Deep Learning (Unit - 1)

Syllabus :- Fundamentals of Deep Learning

1. Artificial Intelligence
2. History of Machine Learning :-
   (a) Probabilistic Modeling
   (b) Early Neural Networks
   (c) Kernel Methods
   (d) Decision Trees
   (e) Random Forests and Gradient Boosting Machines

Fundamentals of Machine Learning :-

3. Four Branches of Machine Learning
4. Evaluating Machine Learning Models
5. Overfitting and Underfitting.

⇒ Fundamentals of Deep Learning :-

(1) Artificial Intelligence :-

⇒ The term "Artificial Intelligence" refers to the Simulation of human intelligence Processes by Machines Especially Computer Systems, Voice recognition, games.

⇒ A.I Programming focuses on three Processes

1. Learning Process
2. Reasoning Process
3. Self Correction Process

1. Learning Process :-

⇒ This part of A.I Programming is Concerned with gathering data and Creating rules for transforming it into Useful information.

⇒ The rules, which are also called as algorithms

2. Reasoning Process :-

⇒ This Part of A.I Programming is Concerned with Selecting the best Algorithm to Produce accurate output.

3. Self - Correction Process :-

⇒ This Part of A.I Programming is designed to Continually fine-tune algorithms and Ensure

they Provide the most accurate results possible.

⇒ Applications :-

1. Google Maps
2. Face detection and Recognition
3. Health Care
4. Smart Home devices
5. Banking and Finance
6. Gaming
7. Social Media
8. Self driving Cars.

} Describe these in your own wo

---

(2) History of Machine Learning :-

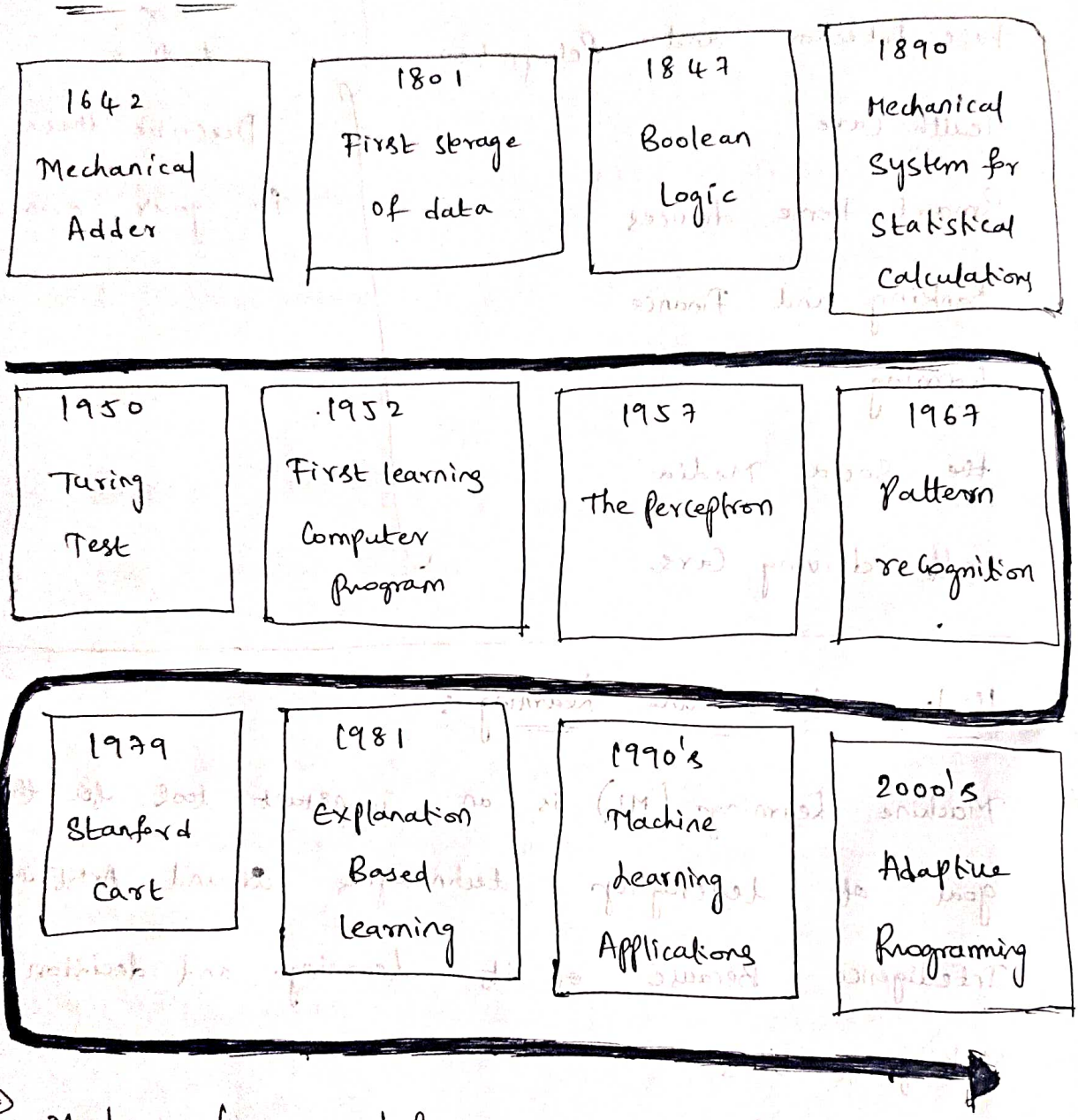
⇒ Machine Learning (ML) is an important tool for the goal of leveraging technologies around Artificial Intelligence Because of its learning and decision Making abilities.

⇒ Machine Learning is often referred to as A-I, though in reality, it is a sub division of A-I, Until the late 1970's it was a part of A-I's Evolution.

⇒ The term Machine Learning was coined in 1959 by Arthur Samuel, an IBM Employee and

Pioneer in the field of computer gaming and artificial Intelligence.

⇒ Architecture :-



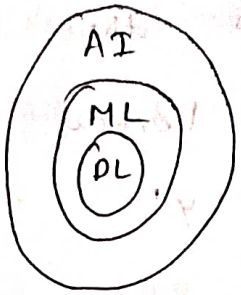
⇒ Machine Learning definition :-

⇒ Machine Learning is an A.I technique that teaches Computers to learn from Experience

⇒ Machine learning Algorithms use computational methods to "learn" information directly from data without relying on a Predetermined Equation as a model.

⇒ A Algorithm that has ability to Learn from Past Experience

1. Supervised Machine Learning Algorithm.
2. Un-Supervised Machine Learning Algorithm
3. Semi-supervised Machine Learning Algorithm.
4. Reinforcement Machine Learning Algorithm.



Ex :- Face detection
Voice recognition
Face recognition | Identification
Self driving cars.

(a) Probabilistic Modelling :-

⇒ Probability is used to find how many success and failure cases will be there i.e how many possibility cases will be there in a model. Some of the models used are

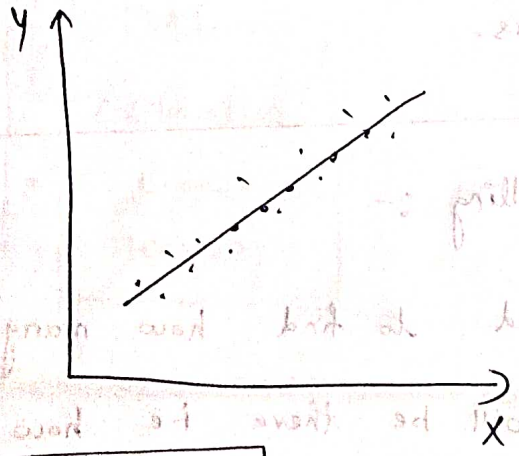
1. Linear model
2. Logistic model
3. Poisson model
4. Bernoulli (or) Binomial model
5. Naive Bayes model

1. Linear Regression model :-

⇒ A Simple Linear Regression model only has two Explanatory Variables one dependent and one independent Variables.

⇒ Think of a two - dimensional space where the horizontal axis represents the independent Variable 'X' and the Vertical axis represents the dependent Variable 'Y'

⇒ It assumes that the relation between the data points is linear i.e, the plot of the data points shows in linear



$$Y = mx + c$$

Y = Independent Variable

X = Dependent Variable

m = Slope

C = Constant

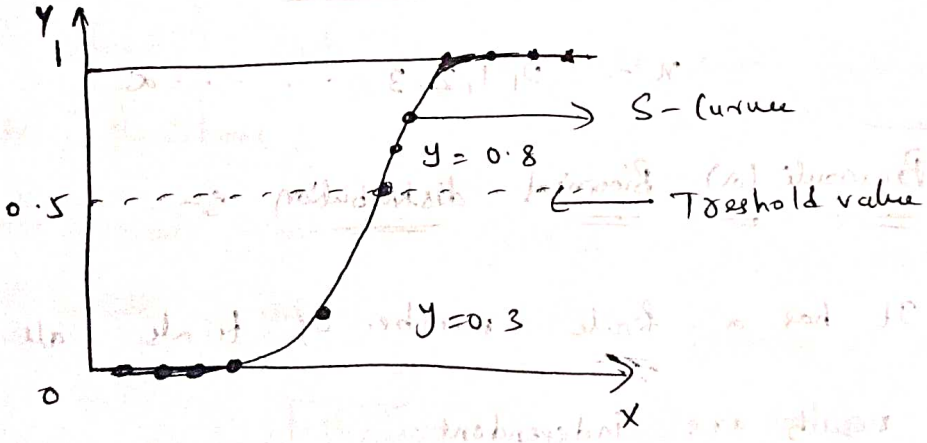
2. Logistic Regression Model / Sigmoid function :-

⇒ The Sigmoid function is a mathematical function used to Map the Predicted Values to Probabilities

⇒ It maps any real value into another value within a range of 0 and 1.

⇒ The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit so it forms curve like "S" form

⇒ The 'S' form curve is called the Sigmoid function (or) the logistic function



Sigmoid function = $\sigma(x) = \frac{1}{1+e^{-x}}$

⇒ Above the threshod value tends to the value 1 and below the threshod value tends to the value 0

3. Poisson Regression (or) distribution model :-

⇒ It has a Infinite number of trials all the trial results are independent

⇒ Probability of Success and Probability of failures are zero.

⇒ 'X' is a discrete random Variable it follows Poisson distribution then Probability mass function of Poisson distribution is

$$P(X=x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

Ex :- rolling a dice

$x = 0, 1, 2, 3 \ldots \ldots \infty$

4. Bernoulli (or) Binomial distribution :-

⇒ It has a finite number of trials all the trial results are independent.

⇒ Probability of Success and Probability of failure are constant for Each trial

⇒ 'X' is a discrete function random Variable and it follows the binomial distribution then the Probability mass function of binomial distribution is

$$P(X=x) = {}^{n}C_x \; p^x \; q^{n-x}$$

$$^{n}C_{x} = \frac{n!}{x!(n-x)!}$$

where

n = no. of trials

x = no. of times success

p = Probability of Success

q = Probability of failure.

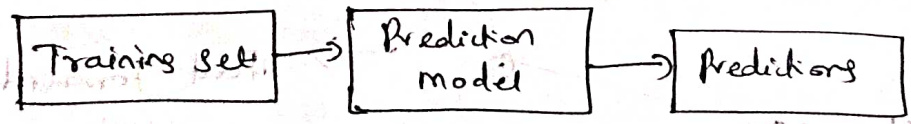
---

5. Naive Bayes theorm :-

⇒ A classifier Under Supervised ML group based on Probabilistic logic

⇒ Set of Probabilities of $\{P_0, P_1, \ldots, P_k\}$ for each attribute from each class set it uses Probabilistic to make predictions.

⇒ The data model which is yielded is called as Predictive model with Probabilistic problems at foundation



Naive Bayes.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A)}$$

(b) Early Neural Networks :-

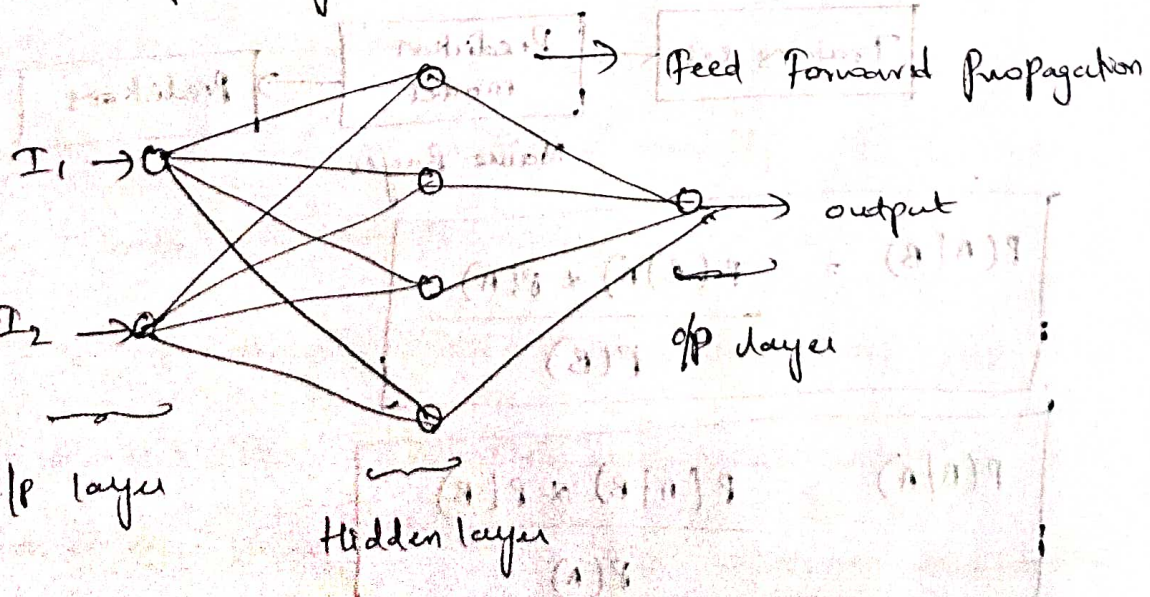
⇒ The First artificial Neural Network (ANN) was invented in 1958 by Psychologist Frank Rosenblatt called Perceptron, it was intended to model how the human brain Processed Visual data and learned to recognize objects. other researchers have Since used Similar ANN's to Study human Cognition

⇒ Artificial Neural Networks (ANN) :-

⇒ Artificial Neural Networks (ANN) are algorithms based on brain function and are used to model complicated patterns and Forecast issues

⇒ There are 3 layers in the network Architecture

1. The Input layer
2. The Hidden layer (It can be more than one)
3. The output layer

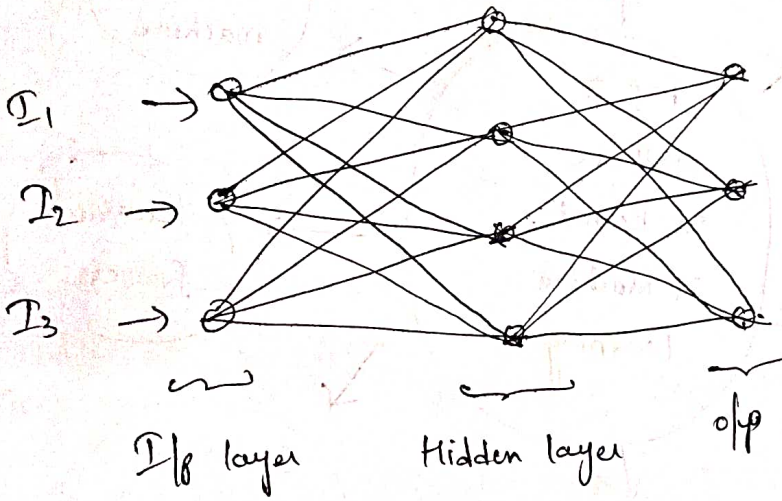


⇒ The ANN Specially there are two types of Networks which we have been used basically

1. SNN (Single Neural Networks)
2. MNN (Multi Neural Networks)

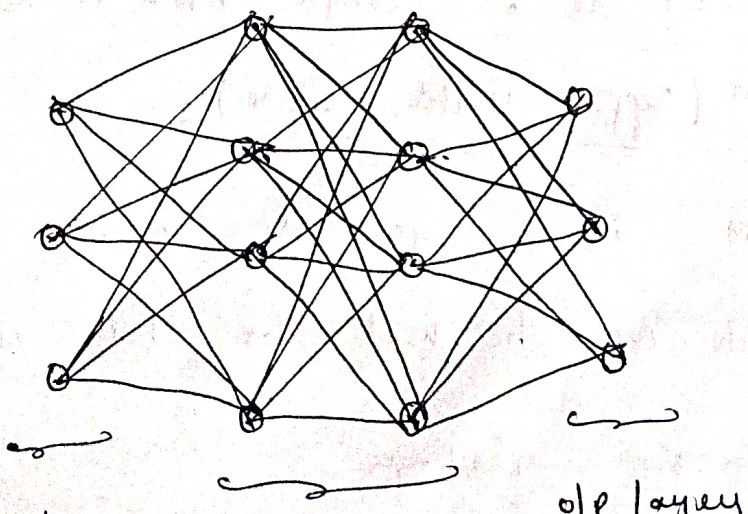
1. SNN (Single Neural Networks) :-

⇒ In this Neural Networks we will be used only one hidden layer to process the output.



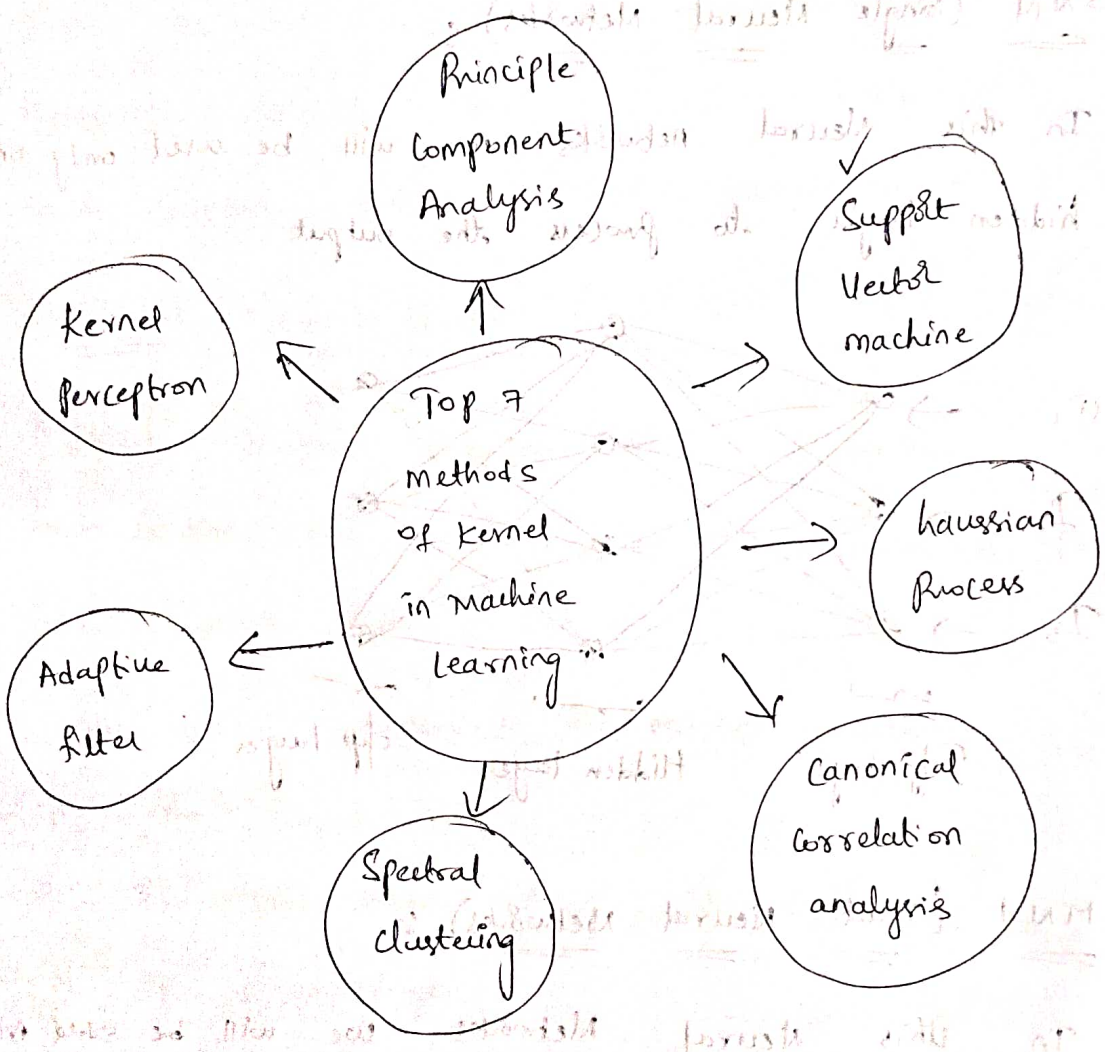
2. MNN (Multi Neural Networks) :-

⇒ In this Neural Networks we will be used more than one hidden layers to process the output.



© Kernel Methods :-

⇒ Kernel methods are types of algorithms that are used for pattern analysis. These methods involve using linear classifiers to solve non-linear problems

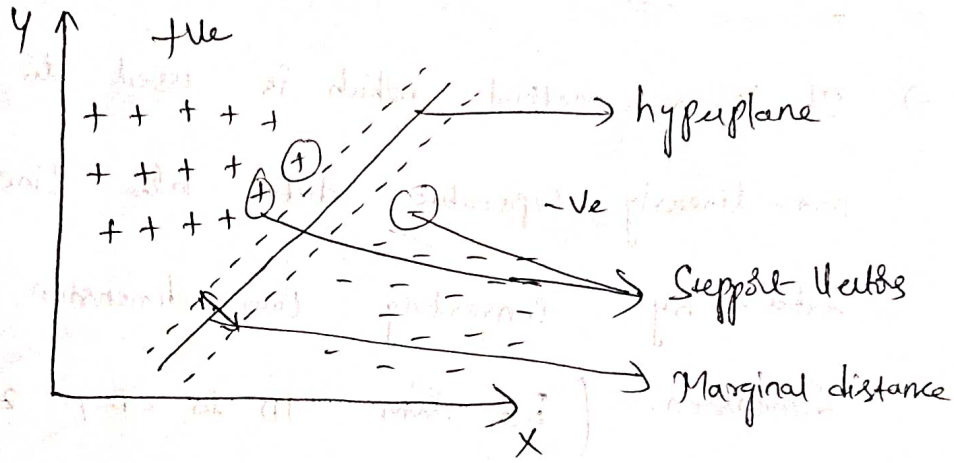


⇒ In the above methods we can SVM (Support Vector Machine) as an Example which is used

⇒ SVM (Support Vector Machine) :-

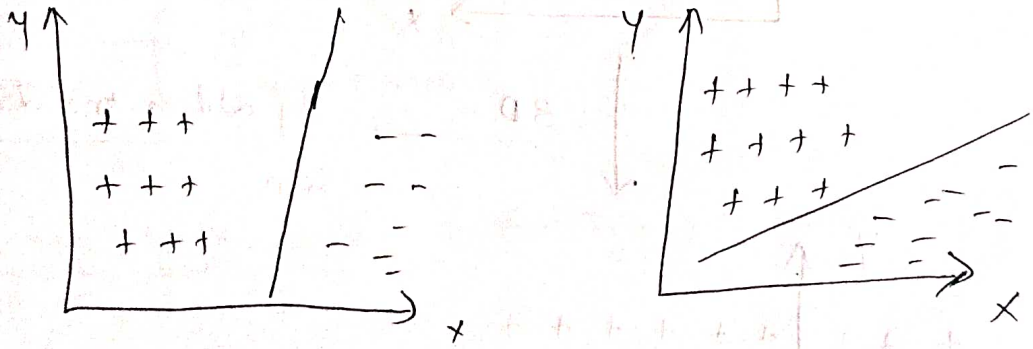
⇒ SVM is a Supervised Machine learning algorithm which can be used for both classification (&) regression challenges

1. Suppst Vectors
2. Hyperplanes
3. Marginal distance
4. Linearly Separable
5. Non-linearly Separable.



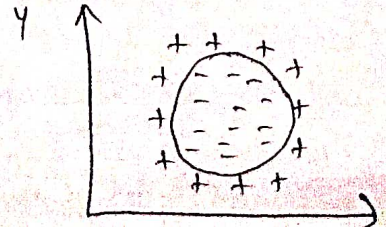
⇒ Linearly Separable :-

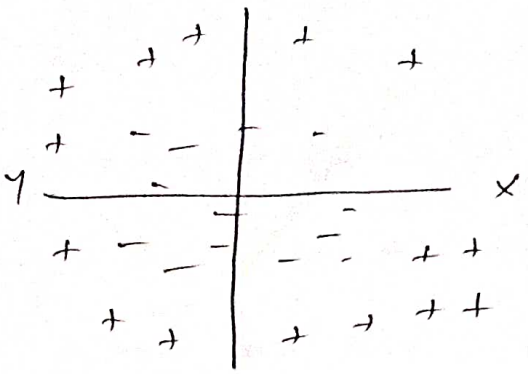
⇒ The data which has been Separated by a Straight line is called linearly Separable SVM



⇒ Non-linearly Separable :-

⇒ The data which has been not Separated by a Straight line is called Non-linearly Separable SVM.

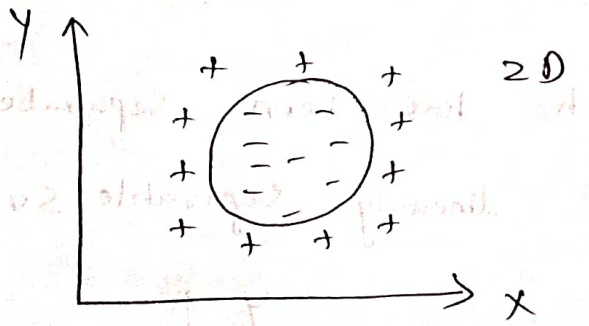




**

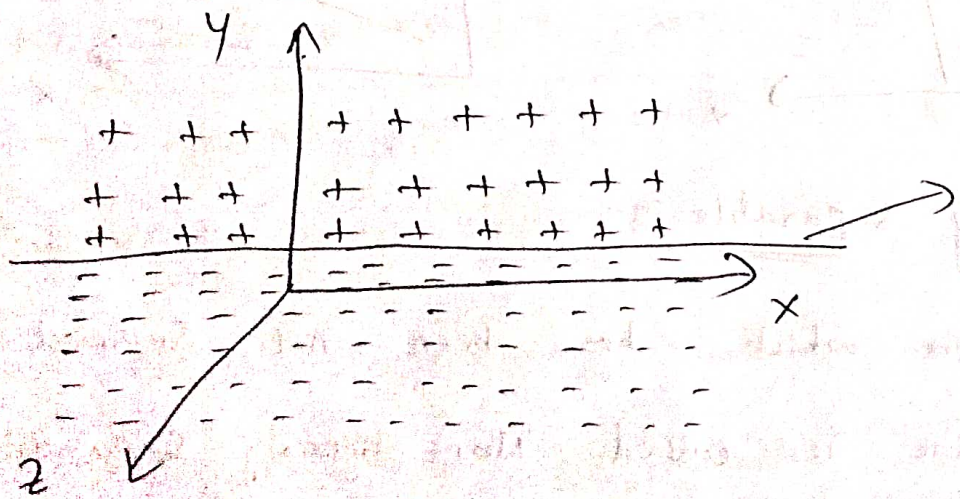
⇒ SVM Kernel :-

⇒ It is a method which is used to convert non-linearly separable data into Linearly Separable data. by converting low dimension into high dimension. (i.e from 1D to 2D, 2D to 3D ...)



→ non-linearly Separable

3D → By adding one dimension

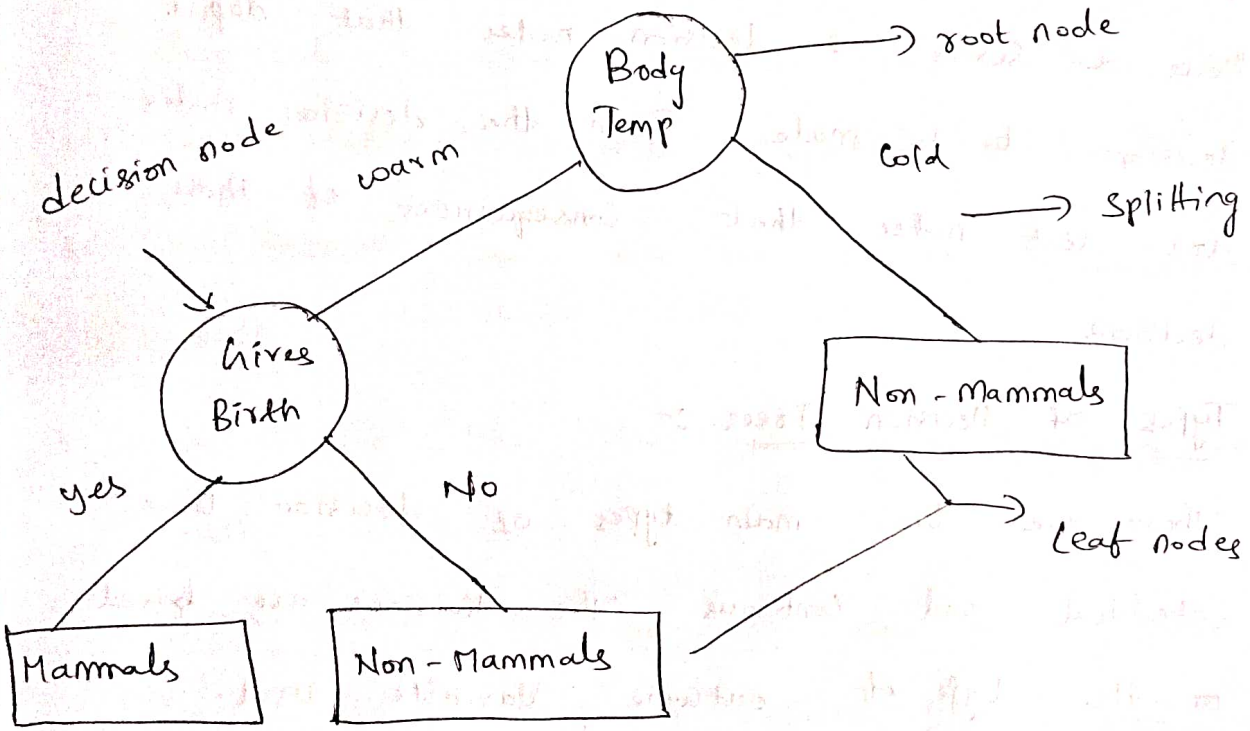


→ linearly Separable

(d) Decision Trees :-

=) A decision Tree is a type of Supervised Machine learning used to categorize & make predictions based on how a previous set of Questions were answered

=) The model is a form of supervised learning, meaning that the model is trained and tested on a set of data that contains the desired categorization.



=) Key terms of a decision Tree :-

1. Root node :-

=) The base (or) starting node of the decision Tree

2. Splitting :-

=) The process of dividing a node into multiple sub-nodes

3 Decision-node :-

=) when a sub node is further split into additional

Sub-nodes

4. Leaf node :-

⇒ When a Sub-node does not further split into additional sub-nodes (or) last nodes

⇒ Decision Tree Working :-

⇒ A decision Tree resembles, well, a tree. The base of the tree is the root node. From the root node flows a series of decision nodes that depict decisions to be made. From the decision nodes are leaf nodes that consequences of those decisions.

⇒ Types of Decision Trees :-

⇒ There are two main types of decision trees Categorical and Continous. The divisions are based on the type of outcome Variables used.

1. Categorical Variable decision Tree :-

⇒ In a Categorical Variable decision Tree, the answer neatly fits into one category (or) another. Was the coin toss heads (or) tails ? Is the animal a reptile (or) mammal ? In this type of decision tree, data is placed into a single category based on the decisions at the nodes throughout the tree.

2. Continous Variable decision Tree :-

⇒ A Continous Variable decision tree is a decision tree with a Continous target Variable

⇒ For Example, the income of an individual whose income is Unknown can be predicted based on available information such as their occupation, age and other Continous Variables.

⇒ Advantages :-

1. Compared to other algorithms decision trees requires less effort for data preparation during pre-processing

2. A decision tree does not require normalization of data

3. A decision tree does not require Scaling of data as well

4. Missing values in the data also do Not affect the process of building a decision tree to any considerable Extent

5. A decision tree model is Very intuitive and easy to Explain to technical teams as well as Stakeholders

⇒ Disadvantages :-

1. A Small change in the data can cause a large change in the Structure of the decision tree causing instability.

2. For a decision tree Sometimes Calculation can go for more Complex Compared to other algorithms.

3. A decision tree often involves higher time to train the model.

4. Decision tree training is relatively Expensive as the Complexity and time has taken are more.

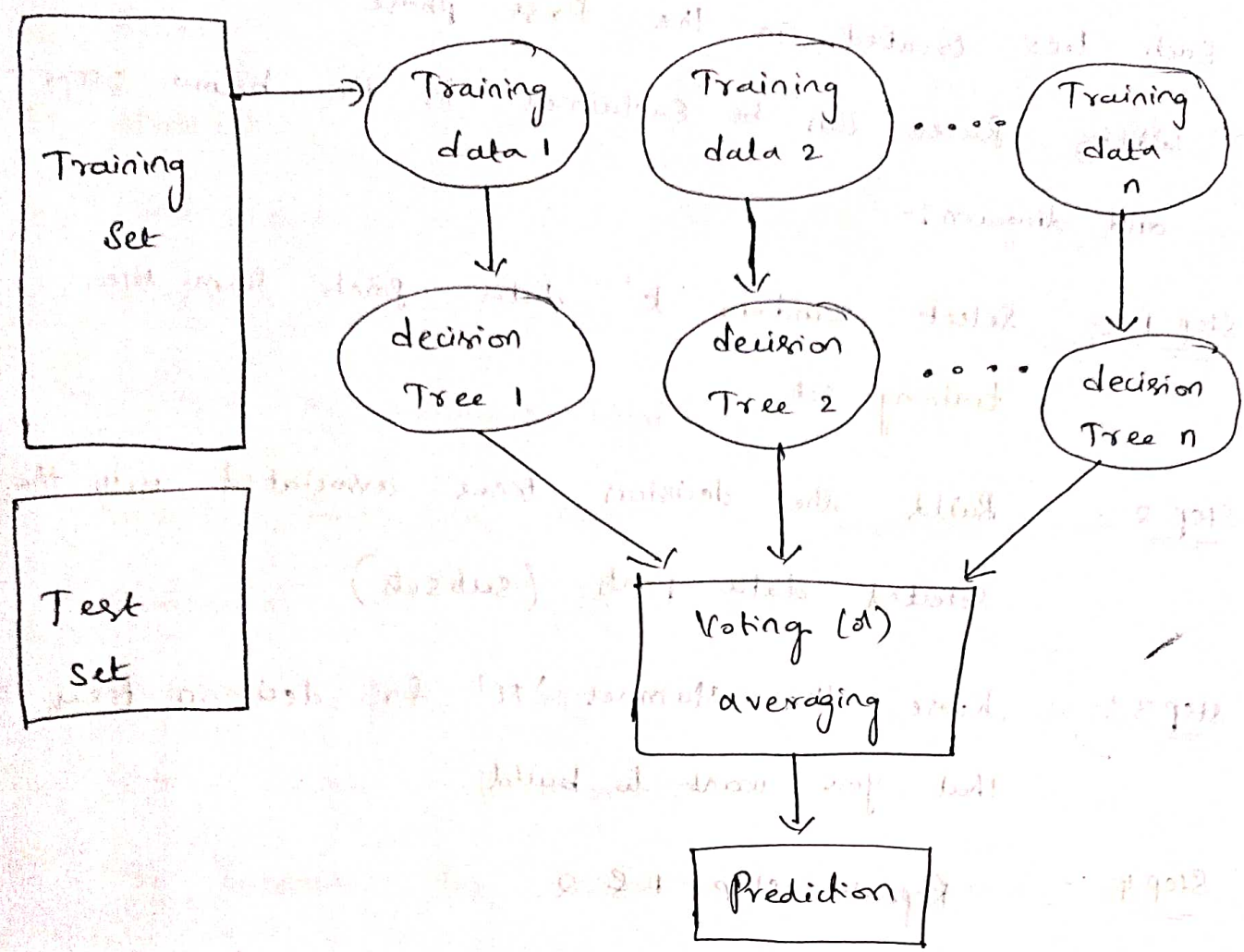
(e) Random Forest Algorithm :-

⇒ Random Forest is a popular machine learning algorithm that belongs to the Supervised learning technique it can be used for both classification and Regression Problems in M.L

⇒ It is based on the concept of Ensemble learning, which is a process of Combining Multiple classifiers to Solve a Complex Problem and to improve the performance of the model

⇒ As the name suggests " Random Forest is a classifier that Contains a number of decision trees on Various Subsets of the given dataset and takes Average | Voting to improve the predictive accuracy of that dataset "

⇒ The below diagram Explains the working of the Random Forest algorithm :



⇒ Why use Random Forest ?

⇒ Below are some points that explain why we should use the Random Forest algorithm :-

1. It takes less training time as compared to other algorithms
2. It predicts output with high accuracy, even for the large dataset it runs efficiently.
3. It can also maintain accuracy when a large proportion of data is missing.

⇒ How does Random Forest algorithm work ?

⇒ Random Forest works in two-phase first is to create the random forest by combining 'N'

decision tree, and second is to make predictions for each tree created in the first phase

⇒ Bagging process can be explained in the below steps and diagram:-

Step 1 :- Select random 'k' data points from the training set

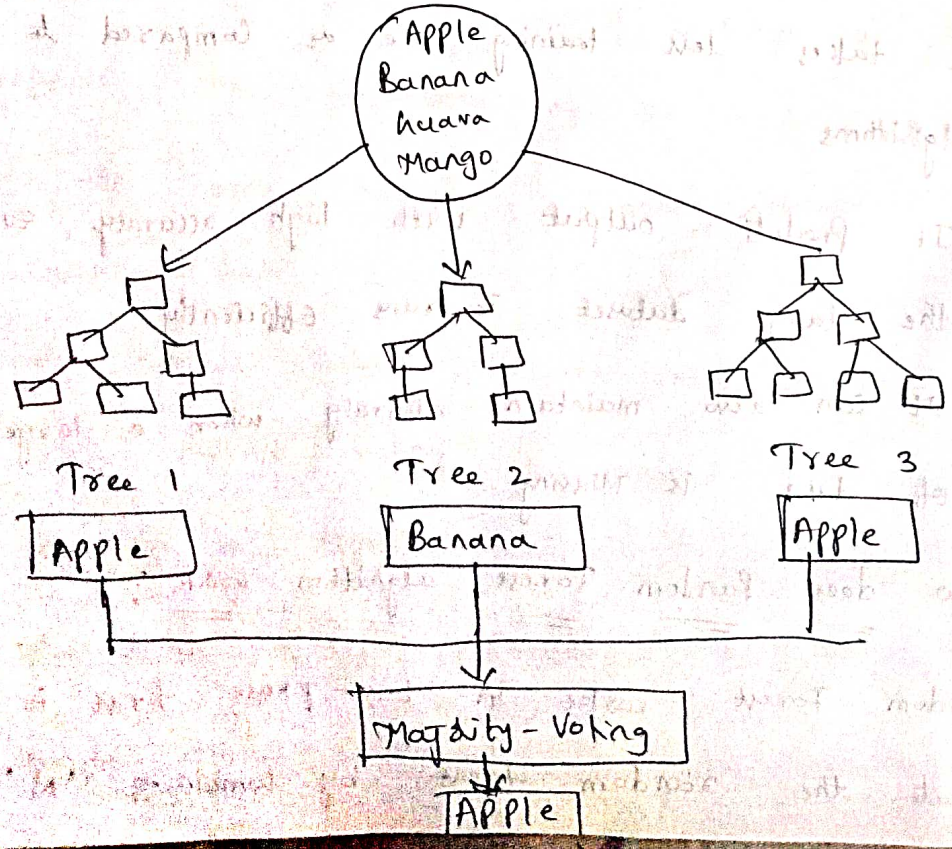
Step 2 :- Build the decision trees associated with the selected data points (subsets)

Step 3 :- choose the number 'N' for decision trees that you want to build

Step 4 :- Repeat step 1 & 2.

Step 5 :- For new data points, find the predictions of each decision tree, that wins the majority votes.

Ex :-



Applications of Random Forest :-

1. Banking
2. Medicine
3. Land use
4. Marketing

⇒ Advantages of Random Forest :-

⇒ Random Forest is capable of performing both classification and Regression tasks.

⇒ It is capable of handling large datasets with high dimensionality

⇒ It enhances the accuracy of the model and prevents the overfitting issue.

⇒ Disadvantages of Random Forest :-

⇒ Although random Forest can be used for both classification and regression tasks, it is not more suitable for Regression tasks.

(f) Gradient - Boosting technique :-

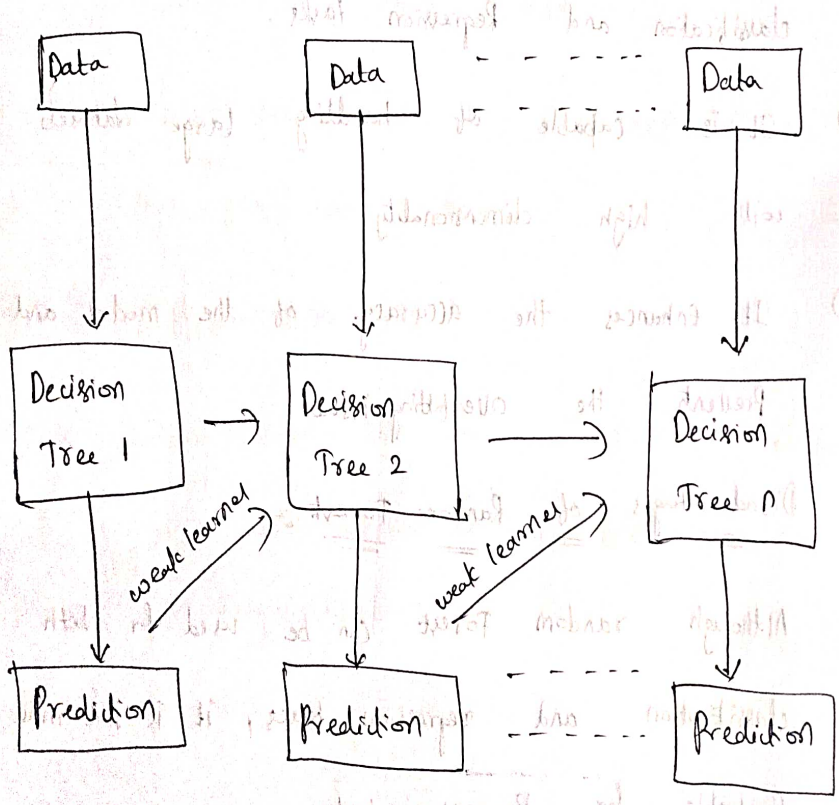
⇒ Gradient Boosting is a popular boosting algorithm. In gradient Boosting, each predictor corrects its predecessor's error.

⇒ Gradient Boosting is a machine learning technique used in regression and classification tasks, among

others, it gives a prediction model in the form of Ensemble of weak prediction models which are typically decision trees.

⇒ when the decision tree is the weak learner, the resulting algorithm is called gradient boosted trees it usually outperforms random forest.

Ex:-



⇒ Working of Gradient Boosting :-

Step 1 :- Load data

Step 2 :- divide that data into subsets and train with decision tree model.

Step 3 :- when the value has been predicted it that value is sent (ie weak learner) to

that model then it will train with another model, Until the strong accurate value occurs

Step 4 :- Step 2 & 3 repeats Until the Accurate value has been predicted

Step 5 :- We can consider that accurated value as final output.

⇒ Fundamentals of Machine learning :-

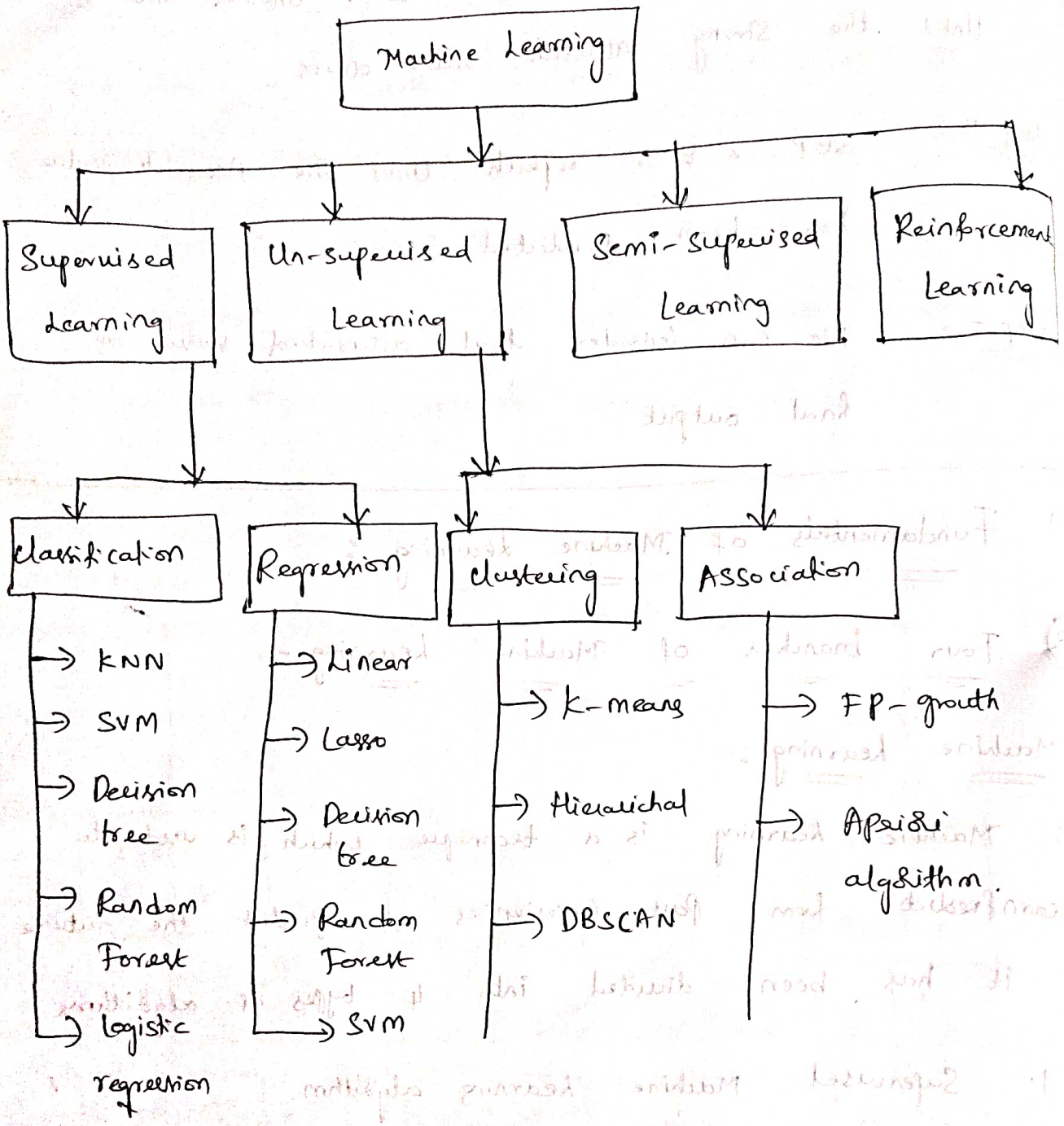
(3) Four branches of Machine Learning :-

⇒ Machine Learning :-

⇒ Machine learning is a technique which is used to Learn Predict from Past Experiences to predict the future it has been divided into 4 types of algorithms.

1. Supervised Machine learning algorithm.
2. UnSupervised Machine learning algorithm
3. Semi - Supervised Machine learning algorithm
4. Reinforcement Machine learning algorithm.

⇒ where the above algorithms has been further divided into sub parts and also it can be further divided into some algorithms in the above diagram it can be described.



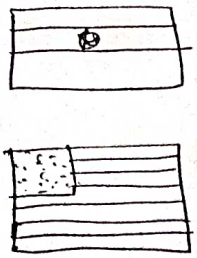
1) Supervised Learning :-

⇒ It is the Machine learning task of learning a function that maps an input to an output

⇒ It refers a function from labeled training data.

⇒ In Supervised learning, each example is a pair consisting of an input object and a desired output value.

Ex :-

Country	Colours :-
Flags  India ✓	Red
USA	Safron ✓
	Green ✓
	White ✓
	Blue

(a) classification :-

⇒ classification algorithms are used to solve the classification problems in which the output variable is categorical such as "Yes" & "No", male (M), Female (F), Red (R) Blue E.t.c

(b) Regression :-

⇒ Regression algorithms are used to solve regression problems which is used to predict continous output variables.

Ex :- Market trends, weather prediction E.t.c.

⇒ Algorithms Supported For Supervised Learning are :-

1. KNN (K-Nearest Neighbour)
2. SVM (Support Vector machine)
3. Decision tree
4. Random Forest
5. Logistic Regression

} classification algorithms

⇒ Regression Supported algorithms are :-

1. Simple Linear Regression algorithm
2. Multivariate Regression algorithm
3. Decision Tree Algorithm
4. Lasso Regression.

2) UnSupervised Machine Learning Algorithm :-

⇒ Ex :- Spam Mail box

⇒ It is an Unlabelled data (i.e it contains useful and Non-Useful information).

⇒ This Algorithm will tell you that any Suspecious mail will sent to spam (i will kill u, terrorist)

⇒ If it goes to inbox then it will find that ok it seems not suspecious for them and it has been learning from past and will not repeat again.

⇒ Categories of Unsupervised Machine Learning :-

1. clustering
2. Association.

1. clustering :-

⇒ The clustering technique is used when we want to find the inherent groups from the data it is a way to group the objects

into a cluster such that the objects with the most similarities

Ex:- grouping the customers by their purchasing behaviour

⇒ Algorithms supported for clustering are:-

1. K-Means clustering algorithm
2. Mean-shift algorithm
3. DBSCAN Algorithm
4. Hierarchical clustering algorithm.

2. Association:-

⇒ Association rule learning is an Unsupervised learning technique, which finds interesting relations among variables within a large dataset

⇒ The main aim of this learning algorithm is to find the dependency of one data item on another data item and map those variables accordingly so that it can generate maximum profit.

Ex:- Market Basket Analysis, Web usage Mining

⇒ Algorithms supported for Association are:-

1. FP-growth Algorithm
2. Apriori algorithm.

3) Semi - Supervised Machine learning Algorithm :-

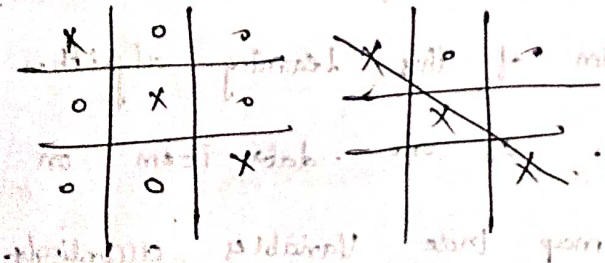
=) Semi - Supervised learning is a type of Machine learning Algorithm that lies between Supervised and Unsupervised Machine learning.

=) It represents the intermediate ground between Supervised (with labelled training data) and Unsupervised learning (with no labelled training data) algorithms and uses the combination of labelled and Unlabelled datasets during the training data.

4) Reinforcement Machine learning algorithm :-

=) It is specially used in games which is Unlabelled dataset

Eg:-



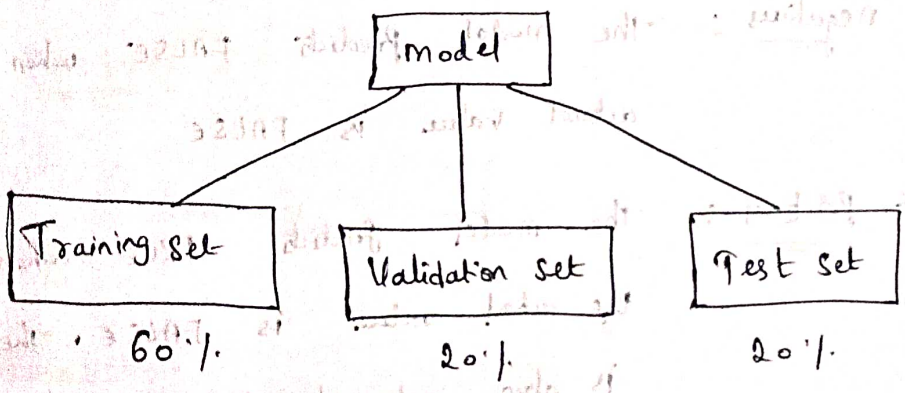
← Machine learning from past

=) Tic-Tac-Toe game if the 'x' will be in diagnal - (or) row (or) column we will cross it. it is the rule of that game then it will identify that do put 'o' b/w 'x' then we can't cancel it, Hence it will learns from past things.

(4) Evaluation of Machine learning Models :-

=> Model creation and training :-

=> once we have split the dataset into three chunks, we can start the training process. We use the training set to construct the machine learning model. Then, we use the validation set to validate the model. once the model has been trained, we use the test set to find the final performance of the model



=> Model Evaluation :-

=> Based on the performance of the test data, we can create a Confusion matrix, this matrix contains four different parameters : true positive (TP), true negative (TN), False positive (FP), False Negative (FN) Consider the following Confusion Matrix by using the above parameters.

⇒ Confusion Matrix :-

	Predictive : Positive	Predictive : Negative
Actual : Positive	True Positive (TP)	False Negative (FN)
Actual : Negative	False Positive (FP)	True Negative (TN)

⇒ This matrix shows four distinct parameters.

1. True positives :- The model predicts TRUE when the actual value is TRUE

2. True negatives :- The model predicts FALSE when the actual value is FALSE

3. False positives :- The model predicts TRUE when the actual value is FALSE. This is also referred to as a Type I Error

4. False Negatives :- The model predicts FALSE when the actual value is TRUE. This is also referred to as Type 2 Error.

⇒ once we know about the confusion matrix, we can compute several accuracies of the model, including precision, negative predictive value (NPV), Sensitivity, Specificity, and accuracy

⇒ Lets take a look at each of them, one by one and learn how they can be computed.

1. Precision :-

=) The Precision is the ratio of True positive and the sum of true positive and False positive the formula as follows

$$\text{Precision} = \frac{TP}{(TP + FP)}$$

2. Negative predictive value (NPV) :-

$$NPV = \frac{TN}{(TN + FN)}$$

3. Sensitivity :-

$$\text{Sensitivity} = \frac{TP}{(TP + FN)}$$

4. Specificity :-

$$\text{Specificity} = \frac{TN}{(TN + FP)}$$

=) Finally the accuracy of the model is given by the following formula.

$$\text{accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)}$$

Example :-

⇒ Consider we built a Supervised classification algorithm that looks at the picture of a window and classifies it as dirty & not dirty. The final Confusion Matrix is as follows.

	Predicted : Dirty	Predicted : Not dirty
Actual : Dirty	TP = 90	FN = 40
Actual : Not dirty	FP = 10	TN = 60

⇒ Now lets compute the accuracy measure as follows :-

1. Precision = $\frac{TP}{(TP + FP)} = \frac{90}{(90+10)} = 90\%$

⇒ This means 90% of the pictures that were classified as dirty were actual dirty.

2. Sensitivity = $\frac{TP}{(TP+FN)} = \frac{90}{(90+40)} = 69.23\%$

⇒ this means 69.23% of the dirty windows were correctly classified and excluded from all non-dirty windows.

3. Specificity :- $\frac{TN}{(TN+FP)} = \frac{60}{(10+60)} = 85.71\%$

⇒ This means that 85.71% of the non-dirty windows were accurately classified, and excluded from the dirty windows

4. Accuracy :- $\frac{(TP+TN)}{(TP+TN+FP+FN)} = 75\%$

⇒ This means 75% of the samples were correctly classified.

---

(5) Overfitting and Underfitting :-

1. Overfitting :-

⇒ It means model performs well on training data but it does not generalize well

⇒ overfitting happens when the model is too complex relative to the amount and noiseness of the training data. The possible solutions are :-

1. To simplify the model by selecting one with fewer parameters (Eg :- a linear model rather than a high-degree polynomial model)
2. To gather more training data

3. To reduce the noise in the training data.

(Eg:- Fix data Errors and remove outliers)

2) Underfitting Training data :-

⇒ when the model is too Simple

1. the main options to fix this problem are :-

Selecting a more powerful model with more Parameters

2. Feeding better Features to the learning algorithm (Feature Engineering) Reducing the Constraints on the model

Example :-

⇒ Considering Complex model for the Simple data like High degree model rather than Linear model.

⇒ Overcomes of overfitting and Underfitting :-

⇒ Firstly

Simple models = linear models

Complex models = High degree Polynomial models

⇒ In overfitting, If we Considered "high degree Polynomial models" for Complex data it will be leads to Accurate results.

2. In Underfitting, if we considered "Simple Linear models" for Simple data it will leads to Accurate results.

Simple models → Linear models

Complex models → High degree Polynomial models

Linear model → $y = mx + c$ (/)

High degree Polynomial models →

Second degree $ax^2 + x + 1$ → (∪)

Third degree $ax^3 + bx^2 + cx + 1$ → (∿)

Fourth degree $ax^4 + bx^3 + cx^2 + dx + 1$ → (∿)