



Course name: **Python programming and analytics by Rahul Sir**

Topic name: **machine learning (linear regression)**

Video name: **machine learning (linear regression)**

Video length: **1 hour 09 minutes 24 seconds**

Content of the video: **Machine Learning**

### **Types of Machine Learning**

#### **Supervised and Unsupervised Machine Learning**

#### **Regression, Classification, Clustering**

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. **Machine learning focuses on the development of computer programs** that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. **The primary aim is to allow the computers learn automatically** without human intervention or assistance and adjust actions accordingly.

Machine learning enables analysis of massive quantities of data. While it generally delivers faster, more accurate results in order to identify profitable opportunities or dangerous risks, it may also require additional time and resources to train it properly. Combining machine

learning with AI and cognitive technologies can make it even more effective in processing large volumes of information.

Machine learning algorithms are often categorized as supervised or unsupervised.

- **Supervised machine learning algorithms** can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.
- In contrast, **unsupervised machine learning algorithms** are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.
- **Semi-supervised machine learning algorithms** fall somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data for training - typically a small amount of labeled data and a large amount of unlabeled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant

resources in order to train it / learn from it. Acquiring unlabeled data generally doesn't require additional resources.

- **Reinforcement machine learning algorithms** is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

Supervised learning classified into two categories of algorithms:

**Classification**: A classification problem is when the output variable is a category, such as "Red" or "blue" or "disease" and "no disease".

**Regression**: A regression problem is when the output variable is a real value, such as "dollars" or "weight".

Supervised learning as the name indicates the presence of a supervisor as a teacher. Basically, supervised learning is a learning in which we teach or train the machine using data which is well labeled that means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples(data) so that supervised learning algorithm analyses the training data(set of training examples) and produces a correct outcome from labeled data.

- **For instance**, suppose you are given a basket filled with different kinds of fruits. Now the first step is to train the machine with all different fruits one by one like this  
If shape of object is rounded and depression at top having color Red then it will be labelled as -**Apple**.

- If shape of object is long curving cylinder having color Green-Yellow then it will be labelled as -**Banana**.

Now suppose after training the data, you have given a new separate fruit say Banana from basket and asked to identify it.

Since the machine has already learned the things from previous data and this time have to use it wisely. It will first classify the fruit with its shape and color and would confirm the fruit name as BANANA and put it in Banana category. Thus, the machine learns the things from training data (basket containing fruits) and then apply the knowledge to test data (new fruit).

Unsupervised learning classified into two categories of algorithms:

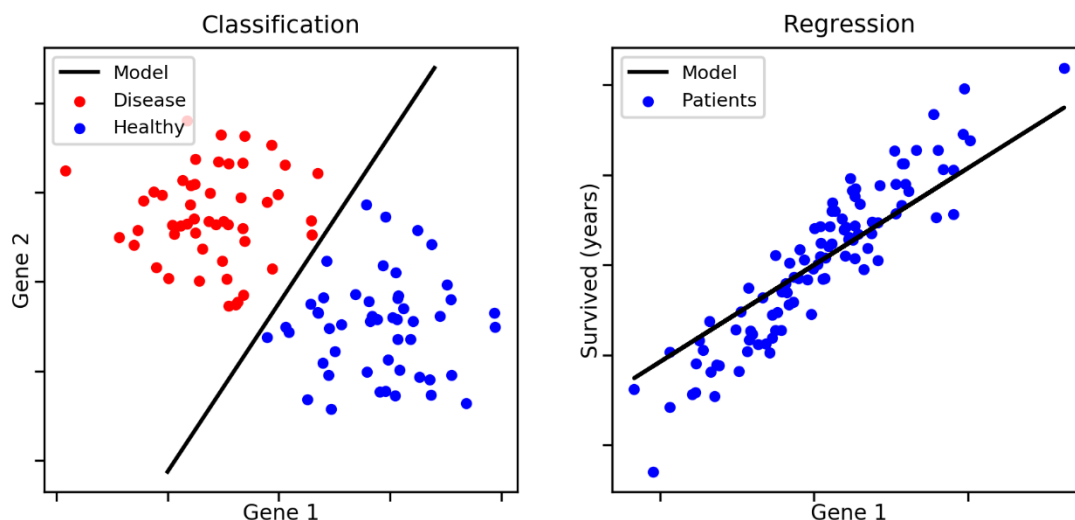
- **Clustering:** A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.
- **Association:** An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Classification and Regression are two major prediction problems which are usually dealt in Data mining. Predictive modelling is the technique of developing a model or function using the historic data to predict the new data. The significant difference between Classification and Regression is that classification maps the input data object to some discrete labels. On the other hand, regression maps the input data object to the continuous real values.

Classification technique provides the predictive model or function which predicts the new data in discrete categories or labels with the help of the historic data. Conversely, the regression method models

continuous-valued functions which means it predicts the data in continuous numeric data.

1. The Classification process models a function through which the data is predicted in discrete class labels. On the other hand, regression is the process of creating a model which predict continuous quantity.
2. The classification algorithms involve decision tree, logistic regression, etc. In contrast, regression tree (e.g. Random forest) and linear regression are the examples of regression algorithms.
3. Classification predicts unordered data while regression predicts ordered data.
4. Regression can be evaluated using root mean square error. On the contrary, classification is evaluated by measuring accuracy.



## Clustering:

It is basically a type of [unsupervised learning method](#). An unsupervised learning method is a method in which we draw references from datasets consisting of input data without labeled responses. Generally, it is used as a process to find meaningful structure, explanatory

underlying processes, generative features, and groupings inherent in a set of examples.

**Clustering** is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.

Clustering is a Machine Learning technique that involves the grouping of data points. Given a set of data points, we can use a clustering algorithm to classify each data point into a specific group. In theory, data points that are in the same group should have similar properties and/or features, while data points in different groups should have highly dissimilar properties and/or features. Clustering is a method of unsupervised learning and is a common technique for statistical data analysis used in many fields.

