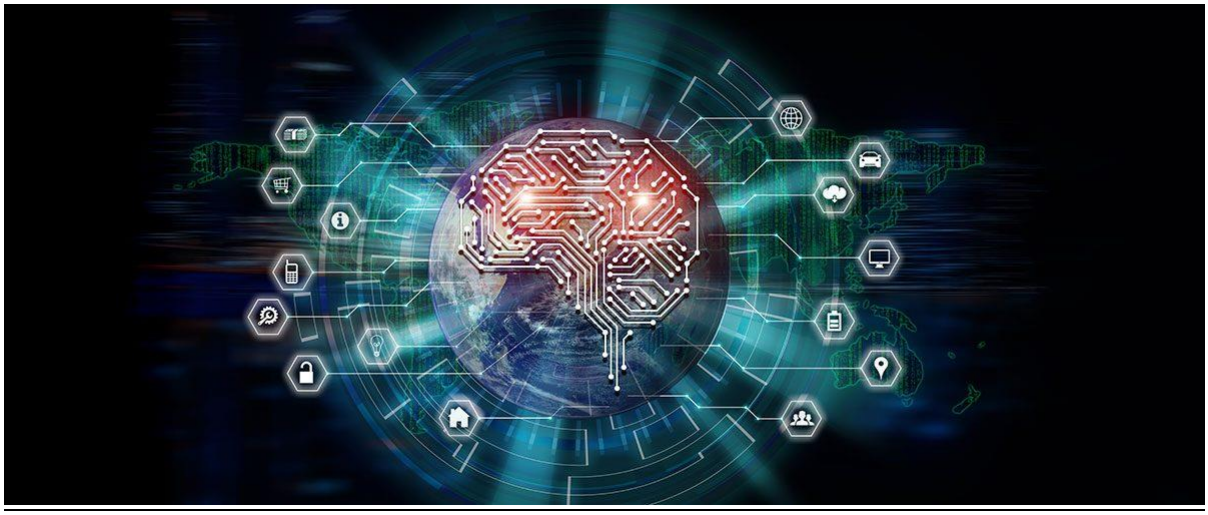

MACHINE LEARNING ALGORITHMS

Machine learning(ML) is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. This is accomplished through building algorithms that can help models learn and respond to their intended use. Machine learning implementations are classified depending on the nature of the learning response available to a learning system. There are a wide range of Machine Learning algorithms in the market. These machine learning algorithms are implemented in python.

Machine Learning Algorithm Types:

List of Machine Learning Algorithms is of a never ending one.
There are 3 types of ML algorithms:

1. Supervised learning:
   Supervised learning uses labeled training data to learn the factor of input that can map the input to the output.
   These algorithms are mainly grouped based on machine learning algorithms for classification and machine learning algorithms for regression.

- Classification: To foresee and predict the result of a given input where the output is a class. For example, labels such as cat and dog, good and bad .

- Regression: To foresee and predict the result of a given input where the output is in the form of real values, machine learning algorithms operating on regression are used. Examples include real-valued labels denoting the amount of rainfall, the height of a person.

- Ensembling: A collaborative approach wherein the predictions of many weak ML models are taken into consideration to predict on a new output.

2. Unsupervised learning:

Unsupervised learning makes use of only the input variables, it doesn't take into account the outputs of the corresponding inputs. It uses unlabeled training data to deploy the structured model.

Machine Learning Algorithms Classification in Unsupervised is as follows,

- Association: Used to predict the interesting relations and probability of co-occurrence between variables in a large collection. It is widely used in market-basket analysis. For example,if a customer purchases notebook, he is 80% likely to also purchase a pen.

- Clustering: Clustering is all about grouping things. Samples that are similar are grouped in same clusters, these groups should in contrast to each other to the widest extent possible. For example, it can be used to characterize & discover customer segments for marketing purposes.

- Dimensionality Reduction: Minimizing the number of input variables of a dataset while guaranteeing that significant data is still passed on. For example, a characterizataion problem that heavily depends on both humidity and rainfall can be broken-down

into just one underlying feature. Hence, we can reduce the number of features in problems of this category.

3. Reinforcement learning:

   Reinforcement learning is a machine learning algorithm, all about taking suitable action to maximize reward in a particular situation. Reinforcement algorithms usually learn optimal actions through trial and error. The machine learning algorithm example for this type is – where a robot can learn to avoid collisions by receiving negative feedback after bumping into obstacles.

Supervised learning algorithms:

1. Linear Regression:

   Linear Regression is used everywhere from stock market to research teams. Some use cases are:

   1. weather data analysis.
   2. In market research studies and customer survey results analysis.
   3. Studying engine performance.

   In Linear algebra, the relationship between the input variables (x) and output variable (y) is expressed as an equation of the form y = mx + c, the equation implies even for linear regression. The goal of linear regression is to find out the values of coefficients a and b. Here, ‘c’ is the intercept and ‘m’ is the slope of the line.

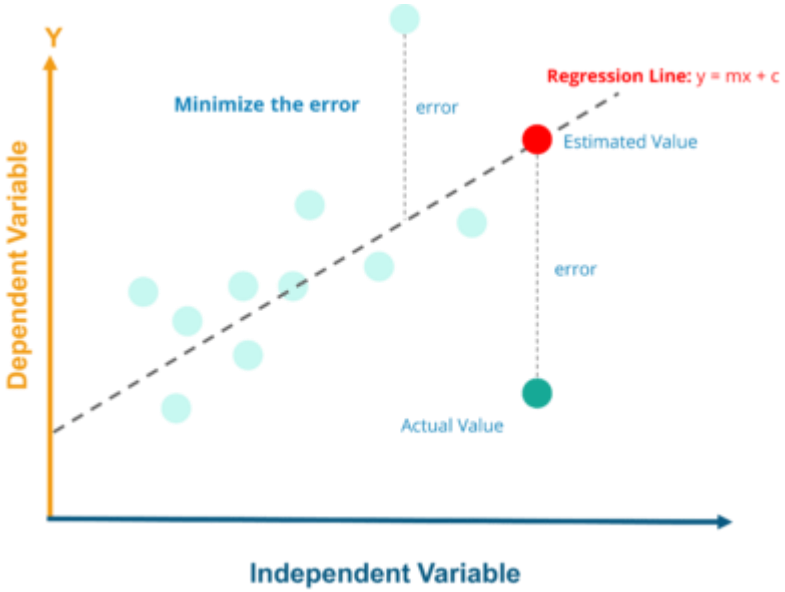

Fig1 : Linear Regression

Figure1 shows the plot between x and y values for a given dataset. The aim is to fit a line that is nearest to most of the data points, this would reduce the distance/error between the y value of a data point and the line.

2. Logistic Regression

It is named after the transformation function used, called the logistic function given by f(x)= 1/ (1 + exponential(-x)), which is an S-shaped curve.

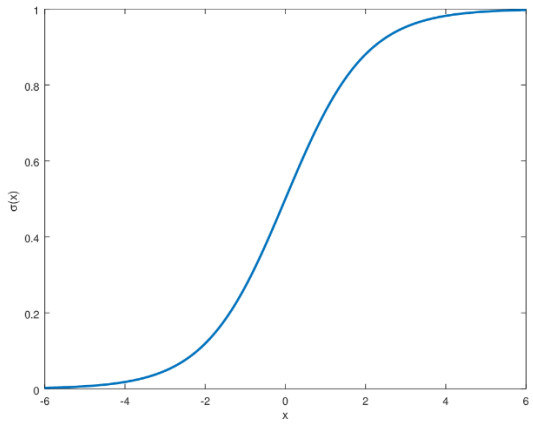

Fig2 : Logistic Regression

Linear regression works on continuous values while logistic regression works on discrete values.

Logistic regression is best suited for binary classification i.e. target variable is either 0 or 1. For example, in predicting whether an event will occur or not, the event that it occurs is classified as 1. In predicting whether a coin toss result would be head or tales, the head instances are denoted as 1).

In logistic regression, the output is in the form of probabilities of the default class. As it depends on probability, the output lies in the range 0-1. The output value is generated by applying an activation/transformation function to x-value, using the logistic mentioned above.

3. Naive Bayes Classification:

This algorithm works on Bayes Theorem. Used to calculate the probability that an event will occur, given that another event has already occurred.

To calculate the probability of an outcome given the value of some variable, that is

$P(x|y) = (P(y|x) * P(x)) / P(y)$

- We are trying to find probability of event x, given the event y is true. Event y is also termed as **evidence**.
- P(x) is the **priori** of x, i.e., probability of event before evidence is seen. The evidence is an attribute value of an unknown instance, here y.
- P(x|y) is a posteriori probability of y, i.e. probability of event after evidence is seen.

This algorithm is named 'Naive' as it assumes that all the variables are independent of each other.

WEATHER	TRAVEL
Windy	No
Cold	Yes
Sunny	Yes

Windy	Yes
Windy	Yes
Cold	Yes
Sunny	No
Sunny	No
Windy	Yes
Sunny	Yes
Windy	No
Cold	Yes
Cold	Yes
Sunny	No

Let us consider the above table an example of machine learning algorithm of this type, we need to find what is the outcome if weather='windy'?

To determine the outcome travel= 'yes' or 'no' given the value of variable weather='windy', calculate P(yes|windy) and P(no|windy) and choose the outcome with higher probability.

->P(yes|windy)= (P(windy|yes) * P(yes)) / P(windy)

= (3/9 * 9/14) / (5/14)

= 0.60

-> P(no|windy)= (P(windy|no) * P(no)) / P(windy)

= (2/5 * 5/14) / (5/14)

= 0.40

Thus, if the weather ='windy', the outcome is travel= 'yes'.

4. Decision Tree:

These algorithms can be used to solve both regression and classification problems. Decision tree uses the tree representation to solve the problem.

- Each leaf node corresponds to a class label and attributes are represented on the internal node of the tree.
- We can represent any boolean function on discrete attributes using the decision tree.
- Each leaf node represents the output variable(y). The internal nodes and root node are the input variables. The model is used as follows to make predictions: walk the splits of the tree to arrive at a leaf node and output the value present at the leaf node.

The decision tree in the figure classifies whether a person is fit or not. It is accomplished by considering the person's routine factors such as age, diet and physical activity.

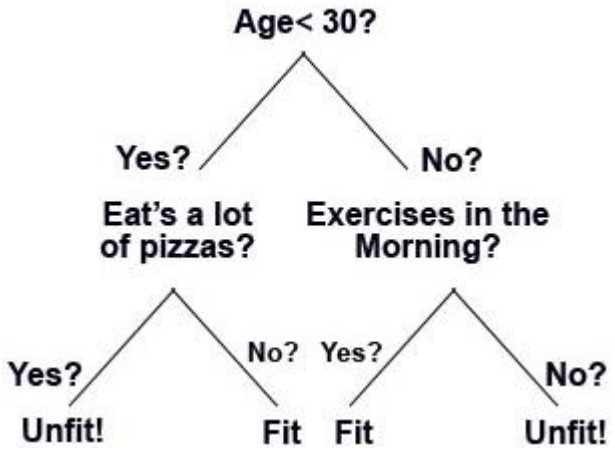

Fig3 : Decision Tree

5. Support Vector Machine:

Support Vector Machine(SVM) can be used for both classification and regression challenges. 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.

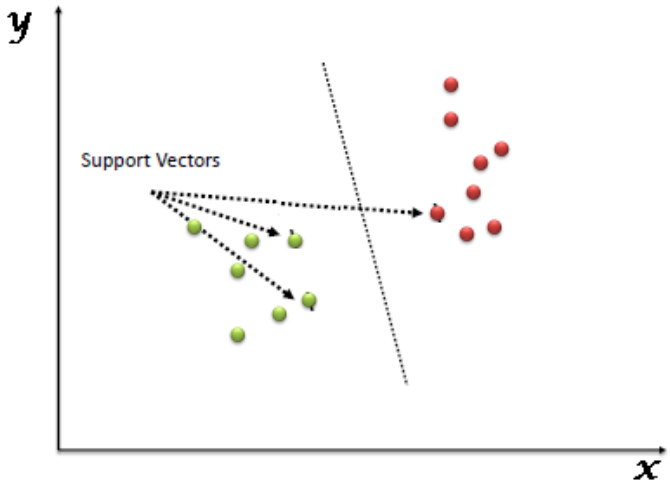

Fig4 : Support Vector Machine

Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line).

- Identifying the right hyper plane: You need to remember a thumb rule to identify the right hyper-plane:
    a. Select the hyper-plane which segregates the two classes better.
    b. Maximize the distances between nearest data point (either class) or hyper-plane will help us to decide the right hyper-plane. This distance is called as Margin.

Unsupervised learning algorithms:

6. Apriori:

It is widely used in market basket analysis. It allows retailers to identify relationships between the items that people buy. Association Rules are

used to analyse transaction data, where one checks for combinations of products that frequently co-occur in the database.

Association rule comes with three factors on which transaction data is analyzed by algorithms. They are; Support, Confidence and Lift.

For example, Let us consider, we need to find the frequency of customers buying sugar along with milk.

- Assume there are 500 customers
- 50 of them bought milk, 30 bought sugar and 20 bought both of them.
- support = P (milk & sugar) = 20/500 = 0.04
- confidence = support/P(sugar) = 0.04/0.16 = 0.25
- lift = confidence/P(Milk) = 0.25/0.10 = 2.5

If an itemset happens to have a very low support, we do not have enough information on the relationship between its items and hence no conclusions can be drawn from such a rule.

The confidence for an association rule having a very frequent consequent will always be high.

Lift is the rise in probability of having {Y} on the cart with the knowledge of {X} being present over the probability of having {Y} on the cart without any knowledge about presence of {X}.

7. K-Means:

The K-means algorithm identifies *k* number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

The 'means' in the K-means refers to averaging of the data; that is, finding the centroid. The objective of K-means is to group similar data points

together and discover underlying patterns. To achieve this objective, K-means looks for a fixed number (*k*) of clusters in a dataset.

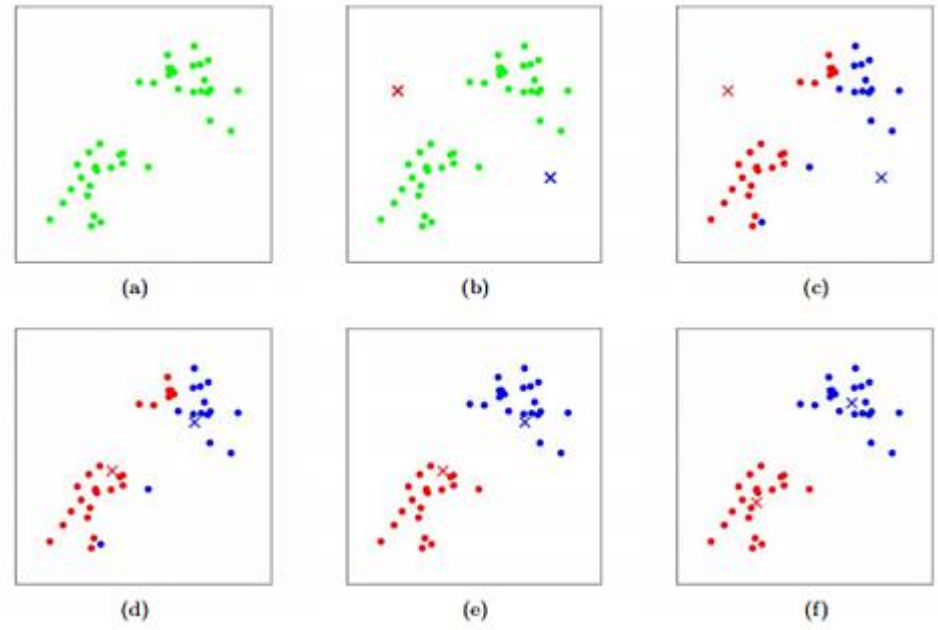

Fig5: K-Means

In Figure 5, you can see a K-means algorithm. Training examples are shown as dots, and cluster centroids (K) are shown as crosses. (a) is an original dataset. (b) is a random initial cluster centroid. (c-f) is an illustration of running two iterations of k-means. In each iteration, we assign each training example to the closest cluster centroid (shown by ”painting” the training examples the same colour as the cluster centroid to which is assigned); then we move each cluster centroid to the mean of the points assigned to it.

8. Principal Component Analysis(PCA):

Let’s say that you want to predict the income of a company in coming years. You have lots of information available, the income in the past years, publicly available economic indicator, efficiency of employees and so on. Despite being an overwhelming number of variables to consider, this is so tiring to take into account.

Principal component analysis is a technique for feature extraction, so it combines our input variables in a specific way, then we can drop the

"least important" variables while still retaining the most valuable parts of all of the variables*!* As an added benefit, each of the "new" variables after PCA are all independent of one another. This is a benefit because the assumptions of linear model require our independent variables to be independent of one another. If we decide to fit a linear regression model with these "new" variables, this assumption will necessarily be satisfied.

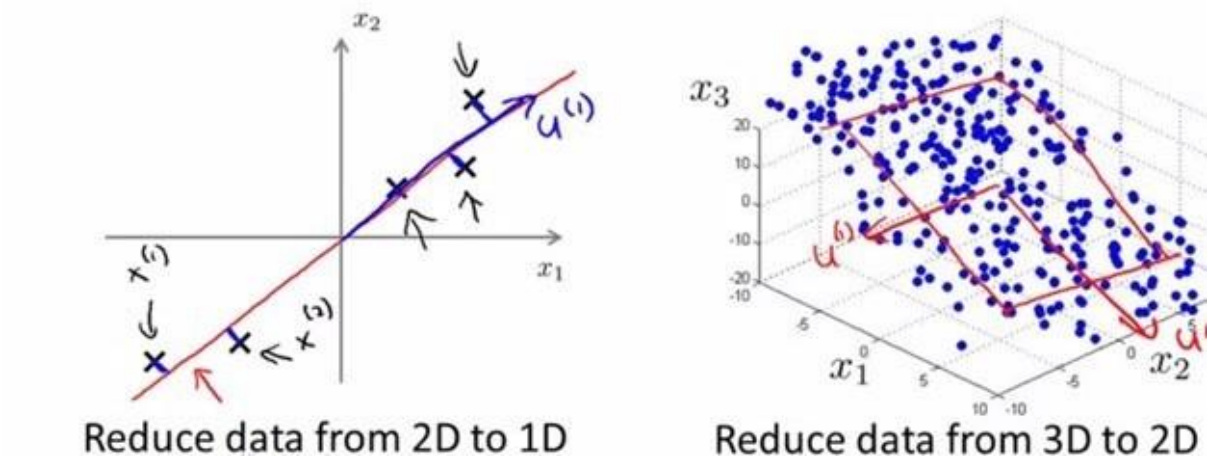

Fig6 : Principal Component Analysis

Machine Learning Algorithm Cheat Sheet

Data Science is an ever-growing field, there are numerous tools available in the market. It is a daunting task to remember all the functions and flow of each concept. To ease our approach towards machine learning algorithms we have something called as cheat sheets. This machine learning cheat sheet will help you choose the appropriate machine learning algorithms for your modeling solution. The cheat sheet takes into account the data nature of your problem and then suggests the best algorithm for the job.

I have listed a few machine learning algorithms cheat sheets which are best in the market,

- SAS Algorithm Flowchart
- Python and Scikit Cheat Sheets
- Microsoft Azure Machine Learning Algorithm Cheat Sheet

Conclusion:

There are a wide range of algorithms available to train and make your systems learn, while it becomes important to choose the best one for your application among these. At the end of the day it's a trial and error approach which can help you choose the right algorithm for your model.