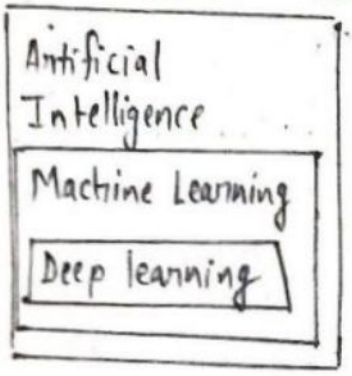
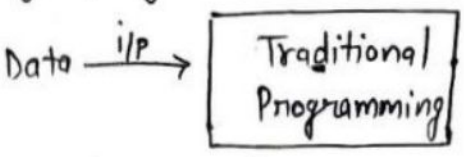
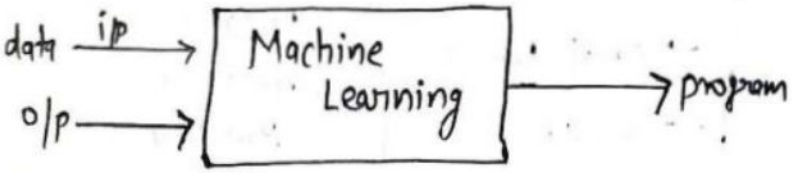
Machine Learning

→ 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.



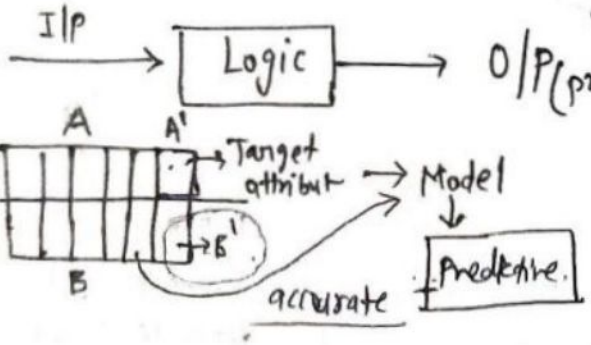




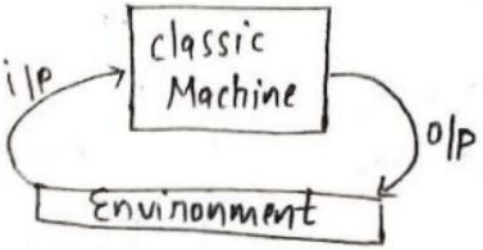
Training & Testing

Training :- I/P → Algo. → Logic / Model

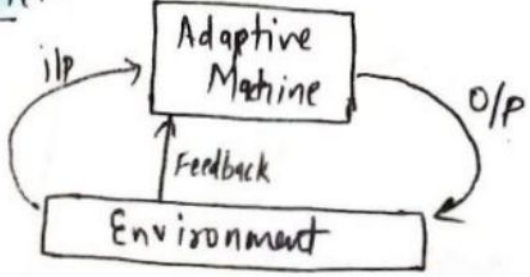
Testing :-



Classic Machine / Non- Adaptive Machine :-



Adaptive Machine :-



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

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

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

- $P(c|x)$: Posterior probability
- $P(x|c)$: Likelihood
- $P(c)$: class prior prob.
- $P(x)$: predictor prior prob.

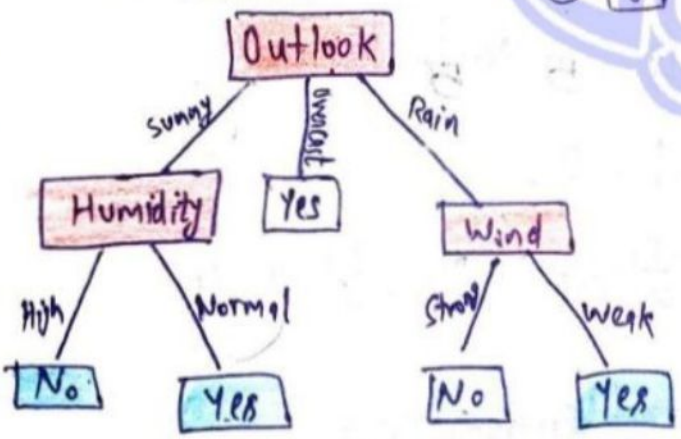
$$P(c|x) = P(x_1|c) \cdot P(x_2|c) \cdots P(x_n|c) \cdot P(c)$$

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.



→ Decision node (Test)

→ Terminal node / Leaf node (classification)

→ Splitting

→ pruning

→ Root node

→ Branch / sub-tree

Decision Tree

Type of DT

- Categorical DT
- Continuous DT

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

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

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

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

Value (outlook) = Sunny, overcast, Rain

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

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

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

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

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

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

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

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

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) ✓

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

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

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

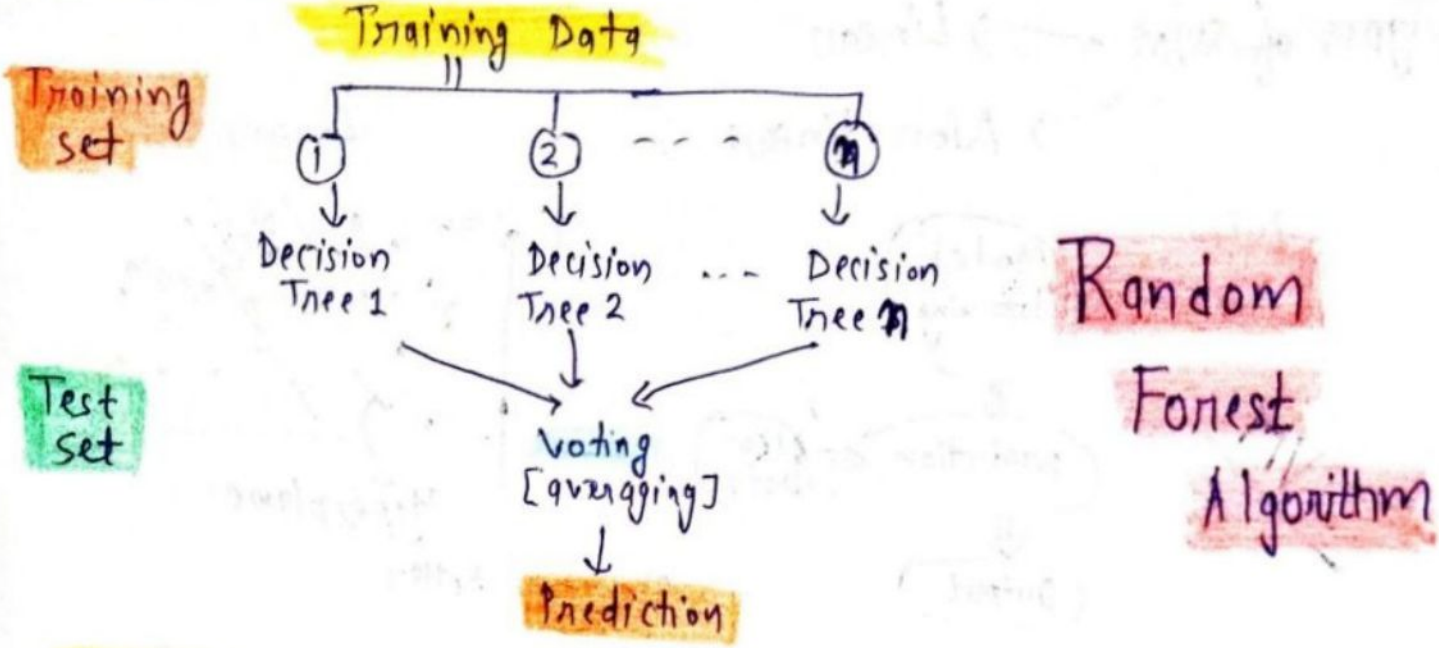
$D(x, iv) = \sqrt{(3-1)^2 + (7-4)^2} = 3.6 \longrightarrow N_2 \rightarrow$ True

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 line [Hyperplane] 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 chooses extreme points that help in creating Hyperplane. These extreme points are called support vector.



Random Forest Algorithm

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

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

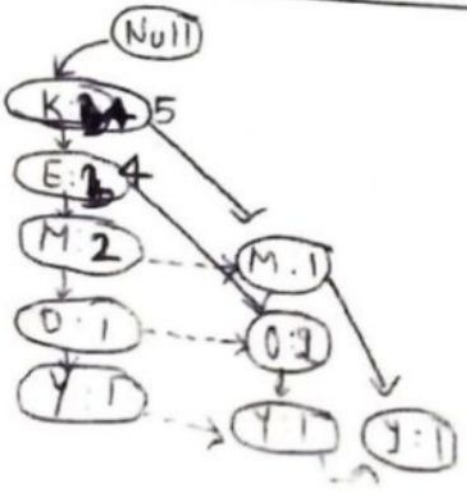
FP Growth Algorithm

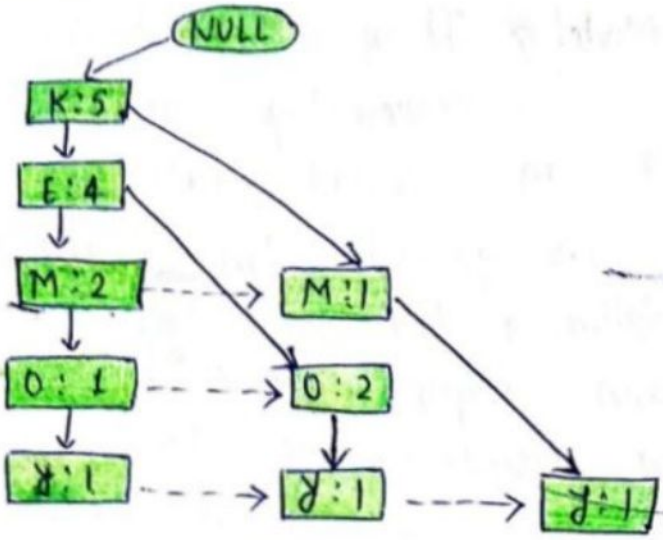
Item	Frequency
A	1
C	2
D	1
E	4
I	1
K	5
M	3
N	2
O	3
U	1
Y	3

⇒

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

Dimensionality Reduction

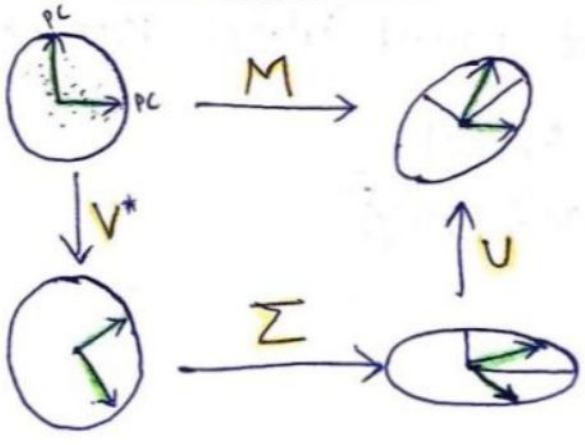
- Feature Selection
  - Filters Methods
  - Wrapper Methods
  - Embedded Methods
- Feature Extraction
  - Principal Component Analysis
  - Factor Analysis
  - Singular Value Decomposition

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.



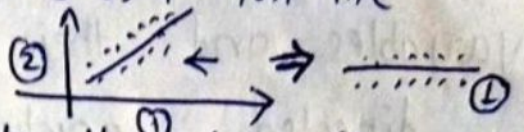
It is a method of decomposing a matrix into 3 other matrices

$$M = U \cdot \Sigma \cdot V^*$$

(Σ: Diagonal matrix; U: orthogonal)

$$AA^T = A^TA$$

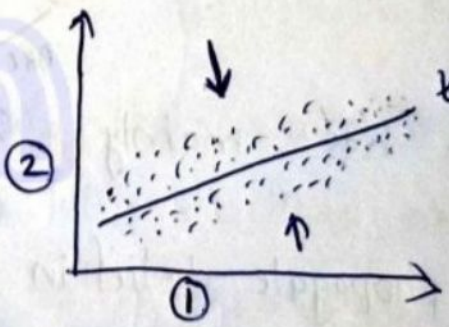
Principal Component Analysis ⇒ It is an unsupervised learning algo. that is used for the dimensionality reduction in ML.



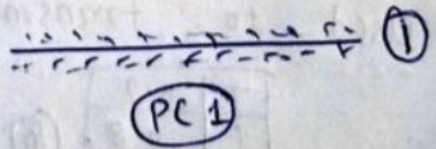
It is a statistical process that converts the observations of correlated features into a set of linear uncorrelated features ~~into~~ with help of orthogonal transformation.

It is used in
- → Image processing
- → Movie Recommendation System
- → Optimizing power allocation in various comm'n channels

It is based on
- → Variance & Covariance
- → Eigenvalues & Eigen factors



PC → views



Purchase the Notes

100₹

Per semester
(All subjects)

Notes (Hand written) ✅
Most Questions ✅

All Branches

Min 100%
amount will go
into charity ⭐

For specific
Subject - 50₹

UPI ID -
sahilkagyan337@ybl

Er Sahil ka Gyan

🔥 Steps for getting NOTES and Most Questions -

👉Do payment using UPI ID -

sahilkagyan337@ybl

👉Take screenshot of transaction and send me on Email -

ersahildrive@gmail.com

Then finally access all Notes and most questions 🔥

Scan & Pay Using PhonePe App

SAHIL KHAN