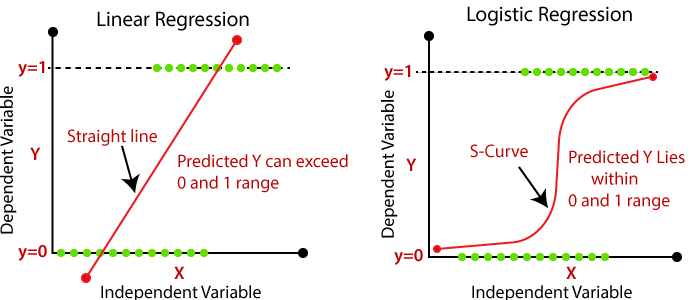
1. Linear Regression

* **Linear Regression** is a machine learning algorithm based on **supervised learning**. It performs a **regression task**.
* Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression.



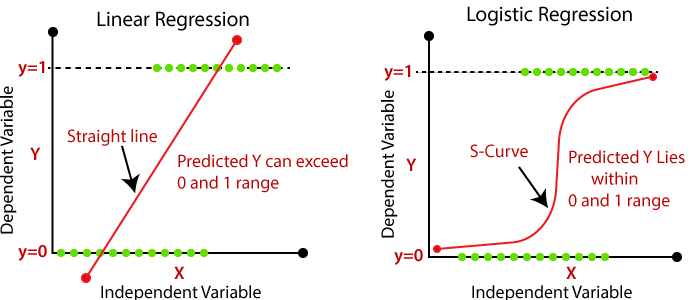
As the simple linear regression equation explains a correlation between 2 variables (one independent and one dependent variable), it is a basis for many analyses and predictions. Apart from business and data-driven marketing, LR is used in many other areas such as analysing data sets in statistics, biology or machine learning projects etc.

* In statistics, linear regression is usually used for predictive analysis. It essentially determines the extent to which there is a linear relationship between a dependent variable and one or more independent variables.

1. **Predicting rainfall for the year**: Weather prediction is not only limited to rainfall but predicting the weather/rainfall is the objective of this use case.
2. **Real estate value prediction**: The house price according to the features of every home is the objective of this use case.
3. **Forest Fire area prediction**: Every year in Africa and Australia there is forest fire spread across. Thus, predicting the area for the year based on the previous data.
4. **Car price prediction**: Either car is being launched or sold the car price prediction helps companies to come up with an appropriate number.
5. **Stock market prediction**: Estimating the rise and fall of stock prices is the objective of this use case.

2.LOGISTIC REGRESSION

* Logistic regression is basically a **supervised classification** algorithm.
* In a classification problem, the target variable (or output), y, can take only discrete values for a given set of features (or inputs), X.
* Logistic regression is a classification algorithm. It is used to predict a binary outcome based on a set of independent variables.
* A **binary outcome** is one where there are only two possible scenarios—either the event happens (1) or it does not happen (0). **Independent variables** are those variables or factors which may influence the outcome (or dependent variable).
* Logistic regression is the correct type of analysis to use when you’re working with binary data. You know you’re dealing with binary data when the output or dependent variable is dichotomous or categorical in nature; in other words, if it fits into one of two categories (such as “yes” or “no”, “pass” or “fail”, and so on).

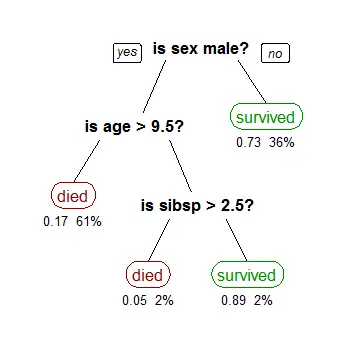


* Predicting weather: you can only have few definite weather types. Stormy, sunny, cloudy, rainy and a few more.
* Medical diagnosis: given the symptoms predict the disease patient is suffering from.

3.DECISION TREE

* Decision trees come under the **supervised learning** algorithms category.
* Decision tree machine learning algorithm in python is helpful in predicting the value in nonlinear situation. Decision tree algorithm divides the data like the tree branches and make it possible to do better prediction.
* A **decision tree** is a graphical representation of possible solutions to a decision based on certain conditions. It's called a decision tree because it starts with a single box (or root), which then branches off into a number of solutions, just like a tree.

1. **Classification trees: -** In this type of decision tree, there is only one outcome from a set of two. The outcome could be either true for a particular dataset or false. The decision variable is categorical. In this decision tree classifier, categorical classifiers are essential.
2. **Regression trees: -**In this type of decision tree for machine learning algorithms, the outcome is continuous, and changes based on the value of variables in the dataset. It is one of the types of decision trees that have continuous target variables. It is also known as a continuous variable decision tree.



* + Screening potential terrorists and drug smugglers at border crossings

4 SVM SUPPORT VECTOR MACHINE

* A support vector machine (SVM) is a **supervised**[**machine learning**](https://monkeylearn.com/machine-learning/)model that uses [classification algorithms](https://monkeylearn.com/blog/machine-learning-algorithms/) for two-group classification problems. After giving an SVM model sets of labelled training data for each category, they’re able to categorize new text.
* The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points.
* **Face detection** – SVM classify parts of the image as a face and non-face and create a square boundary around the face.
* **Text and hypertext categorization** – SVMs allow Text and hypertext categorization for both inductive and transudative models. They use training data to classify documents into different categories. It categorizes on the basis of the score generated and then compares with the threshold value.
* **Classification of images** – Use of SVMs provides better search accuracy for image classification. It provides better accuracy in comparison to the traditional query-based searching techniques.
* **Bioinformatics** – It includes protein classification and cancer classification. We use SVM for identifying the classification of genes, patients on the basis of genes and other biological problems.

5 NAÏVE BAYES

Naive Bayes is a probabilistic algorithm that’s typically used for classification problems. It uses **Conditional probability, which**is a measure of the probability of an event occurring given that another event has (by assumption, presumption, assertion, or evidence) occurred. It is simple, intuitive, and yet performs surprisingly well in many cases. It is based on Bayes’ Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

* **Playing Cards Example**
* **Coin Toss and Fair Dice Example**

**Knn k - nearest neighbours**

KNN is a non-parametric and lazy learning algorithm. Non-parametric means there is no assumption for underlying data distribution. In other words, the model structure determined from the dataset. This will be very helpful in practice where most of the real world datasets do not follow mathematical theoretical assumptions. Lazy algorithm means it does not need any training data points for model generation. All training data used in the testing phase. This makes training faster and testing phase slower and costlier. Costly testing phase means time and memory. In the worst case, KNN needs more time to scan all data points and scanning all data points will require more memory for storing training data.

* In pattern recognition, the k-nearest neighbours algorithm (k-NN) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space.