Machine Learning

Machine learning is a general term which specifies that if a machine can learn from data.

In Machine Learning you provide input, a result and a then a algorithm is applied to it to give a output.

Steps to Machine learning.

1. Data is collected
2. Data is pre-processed

* Analysing the data (seeing what would be useful for the model)
* Dealing the missing, nan data, handling duplicate data, handling abnormal data handling, outliers, extracting relevant feature, scaling data, and other pre-processing techniques in real life.
* Splitting the dataset into training and testing set

1. Model Creation.

* Choosing a learning algorithm to build a model.
* Training the model with the training set.
* Testing the trained model with the testing set

1. Finally, with the results from the testing the model the results are analysed and evaluated.

* Positive improvement required is implemented
* such as:
  + - * + collecting more data
        + adding or removing data or other features
        + choosing another algorithm.
        + Turning hyperparameter etc to improve the model performance.

A good machine learning model should predict the future.

Machine learning algorithm types.

* Supervised
* Each data point is labelled or associated with a category or values of interest.
* The goal of supervised learning is to study many labels like this and then to be able to make prediction about future data points

There are two types of supervised learning.

1. Classification

* Output variable is categorised to two or more classes.

1. Regression

* Output variable is real or continues value.
* Unsupervised
* Data points have no labels associated with them.
* The data needs to be organised in some way that its structure can be described.
* Unsupervised learning groups the data into clusters
* It finds different ways of looking at complex data to make it appear simpler.
* The mechanism of unsupervised learning is to fit the model with an example input without the associated output and repeat the step many times until the model drives structure from the data.
* Now with brand new input the model will find hidden structures in the data.

There are two subcategories of unsupervised learning.

1. Clustering

* Split group data to clusters that are similar or related to each other.

1. Association

* Find structure in a chaotic environment.
* Reinforcement learning
* In reinforcement leaning a algorithm gets to choose an action in response to each data point.
* The algorithm received feed back on how good the action was
* Based on this the algorithm changes its action to receive the highest reward.