Date
Page

1) Supervised learning

→ already tagged data
→ features & labels present
→ classification : classify which label a given set of feature belong to
→ Regression: find value of te label using previous data

11) Unsupervised learning :- All machine knows is the data infront of it - No feature, No label

→ data not already tagged
→ features & labels not present
→ Training not done
→ clustering & association

↓ discover the inherent groupings in the data, such as grouping customers by purchasing behavior

↳ association rule learning problem such as people that by X also tend to buy Y

The Principle That Underpins All Algorithm

→ ML algorithms are described as learning a target function (f) that best maps input variables (X) to an output variable (Y)

$$Y = f(X)$$

This is a general learning task where we would like to make predictions in the future (Y) given new examples of input variable (X). We don't know what the function (f) looks like or its form.

→ most common type of ML is to learn the mapping $Y = f(X)$ to make predictions of Y for new X. This is called predictive modeling or predictive analytics.

Linear and non-linear Algorithm

→ Algorithm that simplify the function to a known form are linear machine learning algorithm.

→ Algorithm involves two steps
1) Select a form for the function
ii) Learn the coefficients for the function from the training data.

→ Ex- linear regression, logistic regression.

→ Algorithm that donot make strong assumptions about the form of the mapping function are called nonlinear machine learning algorithm.
They are free to learn any functional form from the training data.

→ Ex :- Neural networks, Decision Tree, Support Vector Machine

K-Nearest Neighbor (KNN) algorithm for ML

→ is one of the simplest ML algorithms based on Supervised Learning technique
→

How does K-NN work?

→ Select the number K of the neighbour
→ Calculate the Euclidean distance of K number of neighbors

$$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$$

→ Take the K-nearest neighbors as per the Euclidean distance
→ Among these K neighbors, count the number of the data points in each category.
→ Assign the new data points to that category for which the neighbor is maximum
→ Our model is ready.