Machine learning is a science, method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that system can learn from data, identify patterns and make decisions with minimal human intervention. Types of ML:

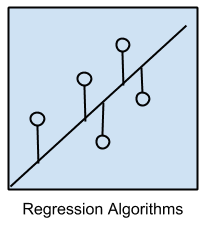
* Supervised Learning;
* Semi-supervised;
* Unsupervised Learning;
* Reinforcement Learning.

**Supervised Learning**

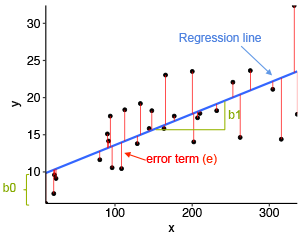
Supervised Learning algorithms model the relationship between features (independent variables) and a label (target) given a set of observations. Then the model is used to predict the label of new observations using the features.

### 1.1 **Regression algorithms**

### Prediction, forecasting, process Optimization, New insights Regression Algorithms

Regression analysis is a predictive modelling technique which investigates the relationship between a dependent (target) and independent variable(s) (predictor). Regression is concerned with modeling the relationship between variables that is iteratively refined using a measure of error in the predictions made by the model.

Regression methods are a workhorse of statistics and have been co-opted into statistical machine learning.

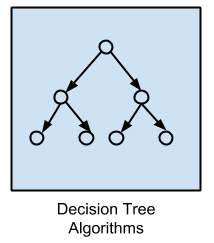
 1.1.1 Linear Regression algorithm.

For statistical technique linear regression is used in which value of dependent variable is predicted through independent variables. Linear regression is a **supervised**learning algorithm and tries to model the relationship between a **continuous**target variable and one or more independent variables by fitting a linear equation to the data.

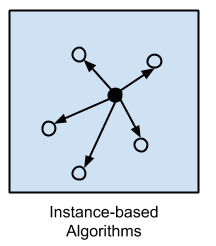
Different regression models differ based on – the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used.

1.1.2 Decision Tree algorithm

The aim of the decision tree algorithm is to increase the predictiveness as much as possible at each partitioning so that the model keeps gaining information about the dataset.

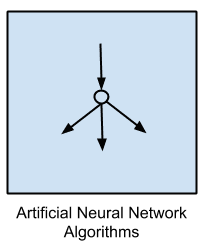
It achieves high accuracy with training set but performs poorly on new, previously unseen data points which indicates overfitting. The depth of a tree is controlled by max\_depth parameter for decision tree algorithm in scikit-learn. It is suitable to work on a mixture of feature data types (continuous, categorical, binary). Decision Tree

is a technique for approximating discrete valued target function which represents the learnt function in the form of a decision tree. Decision trees are built for making a training model which can be used to predict class or the value of target variable.

 1.1.3 Random Forest

Random forest is an ensemble of many decision trees. Random forests are built using a method called bagging in which decision trees are used as parallel estimators. If used for a classification problem, the result is based on majority vote of the results received from each decision tree. Random forests reduce the risk of overfitting and accuracy is much higher than a single decision tree. There are two stages in Random Forest Algorithm, one is to create random forest, and the other is to make a prediction from the random forest classifier created in the first stage

1.1.4 Neural Network



Artificial Neural Networks are models that are inspired by the structure and/or function of biological neural networks.

They are a class of pattern matching that are commonly used for regression and classification problems but are really an enormous subfield comprised of hundreds of algorithms and variations for all manner of problem types.

Note that I have separated out Deep Learning from neural networks because of the massive growth and popularity in the field. Here we are concerned with the more classical methods.

**1.2 Classification** (discrete target variable task) -Diagnostics, Fraud detection, Email Sam Detection, Image Classification. Observations (or data points) in a classification task have labels. Each observation is classified according to some measurements. Classification algorithms try to model the relationship between measurements (features) on observations and their assigned class. Then the model predicts the class of new observations.

1.2.1 Logistic Regression

Risk Assessment, Score Prediction. In logistic regression we have lot of data whose classification is done by building an equation. This method is used to find the discrete dependent variable from the set of independent variables. Its goal is to find the best fit set of parameters and it is commonly used for many binary classification tasks. Customer churn, spam email, website or ad click predictions are some examples of the areas where logistic regression offers a powerful solution.

1.2.2 Support Vector Machine

SVM is mostly used for classification tasks but it is also suitable for regression tasks. SVM distinguishes classes by drawing a decision boundary. How to draw or determine the decision boundary is the most critical part in SVM algorithms.

1.2.3 Naive Bayes classifier

Naïve Bayes classifier is a  simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.It only needs small amount of training data for classification, and all terms can be precomputed thus classifying becomes easy, quick and efficient.

1.2.4 Instance-based Algorithms

Also known as Memory based learning, performs operation after comparing the current instances with the previously trained instances, which have been stored in memory.

Such methods typically build up a database of example data and compare new data to the database using a similarity measure in order to find the best match and make a prediction. For this reason, instance-based methods are also called winner-take-all methods and memory-based learning. Focus is put on the representation of the stored instances and similarity measures used between instances.

The most popular instance-based algorithms are:

* k-Nearest Neighbor (kNN)
* Learning Vector Quantization (LVQ)
* Self-Organizing Map (SOM)
* Locally Weighted Learning (LWL)
* Support Vector Machines (SVM)

2**.Unsupervised**- Unsupervised algorithms try to find the structure in unlabeled data. Input data is not labeled and does not have a known result.

2.1 K-Means Clustering

Recommended system, Targetted Marketing, Customer Segmentation.  It is used to overcome the limitation of clustering. Its goal is to partition ‘n’ observations into ‘k’ clusters in which each observation belongs to the cluster having the nearest mean, serving as a prototype of the cluster. The mean of the observations in a particular cluster defines the center of the cluster. Clustering is a way to group a set of data points in a way that similar data points are grouped together. Therefore, clustering algorithms look for similarities or dissimilarities among data points. Clustering is an unsupervised learning method so there is no label associated with data points. K-means clustering tries to minimize distances within a cluster and maximize the distance between different clusters.

2.2 Dimensionality Reduction

Elicitation, big data visualization, text mining face recognition, image recognition.

It is used to reduce the number of random variables by obtaining some principal variables. Feature extraction and feature selection are types of dimensionality reduction method. It can be done by PCA, Principal component analysis is a method of extracting important variables from large set of variablesDimensionality reduction is a learning technique used when the number of features  (or dimensions) in a given dataset is too high. It reduces the number of data inputs to a manageable size while also preserving the data integrity. Often, this technique is used in the preprocessing data stage, such as when autoencoders remove noise from visual data to improve picture quality.

2.3 Principal Component Analysis (PCA)

PCA is a dimensionality reduction algorithm which basically derives new features from the existing ones features by finding the relations among features within a dataset

PCA is a linear dimensionality reduction algorithm. There are also non-linear methods available.The advantage of PCA is that a significant amount of variance of the original dataset is retained using much smaller number of features than the original dataset. Principal components are ordered according to the amount of variance they explain.

**3.Semi-supervised**

Semi-supervised learning is a medium between supervised and unsupervised learning, where you use a training dataset with both labeled and unlabeled data. It’s particularly useful when it’s difficult to extract relevant features from data — and when you have a high volume of data.Semi-supervised learning is ideal for medical images, where a small amount of training data can lead to a significant improvement in accuracy. For example, a radiologist can label a small subset of CT scans for tumors or diseases so the machine can more accurately predict which patients might require more medical attention.

**4.Renforcement Learning**

Real time decisions, Game AI, Robot Navigation, Finance sector, Inventory, Manufacturing, Management.

**Reinforcement learning** works based on an action-reward principle. An **agent**learns to reach a goal by iteratively calculating the **reward**of its actions. Input data as feedback to the model, emphasizing how to act based on the environment to maximize the expected benefits. The difference between supervised learning is that it does not require the correct input/output pairs and does not require precise correction of sub-optimal behavior. Reinforcement learning is more focused on online planning and requires a balance between exploration (in the unknown) and compliance (existing knowledge).