# Machine Learning - Models

There are various Machine Learning algorithms, techniques and methods that can be used to build **models** for solving real-life problems by using data. In this chapter, we are going to discuss such different kinds of methods.

There are four main types of machine learning methods classified based on human supervision –

- Supervised Learning
- Unsupervised Learning
- Semi-supervised Learning
- Reinforcement Learning

In the next four chapters, we will discuss each of these machine learning models in detail. Here, let's have a brief overview of these methods:

## Supervised Learning

Supervised learning algorithms or methods are the most commonly used ML algorithms. This method or learning algorithm takes the data sample i.e. the training data and its associated output i.e. labels or responses with each data sample during the training process.

The main objective of supervised learning algorithms is to learn an association between input data samples and corresponding outputs after performing multiple training data instances.

For example, we have

**x**: Input variables and

**Y**: Output variable

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

Y=f(x)

Now, the main objective would be to approximate the mapping function so well that even when we have new input data (x), we can easily predict the output variable (Y) for that new input data.

It is called supervised because the whole process of learning can be thought as it is being supervised by a teacher or supervisor. Examples of supervised machine learning algorithms includes **Decision tree, Random Forest, KNN, Logistic Regression** etc.

Based on the ML tasks, supervised learning algorithms can be divided into the following two broad classes –

- Classification
- Regression

## Classification

The key objective of classification-based tasks is to predict categorial output labels or responses for the given input data. The output will be based on what the model has learned in the training phase. As we know the categorial output responses means unordered and discrete values, hence each output response will belong to a specific class or category. We will discuss Classification and associated algorithms in detail in the upcoming chapters also.

## Classification Models

Followings are some common classification models –

- Logistic Regression
- Decision Trees
- Random Forest
- K-nearest Neighbor
- Support Vector Machine
- Naive Bayes
- Linear Discriminant Analysis
- Neural Networks

## Regression

The key objective of regression-based tasks is to predict output labels or responses, which are continuous numeric values, for the given input data. The output will be based on what the model has learned in its training phase. Basically, regression models use the input data features (independent variables) and their corresponding continuous numeric output values (dependent or outcome variables) to learn specific associations between inputs and corresponding outputs. We will discuss regression and associated algorithms in detail in further chapters.

## Regression Models

Followings are some common regression models −

- Linear Regression
- Ridge regression
- Decision Trees
- Random Forest
- K-nearest Neighbor
- Neural Network Regression

# Unsupervised Learning

As the name suggests, unsupervised learning is opposite to supervised ML methods or algorithms in which we do not have any supervisor to provide any sort of guidance. Unsupervised learning algorithms are handy in the scenario in which we do not have the liberty, like in supervised learning algorithms, of having pre-labeled training data and we want to extract useful pattern from input data.

For example, it can be understood as follows −

Suppose we have −

**x: Input variables**, then there would be no corresponding output variable and the algorithms need to discover the interesting pattern in data for learning.

Examples of unsupervised machine learning algorithms includes K-means clustering, **K-nearest neighbors** etc.

Based on the ML tasks, unsupervised learning algorithms can be divided into the following broad classes −

- Clustering

- Association
- Dimensionality Reduction

## Clustering

Clustering methods are one of the most useful unsupervised ML methods. These algorithms used to find similarity as well as relationship patterns among data samples and then cluster those samples into groups having similarity based on features. The real-world example of clustering is to group the customers by their purchasing behavior.

## Clustering Models

Followings are some common clustering models −

- K-Means Clustering
- Hierarchical Clustering
- Mean-shift Clustering
- DBSCAN Clustering
- HDBSCAN Clustering
- BIRCH Clustering
- Affinity Propagation
- Agglomerative Clustering

## Association

Another useful unsupervised ML method is **Association** which is used to analyze large dataset to find patterns which further represents the interesting relationships between various items. It is also termed as **Association Rule Mining** or **Market basket analysis** which is mainly used to analyze customer shopping patterns.

## Association Models

Followings are some common association models −

- Apriori Algorithm
- Eclat algorithm
- FP-growth algorithm

## Dimensionality Reduction

This unsupervised ML method is used to reduce the number of feature variables for each data sample by selecting set of principal or representative features. A question arises here is that why we need to reduce the dimensionality? The reason behind is the problem of feature space complexity which arises when we start analyzing and extracting millions of features from data samples. This problem generally refers to curse of dimensionality. PCA (Principal Component Analysis), K-nearest neighbors and discriminant analysis are some of the popular algorithms for this purpose.

## Dimensionality Reduction Models

Followings are some common dimensionality Reduction models –

- Principal Component Analysis(PCA)
- Autoencoders
- Singular value decomposition (SVD)

## Anomaly Detection

This unsupervised ML method is used to find out the occurrences of rare events or observations that generally do not occur. By using the learned knowledge, anomaly detection methods would be able to differentiate between anomalous or a normal data point. Some of the unsupervised algorithms like clustering, KNN can detect anomalies based on the data and its features.

## Semi-supervised Learning

Semi-supervised learning algorithms or methods are neither fully supervised nor fully unsupervised. They basically fall between the two i.e. supervised and unsupervised learning methods. These kinds of algorithms generally use small supervised learning component i.e. small amount of pre-labeled annotated data and large unsupervised learning component i.e. lots of unlabeled data for training. We can follow any of the following approaches for implementing semi-supervised learning methods –

- The first and simple approach is to build the supervised model based on small amount of labeled and annotated data and then build the unsupervised model by applying the same to the large amounts of unlabeled data to get more labeled samples. Now, train the model on them and repeat the process.
- The second approach needs some extra efforts. In this approach, we can first use the unsupervised methods to cluster similar data samples, annotate these groups and then use a combination of this information to train the model.

## Reinforcement Learning

Reinforcement learning methods are different from previously studied methods and very rarely used also. In this kind of learning algorithms, there would be an agent that we want to train over a period of time so that it can interact with a specific environment. The agent will follow a set of strategies for interacting with the environment and then after observing the environment it will take actions regards the current state of the environment. The following are the main steps of reinforcement learning methods –

- **Step 1** – First, we need to prepare an agent with some initial set of strategies.
- **Step 2** – Then observe the environment and its current state.
- **Step 3** – Next, select the optimal policy regards the current state of the environment and perform important action.
- **Step 4** – Now, the agent can get corresponding reward or penalty as per accordance with the action taken by it in previous step.



optimal policies.

## Reinforcement Learning Models

Following are some common reinforcement learning algorithms –

- Q-learning
- Markov Decision Process (MDP)
- SARSA
- DQN
- DDPG