**ALGORITHM HOME LEARNING.7**

**LINEAR REGRESSION**

1. Linear regression is an unsupervised learning algorithm which means that the output of the model is known in advance.
2. Linear regression is the linear approximation of a variable to find the best fit line that accurately represents the relation of the dependent variable (x) and independent variable (y). Predicting a continuous quantity, ŷ =B0+B1·X1 ( ŷ the output to be predicted) (B0 intercept on the y-axis) (B1 slope of the line X1). Linear regression is to find the best fit line, so a new point can be predicted.
3. The linear Regression algorithm is useful when there is a continuous relationship between the inputs and output.
4. Medical researchers often use linear regression to understand the relationship between drug dosage and blood pressure of patients.
5. For example, researchers might administer various dosages of a certain drug to patients and observe how their blood pressure responds. They might fit a simple linear regression model using dosage as the predictor variable and blood pressure as the response variable. The regression model would take the following form:

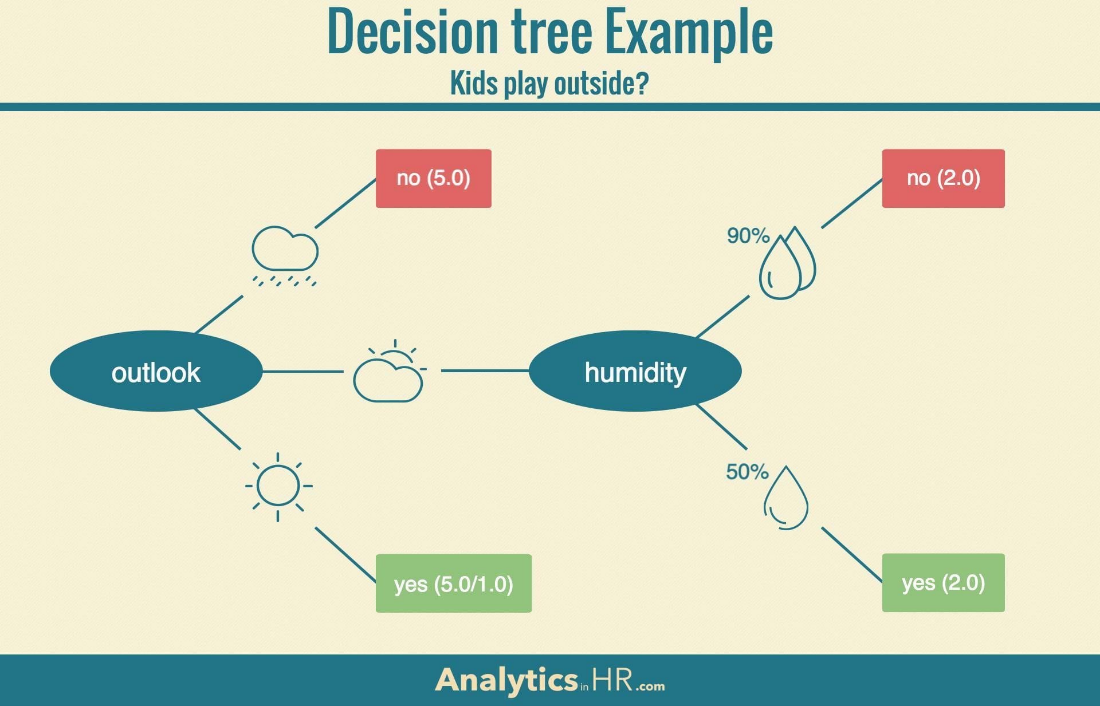
blood pressure = β0 + β1(dosage)

The coefficient β0 would represent the expected blood pressure when dosage is zero. The coefficient β1 would represent the average change in blood pressure when dosage is increased by one unit.

**LOGISTIC REGRESSION**

1. Logistic regression is a supervised learning classification algorithm.
2. It is used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no). Mathematically, a logistic regression model predicts P(Y=1) as a function of X.
3. Logistic regression is most usewhen the data in question has binary output, so when it belongs to one class or another, or is either a 0 or 1.
4. It is one of the simplest ML algorithms that can be used for various classification problems such as spam detection, Diabetes prediction, cancer detection etc.

**DECISION TREE**

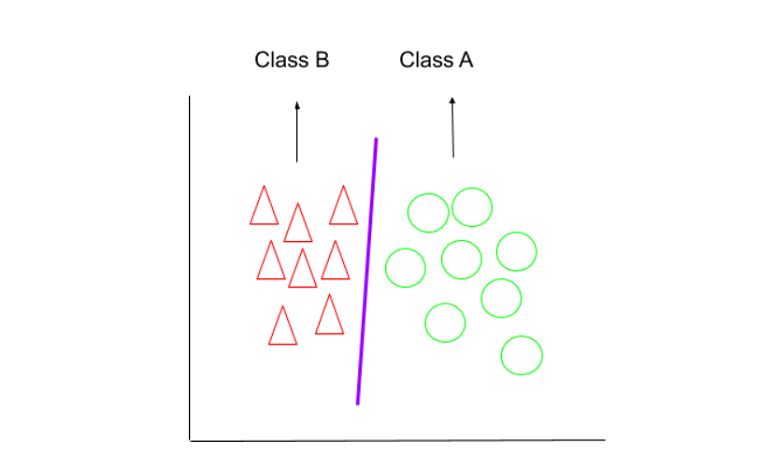
1. Decision trees can be used for supervised AND unsupervised learning. Yes, even with the fact that a decision tree is per definition a supervised learning algorithm where you need a target variable, they can be used for unsupervised learning, like clustering. Decision trees can easily handle unbalanced datasets.
2. It is like tree it also starts with root node and expands further into branches. There are two nodes. Decision Nodes are used for making decisions and have multiple branches while Leaf Nodes have the output of these decisions. Leaf Node does not have any other further branches. It simply asks a question and gets split further into subtree based on answer yes or no. It basically a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.
3. It is used for both linear and non-linear data. The decision tree uses the CART algorithm which stands for Classification and Regression Trees. The aim of using a decision tree is to create a training model which can be used for predicting the class or values of the target variable just by learning a simple decision rule inferred from prior data or training data.
4. 

**SUPPORT VECTOR MACHINE**

1. In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis.
2. Support vector machine is a [machine learning](https://www.analyticssteps.com/blogs/introduction-machine-learning-supervised-and-unsupervised-learning) model that can generalize between two different classes if the set of labelled data is provided in the training set to the algorithm. The main function of the SVM is to check for that hyperplane that can distinguish between the two classes. There can be many hyperplanes that can do this task, but the objective is to find that hyperplane that has the highest margin that means maximum distances between the two classes, so that in future if a new data point comes that is two be classified then it can be classified easily. Let us understand the working of SVM by taking an example where we have two classes that are shown in the below image which are a class A: Circle & class B: Triangle. Now, we want to apply the SVM algorithm and find out the best hyperplane that divides both classes. SVM takes all the data points into consideration and gives out a line that is called ‘*Hyperplane*’ which divides both the classes. This line is termed as ‘*Decision boundary*’. Anything that falls in circle class will belong to class A and vice versa.
3. Support Vector Machine Algorithm Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning

Diagram

Description automatically generated



**NAIVE BAYES**

1. Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.
2. Naïve Bayes Algorithm consists of two words: Naïve and Bayes, Naïve because it assumes occurrence of a certain feature is independent of occurrence of other features and Bayes because it depends on Bayes’ Theorem. This algorithm can effectively solve many complex problems such as text classifiers than the much-hyped neural networks. The model works well with insufficient and mislabeled data. Probability is a field of math that helps us to reason about uncertainty and calculates the likelihood of some events. When we work with a predictive machine learning model, such as Naïve Bayes Algorithm, we must predict an uncertain future. Classification is the most used form of prediction. For binary classification, prediction results in classification to 0 or 1, such as spam or not spam. In the case of multiclass classification, it aims to predict the class of record to wide variety of classes rather than 0 or 1. The Naïve Bayes machine learning algorithm is one of the methods to deal with uncertainty with the help of probabilistic methods. When dealing with classification problems in supervised learning, we use labelled data where the target class of the records is known. We use these data to train the model using these data and apply this trained model to new data where classification must be done.
3. It is used in *text classification* that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building 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. Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

**KNN (K NEAREST NEIGHBOUR)**

1. K Nearest Neighbor (KNN) algorithm is basically a classification algorithm in Machine Learning which belongs to the supervised learning category. However, it can be used in regression problems as well. KNN algorithms have been used since 1970 in many applications like pattern recognition, data mining, statistical estimation, and intrusion detection, and many more. It is widely disposable in real-life scenarios since it is non-parametric, i.e., it does not make any underlying assumptions about the distribution of data.
2. In K-NN the whole data is classified into training and test sample data. In a classification problem, k nearest algorithm is implemented using the following steps. Pick a value for k, where k is the number of training examples in the feature space. Calculate the distance of unknown data points from all the training examples. Search for the k observations in the training data that are nearest to the measurements of the unknown data point. Calculate the distance between the unknown data point and the training data. The training data which has the smallest value will be declared as the nearest neighbor. In the KNN-regression problem, the only difference is that the distance between training points and sample points is evaluated and the point with the lowest average distance is declared as the nearest neighbor. It predicts the result based on the average of the total sum.
3. KNN is most useful when labeled data is too expensive or impossible to obtain, and it can achieve high accuracy in a wide variety of prediction-type problems. KNN is a simple algorithm, based on the local minimum of the target function which is used to learn an unknown function of desired precision and accuracy.
4. Example: Suppose, we have an image of a creature that looks like cat and dog, but we want to know either it is a cat or dog. So, for this identification, we can use the KNN algorithm, as it works on a similarity measure.

**K MEANS**

1. K-means clustering is one of the unsupervised algorithms where the available input data does not have a labeled response.
2. K-Means performs the division of objects into clusters that share similarities and are dissimilar to the objects belonging to another cluster. The term ‘K’ is a number. You need to tell the system how many clusters you need to create. For example, K = 2 refers to two clusters. There is a way of finding out what is the best or optimum value of K for a given data.

1. K-means algorithm is extremely popular and used in a variety of applications such as market segmentation, document clustering, image segmentation and image compression, etc. The goal usually when we undergo a cluster analysis is either: Get a meaningful intuition of the structure of the data we are dealing with. Cluster-then-predict where different models will be built for different subgroups if we believe there is a wide variation in the behaviors of different subgroups. An example of that is clustering patients into different subgroups and building a model for each subgroup to predict the probability of the risk of having a heart attack.
2. For a better understanding of k-means, let's take an example from cricket. Imagine you received data on a lot of cricket players from all over the world, which gives information on the runs scored by the player and the wickets taken by them in the last ten matches. Based on this information, we need to group the data into two clusters, namely batsman and bowlers.

**RANDOM FOREST**

1. Random forest is a supervised learning algorithm. it uses labeled data to “learn” how to classify unlabeled data.
2. The general idea of the bagging method is that a combination of learning models increases the overall result. Put simply: random forests build multiple decision trees and merge them together to get a more accurate and stable prediction.
3. Although it is not as easily explainable as its underlying algorithm decision tree regression, random forests can be inspected to output the decision trees which were used in the final decision. The individual trees can be used to understand what the important decision nodes were, as well as prompt questions around what led to the final prediction
4. Decision tree algorithms are perfect for busy companies. Random forest algorithms rely heavily on decision trees so if your business depends on decision trees, then random forest algorithms are perfect for you. Random forest is well suited to basic regression problems and basic classification problems