

MACHINE LEARNING

ARTIFICIAL INTELLIGENCE



Machine Learning

- Machine Learning is making the computer learn from studying data and statistics.
- Machine learning is the concept that a computer program can learn and adapt to new data without human intervention.
- Machine Learning is a program that analyses data and learns to predict the outcome.



Machine Learning

- Machine Learning (ML) is a field of computer science where computer systems provide pattern to data.
- ML extracts patterns out of raw data by using algorithms.
- ML allow computer systems learn from experience without being explicitly programmed.



Supervised vs. Unsupervised Learning

- Supervised learning finds patterns for a prediction task

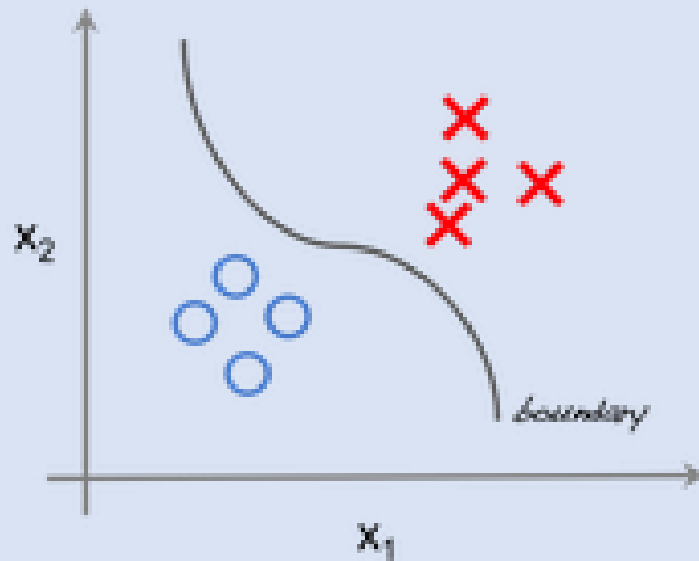
E.g., classify tumors as benign or cancerous (labels)

- Unsupervised learning finds patterns in data

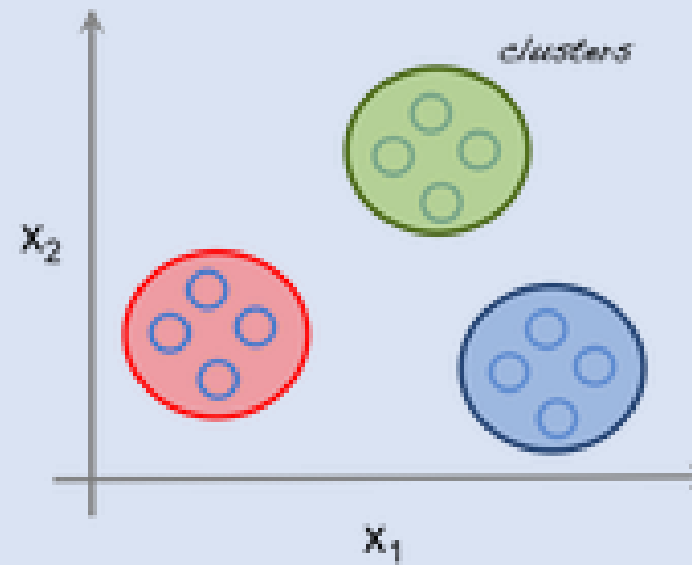
... but without a specific prediction task in mind

- The left image an example of supervised learning (regression techniques are used to find the best fit line between the features).
- In unsupervised learning the inputs are segregated based on features and the prediction is based on which cluster it belonged to.

Supervised learning



Unsupervised learning



Supervised Learning

- Supervised learning algorithms are the most commonly used ML algorithms.
- It is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately.
- The main objective is to learn an association between input data samples and corresponding output.



For example

x: Input variables

Y: Output variable

Now, apply an algorithm to learn the mapping function from the input to output as follows:

$$Y = f(x)$$

Logistic Regression

- A logistic regression model predicts $P(Y=1)$ as a function of X . It is one of the simplest ML algorithms that can be used for various classification problems such as spam detection, Diabetes prediction, cancer detection etc.
- Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable.
- The dependent variable is binary having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no).

- In case of logistic regression, the linear function is basically used as an input to another function such as g in the following relation –

$$h_{\theta}(x) = g(\theta^T x) \text{ where } 0 \leq h_{\theta} \leq 1$$

- Here, g is the logistic or sigmoid function which can be given as follows –

$$g(z) = \frac{1}{1 + e^{-z}} \text{ where } z = \theta^T x$$

- Plotting growth rate against time, a typical S-shaped **curve** is obtained. The sigmoid curve can be represented by the graph –

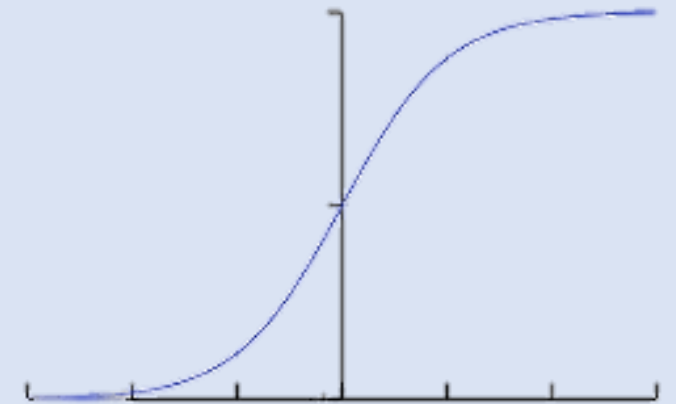
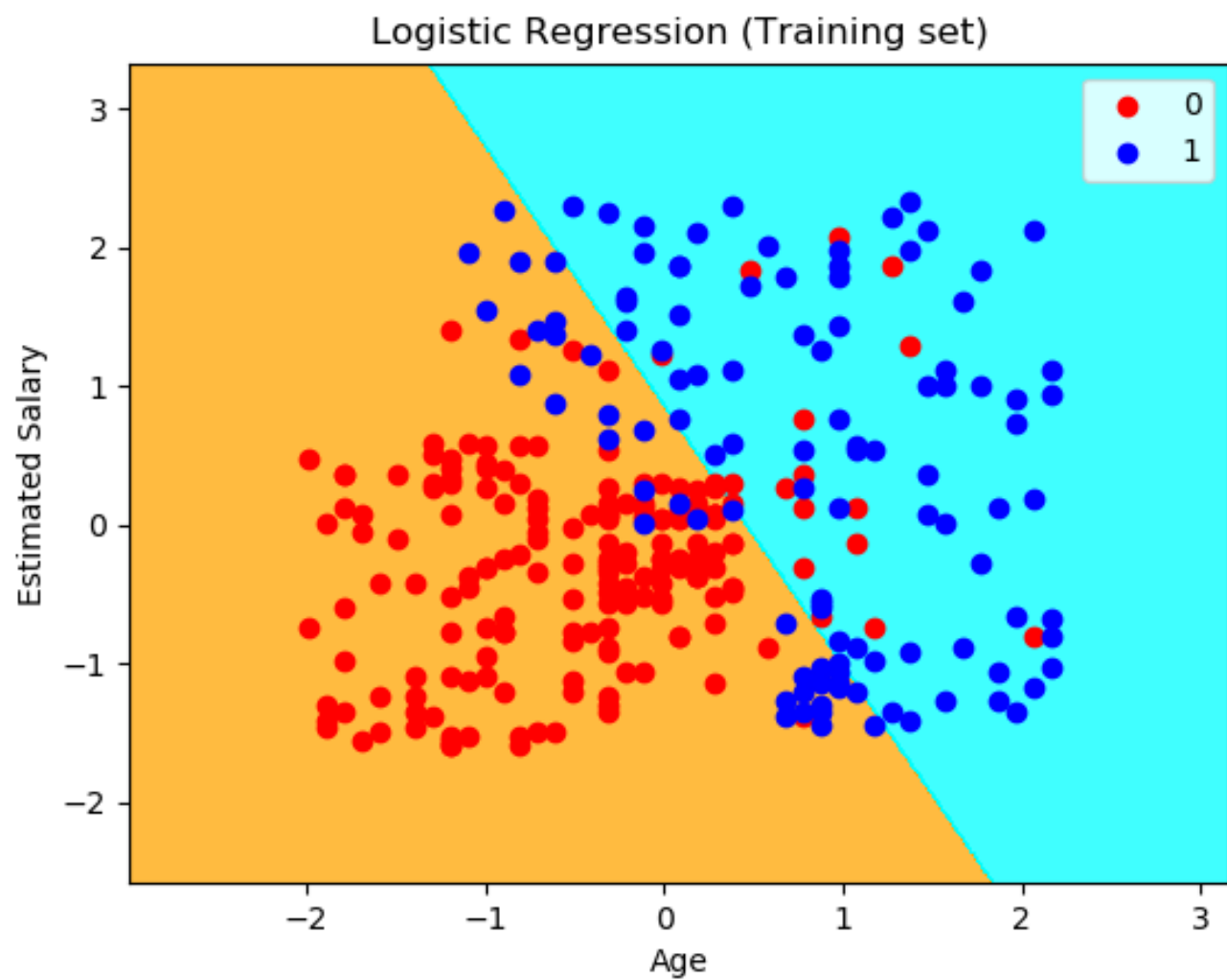


Figure 1



Unsupervised Learning

- Unsupervised machine learning is the machine learning task of inferring a function to describe hidden structure from “unlabeled” data.
- Common scenarios for using unsupervised learning algorithms include:
 - Data Exploration
 - Outlier Detection
 - Pattern Recognition



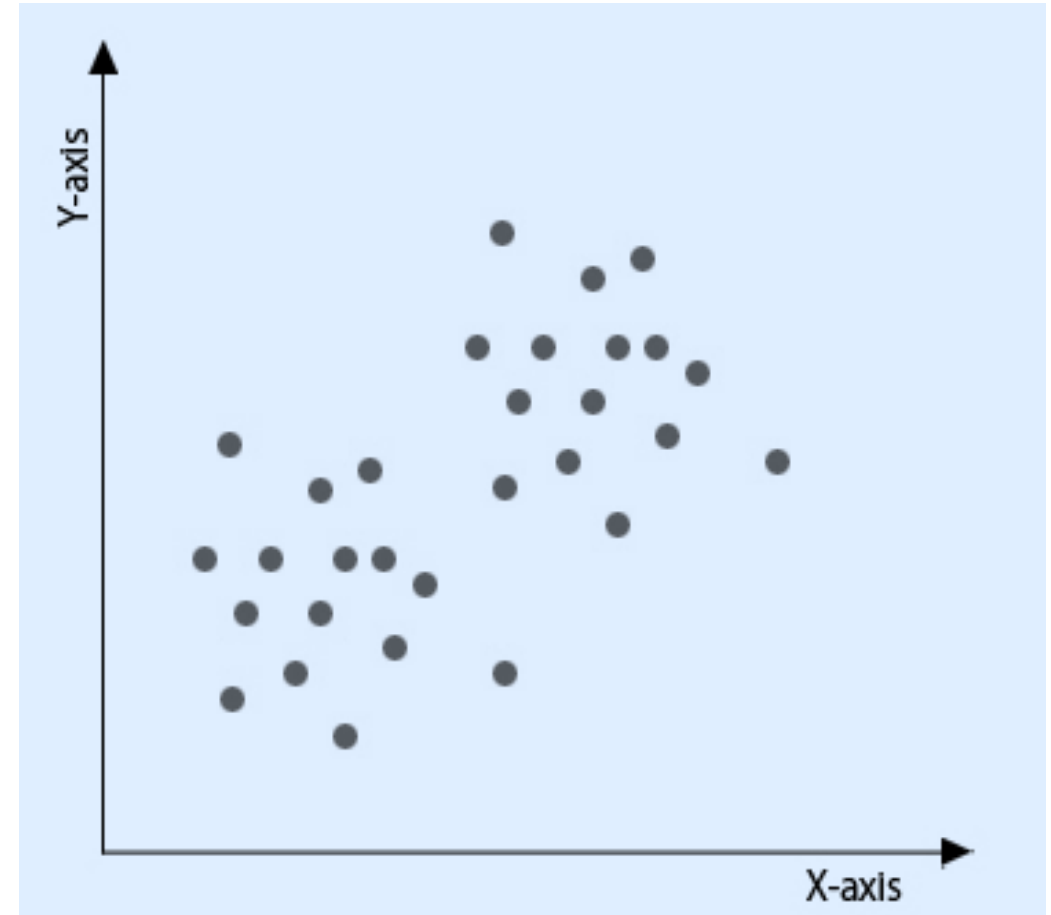
- For example:

x: Input variables, there is no corresponding output variable

The algorithms need to discover the interesting pattern in data for learning.

K-Means Clustering

- K-Means clustering algorithm is a unsupervised learning method that iterates over a dataset
- The data is grouped into k number of predefined non-overlapping clusters or subgroups making the inner points of the cluster as similar as possible
- The clusters at distinct space it allocates the data points to a cluster the sum of the squared distance between the clusters centroid and the data point is at a minimum



Cluster Iterating

