**1.Introduction**

Supervised learning is the task of inferring a function from labeled training data. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object(X) (typically a vector) and a desired output value(Y). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. What we try to achieve with machine learning is to find the **true** relationship between them, what we usually call **the model** in math. There are many different algorithms in machine learning that allow us to obtain a model of the data.

Supervised learning problems are mainly classified into two types namely classification and regression

* **Classification:** A classification problem is when the output variable(Y) is categorical variable, such as “gender” or “country”. Classification tasks further classified as follows
* Binary Classification: output variable has only two possible values Eg: Gender classification (Male / Female)
* Multi-class classification: output variable has two or more possible values. Eg: mother tongue(telugu/tamil/hindi/English)
* Multi-label classification: Classification task where each sample is mapped to more than one class. Eg: A news article can be about sports, a person, and location at the same time.
* **Regression:** A regression problem is when the output variable is a real variable, such as “salary” or “weight”.

Different types of classifiers are Perceptron, Naive Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbor, Artificial Neural Networks/Deep Learning, Support Vector Machine

Different types of regression are Linear regression, Logistic regression, Polynomial regression, Stepwise regression, Ridge regression, Lasso regression, ElasticNet regression

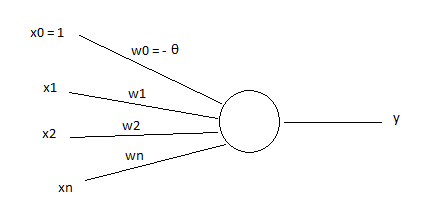
**2. Classification**

A classification problem is when the output variable(Y) is categorical variable, such as “gender” or “country”. Different types of classifiers are Perceptron, Naive Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbor, Artificial Neural Networks/Deep Learning, Support Vector Machine.

**2.1.Perceptron**

**2.1.1.What is perceptron**

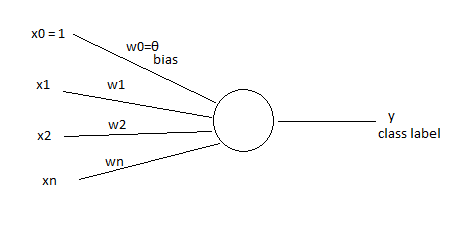
* Perceptron is a computational model which takes an input, aggregates it (weighted sum) and returns 1 only if the aggregated sum is more than some threshold (θ) else returns 0.
* Let x = [x1,x2,x3,------,xn] be an input vector, w = [w1,w2,w3,--------,wn] be a weight vector (initialized randomly), then perceptron will produce
  + 1 if
  + 0 if



* + Y=1 if
  + Y=0 if where x0 = 1, w0 = -, w0 is called as bias.
* For a Perceptron, input values can be real or ordinary.
* A single Perceptron can be used to implement only linearly separable function.

**2.1.2.Perceptron for classification problem**

* Let the training data consists of ‘m’ number of inputs and the corresponding class labels
* Let the class labels be
  + 1- positive class
  + 0- negative class
* Let each input has independent attributes that is x = [x1,x2,------,xn].
* Design a perceptron with initially random weights w1, w2, -------, wn and bias θ



* Now, our goal is to find w = [w1, w2, ---------, wn] , which classifies all samples in training data correctly.
* We update w by using the following perceptron learning algorithm.

**2.1.3.Perceptron Learning Algorithm**

Let P – set of all inputs with class label 1 and

N - set of all inputs with class label 0

While (! convergence) do

{

Pick a random x ∈P∪N

If (x ∈ P & w . x <0) then

w=w+x;

if(x ∈ N & w . x >= 0) then

w=w-x;

}

Note:If the sets P, N are finite and linearly separable, then the perceptron learning algorithm converges after a finite number of iterations.

**2.2 Naive-Bayes**

**2.2.1.What is a bayes theorem?**

For any two random variables X, Y

P(Y/X) =

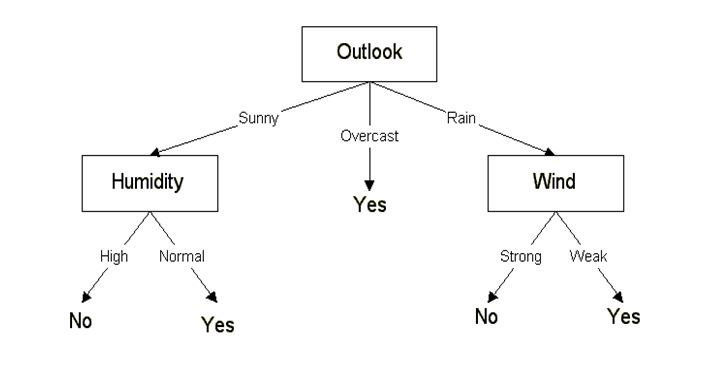
2.2.2.Baye’s classifier for simple classification problem

* Let x be a independent variable and y be a dependent variable
* Let y can take values y1,y2, -------,yk and x can take values x1,x2,--------,xd
* From the training data, first we calculate P(y=yj) for j=1,2,---,k and P(x=xi/y=yj) for i=1,2,-----,d and j=1,2,-----,k
* Now for any test record x for a given value , we find the corresponding class label as follows
* Calculate P(y=yj/x=xi) = (P(y=yj)P(x=xi/y=yj))/P(x=xi) for every j= 1,2,------,k
* Choose the class label j which maximizes P(y=yj/x=xi)

**2.3.Decision Tree**

A decision tree is a [decision support](https://en.m.wikipedia.org/wiki/Decision_support_system) tool that uses a [tree-like](https://en.m.wikipedia.org/wiki/Tree_(graph_theory)) [model](https://en.m.wikipedia.org/wiki/Causal_model) of decisions and their possible consequences, including [chance](https://en.m.wikipedia.org/wiki/Probability) event outcomes, resource costs, and [utility](https://en.m.wikipedia.org/wiki/Utility).

Decision tree learning is a method commonly used in classification task. The goal is to create a model that predicts the value of a target variable based on several input variables. An example is shown in the following diagram.



Each [interior node](https://en.m.wikipedia.org/wiki/Interior_node) corresponds to one of the input variables and there are edges to children for each of the possible values of that input variable. Each leaf represents a value of the target variable given the values of the input variables represented by the path from the root to the leaf.

**2.4.Logistic Regression**

**2.4.1.What is logistic function?**

* A Logistic function is a common S shape with equation f(x)=L/(1+e-k(x-x0)), where
  + L – The curves maximum value
  + K – The steepness of the curve
  + – The x- value of the sigmoid’s point
* Standard logistic function is given by f(x) = 1/(1+e-x) = /(1+ex)

**2.4.2.Logistic Regression for Classification**

* Let the training data consists of ‘m’ number of inputs and the corresponding class labels
* Let each input x has n independent attributes that is x = [x1, x2, ------,xn]
* Let the class labels be
  + 1 – Positive Class
  + 0 – Negative Class
* Design a sigmoid (logistic) neuron with initially random weights w1, w2, ------, wn and bias b which implements f(x) = b+w1x1+w2x2+---------+wnxn and = 1/(1+)
* Assign classes 1 and 0 to y^ as follows
  + =1 if >= 0.5
  + =0 if <0.5

**Learning Rule for Sigmoid (Logistic)**

Neuron:

* We update w by minimising the loss function

L(w) = -1/m

Where

y(i) – Actual class label ith input

– estimated class label of ith input

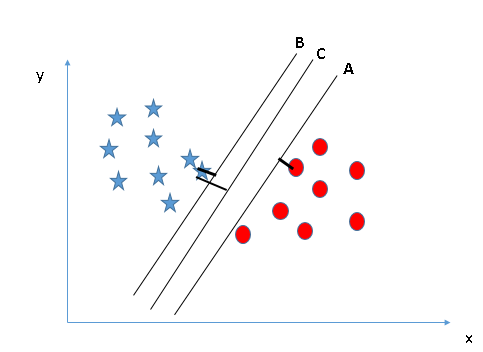
* Logistic regression can be applied only to binary classification problem

**2.5. k-nearest neighbour**

In k-nearest neighbour classification, for the given input vector, we compute the k nearest data points and we choose the class label which mostly appears among those k nearest neighbours

**2.6. Support vector machine**

Support Vector Machine” (SVM) is a supervised [machine learning algorithm](https://courses.analyticsvidhya.com/courses/introduction-to-data-science-2?utm_source=blog&utm_medium=understandingsupportvectormachinearticle) which can be used for classification. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the following image).



**3. Regression**

A regression problem is when the output variable is a real variable, such as “salary” or “weight”. Different types of regression are Linear regression, Logistic regression, Polynomial regression, Stepwise regression, Ridge regression, Lasso regression, ElasticNet regression

**3.1.Linear Regression**

Linear regression is used for finding linear relationship between target variable and one or more independent variables(predictors). There are two types of linear regression. There are two types of linear regression- Simple and Multiple.

Simple linear regression is useful for finding relationship between two continuous variables. One is predictor or independent variable and other is response or dependent variable by using the parametric equation y=a+bx

Multiple linear regression is useful for finding relationship between one dependent variable and two or more independent variable by using the parametric equation y=a+++.....+.

**3.2. Logistic Regression**

Logistic regression is used to find the probability of target class given the input feature values. logistic regression is mostly used when the dependent variable is binary (0/ 1, True/ False, Yes/ No) in nature, which is in generally called a classification problem. If the values of dependent variable is ordinal, then it is called as Ordinal logistic regression. If dependent variable is multi class then it is known as Multinomial Logistic

**3.3. polynomial Regression**:

In ploynomial regression, we use the parametric regression equation in which the power of independent variable is more than one like y=a+bx+c

**3.4. Stepwise regression**

Stepwise regression is used when we deal with multiple independent variables. The choice of dependent variables is carried out by an automatic procedure. In each step, a variable is considered  based on some pre specified criterion. Usually, this takes the form of a sequence of [F-tests](https://en.wikipedia.org/wiki/F-test) or [t-tests](https://en.wikipedia.org/wiki/T-test)