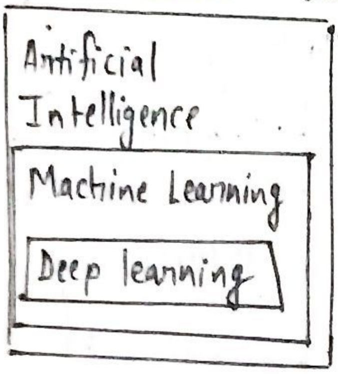
Machine Learning

→ Machine learning is study of computer algorithms that improve automatically through experience & by the use of data.

→ It is field of study that gives computers the capability to learn without being explicitly programmed.



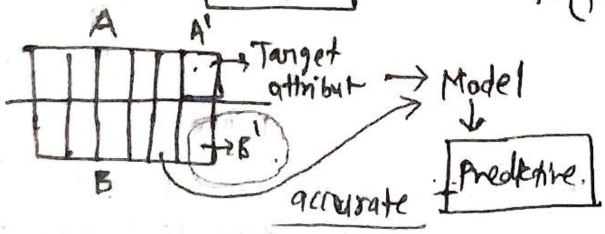
Data $\xrightarrow{\text{i/p}}$ Traditional Programming → O/P

data $\xrightarrow{\text{i/p}}$ Machine Learning → program
O/P → Machine Learning

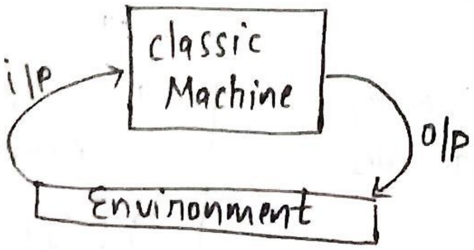
Training & Testing

Training :- $\xrightarrow{\text{I/P}}$ Algo. → Logic / Model

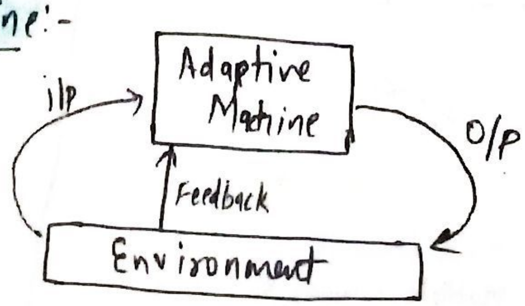
Testing:- $\xrightarrow{\text{I/P}}$ Logic → O/P (predictive)



Classic Machine / Non- Adaptive Machine :-

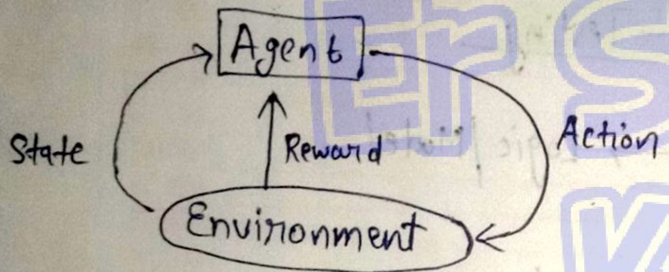


Adaptive Machine:-



Type of ML

- **Supervised**
  - Regression
    - Linear Regression
    - Decision Tree
    - Neural Network
    - SVR, GPR
  - Classification
    - Naive Bayes
    - Discriminant Analysis
    - Support Vector Machines
- **Unsupervised**
  - Association Rule Mining
  - Clustering
    - K-mean
    - K-medoids
    - Hidden Markov
    - Gaussian mixture
- **Semi-supervised**
- **Reinforcement** (Learn from mistakes)
  - Gaming
  - finance Sector
  - Manufacturing
  - Robot Navigation



Linear Regression

→ It is used for solving regression problem.

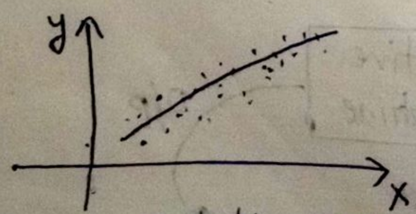
→ It is used to predict the continuous dependent variable using IV.

→ We find best fit line.

→ Least square estimation method for accuracy

→ There may be collinearity b/w independent variables.

$$y = b_0 + b_1 x + \varepsilon$$



1→I, 1→D ⇒ Simple Linear R

~~M~~→I 1→D ⇒ Multiple Linear R

Logistic Regression

→ It is used for solving classification problem.

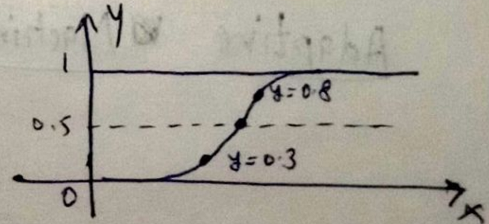
→ To predict categorical dependent variable using IV.

→ We find S-curve.

→ Maximum likelihood estimation method for accuracy.

→ There should not be collinearity b/w IVs.

$$\log\left[\frac{y}{1-y}\right] = b_0 + b_1 x_1 + \cdots b_n x_n$$



$$P(y^{(i)}=1) = \frac{1}{1+\exp\left(-(\beta_0+\beta_1 x_1^{(i)} + \cdots \beta_p x_p^{(i)})\right)}$$

Cost Function :- $$\frac{\sum_{i=1}^{n} \left((\beta_1 x_i + \beta_0) - y_i\right)^2}{2n}$$

Naive Bayes ⇒ It assumes that presence of particular features in a class is unrelated to presence of any other features.

$$\boxed{P(c|x) = \frac{P(x|c)\,P(c)}{P(x)}}$$

Posterior probability ← $P(c|x)$

Likelihood ← $P(x|c)$

class prior prob. ← $P(c)$

predictor prior prob. ← $P(x)$

$$P(c|x) = P(x_1|c) \cdot P(x_2|c) \cdots\cdots P(x_n|c) \cdot P(c)$$

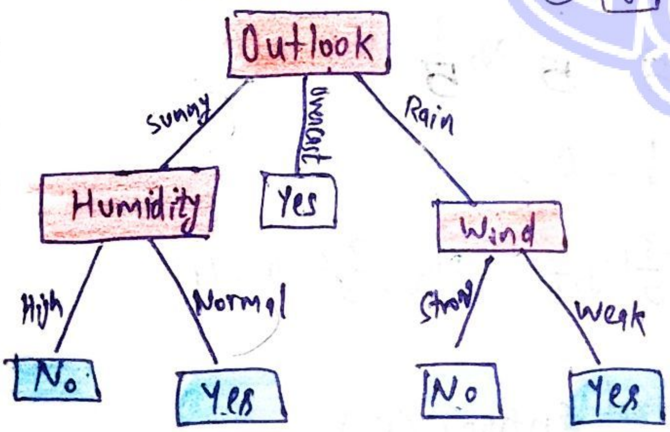
Decision Tree ⇒ It is one of the predictive modelling approaches used in ML.

→ Decision Trees are constructed via an algo. approach that identifies way to split a data set based on different conditions.

→ Decision Trees are used for both classification & regression tasks.

→ In decision tree, each internal node represents a test on a feature.

D → Algo. → Model



→ Decision node (Test)

→ Terminal node / Leaf node (classification)

→ Splitting

→ pruning

→ Root node

→ Branch / sub-tree

Decision Tree

Type of DT

→ Categorical DT

→ Continuous DT

Naive algorithm

No	Color	Type	Origin	Stolen
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

$X = \{Red, SUV, Domestic\}$

$$\boxed{P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)}}$$

$P(x|Yes) = ?$

$P(x|No) = ?$

$$P(Red|Yes) = \frac{P(Yes|Red) \cdot P(Red)}{P(Yes)} = \frac{3/5 \cdot \cancel{5}/\cancel{10}}{\cancel{5}/\cancel{10}} = \frac{3}{5}$$

$$P(SUV|Yes) = \frac{P(Yes|SUV) \cdot P(SUV)}{P(Yes)} = \frac{\frac{1}{\cancel{4}} \cdot \frac{\cancel{4}}{\cancel{10}}}{\frac{5}{10}} = \frac{1}{5}$$

$$P(Domestic|Yes) = \frac{P(Yes|Domestic) \cdot P(Domestic)}{P(Yes)} = \frac{\frac{2}{5} \cdot \frac{\cancel{5}}{\cancel{10}}}{\frac{\cancel{5}}{\cancel{10}}} = \frac{2}{5}$$

$(\because p + q = 1 \Rightarrow q = 1 - p)$

$P(Red|No) = 1 - \frac{3}{5} = \frac{2}{5}$, $P(SUV|No) = 1 - \frac{1}{5} = \frac{4}{5}$

$P(Domestic|No) = 1 - \frac{2}{5} = \frac{3}{5}$

$P(x|Yes) = P(Yes) \cdot P(Red|Yes) \cdot P(SUV|Yes) \cdot P(Domestic|Yes)$

$$\frac{1}{\cancel{2}} \cdot \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{\cancel{2}}{5} = \frac{3}{125} = 0.024$$

$P(x|No) = P(No) \cdot P(Red|No) \cdot P(SUV|No) \cdot P(Domestic|No)$

$$= \frac{1}{\cancel{2}} \cdot \frac{\cancel{2}}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} = \frac{12}{125} = 4 \times 0.024 = 0.096$$

$P(x|No) > P(x|Yes)$

Therefore No ✓

Fruit	Yellow	Sweet	long	Total
Mango	350	450	0	650
Banana	400	300	350	400
Others	50	100	50	150
Total	800	850	400	1200

$X = \{ \text{Yellow}, \text{Sweet}, \text{long} \}$

Naive bayes

$$\boxed{P(x|y) = \frac{P(y|x) \cdot P(x)}{P(y)}}$$

find

$P(X|\text{Mango}) = P(\text{mango}) \cdot P(\text{Yellow}|\text{Mango}) \cdot P(\text{Sweet}|\text{Mango}) \cdot P(\text{long}|\text{Mango})$

$P(X|\text{Banana}) = P(\text{Banana}) \cdot P(\text{Yellow}|\text{Banana}) \cdot P(\text{Sweet}|\text{Banana}) \cdot P(\text{long}|\text{Banana})$

$P(X|\text{others}) = P(\text{others}) \cdot P(\text{Yellow}|\text{others}) \cdot P(\text{Sweet}|\text{others}) \cdot P(\text{long}|\text{others})$

$P(\text{Mango}) = \frac{650}{1200} = 0.541$, $P(\text{Banana}) = \frac{400}{1200} = 0.33$, $P(\text{others}) = \frac{150}{1200} = 0.125$

$P(\text{Yellow}) = \frac{800}{1200} = 0.66$, $P(\text{Sweet}) = \frac{850}{1200} = 0.70$, $P(\text{long}) = \frac{400}{1200} = 0.33$

$P(\text{Yellow}|\text{Mango}) = \frac{P(\text{Mango}|\text{Yellow}) \cdot P(\text{Yellow})}{P(\text{Mango})} = \frac{\frac{350}{800} \cdot 0.66}{0.541} = 0.533$

$P(\text{Sweet}|\text{Mango}) = \frac{\frac{450}{850} \cdot 0.70}{0.541} = 0.685$, $P(\text{long}|\text{Mango}) = \frac{0 \cdot 0.33}{0.541} = 0$

$\Rightarrow P(X|\text{Mango}) = 0.541 \times 0.533 \times 0.685 \times 0 = 0$

$\boxed{P(X|\text{Mango}) = 0}$

$P(\text{Yellow}|\text{Banana}) = \frac{\frac{400}{800} \times 0.66}{0.33} = 1$, $P(\text{Sweet}|\text{Banana}) = 0.74$, $P(\text{long}|\text{Banana}) = 0.875$

$\Rightarrow P(X|\text{Banana}) = 0.33 \times 1 \times 0.74 \times 0.875 = 0.21$ $\boxed{P(X|\text{Banana}) = 0.21}$

$P(\text{Yellow}|\text{others}) = \frac{\frac{50}{800} \cdot 0.66}{0.125} = 0.33$, $P(\text{Sweet}|\text{others}) = \frac{\frac{100}{850} \cdot 0.70}{0.125} = 0.658$

$P(\text{long}|\text{others}) = \frac{\frac{50}{400} \cdot 0.33}{0.125} = 0.33$

$\Rightarrow P(X|\text{others}) = 0.125 \times 0.33 \times 0.33 \times 0.658 = 0.0089$ $\boxed{P(X|\text{others}) = 0.0089}$

$\boxed{P(X|\text{Banana}) \text{ has large value.}}$

Q/

Outlook	Temp	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Decision Tree

ID3 algorithm

$$\boxed{\text{Entropy} = \sum -P_i \log_2 P_i}$$

$$\boxed{\text{Entropy}(x,c) = \sum P(c)\, E(c)}$$

$$\boxed{\text{Gain} = \text{Entropy}(\text{before}) - \sum_{j=1}^{K} \text{Entropy}(j, \text{after})}$$

Value (outlook) = Sunny, overcast, Rain

$S = [+9, 5-] \Rightarrow \text{Entropy} = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$

$S_{sunny} = [+2, 3-] \Rightarrow \text{Entropy} = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.971$

$S_{overcast} = [+4, 0-] \Rightarrow \text{Entropy} = -\frac{4}{4}\log_2\frac{4}{4} = 0$

$S_{Rain} = [+3, 2-] \Rightarrow \text{Entropy} = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.971$

$$\text{Gain}(S, \text{outlook}) = \text{Entropy} - \sum \frac{|S_v|}{|S|}\,\text{Entropy}(S_v)$$

$$= 0.94 - \frac{5}{14} \times \text{Entropy(sunny)} - \frac{4}{14}\,\text{Entropy(overcast)} - \frac{5}{14}\,\text{Entropy(Rain)}$$

$$= 0.94 - \frac{5}{14} \times 0.971 - \frac{4}{14} \times 0 - \frac{5}{14} \times 0.971 = 0.2464$$

$$\boxed{\text{Gain}(S, \text{outlook}) = 0.2464}$$

Value(Temp) = Hot, mild, cool

$S = [+9, 5-] = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{9}{14} = 0.94$

$S_{Hot} = [2+, 2-] = \frac{-2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = 1.0$

$S_{mild} = [4+, 2-] = -\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{4}{6} = 0.9183$

$S_{cool} = [3+, -1] = \frac{-3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4} = 0.8113$

$G(S, temp) = \text{Entropy} - \frac{4}{14}\text{Entropy(Hot)} - \frac{6}{14}\text{Entropy(mild)} - \frac{4}{14}\text{Entropy(cool)}$

$= 0.94 - \frac{4}{14}\times 1 - \frac{6}{14}\times 0.9183 - \frac{4}{14}\times 0.8113$

$\boxed{G(S, temp) = 0.0289}$

Value (Humidity) = High, Normal

$S = [+9, 5-] = E = 0.94$

$S_{High} = [+3, 4-] = \text{Entropy} = \frac{-3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7} = 0.9852$

$S_{Normal} = [6+, 1-] = \text{Entropy} = -\frac{6}{7}\log_2\frac{6}{7} - \frac{1}{7}\log_2\frac{1}{7} = 0.5916$

$G(S, Humidity) = \text{Entropy} - \frac{7}{14}\text{Entropy(High)} - \frac{7}{14}\text{Entropy(Normal)}$

$= 0.94 - \frac{7}{14}\times 0.9852 - \frac{7}{14}\times 0.5916$

$\boxed{G(S, Humidity) = 0.1516}$

Value(wind) = Weak, strong

$S = 0.94$

$S_{weak} = [+6, 2-] = -\frac{6}{8}\log_2\frac{6}{8} - \frac{2}{8}\log_2\frac{2}{8} = 0.8113$

$S_{strong} = [+3, 3-] = \frac{-3}{6}\log_2\frac{3}{6} - \frac{3}{6}\log_2\frac{3}{6} = 1$

$G(S, wind) = 0.94 - \frac{8}{14}\times 0.8113 - \frac{6}{14}\times 1 = 0.0478$

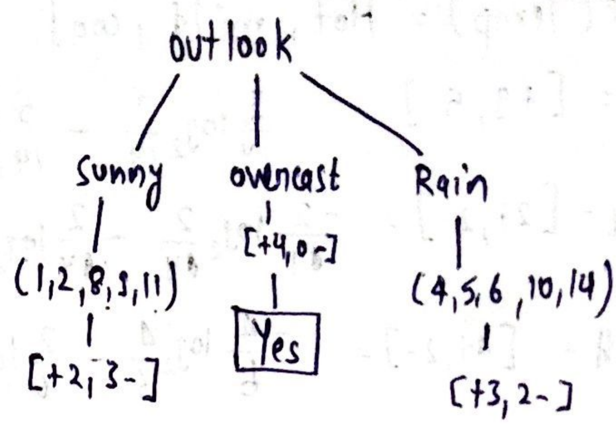
$\boxed{G(S, wind) = 0.0478}$

Gain(S, outlook) = 0.2464

Gain(S, temp) = 0.0289

Gain(S, Humidity) = 0.1516

Gain(S, Wind) = 0.0478



Day	Temp	Humidity	Wind	Play Tennis
1	Hot	High	Weak	N
2	Hot	High	Strong	N
8	mild	High	Weak	N
9	Cool	Normal	weak	Y
11	Mild	Normal	Strong	Y

Value(Temp) = Hot, mild, cool

$S_{sunny} = [2+, 3-] = 0.97$

$S_{Hot} = [0+, 2-] = E = 0$ $-\frac{2}{2}\log\frac{2}{2}$

$S_{mild} = [1+, 1-] = E = 1$ $2\left(-\frac{1}{1}\log\frac{1}{1}\right)$

$S_{cool} = [1+, 0-] = E = 0$

$G(S_{sunny}, Temp) = Entropy - \frac{2}{5} Entropy(Hot) - \frac{2}{5} Entropy(mild) - \frac{1}{5} Entropy(cool)$

$= 0.97 - \frac{2}{5}\times 0 - \frac{2}{5}\times 1 - \frac{1}{5}\times 0 = 0.570$

Value(Humidity) = High, Normal

$S_{sunny} = 0.97$, $S_{High} = [0+, 3-] = E = 0$, $S_{Normal}\,[2+, 0-] \Rightarrow E = 0$

$Gain(S_{sunny}, Humidity) = Entropy - \frac{3}{5} Entropy(High) - \frac{2}{5} Entropy(Normal)$

$= 0.97$

Value(Wind) = weak, strong

$S_{sunny} = 0.97$, $S_{weak} = [+1, 2-] \Rightarrow E = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.9183$

$S_{strong} = [+1, 1-] \Rightarrow E = 1.0$

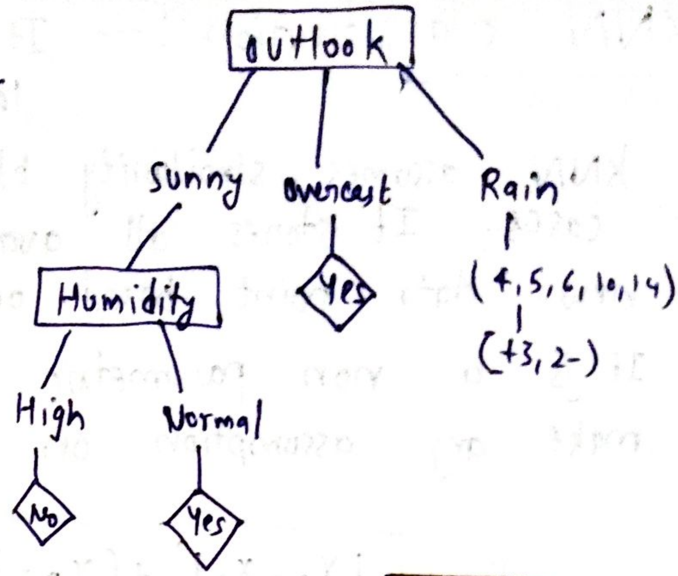
$Gain(S_{sunny}, wind) = Entropy - \frac{2}{5} E(strong) - \frac{3}{5} Entropy(weak)$

$= 0.97 - \frac{2}{5}\times 1 - \frac{3}{5}\times 0.9183 = 0.0192$

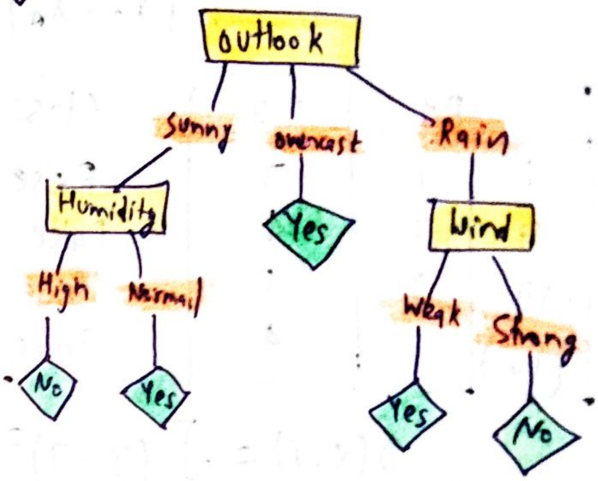
Gain $(S_{sunny}, \text{Temp}) = 0.570$

Gain $(\text{Sunny}, \text{Humidity}) = 0.97$ ✓

Gain $(S_{sunny}, \text{Wind}) = 0.0192$



Day	Temp	Humidity	Wind	Play Tennis
4	Mild	High	Weak	Y
5	Cool	Normal	Weak	Y
6	Cool	Normal	Strong	N
10	Mild	Normal	Weak	Y
14	Mild	High	Strong	N



V(Temp) = Hot, mild, Cool

$S_{Rain} = [+3, 2-] = E = 0.97$

$S_{Hot} = [+0, 0-] = E = 0$

$S_{mild} = [2+, 1-] = E = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{2}{3} = 0.9183$

$S_{cool} = [+1, 1-] = E = 1$

Gain $(S_{Rain}, \text{temp}) = \text{Entropy} - \frac{0}{5}E(\text{Hot}) - \frac{3}{5}E(\text{mild}) - \frac{2}{5}E(\text{cool})$

$= 0.97 - 0 - \frac{3}{5} \times 0.9183 - \frac{2}{5} \times 1 = 0.0192$

V(Humidity) = High, Normal

$S_{High} = [+1, 1-] = E = 1$

$S_{Normal}\ [2+, 1-] = E = 0.9183$

$G(S_{Rain}, \text{Humidity}) = 0.97 - \frac{2}{5} \times 1 - \frac{3}{5} \times 0.9183$
$= 0.0192$

V(Wind) = Strong, weak

$S_{strong} = [0, 2-], E = -\frac{2}{2}\log_2\frac{2}{2} = 0$

$S_{weak} = [3+, 0], E = -\frac{3}{3}\log_2\frac{3}{3} = 0$

$G(S_{Rain}, \text{Wind}) = 0.97 - \frac{2}{5} \times 0 - \frac{3}{5} \times 0 = 0.97$ ✓

KNN classification :- It is based on Supervised learning technique.

→ KNN assumes similarity b/w new cases and available cases. It stores all available data & classifies a new data point based on similarity.

→ It is a non parametric algo., which means it doesn't make any assumption on underlying data.

$$D = \sqrt{(X_{P_1} - X_{A_1})^2 + (X_{P_2} - X_{A_2})^2}$$

	P1	P2	Class
i	7	7	False
ii	7	4	False
iii	3	4	True
iv	1	4	True

Perform KNN Classification
$X(P1=3, P2=7)$, $K=3$.
(True) ✓

$D(x,i) = \sqrt{(3-7)^2+(7-7)^2} = 4 \longrightarrow N_3 \longrightarrow$ false

$D(x,ii) = \sqrt{(3-7)^2+(7-4)^2} = 5$

$D(x,iii) = \sqrt{(3-3)^2+(7-4)^2} = 3 \longrightarrow N_1 \longrightarrow$ True

$D(x,iv) = \sqrt{(3-1)^2+(7-4)^2} = 3.6 \longrightarrow N_2 \longrightarrow$ True

2 True > 1 false

Support Vector Machine ⇒ It is a supervised learning algorithms, which is used for classification as well as Regression problems.

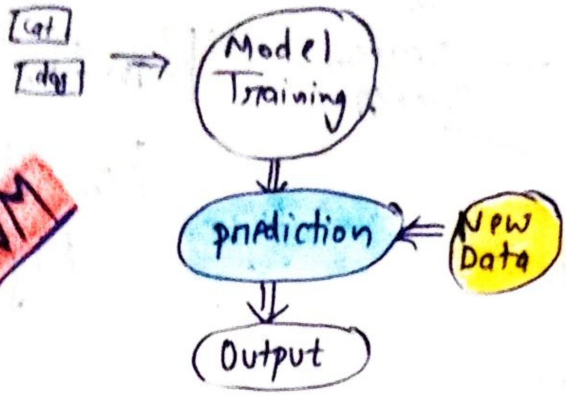
→ It is to create best [Hyperplane] line or decision boundary that can segregate n- dimensional space into classes so that we can easily put new data point in correct category in future.

→ SVM chooes extreme points that help in creating Hyperplane. These extreme points are called support vector.

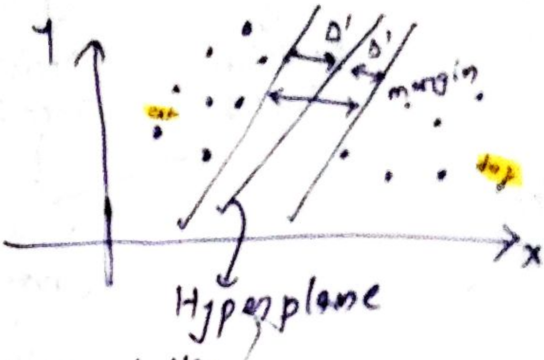
Types of SVM

Types of SVM → Linear
→ Non-linear

SVM

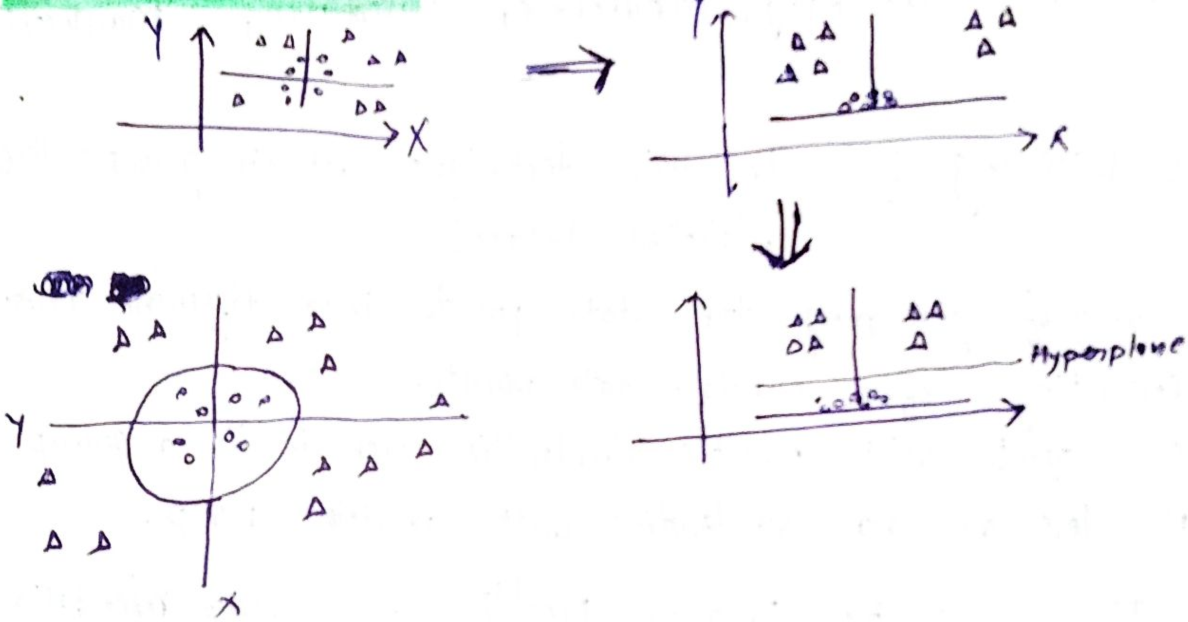


Linear SVM



$m_2 >> m_1$ better

Non Linear SVM



Random Forest Algorithm:-

It belongs to supervised learning techniques.

→ It can be used for both classification & regression Problem.

→ It is based on concept of ensemble learning, which is a process of combining multiple classifier to solve a complex problem & to improve performance of the model.

→ Random forest is a classifier that contains a no. of decision trees on various subsets of given dataset & takes the average to improve the predictive accuracy of that dataset

Random Forest Algorithm

Training Data

Training set

1 → Decision Tree 1

2 → Decision Tree 2

... Decision Tree n

Test set

Voting [averaging]

↓

Prediction

Application - Banking, Medicine, Land use, Marketing

Clustering :-

In this technique, which groups the unlabelled dataset.

A way of grouping the data points into different clusters, consisting of similar data points.

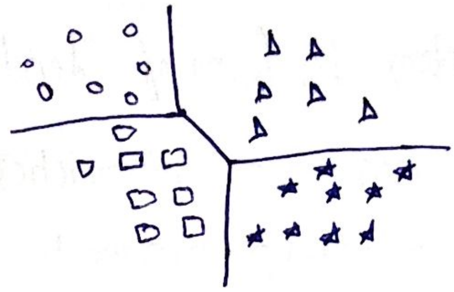
The objects with possible similarities remain in a group that has less or no similarities with another group.

→ It is used by Amazon, Netflix in its recommendation system to provide recommendations as per the past search of products, movies respectively.

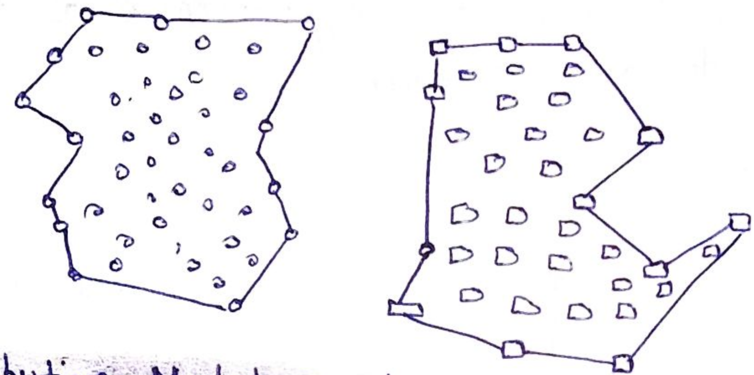
Type of clustering methods

- Partitioning Clustering
- Density Based Clustering
- Distribution Model Based Clustering
- Hierarchical Clustering
- Fuzzy Clustering

Partitioning Clustering :- It divides the data into non-hierarchical groups. It is also known as centroid Based method.
Eg - k-means clustering algo.



Density Based :- It connects highly dense area into clusters, and arbitrarily shaped distributions are formed as long as dense region can be connected.



Distribution Model Based :- The data is divided based on probability of how a dataset belongs to particular distribution. The grouping is done by assuming some distribution commonly Gaussian Distribution.

Hierarchical Clustering :- It Can be Used as alternative for partitioned clustering as there is no requirement of pre-specifying no. of clusters to be created.
Dataset is divided into clusters to create dendrogram (Tree like).

Fuzzy Clustering :- It is type of soft method in which data object may belong to more than one cluster.
Each dataset has a set of membership coefficients, which depend on degree of membership to be in a cluster.

Hierarchical Clustering ⟹ It is another unsupervised machine learning algo., which is used to group the unlabeled datasets into a cluster and also know as HCA.

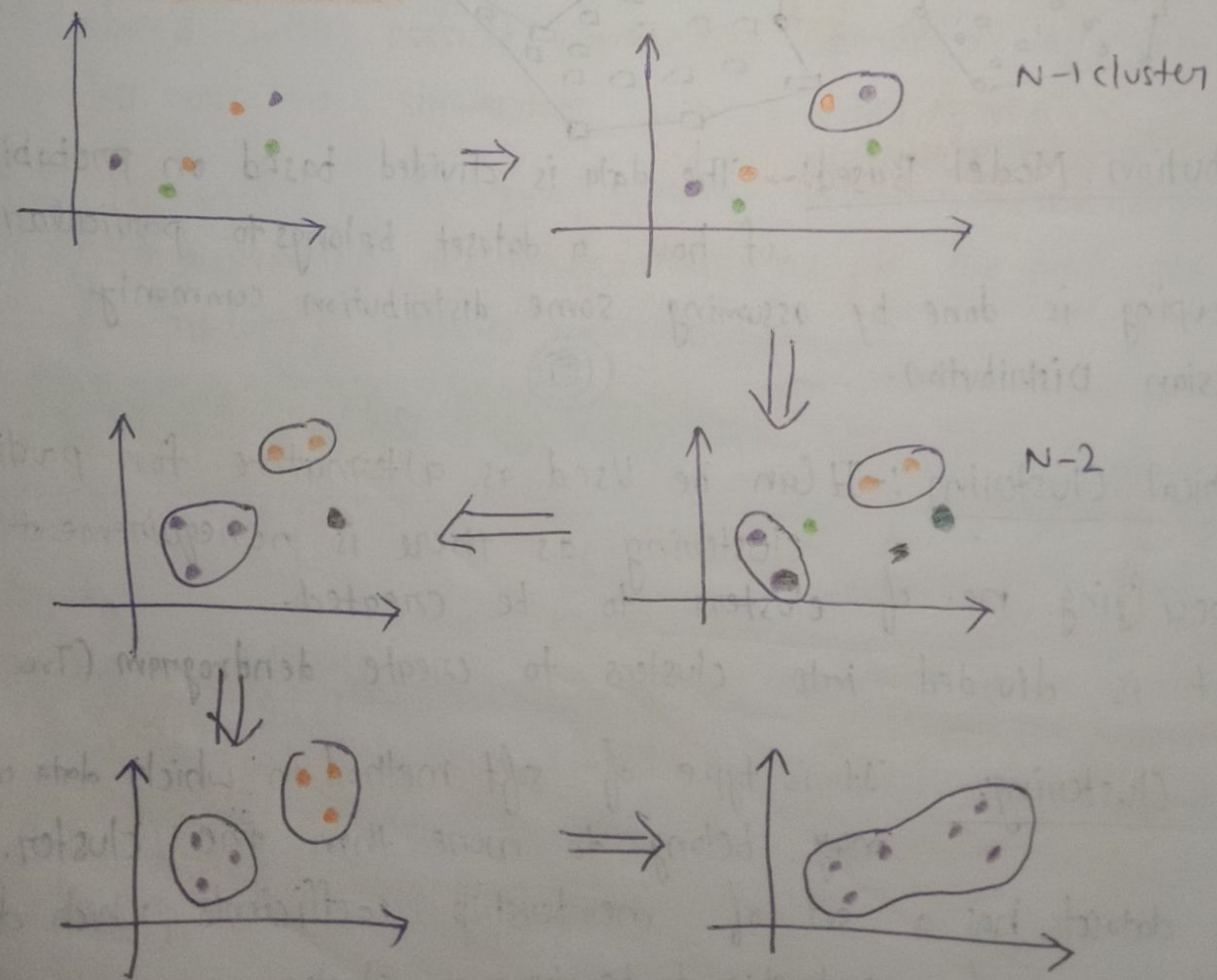
We develop hierarchy of clusters in form of dendrogram.

Hierarchical clustering technique has two approaches:-

(i) Agglomerative :- It is bottom-up approach, in which algo. starts with taking all data points as single clusters and merging them until one cluster is left.

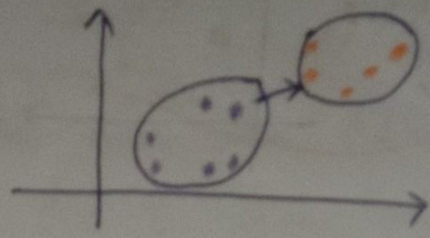
(ii) Divisive:- It is reverse of agglomerative as it is top-down approach.

Agglomerative :-

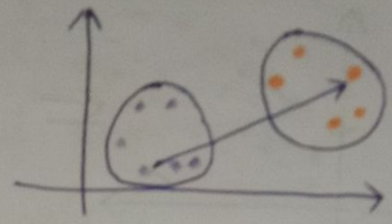


Measure for distance b/w two clusters:—

(i) Single Linkage →

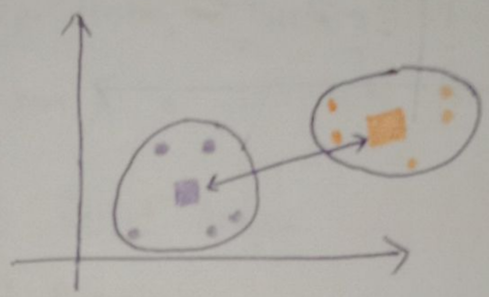


(ii) Complete Linkage:-



(iii) Average Linkage:- Each pair of datasets is added up & then divided by total no. of datasets

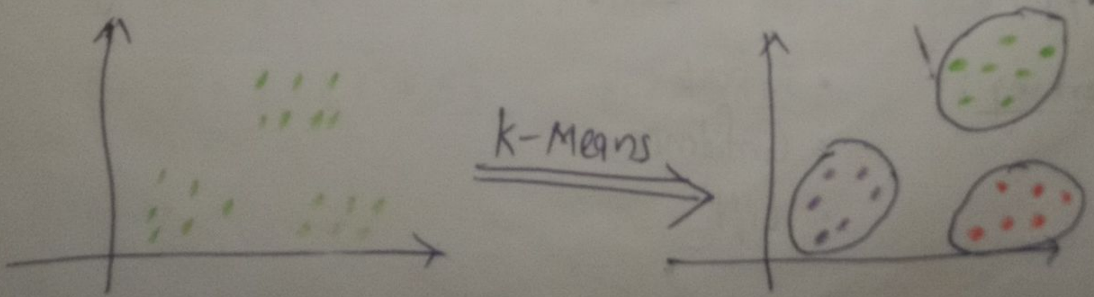
(iv) Centroid Linkage:-



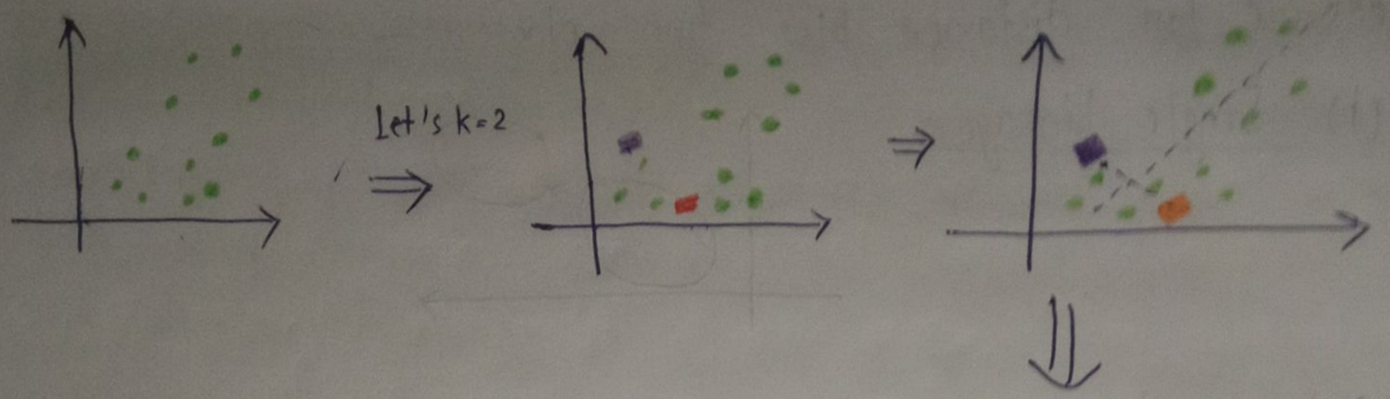
k-Means Clustering → It is an unsupervised learning algo. that is used to solve the clustering problem in ML.

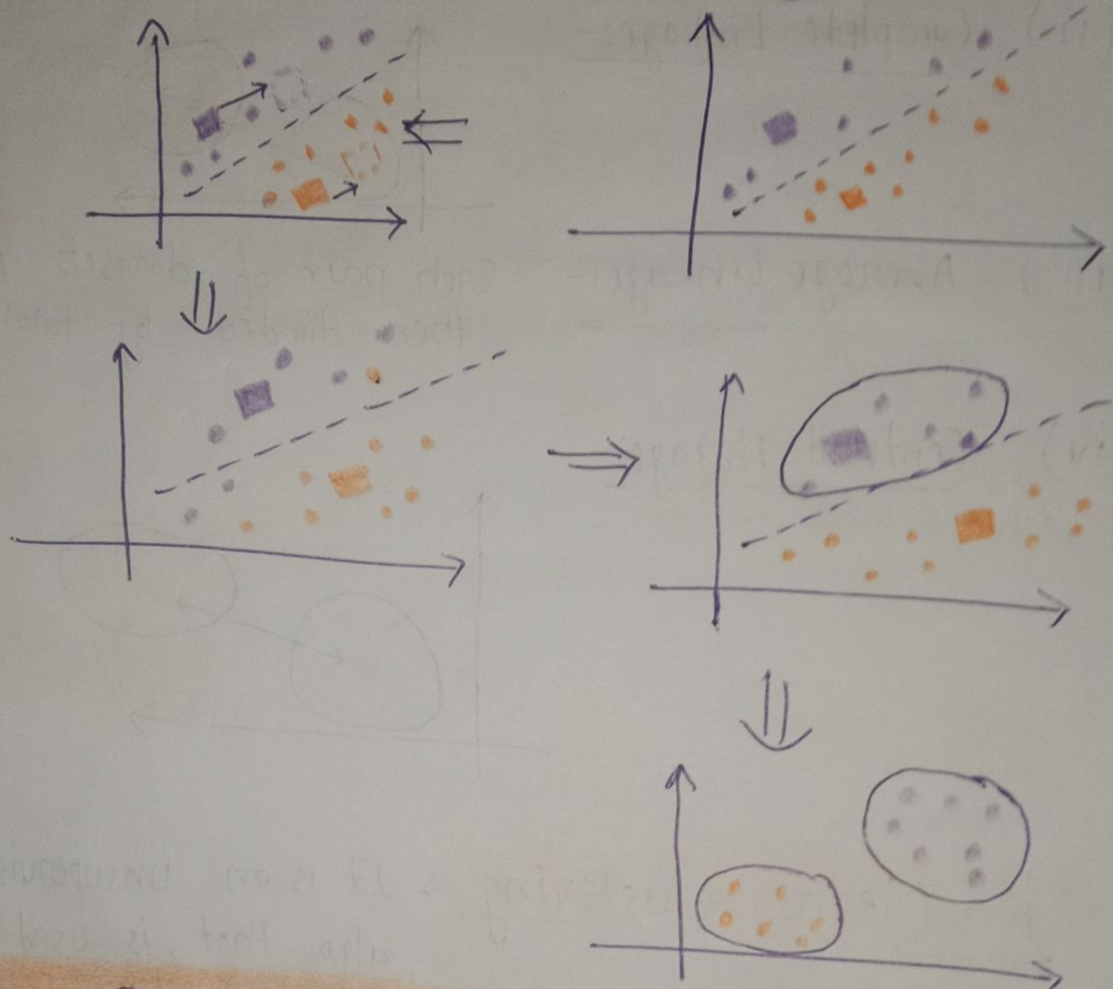
k → no. of pre-defined clusters that need to be created in process.

→ It allows us to cluster data into different groups and a convenient way to discover categories of groups in unlabeled dataset on its own without need for any training.



K-Mean Clustering





Association Rule Learning

Association Rule Learning:- It is a type of unsupervised learning technique that checks for dependency of one data item on another data item & maps accordingly so that it can be more profitable.

ARL

- Apriori
- Eclat
- F-p Growth Algorithm

Working :-

- Support
- Confidence
- Lift

Support :- It is the frequency of A.

$$Supp(x) = \frac{Freq(x)}{T}$$

$\boxed{\text{If } A} \Rightarrow \boxed{\text{then } B}$

Confidence :- It indicates how often the items X & Y occur together in dataset when the occurence of X is given.

$$Confidence = \frac{Freq(X,Y)}{Freq(X)}$$

Lift :- It is strength of any rule.

$$Lift = \frac{Supp(X,Y)}{Supp(X) \cdot Supp(Y)}$$

If Lift = 1, independent to each other

If Lift > 1, dependent to each other

If Lift < 1, one item is a substitute for other items.

Application :- Market Basket Analysis

Medical Diagnosis

Protein Sequence

Apriori Algorithm :- It uses frequent itemsets to generate association rules & it is designed to work on databases that contain transactions.

→ This algo. uses a BFS & Hash Tree to calculate the itemset associations efficiently.

Frequent itemset :- FI are those items whose support is greater than threshold value.

Eg- A = {1,2,3,4,5}   B = {2,3,7}   , 2,3 are FI.

Step ① Determine support of itemsets & select min support & confidence.

Step-② Take all supports in transaction with higher support value than min or selected support value.

Step-③ Find all rules of these sub ~~items~~sets that have higher confidence value than threshold or min confidence.

Step-④ Sort the rules as decreasing order of lift.

Eg-

Tid	Itemsets
T_1	A, B
T_2	B, D
T_3	B, C
T_4	A, B, D
T_5	A, C
T_6	B, C
T_7	A, C
T_8	A, B, C, E
T_9	A, B, C

Given:-
min support = 2
min confidence = 50%

Aprioni Algorithm

Step-① Calculating C1 and L1:-

C1:

Itemset	Support_count
A	6
B	7
C	5
D	2
E	1

Candidate set or C1

L1:

Itemset	support count
A	6
B	7
C	5
D	2

All the itemsets have greater or equal than min support

Frequent itemset L1

Step-② Candidate Generation C2 & L2:-

C2:

Itemset	Support_Count
{A, B}	4
{A, C}	4
{A, D}	1
{B, C}	4
{B, D}	2
{C, D}	0

L2:-

Itemset	Support_count
{A, B}	4
{A, C}	4
{B, C}	4
{B, D}	2

Step-③ Candidate Generation C3 & L3:-

C3:-

Itemset	Support_count
{A, B, C}	2
{B, C, D}	0
{A, C, D}	0
{A, B, D}	1

L3:-

Itemset	Supportcount
{A, B, C}	2

Step-④ Finding Association Rules for subsets:-

→ we will calculate the confidence using $A \rightarrow B$ $\frac{sup(A \wedge B)}{sup(A)}$

→ After calculating we will exclude rules that have less confidence than min. threshold (50%).

Rules	Support	Confidence
$A \wedge B \rightarrow C$	2	$Sup((A \wedge B) \wedge C)/sup(A \wedge B) = \frac{2}{4} = 50\%$
$B \wedge C \rightarrow A$	2	$sup((B \wedge C) \wedge A)/sup(B \wedge C) = 2/4 = 50\%$
$A \wedge C \rightarrow B$	2	$sup((A \wedge C) \wedge B)/sup(A \wedge C) = 2/4 = 50\%$
$C \rightarrow A \wedge B$	2	$sup(C \wedge (A \wedge B))/sup(C) = 2/5 = 40\%$
$A \rightarrow B \wedge C$	2	$sup(A \wedge (B \wedge C))/sup(A) = 2/6 = 33.33\%$
$B \rightarrow A \wedge C$	2	$sup(B \wedge (A \wedge C))/sup(B) = 2/7 = 28\%$

So the first 3 rules $A \wedge B \rightarrow C$, $B \wedge C \rightarrow A$, $A \wedge C \rightarrow B$ can be considered as the strong association rules for given problem.

Frequent Pattern Growth Algorithm ⇒ It is an improvement of apriori algorithm.

Tid	Items
T_1	E, K, M, N, O, Y
T_2	D, E, K, N, O, Y
T_3	A, E, K, M.
T_4	C, K, M, U, Y
T_5	C, E, I, K, O, O

min. support = 3

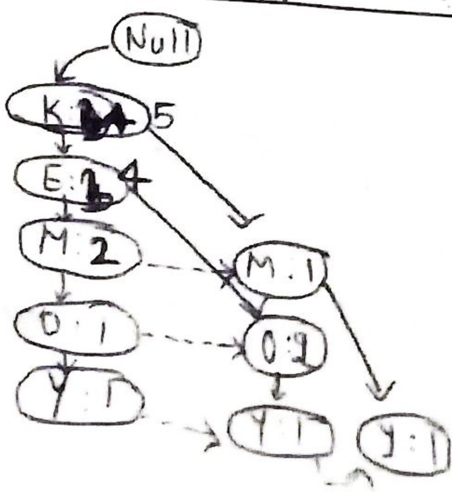
FP Growth Algorithm

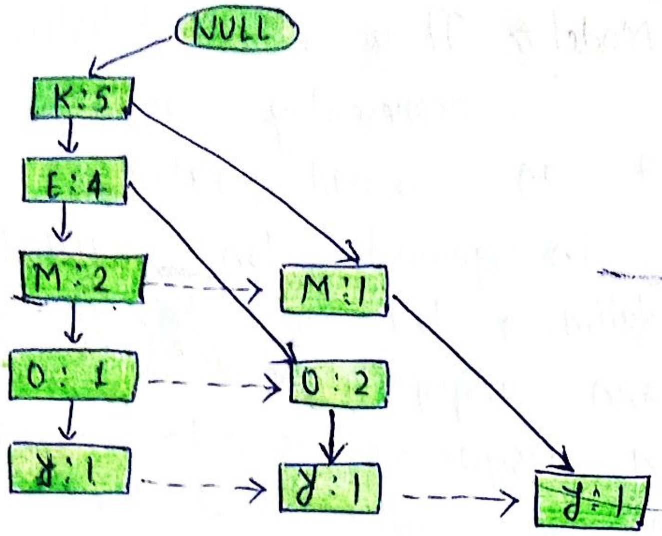
Item	Frequency
A	1
C	2
D	1
E	4
I	1
K	5
M	3
N	2
O	3
U	1
Y	3

⇒

Item	Frequency
K	5
E	4
M	3
O	3
Y	3

Tid	Items	Ordered-Item set
T_1	E, K, M, N, O, Y	K, E, M, O, Y
T_2	D, E, K, N, O, Y	K, E, O, Y
T_3	A, E, K, M	K, E, M
T_4	C, K, M, U, Y	K, M, Y
T_5	C, E, I, K, O, O	K, E, O





Items	Conditional Pattern Base	Conditional Frequent Pattern Tree
Y	{{K,E,M,O:1}, {K,E,O:1}, {K,M:1}}	{K:3}
O	{{K,E,M:1}, {K,E:2}}	{K,E:3}
M	{{K,E:2}, {K:1}}	{K:3}
E	{K:4}	{K:4}
K	–	–

Items	Frequent Pattern Generated
Y	{<K,Y:3>}
O	{<K,O:3>, <E,O:3>, <E,K,O:3>}
M	{<K,M:3>}
E	{<E,K:3>}
K	–

K→Y , Y→K

K→O E→O , O→K, O→E, E→K, K→E

K→M, M→K

E→K, K→E

Gaussian Mixture Model ⇒ It is a probabilistic model for representing normally distributed subpopulations within an overall population.

Mixture models in general don't require knowing which subpopulation a data point belongs to, allowing the model to learn subpopulations automatically. Since subpopulation assignment is not known, this constitutes a form of unsupervised learning.

$$p(x) = \sum_{k=1}^{K} \pi_k G(x \mid \mu_k, \Sigma_k)$$

π_k is mixing coefficient for kth distribution.

Gaussian density function:-

$$N(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

Dimensionality Reduction ⇒ The no. of input features, variables or columns present in a given dataset is known as dimensionality and the process to reduce these features is called Dimensionality Reduction.

→ It is a way of converting the higher dimensions dataset into lesser dimensions dataset ensuring that it provides similar information.

→ It is used in

- Speech Recognition
- Signal Processing
- Bioinformatics
- Data Visualization
- Noise Reduction

Dimensionality Reduction

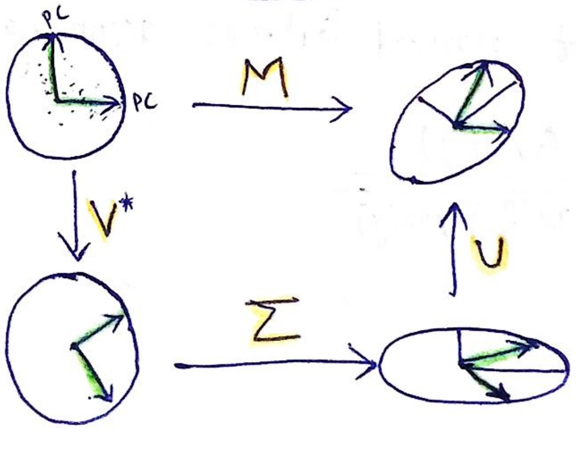
- **Feature Selection**
  - Filters Methods
  - Wrapper Methods
  - Embedded Methods
- **Feature Extraction**
  - Principal Component Analysis
  - Factor Analysis
  - Singular Value Decomposition

Feature Selection:- It is the process of selecting the subset of relevant features and leaving out the irrelevant features present in a dataset to build a model of high accuracy.

Feature Extraction:- It is the process of transforming space containing many dimensions into space with fewer dimensions.

Singular Value Decomposition [SVD] ⇒ It provides another way to factorize a matrix, into singular vectors & singular values.

The SVD is used widely both in calculation of matrix operations such as matrix inverse, but also as a Data Reduction method.



It is a method of decomposing a matrix into 3 other matrices

Diagonal

$$\boxed{M = U \cdot \Sigma \cdot V^*}$$

matrix orthogonal

$AA^T = A^TA$

→ $a_{ij} = \sum_{k=1}^{n} U_{ik} \Sigma_k V^{*}_{jk}$

→ $\Sigma_{i+1} \leq \Sigma_i$, $U^T U = V V^T = I$ (orthogonal)

→ $A A^T U = U S^2$, $A^T A V = V S^2$

→ $U = A V S^{-1}$, $A^T = V S U^T$

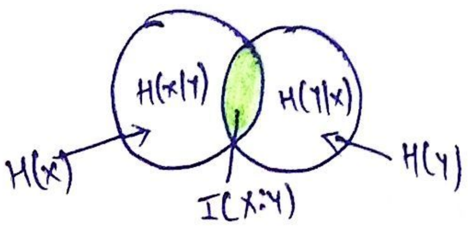
$\boxed{A A^T = A^T A}$

Feature Selection ⇒ It is the processing of selecting the subset of relevant features and leaving out the irrelevant features present in a dataset to build a model of high accuracy.

Feature Selection
- **Variable Ranking**
- **Feature Subset Selection**
  - Filter
  - Wrapper
  - Embedded

Variable Ranking ⇒ It is the process of ordering features by the value of some scoring function $s(f_i)$, which usually measures feature-relevance.
It is to select k highest ranked features according to s.

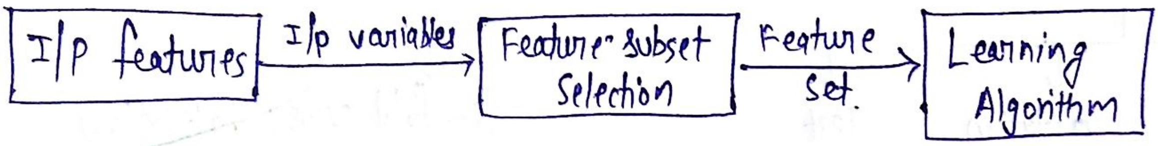
Correlation $R(f_i, y) = \dfrac{Cov(f_i, y)}{\sqrt{var(f_i)\, var(y)}}$



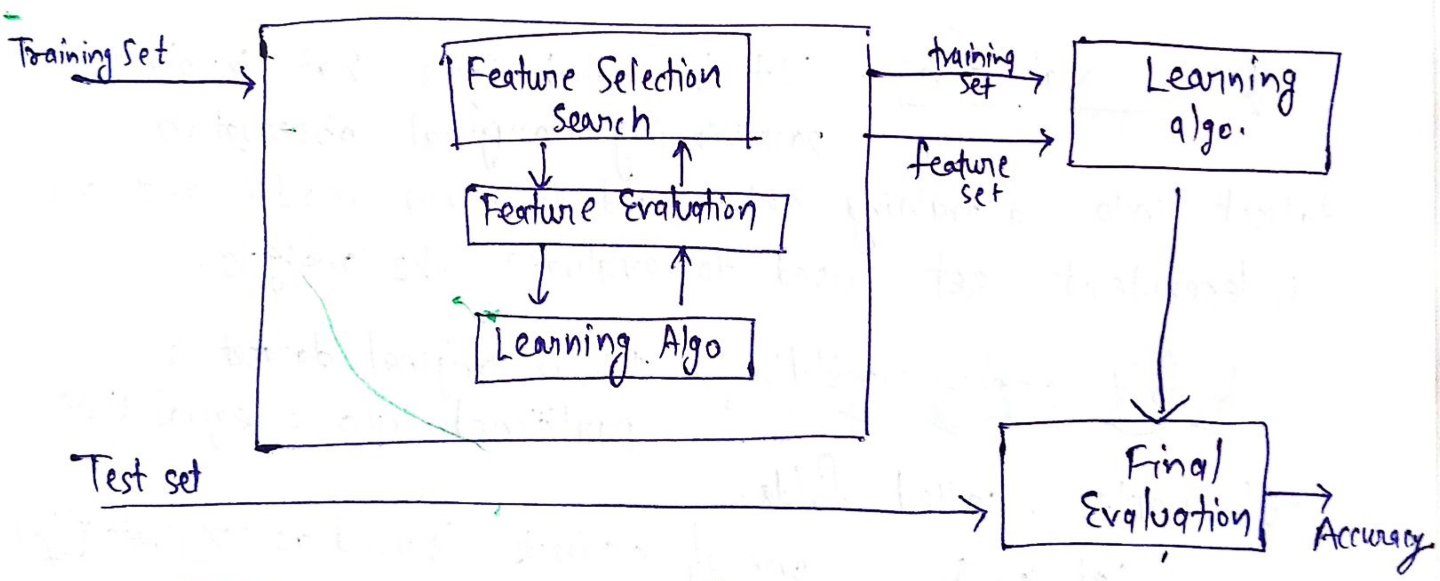
Feature Subset Selection :- It is to find optimal feature subset.

(i) **Filter Methods** :- Select subsets of variables as a pre-processing step, independently of used classifier.

→ It would be worthwhile to note that variable Ranking feature selection is a filter Method.



(ii) **Wrapper Methods** :- The learner is considered a black-box. Interface of black-box is used to score subsets of variable according to predictive power of the learner when using subsets.



(iii) **Embedded** :- It is specific to a given learning machine. Performs variable selection in process of training.

Eg- Window-algo.

Evaluating Machine learning Algorithms & Models ⇒

Methods for evaluating a model's performance are divided into 2 categories:– Holdout & Cross validation.

- Holdout
  - Training set
  - Validation set
  - Test set
- Cross-validation
  - k-fold cross validation

Holdout:– The purpose of holdout evaluation is to test a model on different data than it was trained on. This provides an unbiased estimate of learning performance.

Cross-validation:– It is a technique that involves partitioning original observation dataset into a training set, used to train model and an independent set used to evaluate the analysis.

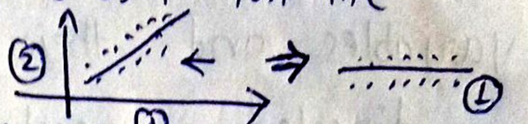
k-fold cross-validation ⇒ where original dataset is partitioned into k equal size subsamples, called folds.

Let $k = 5$, one of k subsets is used as test/validation set, k-1 to form training set.

→ The error estimation is averaged over all k trials to get total effectiveness of our model.

Principal Component Analysis → It is an unsupervised learning algo. that is used for the dimensionality reduction in ML.

ML



It is a statistical process that converts the observations of correlated features into a set of linear uncorrelated features ~~into a~~ with help of orthogonal transformation.

It is used in

→ Image processing

→ Movie Recommendation System

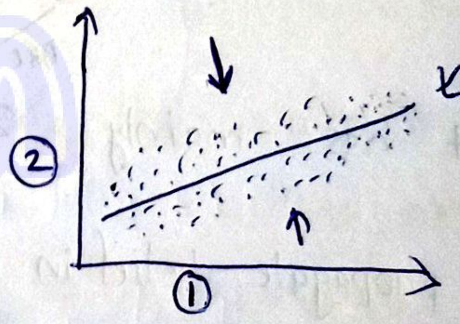
→ Optimizing power allocation in various commn channels.

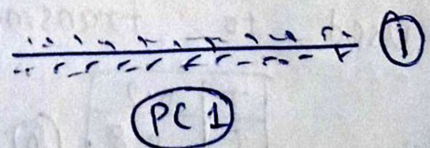
It is based on

→ Variance & Covariance

→ Eigenvalues & Eigen factors

PC → views





ANN :- Artifical Neural Network provides basic & advanced concepts of ANNs. The term "ANN" refers to a biologically inspired sub-field of artifical intelligence modeled after the brain. An ANN is usually a computational network based on biological neutral networks that construct the structure of human brain.

→ ANN is derived from biological neural networks that develop the structure of human brain.

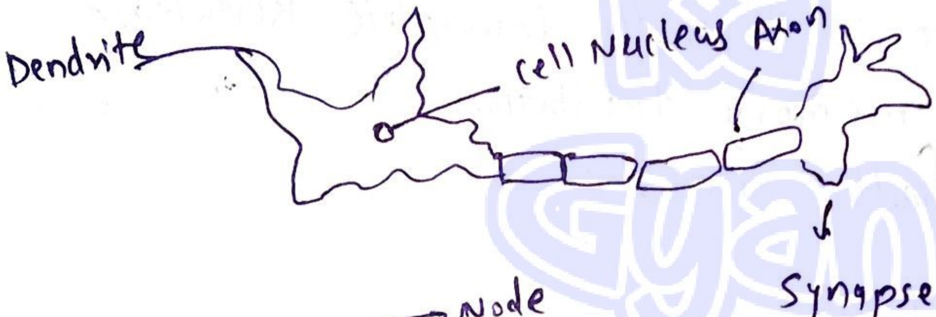


fig:- BNN

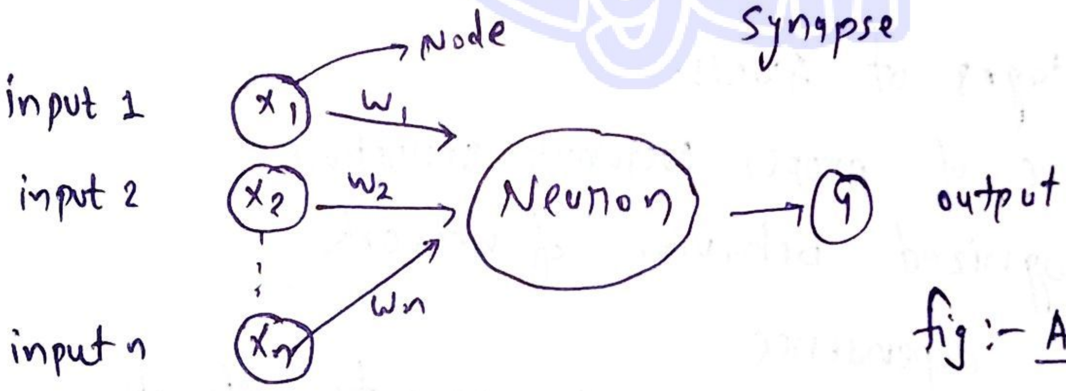
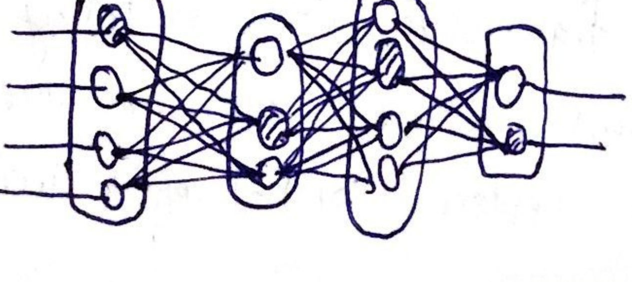


fig:- ANN

Architecture of an ANN :-

In order to define a neural network that consists of a large no. of artifical neurons, which are termed units arranged in a sequence of layers.



- input layer
- Hidden L_1
- Hidden L_2
- O/P Layer

Input layer:- It accepts inputs in several different formats provided by programmer.

Hidden Layer:- The hidden layer presents in b/w i/p & o/p ②

o/p Layer:- The i/p goes through a series of transformations using hidden layer, which finally results in o/p

$$\sum_{i=1}^{n} w_i * x_i + b$$

Advantages of ANN:-

(i) parallel processing capability

(ii) Storing data on entire network

(iii) Capability to work with incomplete knowledge

(iv) Having a memory distribution

(v) Having fault tolerance

Disadvantages of ANN:-

(i) Assurance of proper Network structure

(ii) Unrecognized behavior of network

(iii) H/w dependence

(iv) Difficulty of showing the issue to network

(v) The duration of network is unknown.