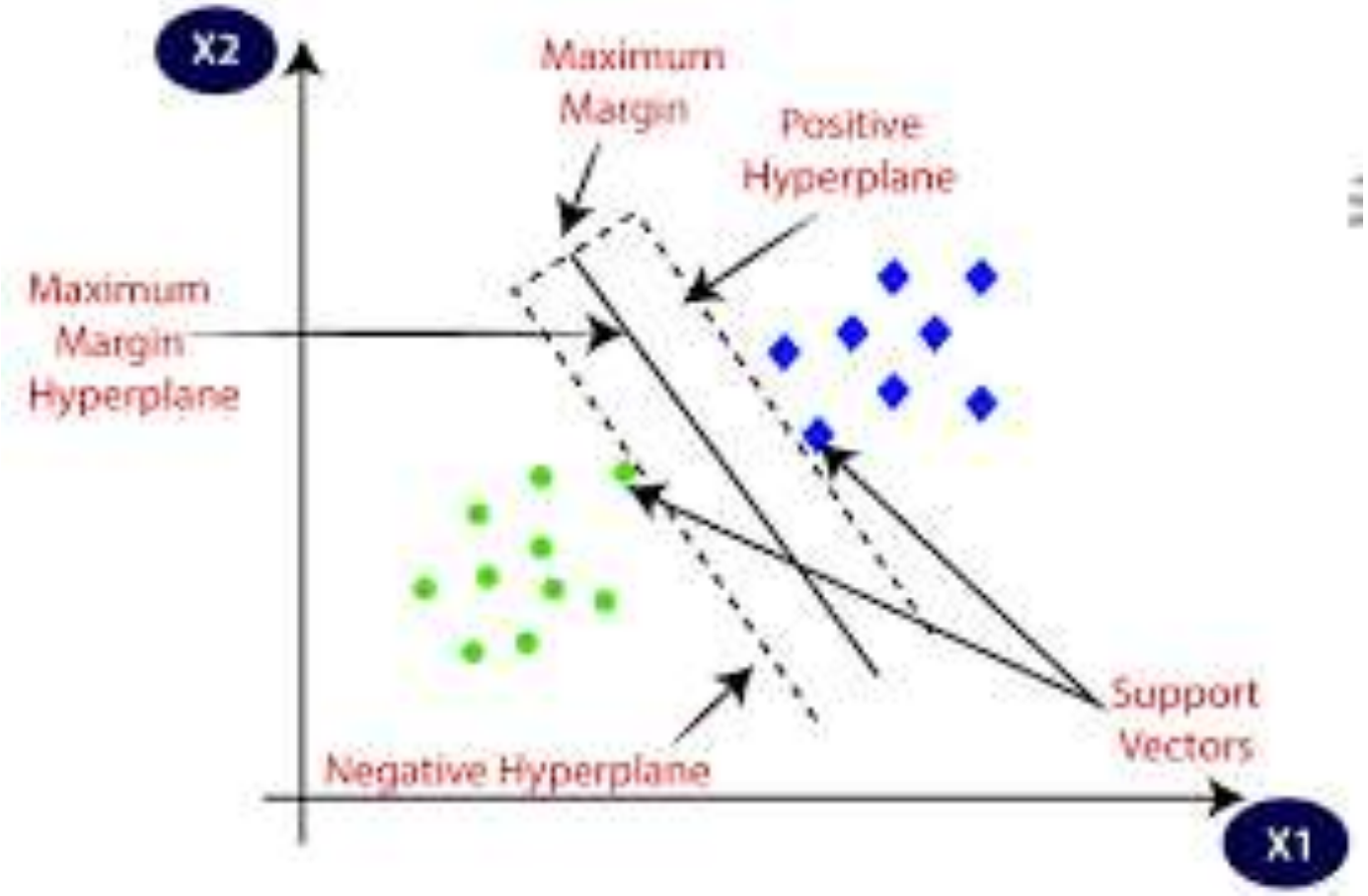
K Nearest Neighbourhood Steps

- Load the data and Initialize the value of k
- For getting the predicted class, iterate from 1 to total number of training data points
  - Calculate the **distance between test data and each row of training data using Euclidean distance** as our distance metric since it's the most popular method. The other metrics that can be used are Chebyshev, cosine, etc.
  - Sort the calculated distances in ascending order based on distance values
  - Get top k rows from the sorted array
  - Get the most frequent class of these rows
  - Return the predicted class

Support Vector Machine

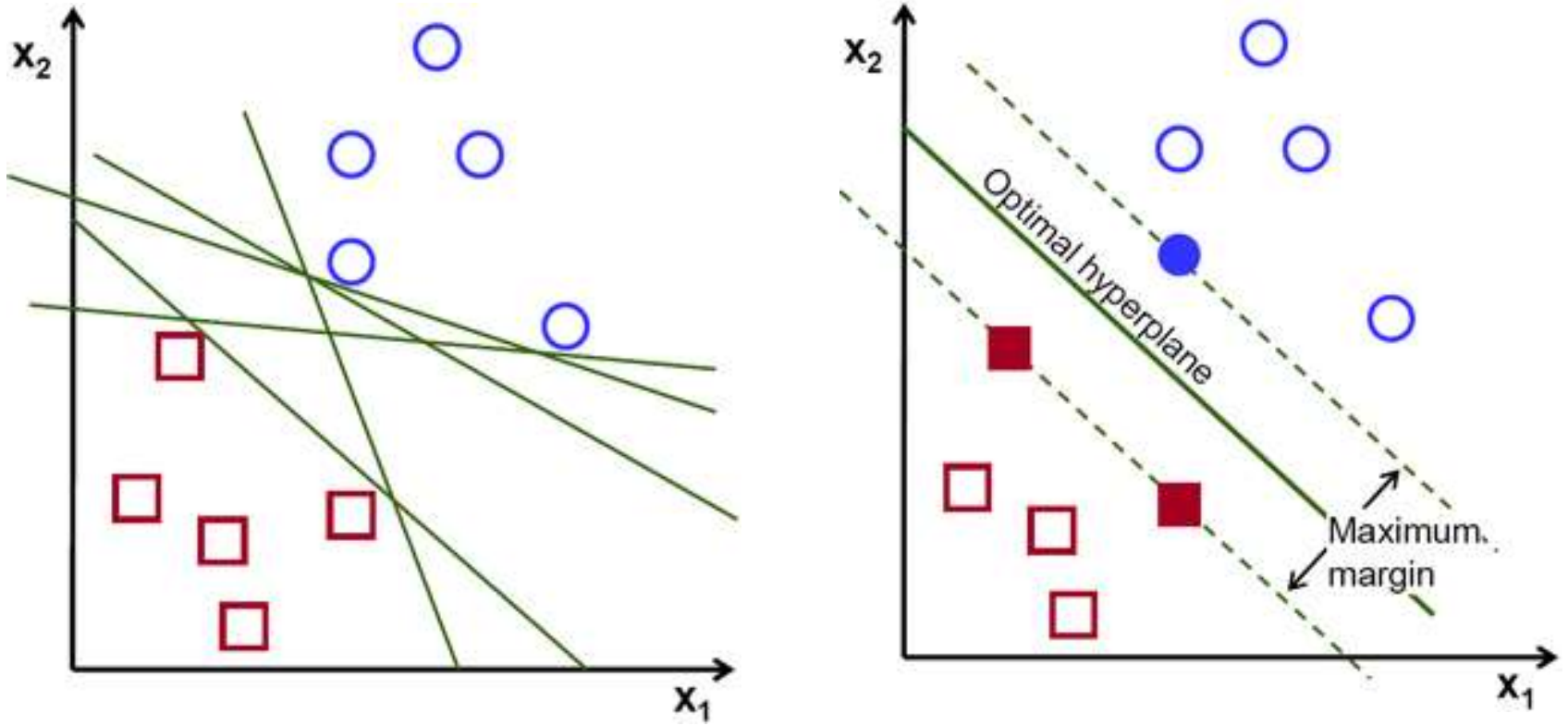
- The technique **generates multiple hyper planes** to **divide the data points** into various classes.
- The **objective is to choose the ideal hyper plane** that classifies the data sets accurately.
- Once the multiple possible hyper planes are created, the one hyper plane is chosen **which is having maximum distance from the nearest data sets**.
- To identify an efficient hyper plane that divides the data into clear and accurate classes is the basic objective of the support vector machine. **For this, support vectors are used.**
- **The support vectors are the vectors that are nearest to the hyper plane.**
- The **margin** between the support vectors and the hyper plane **needs to be as large as possible** for the effective classification.

Support Vector Machine



https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

Support Vector Machine



https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

Deep Learning

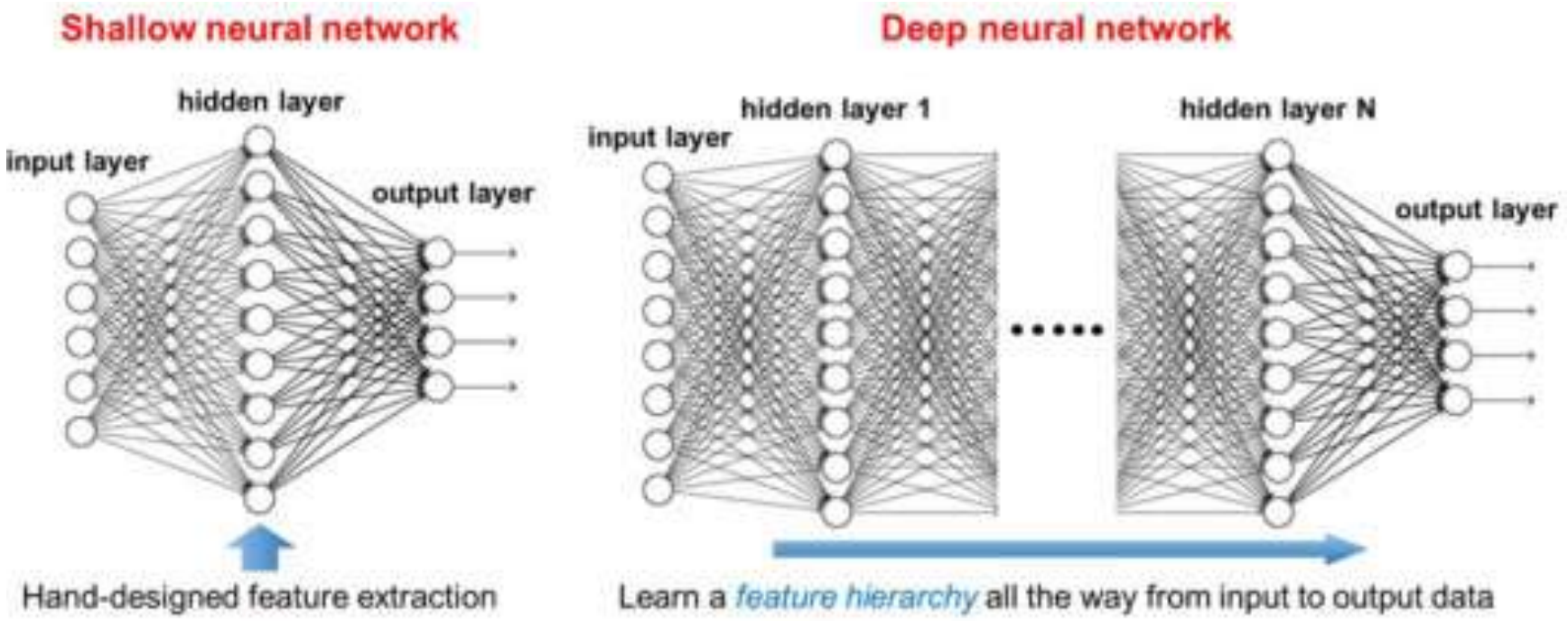
- Deep learning is a **subset of artificial intelligence**.
- Since neural networks imitate the human brain and so deep learning will do.
- In deep learning, **nothing is programmed** explicitly.
- The output from each **preceding layer is taken as input by each one of the successive layers**.
- Deep learning models are capable enough **to focus on the accurate features themselves** by requiring a little guidance from the programmer and are very helpful in solving out the problem of dimensionality.
- Deep learning algorithms are used, especially when we have **a huge no of inputs and outputs**.
- Deep learning is **implemented with the help of Neural Networks**, and the idea behind the motivation of Neural Network is the biological neurons, which is nothing but a brain cell.

Deep Learning Architectures

- ***Deep Neural Network***
- ***Deep Belief Networks***
- ***Recurrent Neural Networks***
- **Generative Adversarial Networks - GANs**
- **Convolutional Deep Neural Networks - CNNs**
- Etc.

Deep Learning Architectures

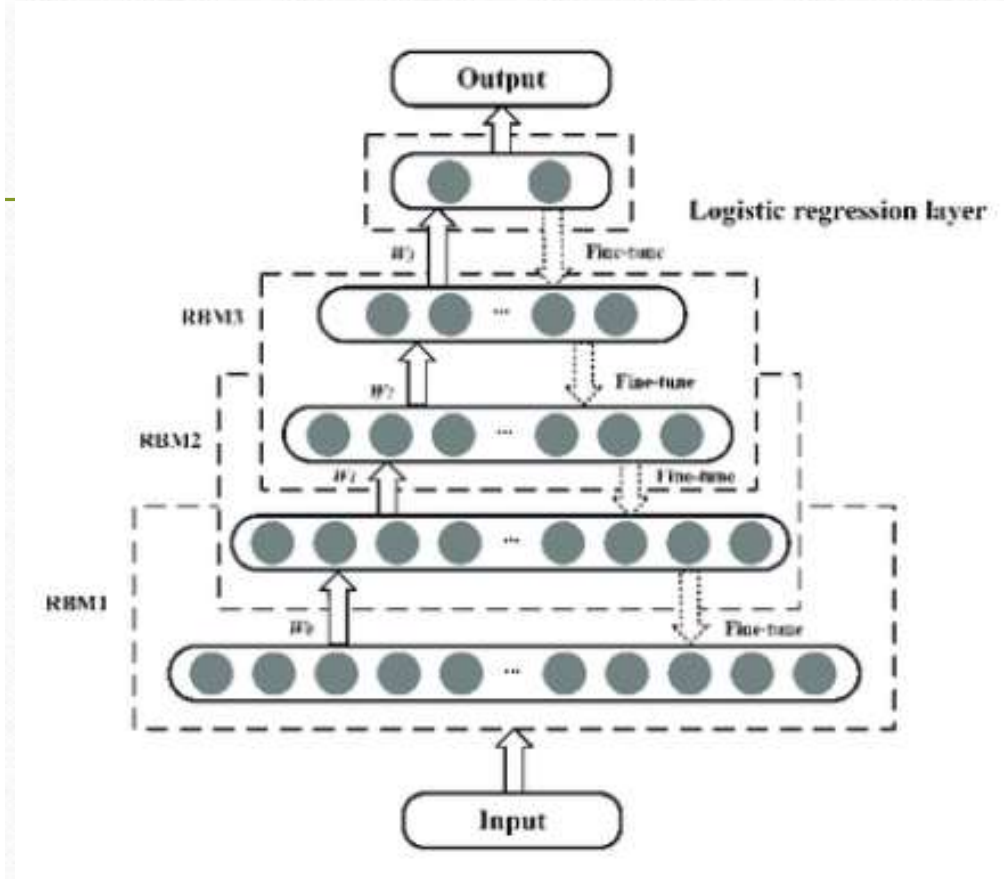
- ***Deep Neural Network***



ttps://www.i2tutorials.com/explain-deep-neural-network-and-shallow-neural-networks/

Deep Learning Architectures

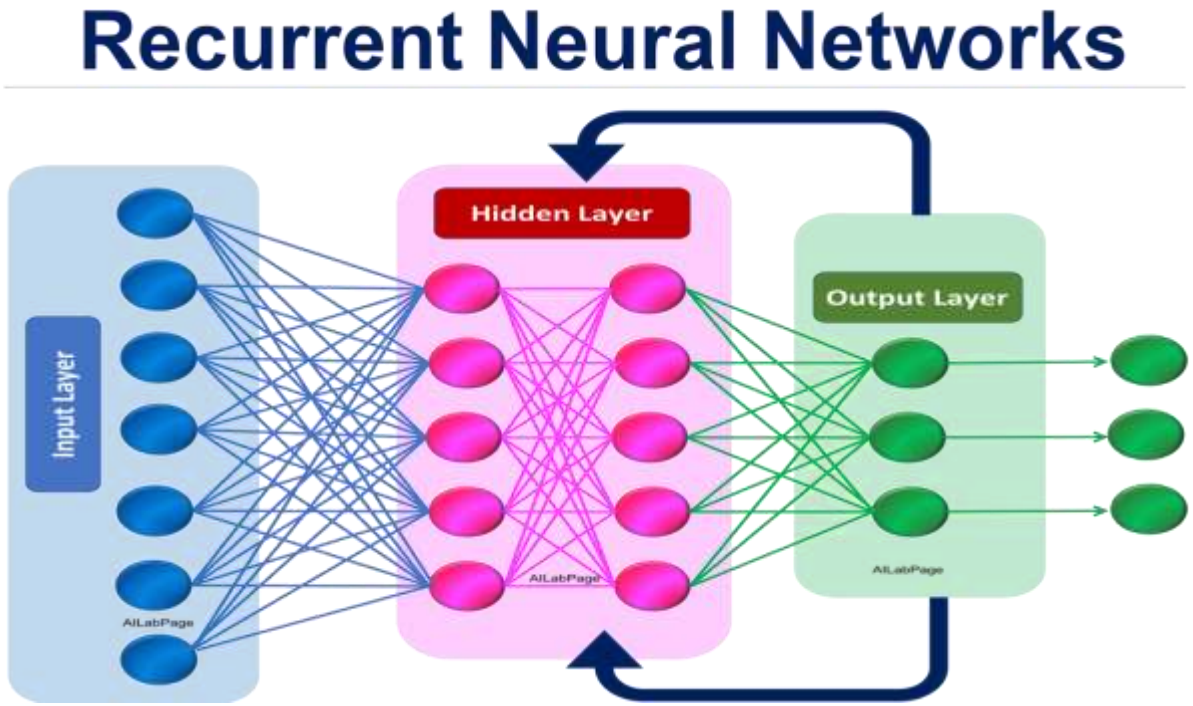
- ***Deep Belief Network***
- Deep Belief Networks (DBNs) is the technique of **stacking many individual unsupervised networks** that use each network's hidden layer as the input for the next layer.



https://deepai.org/machine-learning-glossary-and-terms/deep-belief-network

Deep Learning Architectures

- ***Recurrent Neural Network***

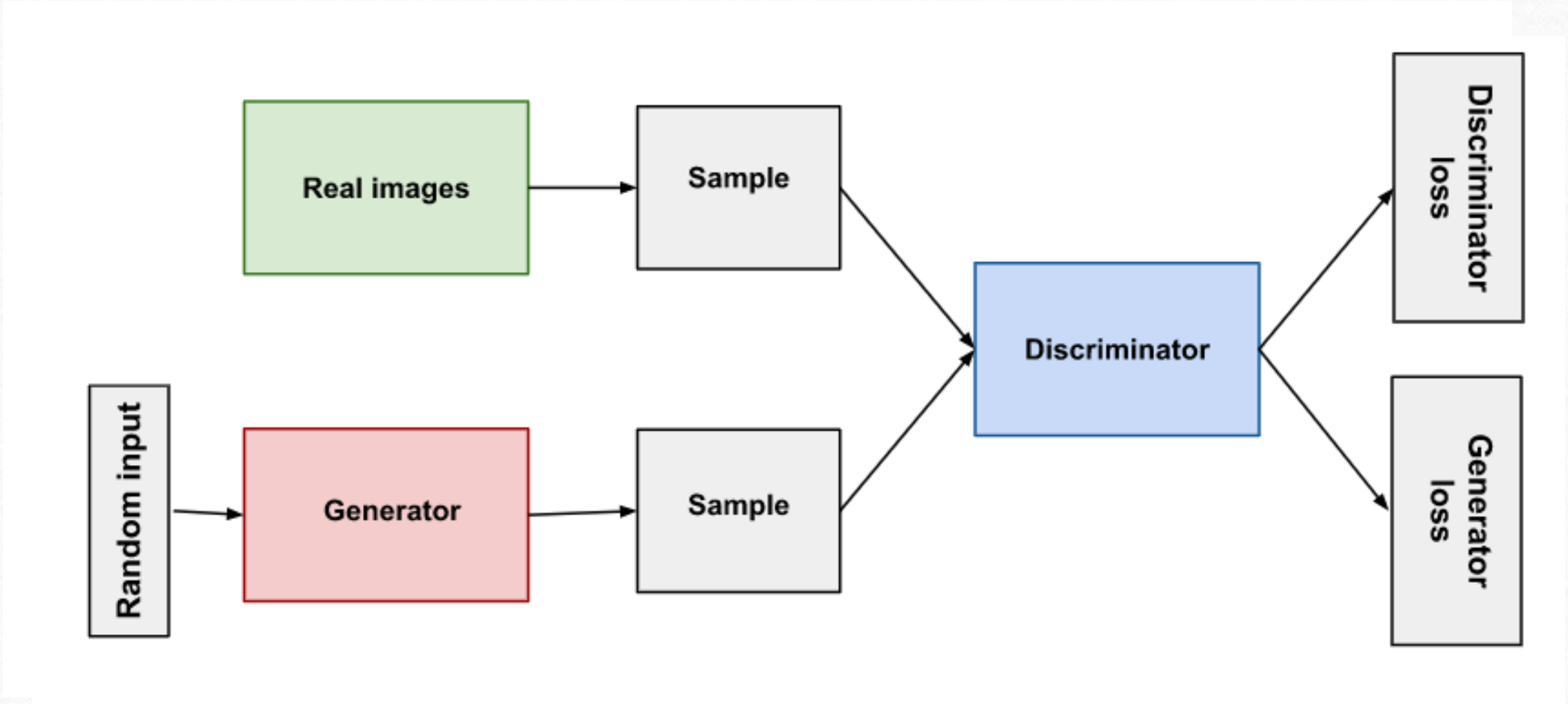


https://morioh.com/p/1bc305d7dbdf

Generative Adversarial Network (GAN)

- A generative adversarial network (GAN) has two parts:
- The generator learns to generate plausible data. The generated instances become negative training examples for the discriminator.
- The discriminator learns to distinguish the generator's fake data from real data. The discriminator penalizes the generator for producing implausible results.
- Both the generator and the discriminator are neural networks. The generator output is connected directly to the discriminator input.
  Through backpropagation, the discriminator's classification provides a signal that the generator uses to update its weights.

Generative Adversarial Network (GAN)



https://medium.com/@devnag/generative-adversarial-networks-gans-in-50-lines-of-code-pytorch-e81b79659e3f

Convolutional Neural Network

- Convolutional neural networks are distinguished from other neural networks by their superior performance with image, speech, or audio signal inputs. They have three main types of layers, which are:
- Convolutional layer
- Pooling layer
- Fully-connected (FC) layer

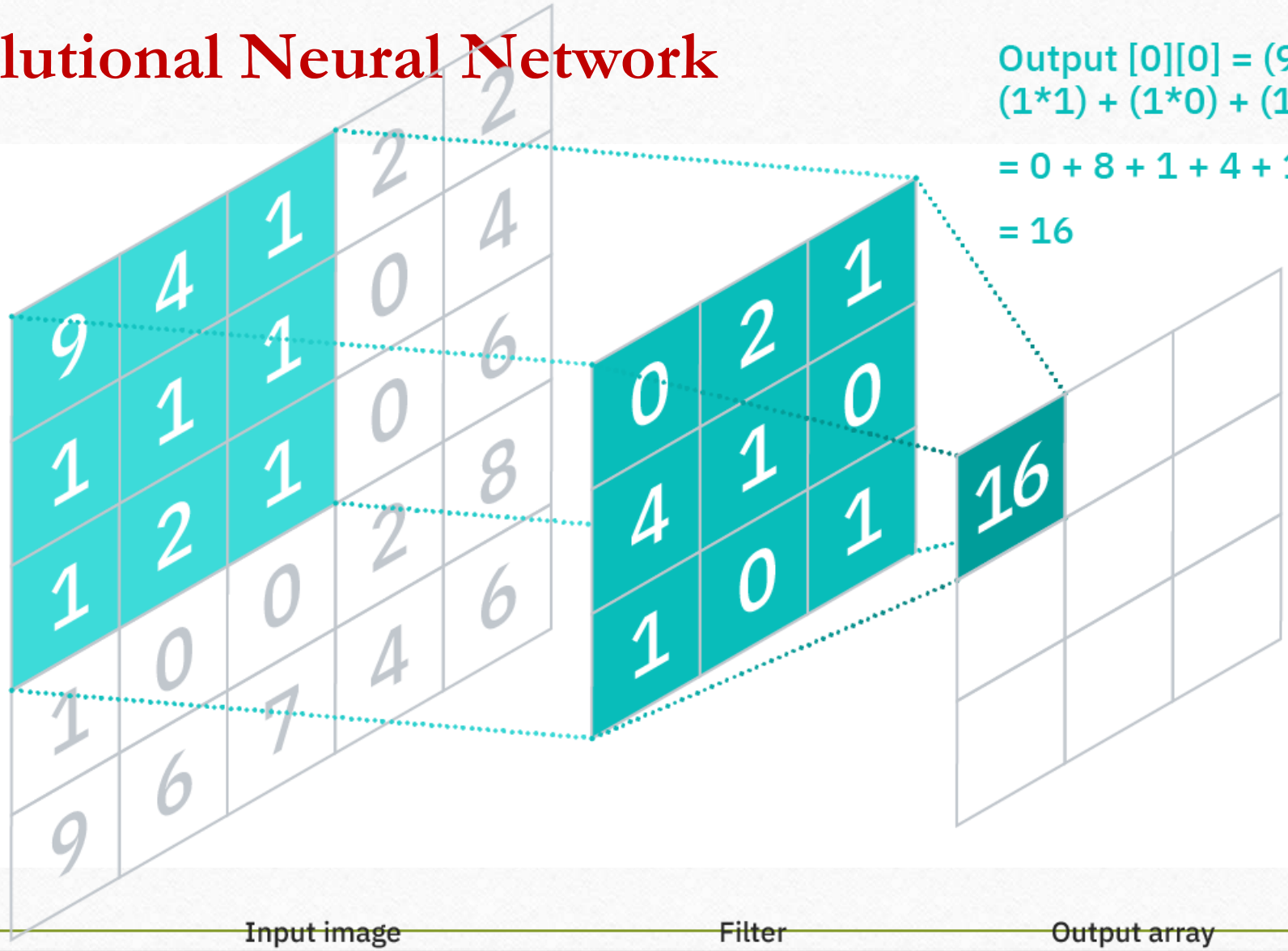
https://www.ibm.com/cloud/learn/convolutional-neural-networks

Convolutional Neural Network

Output [0][0] = (9*0) + (4*2) + (1*4) + (1*1) + (1*0) + (1*1) + (2*0) + (1*1)

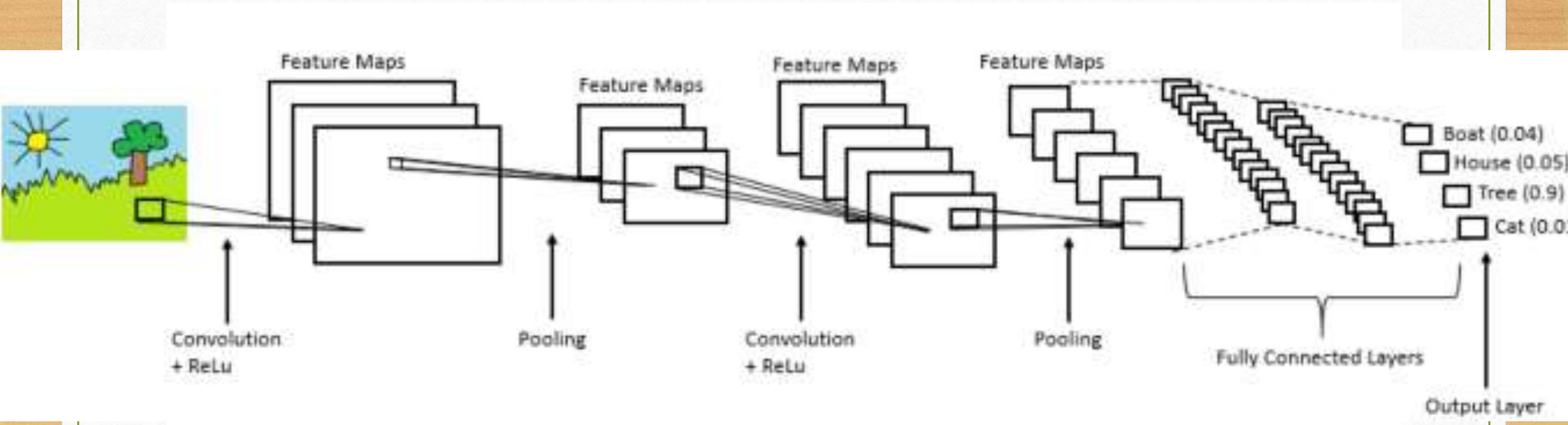
= 0 + 8 + 1 + 4 + 1 + 0 + 1 + 0 + 1

= 16



Input image Filter Output array

Convolutional Neural Network



References

- *Shutterstocks.com, Microsoft.com,*
- *Amazinganimations.com*
- *illustrationof.com, Clipartof.com*
- *towardsdatascience.com*
- *Javapoint.com, tutorialspoint.com,*
- *Medium.com, section.io,*
- *Dotnetlovers.com,*
- *Machinelearningmastery.com*
- *Sajja, P.S. "Illustrated computational intelligence: Examples and applications", Springer International Publishing, Singapore (Nov'20)*