University Kasdi Merbah Ouargla

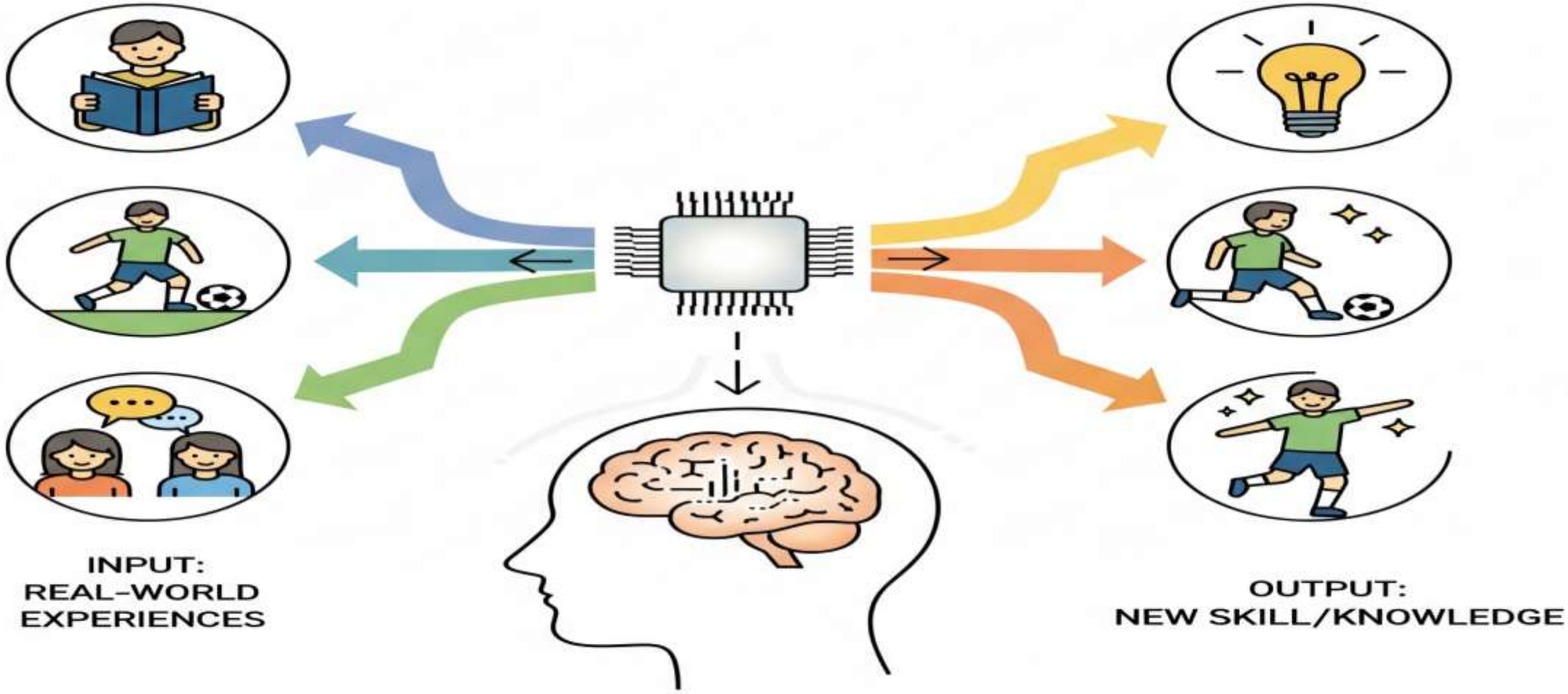
Machine Learning

2025-2026

Prepared by:

Mrs Leila AMRANE

Machine Learning



MACHINE LEARNING

Machine Learning (ML) is a subfield of **Artificial Intelligence (AI)** that enables computers to **identify patterns, learn from data, and make predictions or decisions** without being explicitly programmed for every scenario.

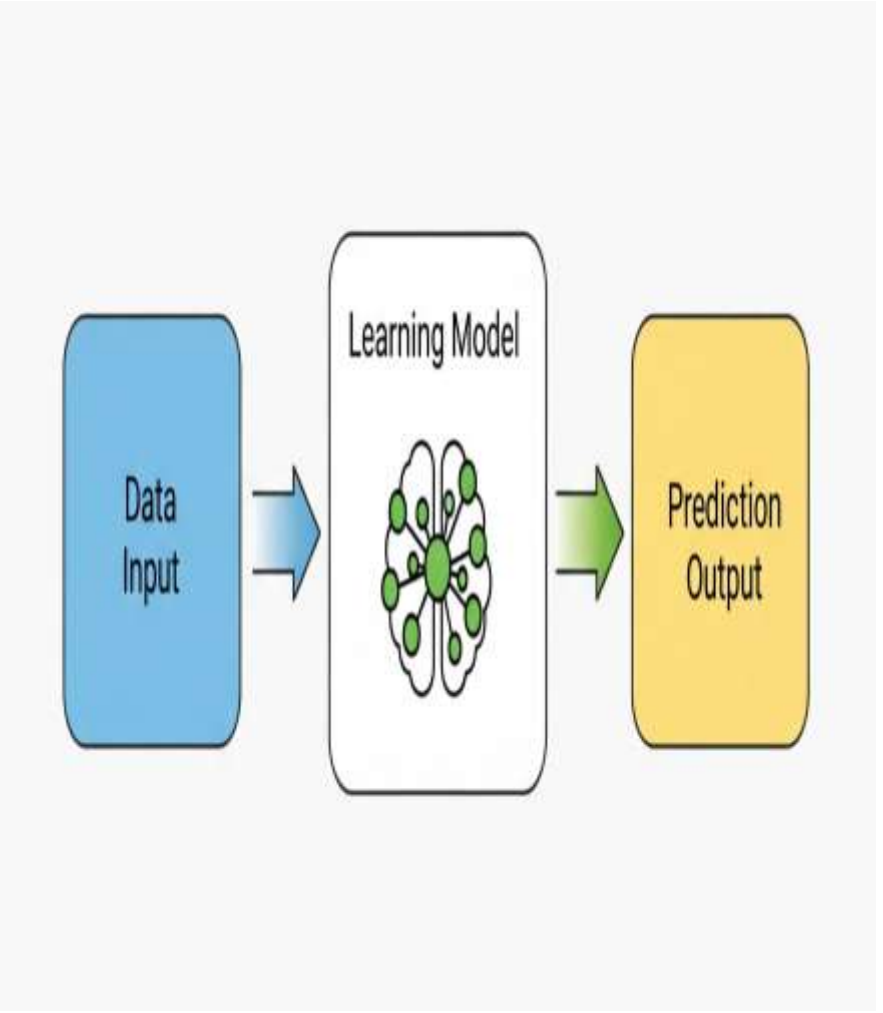
Key Idea:

Instead of writing rules for every situation, we give the computer **examples** and let it discover the rules by itself.

Analogy:

Like teaching a child: You show them hundreds of pictures of cats and dogs, and they learn to recognize each, even without memorizing every picture.

The computer does the same: it looks for patterns (shapes, colors, features) and creates a *model* to classify or predict new inputs.



Machine learning workflow

Training set

Unsupervised

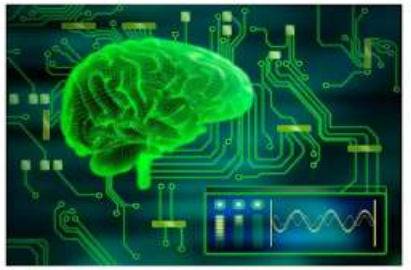
Supervised

Feature extraction

Machine learning algorithm

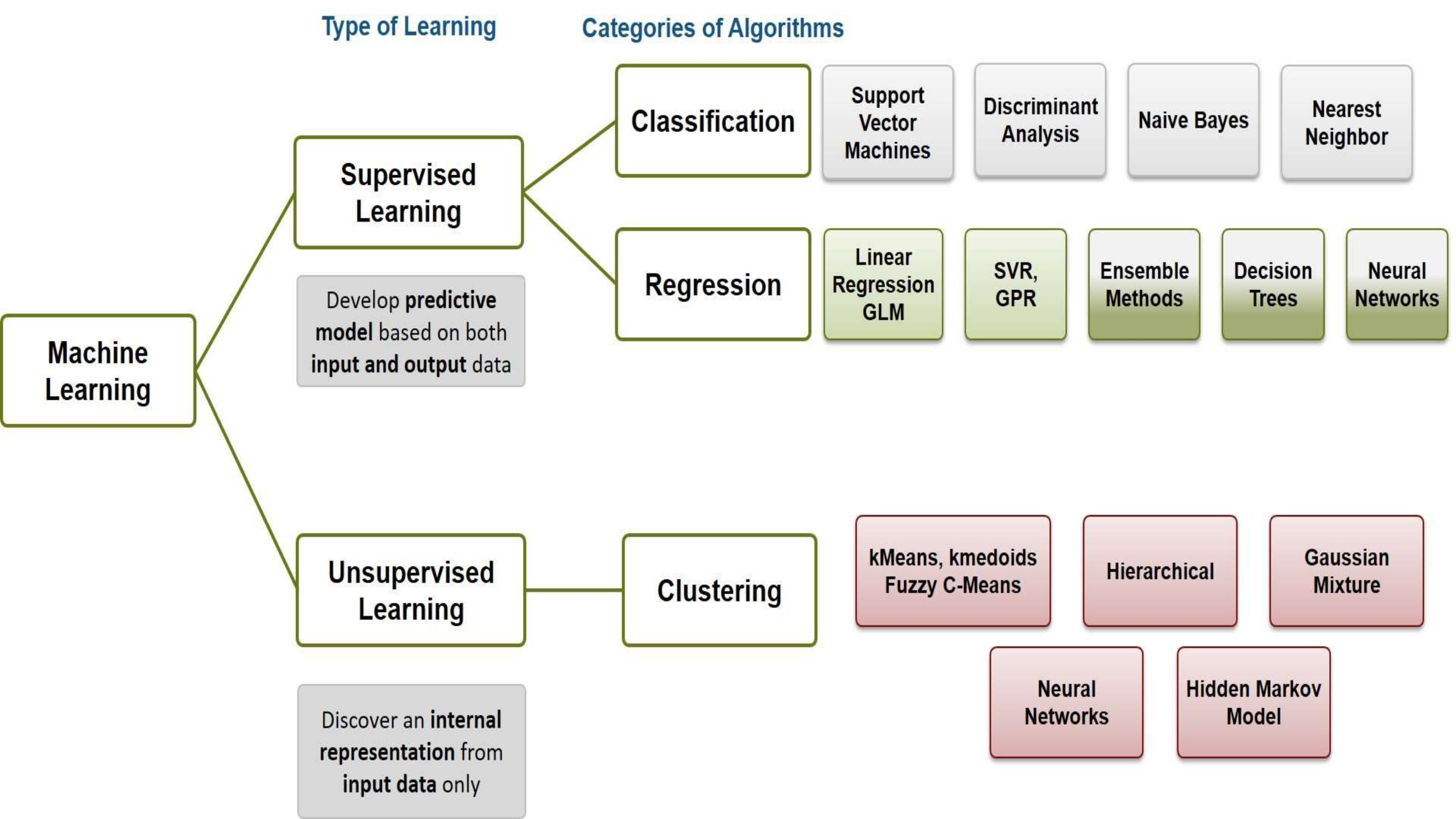
Grouping of objects

Predictive model



New Data

Annotated data


Type of Learning
Categories of Algorithms
Machine Learning
Supervised Learning
Develop predictive model based on both input and output data
Classification
Support Vector Machines
Discriminant Analysis
Naive Bayes
Nearest Neighbor
Regression
Linear Regression GLM
SVR, GPR
Ensemble Methods
Decision Trees
Neural Networks
Unsupervised Learning
Discover an internal representation from input data only
Clustering
kMeans, kmedoids Fuzzy C-Means
Hierarchical
Gaussian Mixture
Neural Networks
Hidden Markov Model

Supervised

- In Supervised learning, you train the machine using data which is well **"labeled**."
- It means data is already tagged with the correct answer.
- It can be compared to learning which takes place in the presence of a supervisor or a teacher.
- A supervised learning algorithm learns from labeled training data, helps you to predict outcomes for unforeseen data.
- One disadvantage of this learning method is that the dataset has to be hand-labeled either by a Machine Learning Engineer or a Data Scientist. This is a very *costly process*, especially when dealing with large volumes of data.

Example

- A person who have never seen a game of football in his entire life.
- Learning with a coach.
- A coach explains the rules, show the videos of football games and clearly labels things. “That’s a goal keeper, that’s a striker, this is offside, that’s a goal.”
- After enough labeled examples and corrections, the person can join a game and play correctly from the start.

Unsupervised

- ❑ Unsupervised Learning is a machine learning technique in which the users do not need to supervise the model.
- ❑ Instead, it allows the model to work on its own to discover patterns and information that was previously undetected.
- ❑ It mainly deals with the unlabeled data.
- Unsupervised learning problems are grouped into clustering and association problems.

- The most basic disadvantage of any Unsupervised Learning is that its application spectrum is limited.

Example

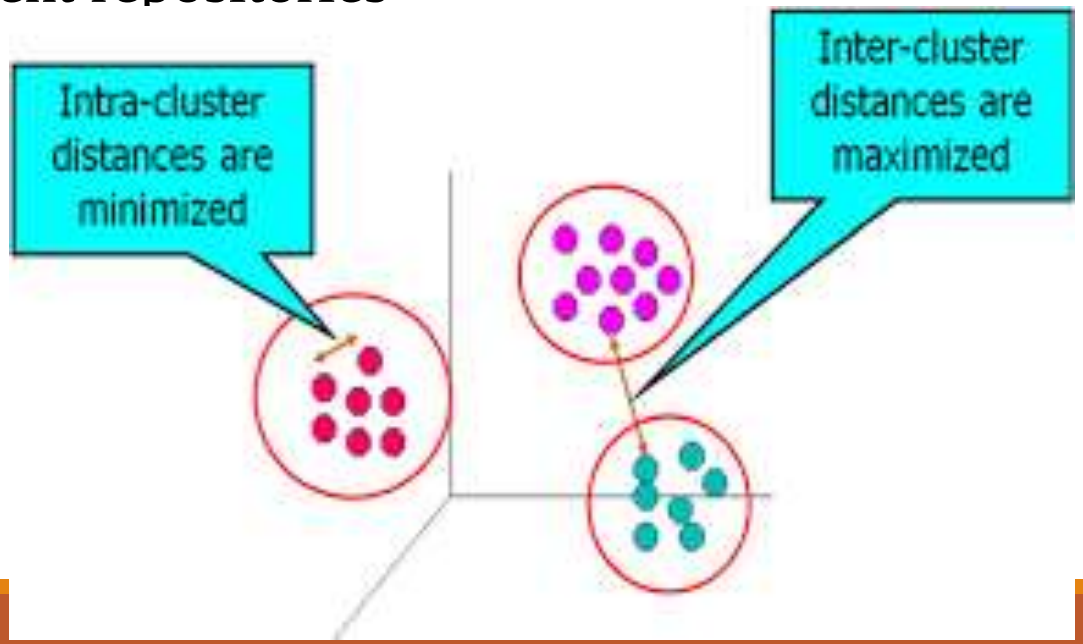
- Watching football without guidance
- No one tells him the rules, the name of the players, or even what the game is called.
- All he can do is watch carefully noticing that some players wear red some blue(finding group and clusters).
- Hey might observe that the players with gloves stand near the goalposts(finding pattern).

Clustering:

- Finding groups of data. Basically, put similar ROWS in data under one group

EXAMPLES:

- Finding the types of customers using shopping behavior data
- Finding the types of users in telecom
- Organize news articles, manage large document repositories
- Detect influencers
- Detect credit card fraud



Good clustering allows you to:
Maximize inter-cluster distance
Minimize intra-cluster distance.

K-means:

- Finding groups of data. Basically, put similar ROWS in data under one group

Algorithm:

1. Initialize k cluster centers randomly among the n individuals: v1, v2,…. vk
2. Evaluate (calculate) the distances between each individual aj and each centroid vi
3. Assign each individual aj to the cluster center vi whose distance || aj -vi|| is minimal
4. Recalculate each centroid using the average of the individuals in the cluster.
5. Repeat steps 2-5 until it converges to k clusters (no reassignment, the centroids do not change)

EXERCISE1:

x	**1**	**2**	**2**	**3**	**4**	**5**
y	**1**	**1**	**3**	**2**	**3**	**5**

Randomly choose cluster centers :V1(2, 1), V2(2, 3)

The distance function is Euclidean distance. Use the k-means algorithm

EXERCISE2:

Suppose that the data mining task is to cluster points into three clusters, where the points are :
A1(2, 10), A2(2, 5), A3(8, 4), B1(5, 8), B2(7, 5), B3(6, 4), C1(1, 2), C2(4, 9).

The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively.

What is Knn?

KNN is a type of supervised ML algorithm which can be used for both classification as well as regression predictive problems. However, it is mainly used for classification predictive problems in industry.

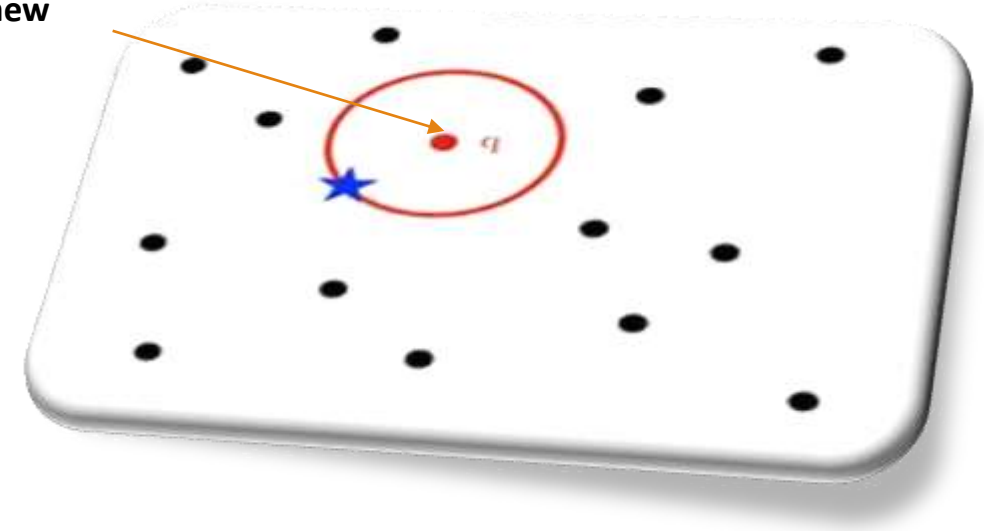
It is a **Lazy Learning Algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

Nearest Neighbor?

One of the simplest of all machine learning classifiers.

Simple Idea: Label a new point the same as the closest known point.

Red is nearest to this new instance



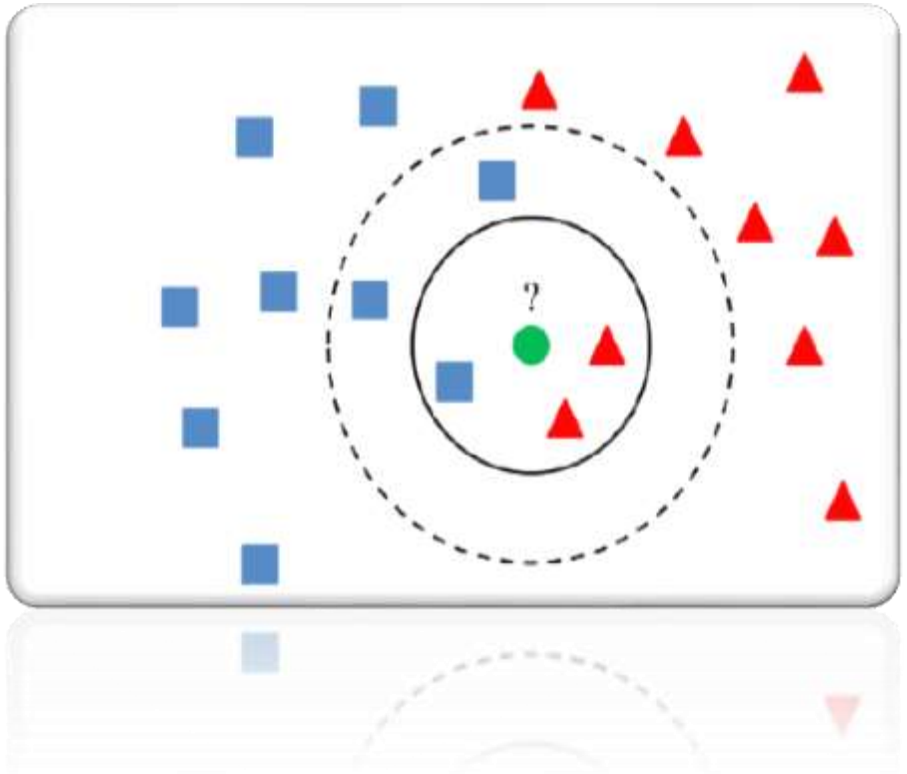
Principle of KNN - Classifier

This technique implements classification by considering majority of vote among the "k" closest points to the unlabeled data point.

K here describes the Numbers of instance point that should be taken into the consideration.

Euclidean Distance or Manhattan Distance or Hamming distance are used as metric for calculating the distance between points.

K Nearest neighbor



Green circle is the unlabeled data point.

K=3 in this problem

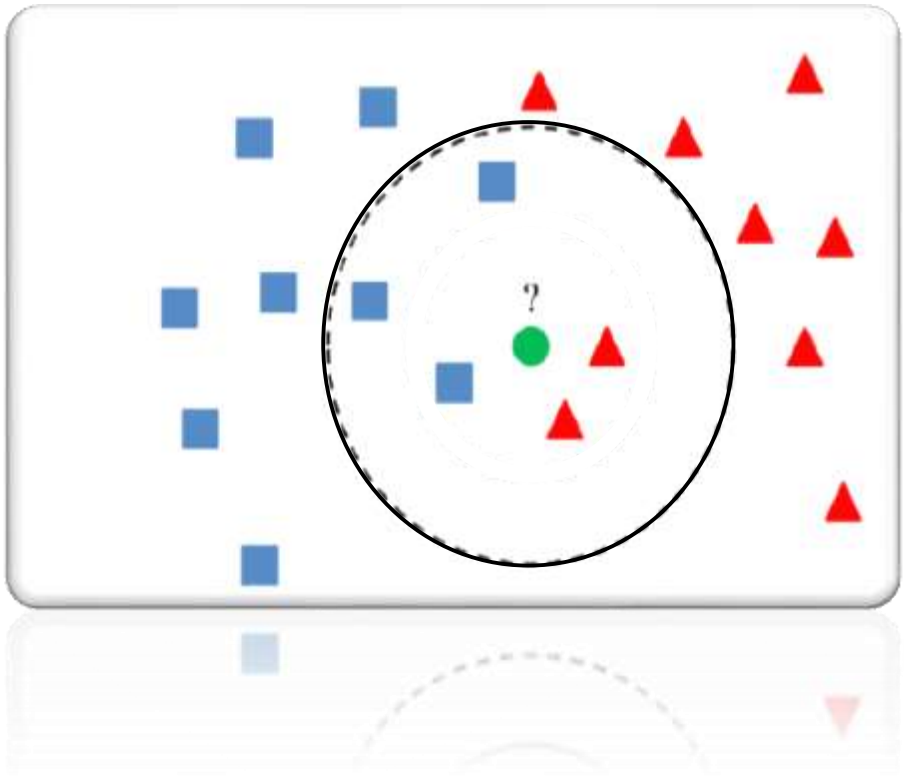
Closest 3 points taken

2 are red 1 is blue

Votes = 2 Red > 1 Blue

So, Here the Green circle will be red triangle.

K Nearest neighbor



Green circle is the unlabeled data point.

K=5 in this problem

Closest 5 points taken

2 are red & 3 are blue

Votes = 2 Red < 3 Blue

So, Here the Green circle will be Blue rectangle.

EXAMPLE

Sepal length	Sepal width	species
5,3	3,7	Setosa
5,1	3,8	Setosa
7,2	3	Virginica
5,4	3,4	Setosa
5,1	3,3	Setosa
5,4	3,9	Setosa
7,4	2,8	Virginica
6,1	2,8	verscicolor
7,3	2,9	Virginica
6	2,7	Verscicolor
5,8	2,8	Virginica
6,3	2,3	verscicolor
5,1	2,5	verscicolor
6,3	2,5	verscicolor
5,5	2,4	verscicolor

Using KNN to find the classe o this sepal (5,2 ; 3,1).
K=4;

Euclidien distance:

$$d((x,y),(5{,}2,3{,}1)) = \sqrt{(x-5{,}2)^2+(y-3{,}1)^2}.$$

EXAMPLE

Euclidean Distance	Sepal length	Sepal width	species
0.608	5,3	3,7	Setosa
0.707	5,1	3,8	Setosa
2.002	7,2	3	Virginica
0.361	5,4	3,4	Setosa
0.224	5,1	3,3	Setosa
0.825	5,4	3,9	Setosa
2.22	7,4	2,8	Virginica
0.949	6,1	2,8	verscicolor
2.11	7,3	2,9	Virginica
0.894	6	2,7	Verscicolor
0.671	5,8	2,8	Virginica
1.36	6,3	2,3	verscicolor
0.608	5,1	2,5	verscicolor
1.253	6,3	2,5	verscicolor
1,315	5,5	2,4	verscicolor

The sepal (5,2 ; 3,1) is the class Setosa