

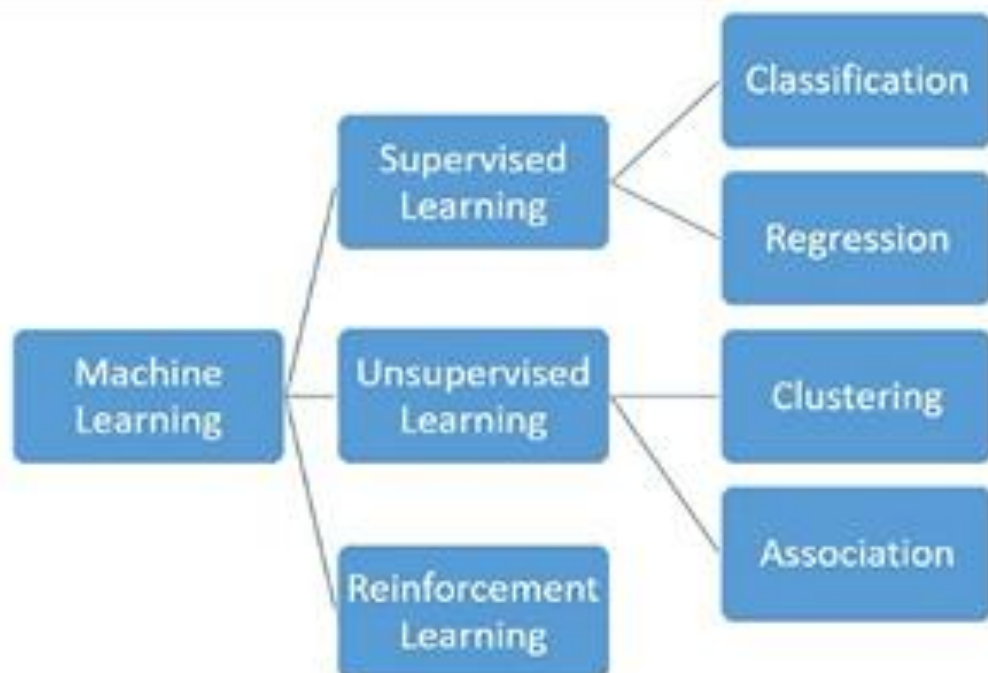
# Machine Learning

## Lecture 2: Types of Machine Learning

COURSE CODE: CSE451

2021

## Types of Machine Learning



## Supervised Learning (Cont.)

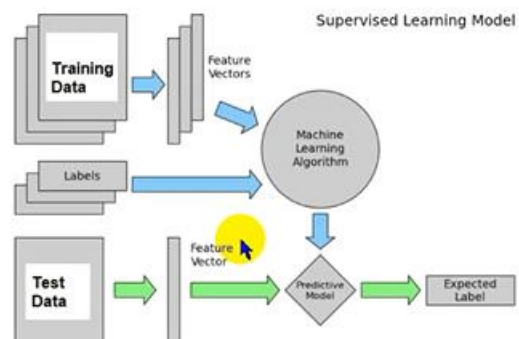
- Supervised learning is used to learn a mapping function  $f$  from the input ( $X$ ) to the output  $Y$  as follows:

$$Y = f(X)$$

- The goal is to approximate the mapping function  $f$  so well that when you have new input data ( $X$ ) that it can predict the output variables ( $Y$ ) for that data.
- It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process

## Supervised Learning

- Also called Predictive Modeling



## Some Classification Problems

- Spam filtering
- Language detection
- A search of similar documents
- Sentiment analysis
- Recognition of handwritten characters and numbers
- Fraud detection etc.

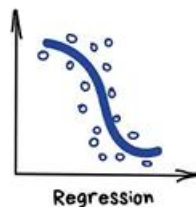


## Classification

- When the output variable / target variable is a category or a set of categories, such as “disease or no disease”, “red or green or blue” etc.
- Predicts one or more categories/labels/classes for each input/instance
- Binary Classification: Classifying instances into one of two classes/categories
- Multiclass Classification: Classifying instances into one of three or more classes/categories
- Multi-Label Classification: Multiple labels/classes are to be predicted for each instance

## Regression

- When the output variable is a real/continuous value, such as “dollars” or “weight” or “Score”
- Predicts a single output value
- Why do we use Regression Analysis?
  - Forecasting
  - Demand and sales volume analysis
  - Time series modelling
  - Medical diagnosis etc.
- Linear Regression
- Polynomial Regression



Source: [AnalyticsVidhya](#)

# List of Commonly Supervised Learning Algorithms

---

- Linear Regression
- Logistic Regression
- k-Nearest Neighbors (kNN)
- Decision Trees
- Random Forest
- Support Vector Machines (SVM)
- Gradient Boosting Machines (GBM)
- LightGBM
- XGBoost
- Neural Networks

## What is Unsupervised Learning?

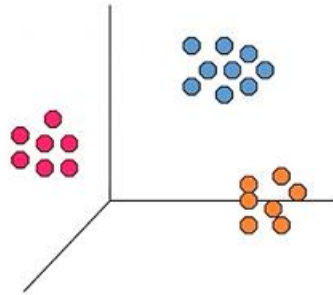
---

- Unsupervised learning is where you only have unlabeled input data (X) and allow the algorithm to work on its own to discover the underlying groupings, structure or pattern in the data.
- These are called unsupervised learning because unlike supervised learning there is no correct answers and there is no teacher (i.e., learning from the labeled training data).
- Unsupervised learning problems can be further grouped into **clustering** and **association** rule mining.

# What is Cluster Analysis / Clustering?

---

Finding groups of objects such that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups



## Applications of Clustering

---

- **Customer Segmentation:** This strategy is across functions, including banking, telecom, e-commerce, sports, advertising, sales, etc.
- **Document Clustering:** Cluster similar documents together
- **Image Clustering:** You can group similar images together.
- **Image Segmentation:** You can apply clustering to create clusters having similar pixels in the image together.
- **Recommendation Engines:** You can look at the songs liked by a person and then use clustering to find similar songs and finally recommend the most similar songs to him.

# What is Association Rule Mining?

---

- Association Rule Mining is a rule-based machine learning method for discovering frequently occurring **patterns, correlations, or associations** between variables in large databases.
- It is intended to identify strong rules discovered in databases.
- A typical example is Market Based Analysis: for example, if a customer buys bread, he most likely can also buy butter, eggs, or milk, so these products are stored within a shelf or mostly nearby

## Applications of Association Rule Mining

---

- **Market Basket Analysis:** It is one of the popular examples and applications of association rule mining. This technique is commonly used by big retailers to determine the association between items.
- **Medical Diagnosis:** With the help of association rules, patients can be cured easily, as it helps in identifying the probability of illness for a particular disease.
- **Protein Sequence:** The association rules help in determining the synthesis of artificial Proteins.
- **Census Data:** The association rule mining has immense potential in census data in supporting sound public policy and bringing forth an efficient functioning of a democratic society.
- It is also used for the **Catalog Design** and **Loss-leader Analysis** and many more other applications.

Source: [javaTpoint](#), [upGrad](#)

---



# List of Commonly Used Unsupervised Learning Algorithms

---

- K-means, Hierarchical for clustering problems
- Apriori algorithm for association rule learning problems
- Principal Component Analysis
- Singular Value Decomposition
- Independent Component Analysis

## Supervised vs Unsupervised Learning

Supervised Learning	Unsupervised Learning
Algorithms are trained using labeled data.	Algorithms are trained using unlabeled data.
The goal is to train the model so that it can predict the output when it is given new data.	The goal is to find underlying groupings, structure or pattern in the data.
Supervised learning model takes direct feedback to check if it is predicting correct output or not.	Unsupervised learning model does not take any feedback.
Supervised learning model produces an accurate result.	Unsupervised learning model may give less accurate result as compared to supervised learning.
Supervised learning is a simpler method.	Unsupervised learning is computationally complex.

---