20/10.

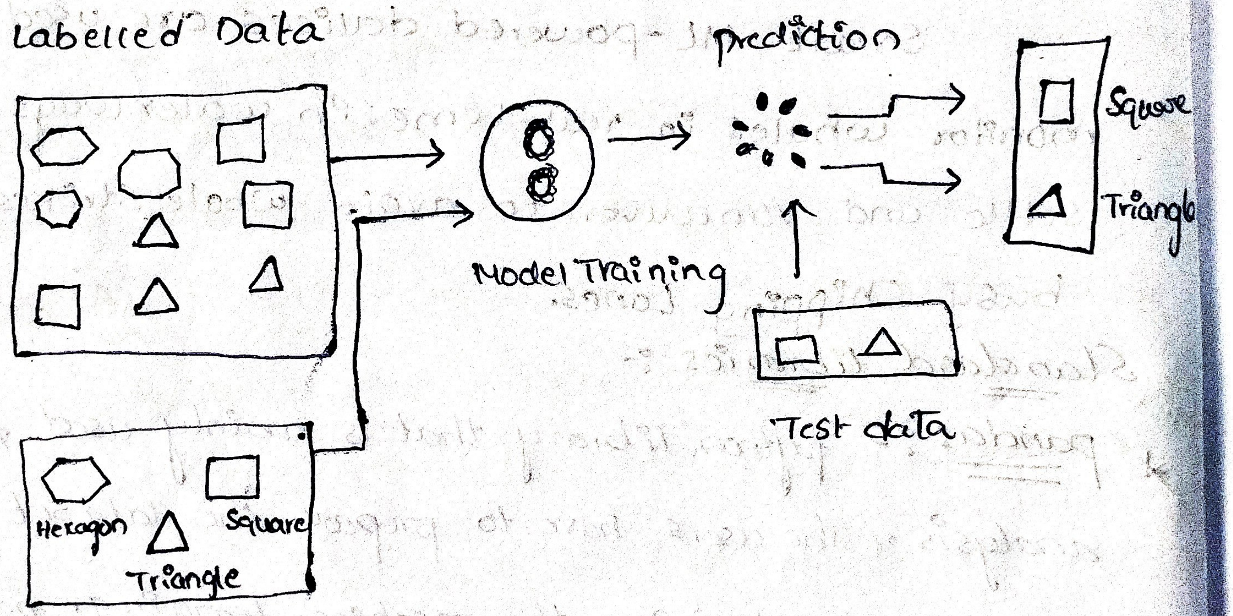
UNIT-II

Supervised Learning

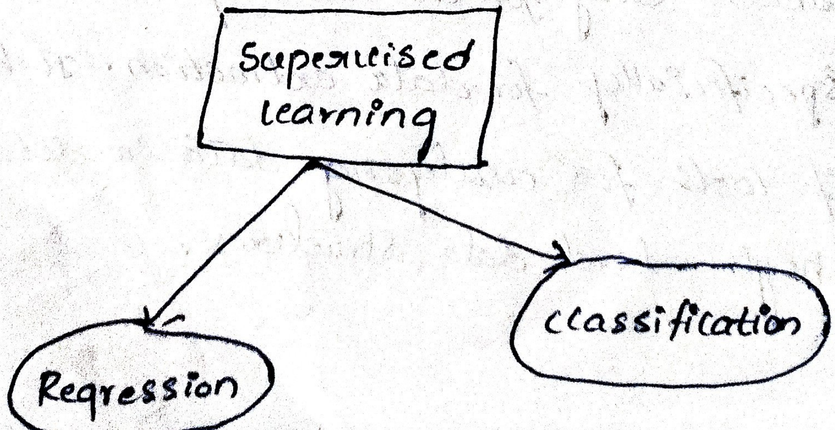
* Supervised learning :- It is a process of providing input as well as correct output data to the machine learning model. The aim of supervised learning algorithm is to find a mapping function to map the input variable (x) with the output variable (y).

In the real-world, supervised learning can be used for Risk Assesment, Image classification, Fraud Detection, Spam Filtering etc.

Working :-



Types of supervised learning :-



1. Regression :-

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as weather forecasting, Market Trends etc. Some popular regression algorithms which come under Supervised learning.

- Linear Regression
- Regression Trees
- Non-linear Regression
- Bayesian linear Regression
- polynomial Regression.

linear :- It establishes the linear relationship between two variables on a line of best fit.

Regression :- Decision trees where the target variable can take continuous values. Classification and Regression Tree (CART) is general term.

Non-linear regression :- It is a type of polynomial regression. It is the method to model a non-linear relationship between the dependent and independent variables.

Bayesian linear Regression :- It allows a fairly natural mechanism to survive insufficient data, or poor distributed data.

Polynomial regression :- It is a form of linear regression in which the relationship between the independent variable x and dependent variable y is modeled as an nth degree polynomial .

2. Classification :- Algorithms are used when the output variable is categorical which means there are two classes such as Yes-No, Male-female, True - False etc .

- Random Forest
- Decision Trees
- Logistic Regression
- Support vector Machines .

Random Forests :- It is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees .

Decision Tree :- It is a decision suppor tool that uses a tree-like model of decision and their possible consequences includes outcomes, costs & utility .

Logistic regression :- It is used for predicting the categorical dependent variable using a given set of independent variables .

Support vector Machine :- It is used for classification as well as Regression problems. And also it is mainly used for Machine learning.

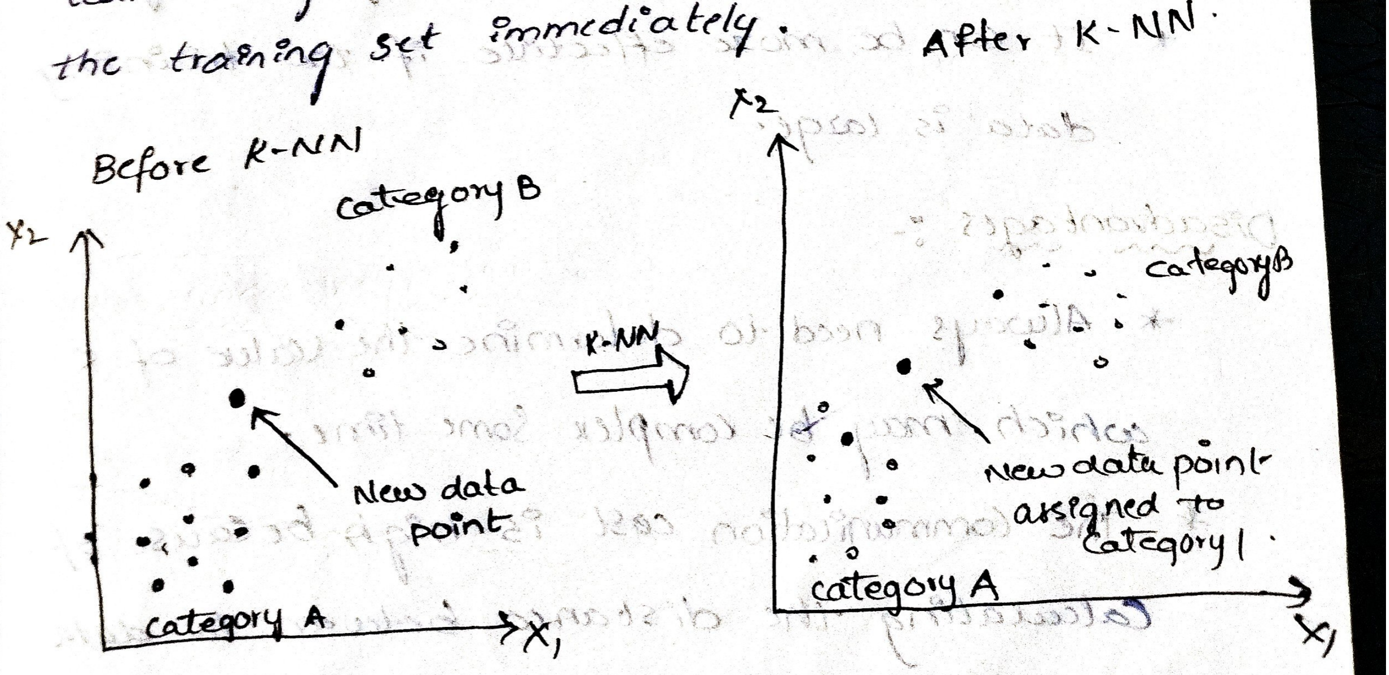
* Supervised Learning algorithms :-

~~i)~~ K- Nearest Neighbour algorithms :-

3a) K-Nearest Neighbour is one of the simplest Machine learning algorithms based on Supervised Learning Technique.

K-NN algorithm stores all the available data and classifies a new data point based on the similarity. K-NN algorithm can be used for Regression as well as for classification but mostly it is used for classification problems.

K-NN is a non-paramctric algoaithm which makes it does not make any assumption on underlying data. It is also called Lazy learner algorithm because it does not learn from the training set immediately.



Algorithm :-

The K-NN algorithm can be explained on the bases of the beloco algorithm.

Step 1 : Select the number of K of the

Step 2 : Calculate the Eucledian distance of K number of neighbors.

Step 3 : Take the K nearest neighbours as per the calculated Eucledian distance.

Step 4 : Among these K neighbours, count the number of the data points in each category.

Step 5 : Assign the new data points to that category for which the number of the neighbour is maximum.

Step 6 : our model is ready.

Advantages :-

* It is simple to implement (or) understand.
* It is the robust to the noisy training data.
* It can be more effective if the training data is large.

Disadvantages :-

* Always need to determine the value of K which may be complex some time.
* The Communication cost is high because of calculating the distance between the data points for all training samples.
* High memory required.
* Not preferred for Internet application.
* Lazy learner (slow in training and predictions).

Regression Models

Example :-

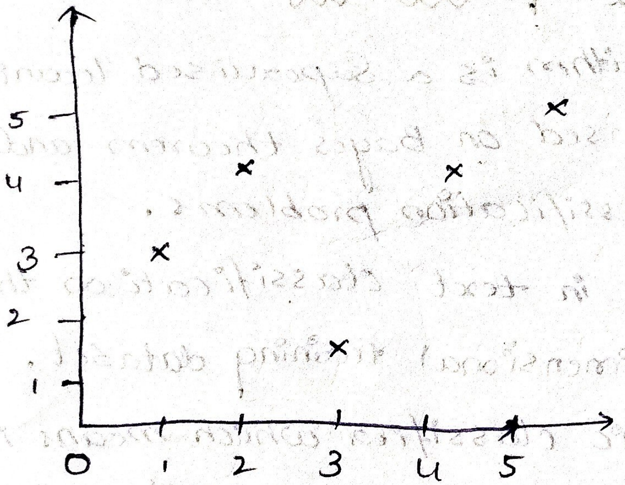
Step 1 :-

$y-\bar{y}$	x	y	$x-\bar{x}$
-0.6	1	3	-2
0.4	2	4	-1
-1.6	3	2	0
0.4	4	4	1
1.4	5	5	2

$y = f(x)$

$y = mx + c.$

$(x-\bar{x})\ (y-\bar{y})$	$(x-\bar{x})^2$
1.2	4
-0.4	1
0	0
0.4	1
2.8	4
4.0	10



Step 2 :- $m = \dfrac{(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$ $\Rightarrow m = \dfrac{4}{10} = 0.4$

Step 3 : $y = mx + c.$

$3.6 = (0.4)(3) + c$

$c = 3.6 - 1.2$

$\boxed{c = 2.4}$

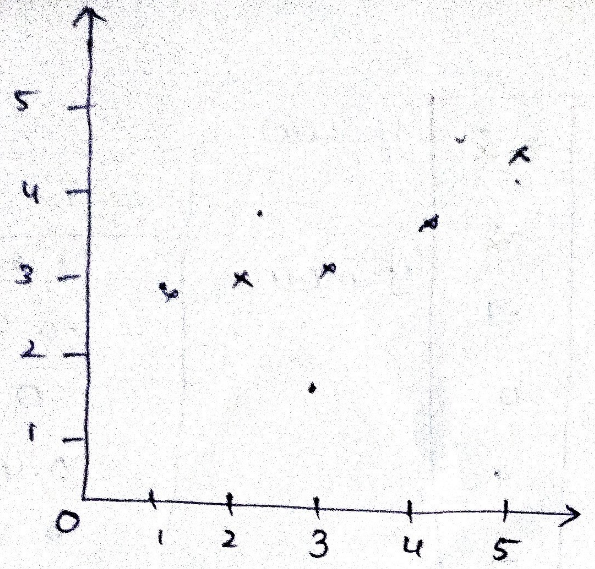
Step 4 : $p(1) = (0.4)1 + 2.4 = 2.8$

$p(2) = (0.4)2 + 2.4 = 3.2$

$p(3) = (0.4)3 + 2.4 = 3.6$

$p(4) = (0.4)4 + 2.4 = 4.0$

$p(5) = (0.4)5 + 2.4 = 4.4$



* Naives Bayes Algorithm / classifiers :-

- * Naives Bayes algorithm is a Supervised learning algorithm which is based on Bayes theorem and used for solving classification problems.
- * It is mainly used in text classification that includes a high-dimensional training dataset.
- * It is a probabilistic classifier which means it predicts on the basis of the probability of an object.
- * Examples of Naives Bayes algorithm are spam filtration. Sentimental analysis and classifying articles.

Naive :- It assumes that the occurrence of a Certain feature is independent of the occurence of others features.

Bayes :- It depends on the principle of Bayes theorem.

Bayes theorem :- Also known as Baye's Rule (or) Baye's Law, which is used to determine the probability of hypothesis with prior knowledge.

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

Here $P(A|B)$ = posterior probability

$P(B/A)$ = likelihood probability

$P(A)$ = prior probability.

$P(B)$ = Marginal probability

Working :-

1. convert the given data set into frequency tables
2. Generate likelihood table by finding the probabilities of given features.
3. use Bayes theorem to calculate the posterior probability.

Day	outlook	Temp	Humidity	wind	play
1	Sunny	Hot	High	weak	NO
2	Sunny	Hot	High	Strong	NO
3	overcast	Hot	high	weak	yes
4	Rainy	mild	high	weak	Yes
5	Rainy	cool	Normal	weak	yes
6	Rainy	cool	Normal	Strong	NO
7	overcast	cool	Normal	Strong	yes
8	Sunny	mild	high	weak	NO
9	Sunny	cool	normal	Weak	yes
10	Rainy	mild	normal	weak	yes
11	Sunny	mild	normal	Strong	Yes
12	overcast	mild	high	strong	Yes

13 overcast hot normal weak yes

14 rainy mild high strong No.

Q) find the probability to play cricket on 15th day where conditions are outlook =sunny, temp=cool, humidity=high, wind=strong.
p(play) =yes=9/14
p(play) =no=5/14

p(play) = NO = 5/14.

outlook	yes	NO
sunny	2/9	3/5
overcast	4/9	0
rain	3/9	2/5

temp	yes	NO
hot	2/9	2/5
cool	3/9	1/5
mild	4/9	2/5

humidity	yes	NO
high	3/9	4/5
normal	6/9	1/5

wind	yes	NO
weak	6/9	2/5
Strong	3/9	2/5

let x = { sunny, cool, high, strong }

p(x|yes) = p(yes) * p(sunny|yes) * p(cool|yes) * p(high|yes) * p(strong|yes).

$= \frac{9}{14} * \frac{2}{9} * \frac{3}{9} * \frac{3}{9} * \frac{3}{9}$

$= \frac{1}{189} \approx 0.005$

$p(x/No) = p(No) * p(Sunny/No) * p(cool/No) * p(high/No) * p(strong/No)$.

$= \frac{5}{14} * \frac{3}{5} * \frac{1}{5} * \frac{4}{5} * \frac{3}{5}$

$= \frac{36}{125 \times 14} = 0.020$

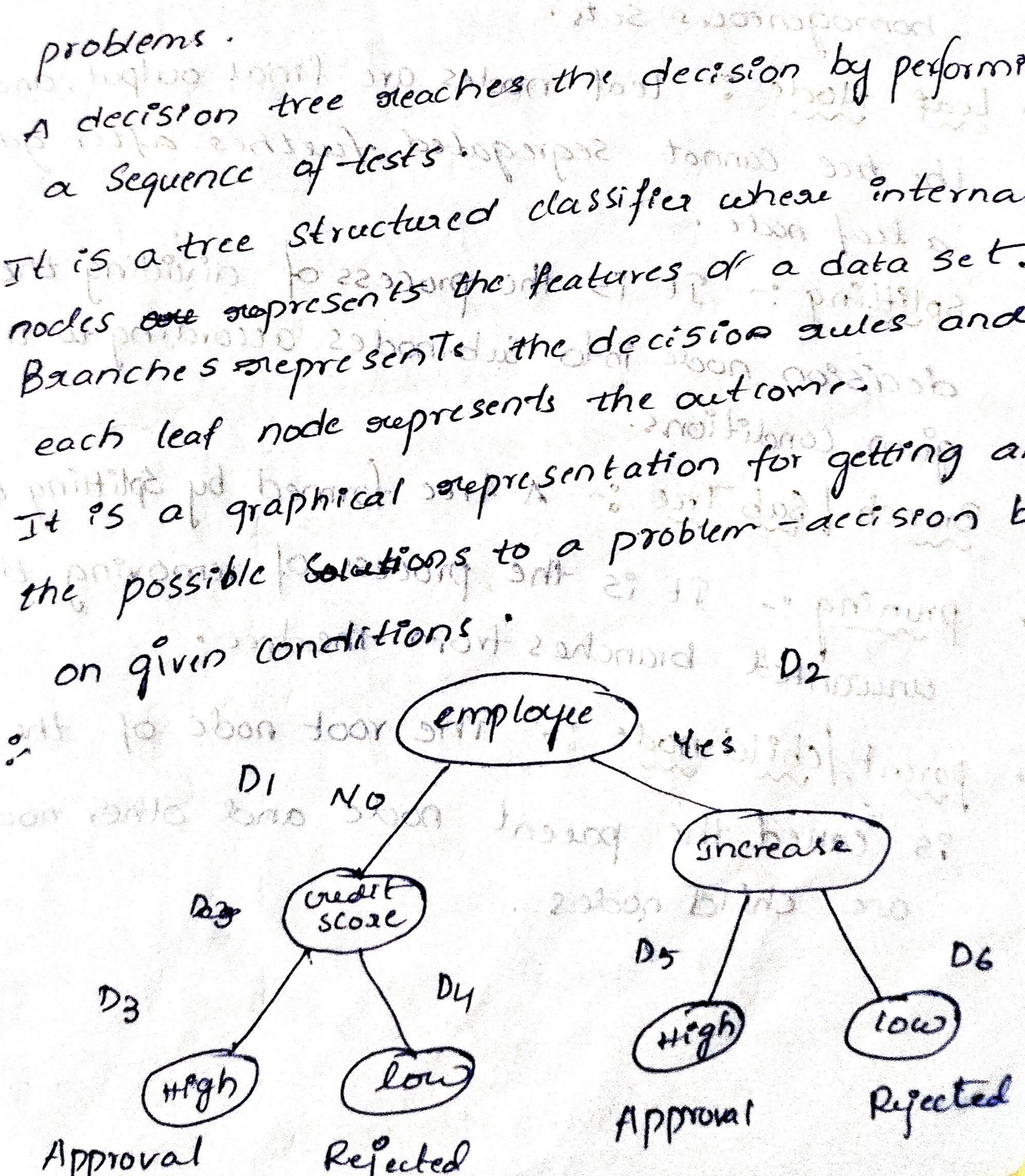
Since $0.005 < 0.020$

∴ Probability to play Cricket on 15th day is No.

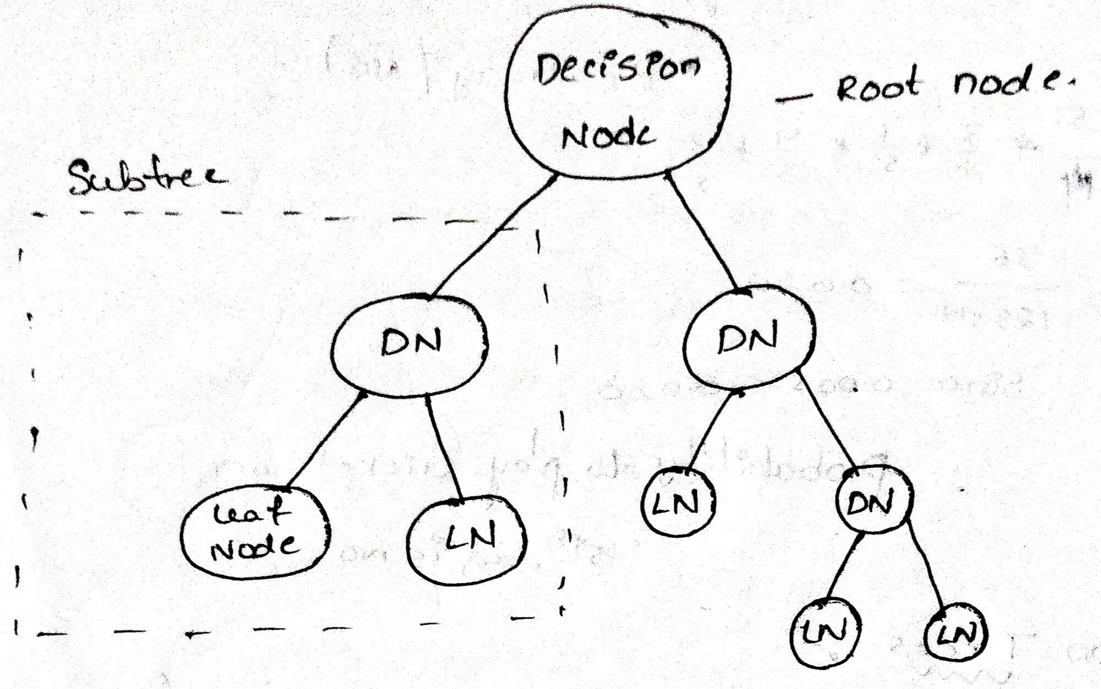
Decision Trees :-

* This algorithm falls under the category of Supervised learning.
* It can be solved both regression and classification problems.
* A decision tree reaches the decision by performing a Sequence of tests.
* It is a tree structured classifier where internal nodes ~~are~~ represents the features of a data set.
* Branches represents the decision rules and each leaf node represents the outcome.
* It is a graphical representation for getting all the possible Solutions to a problem - decision based on given conditions.

Ex:-



Decision Trees :-



Terminologies :-

- Root Node :- It represents the entire data set which further get divided into two or more homogeneous sets.
- Leaf Node :- Leaf nodes are final output, and the tree cannot segregated further after getting a leaf node.
- Splitting :- It is the process of dividing the decision node into sub-nodes according to the given conditions.
- Branch | Sub Tree :- A tree formed by splitting the tree.
- pruning :- It is the process of removing the unwanted branches from the tree.
- parent / child node :- The root node of the tree is called the parent node and other nodes are child nodes.

Algorithm :-

Step 1 : Begin the tree with the root node says s. which contains the complete data set.

Step 2 : Find the best attribute in the dataset using Attribute Selection Measure (ASM).

Step 3 : Divide the S into Subsets that contains possible values for the best attributes'

Step 4 : Generate the decision tree node which contains the best attribute.

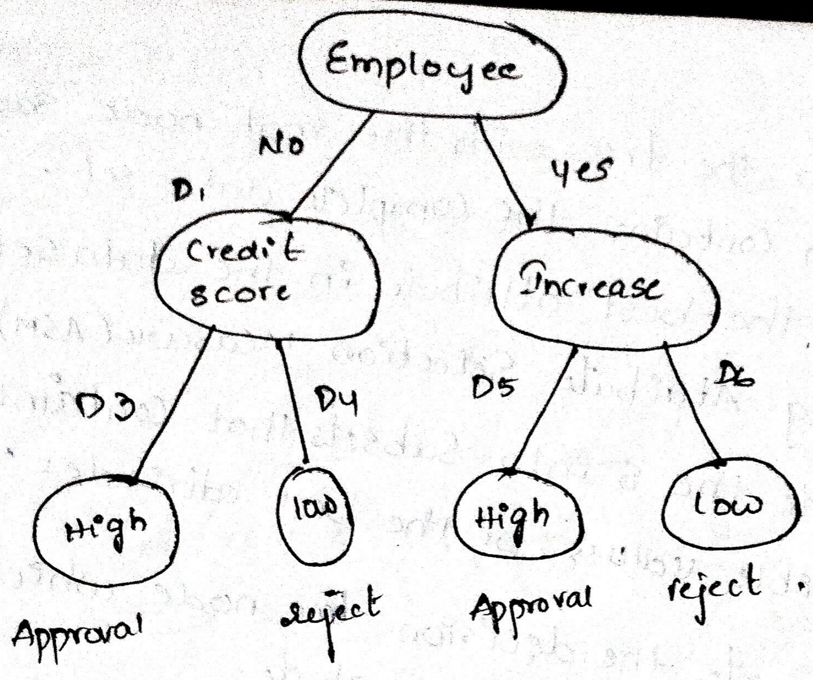
Step 5 : Recursively make new decision trees using the Subsets of the data set created in the step-3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

Advantages :-

- It can be very useful for solving decision-related problems.
- It helps to think about all possible outcomes for a problem.
- There is less requirement of data cleaning compared to other algorithms.

Disadvantages :-

- The decision tree contains lots of layers which makes it complex.
- It may have an overfitting issue, which can be resolved using Random Forest algorithm.
- For more class labels, the computational Complexity of decision tree may increase.



* Attribute : outlook.

Values (outlook) = Sunny, Overcast, Rainy

$$\text{Entropy}(S) = \frac{-P}{P+N}\log\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$$

$$S = [9+\ 5-]$$

$$= -\frac{9}{14}\log_2\left(\frac{9}{9+5}\right) - \frac{5}{9+5}\log_2\left(\frac{5}{14}\right)$$

$$= \frac{-9}{14}\left(\log\frac{9}{14} \div \log 2\right) - \frac{5}{14}\left(\log\frac{5}{14} \div \log 2\right)$$

$$= 0.40917 + 0.5305$$

$$[\log_2 n = \log n \div \log 2]$$

$$= 0.94$$

$$\text{Entropy}(\text{Sunny}) = \frac{-P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$$

$$[2+3-]$$

$$= \frac{-2}{5}\log_2\left(\frac{2}{5}\right) - \left(\frac{3}{5}\right)\log_2\left(\frac{3}{5}\right)$$

$$= -\frac{2}{5}\left(\log\frac{2}{5} \div \log 2\right) - \frac{3}{5}\left(\log\frac{3}{5} \div \log 2\right)$$

$$= 0.528 + 0.443$$

$$= 0.971$$

Entropy $(S_{overcast})$ =

$[4+, 0-]$

$= 0+0 = 0;$

Entropy $(S_{rainy}) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)$

$[3+, 2-]$

$= -\frac{3}{5}\left(\log\frac{3}{5} \div \log 3\right) - \frac{2}{5}$

$\left(\log\frac{2}{5} \div \log 2\right)$

$= 0.442 + 0.528$

$= 0.971.$

No we need to calculate the gain

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

(VE, sunny, overcast, rainy)

$$\text{Gain}(S, \text{outlook}) = \text{Entropy}(S) - \frac{5}{14}\,\text{Entropy}(S_{sunny}) -$$

$$- \frac{4}{14}\,\text{Entropy}(S_{overcast}) - \frac{5}{14}\,\text{entropy}(S_{rainy})$$

$$\text{Gain}(S_{outlook}) = 0.94 - \frac{5}{14}(0.971) - 0 - \frac{5}{14}(0.971)$$

$= 0.2464.$

* Attribute : Temperature :-

values (Temperature) = Hot, Mild, cool

$$\text{entropy}(S) = -\frac{P}{P+N}\log\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$$

$S = [9, -5]$

$$= -\frac{9}{14}\log_2\left(\frac{9}{9+5}\right) - \frac{5}{9+5}\log_2\left(\frac{5}{14}\right)$$

$= 0.94$

$$\text{Entropy (Hot)} = \frac{-2}{2+2} \log_2 \left(\frac{2}{2+2}\right) - \frac{2}{2+2} \log \frac{2}{2+2}$$

$$= \frac{-2}{4} \left(\log 2/4 \div \log 2 \right) - \frac{2}{4} \left(\log 2/4 \div \log 2 \right)$$

$$= 1.0.$$

$[4+, 2-]$

$$\text{Entropy (Smild)} = -\frac{4}{6} \log_2 \left(\frac{4}{6}\right) - \frac{2}{6} \log \left(\frac{2}{6}\right)$$

$$= -\frac{4}{6} \left(\log \frac{4}{6} \div \log 2 \right) - \frac{2}{6} \left(\log \frac{2}{6} \div \log 2 \right)$$

$$= \frac{-2}{3} \left(\log \left(\frac{2}{3}\right) \div \log 2 \right) - \frac{1}{3} \left(\log \left(\frac{1}{3}\right) \div \log 2 \right)$$

$$= 0.389975 + 0.52832$$

$$= 0.9182$$

$$\text{Entropy (Scold)} = [3+, 1-]$$

$$= 3/4 \log_2 \left(\frac{3}{4}\right) - \frac{1}{4} \left(\log \left(\frac{1}{4}\right) \div \log 2 \right)$$

$$= 0.31127812 + 0.5000$$

$$= 0.8113.$$

$$G (S, \text{temp}) = \text{entropy} (S) - \frac{4}{14} \text{entropy (Shot)} - \frac{4}{14} \text{entropy (Scool)} - \frac{6}{14} \text{entropy (Smild)}$$

$$= 0.94 - \frac{4}{14} (1.0) - \frac{4}{14} (0.8113) - \frac{6}{14} (0.9182)$$

$$= 0.94 - 0.2857 - 0.2318 - 0.3935$$

$$= 0.02897$$

* Attribute : Humidity :

$\text{entropy}(S) = \frac{-P}{P+N} \log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N} \log_2\left(\frac{N}{P+N}\right)$

$= 0.94$

$\text{entropy}(S_{high}) = [3+, 4-]$

$= -\frac{3}{7}\left(\log\left(\frac{3}{7}\right) \div \log 2\right) - \frac{4}{7}\left(\log\left(\frac{4}{7}\right) \div \log 2\right)$

$= 0.9852$

$\text{entropy}(S_{Noamal}) = [\underset{P}{\overset{\downarrow}{6+}}, \underset{N}{\overset{\downarrow}{1-}}]$

$= -\frac{6}{7}\left(\log(6/7) \div \log 2\right) - \frac{1}{7}\left(\log(1/7) \div \log 2\right)$

$= 0.5916$

$\text{Gain}(S, \text{Humidity}) = 0.94 - \frac{7}{14}(0.9852) - \frac{7}{14}(0.502)$

$= 0.94 - 0.4926 - 0.2958$

$= 0.1516$

* Attribute : Wind.

$\text{entropy}(S) = 0.94$

$\text{entropy}(S_{strong}) = [3+, 3-]$

$= -\frac{3}{6}\left(\log\left(\frac{3}{6}\right) \div \log 2\right) - \frac{3}{6}\left(\log\left(\frac{3}{6}\right) \div \log 2\right)$

$= -\frac{1}{2}\left(\log\left(\frac{1}{2}\right) \div \log 2\right) - \frac{1}{2}\left(\log\left(\frac{1}{2}\right) \div \log 2\right)$

$= 0.5283 + 0.5283$

$= 1.056$

$\text{entropy}(S_{weak}) = [6+, 2-]$

$= -\frac{6}{8}\left(\log\left(\frac{6}{8}\right) \div \log 2\right) - \frac{2}{8}\left(\log\left(\frac{2}{8}\right) \div \log 2\right)$

$- 3\left(\log\left(\frac{3}{}\right) \div \log 2\right) - \frac{1}{4}\left(\log\left(\frac{1}{4}\right) \div \log 2\right)$

$= 0.311278 + 0.500$

$= 0.81127.$

$$\text{Gain}(S, \text{wind}) = 0.94 - \frac{6}{14}\,\text{entropy}(S_{strong}) - \frac{8}{14}\,\text{entropy}(S_{weak})$$

$$= 0.94 - \frac{6}{14}(1.056) - \frac{8}{14}(0.81127)$$

$$= 0.94 - 0.42857 - 0.4636$$

$$= 0.04783$$

Sunny:

Day	outlook	temp	Humidity	Wind	play
1	Sunny	Hot	High	Weak	NO
2	Sunny	Hot	High	strong	NO
3	Sunny	mild.	High	weak	NO
8	Sunny	mild	High	weak	No
9	Sunny	cold	Normal	weak.	Yes
11	Sunny	mild	Normal	Strong	Yes

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

$$\text{entropy}(S) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)$$

$$= -\frac{2}{5}\left(\log\left(\frac{2}{5}\right) \div \log 2\right) - \frac{3}{5}\left(\log\left(\frac{3}{5}\right) \div \log 2\right)$$

$= 0.5 + 0.44$

$= 0.97$

temp = {hot, cold, mild}

entropy (Hot) = [0+, 2-] = 0.0

entropy (Mild) = [1+, 1-] = 1.0

entropy (cool) = [1+, 0-] = 0

Gain (s, temp) = 0.97 - 0 - $\frac{2}{5}$ (1)

= 0.57.

Humidity = {High, Normal}

entropy (High) = 0 [0+, 3-]

entropy (Normal) = 0.0 [2+, 0-]

Gain (s, Humidity) = 0.97 - 0 - 0 = 0.97

wind = {strong, weak}

entropy (strong) = 1 [1+, 1-]

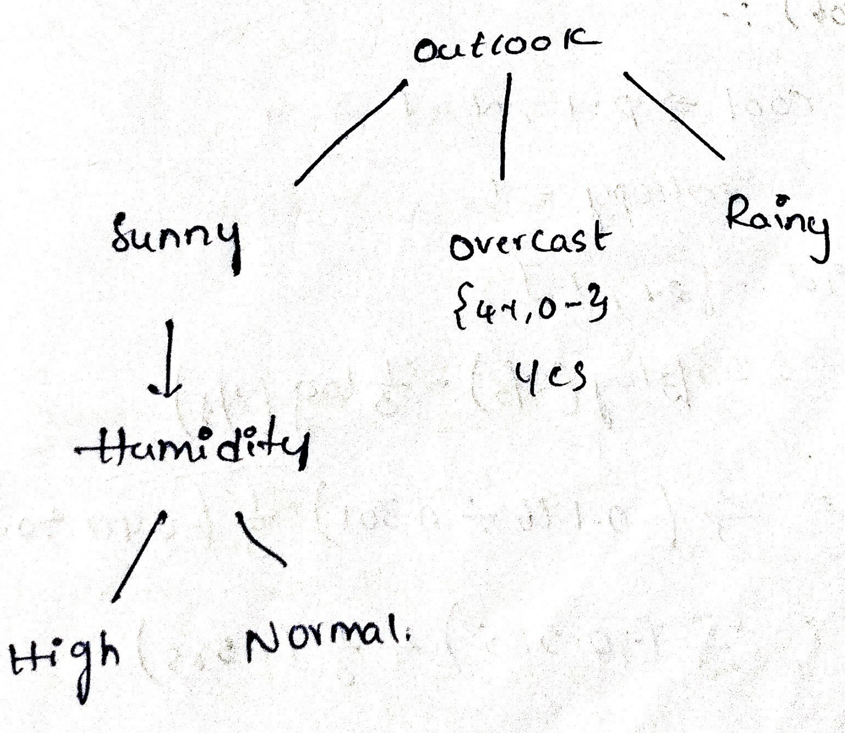
entropy (weak) = $-\frac{1}{3}(\log(\frac{1}{3}) \div \log 2)$

= 0.9182

Gain (s, wind) = entropy (S) - $\frac{2}{5}$ entropy (Sstrong) - $\frac{3}{5}$ entropy (Sweak)

= 0.97 - $\frac{2}{5}$(1) - $\frac{3}{5}$ (0.9182)

= 0.0192



Rainy:-

Day	Mild	high	Weak	play
5	cool	normal	weak	yes
6	cool	normal	strong	NO
10	Mild	normal	weak	yes
14	mild	high	strong	NO

$$\text{entropy}(S) = \frac{-P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$$

$$= \frac{-3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right)$$

$$= -\frac{3}{5}\left(\log\left(\frac{3}{5}\right)\div\log 2\right) - \frac{2}{5}\left(\log\left(\frac{2}{5}\right)\div\log 2\right)$$

$$= -\frac{3}{5}(0.737) - \frac{2}{5}(-1.322)$$

$$= 0.442 + 0.528$$

$$= 0.9708$$

Temp (mild, cool) :-

cool ⇒ $P=1$, $N=1$

entropy = 1

mild = $[2+, 1-]$

$$= -2/3\log(2/3) - \frac{1}{3}\log(1/3)$$

$$= -\frac{2}{3}(-0.176 \div 0.301) - \frac{1}{3}(-0.477 \div 0.301)$$

$$= -\frac{2}{3}(-0.585) \div \frac{1}{3}(1.585)$$

$$= 0.918$$

$\text{Gain}(S\ \text{Temp}) = 0.9708 - \frac{2}{5}(1.0) - \frac{3}{5}(0.918).$

$= 0.9708 - 0.4 - 0.55$

$= 0.02$

Humidity:-

high => $p=1, N=1$

entropy $= [1-, +1]$

$= 1.0$

normal = $p=2, N=1$

$= -\frac{2}{3}\log\left(\frac{2}{3}\right) - \frac{1}{3}\log(1/3)$

$= 0.918$

$\text{Gain}(S, \text{humidity}) = 0.9708 - \frac{2}{5}(1.0) - \frac{2}{5}(0.918)$

$= 0.9708 - 0.9508$

$= 0.62$

Wind :

(weak, Strong)

Strong = $[0, 2-]$

entropy $= 0.$

Weak :-

$[3+, 0-]$

Entropy $= 0$

$\text{Gain}(S\text{wind}) = 0.9708 - 0$

$= 0.9708$

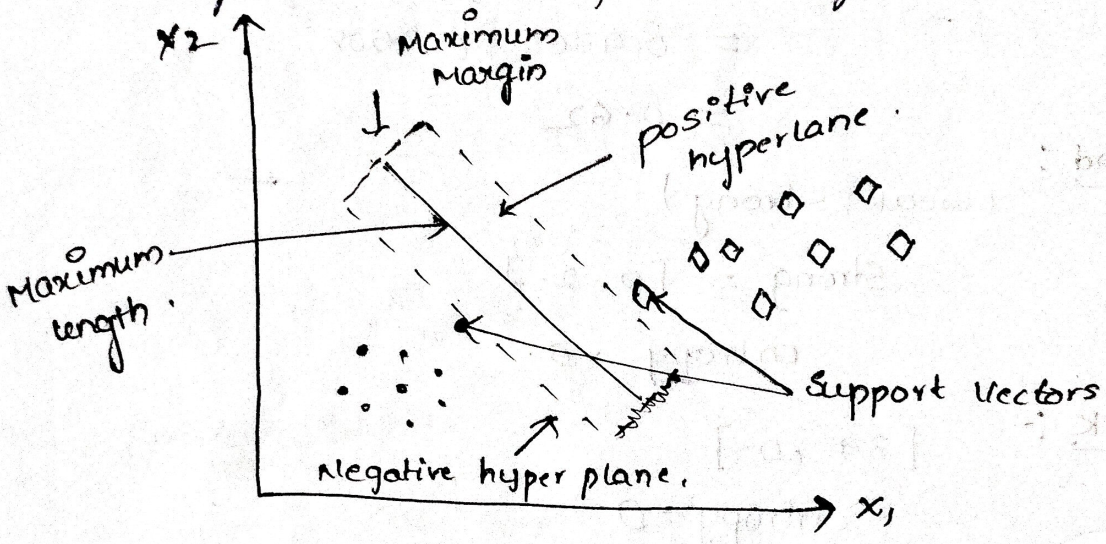
* Support Vector Machine (SVM) :-

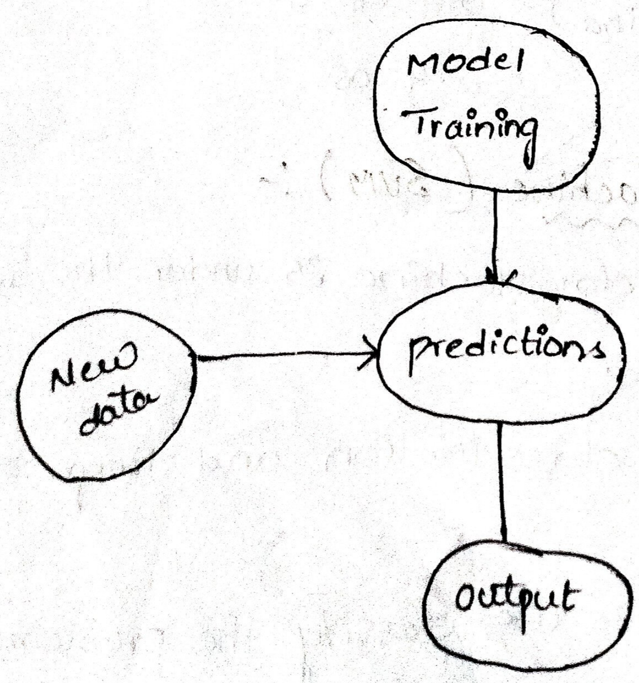
* Support vector Machine is under the Supervised leaning algorithm.

* It is used for Classification and Regression problems.

* Most probably we use classify the problems using SVM.

- * Maximum we used SVM in binary classifications
- * Hyperplane Separates our given data set into classes.
- * The goal of the SVM algorithm is to create the best line or decision boundary that can n-dimensional space into classes. This best decision boundary is called the hyperplane.
- * The data points which are nearer to the hyperplane are called "Support vector".
- * The margin is the distance between 2 dotted lines that is two support vectors and hyperplane place exactly at the center of the margin.





* SVM can be of 2 types - i. linear SVM
ii. Non linear SVM

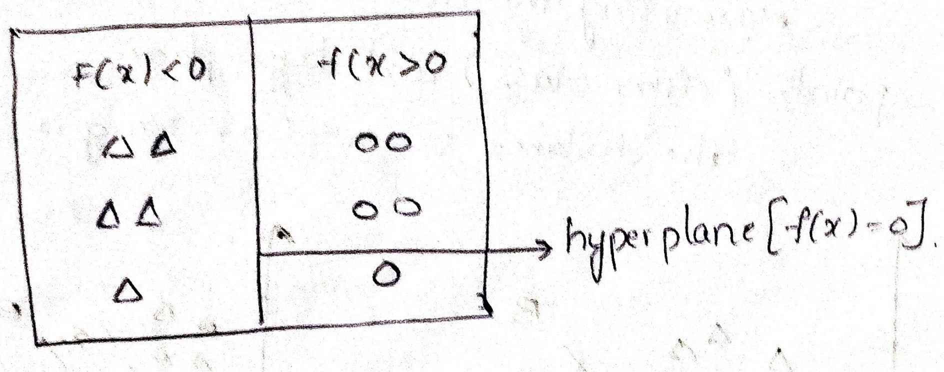
i. linear SVM :-

It is used for linearly Separable data, which means if a dataset can be classified into two classes by using Single Straight line. Termed as linear Separable data.

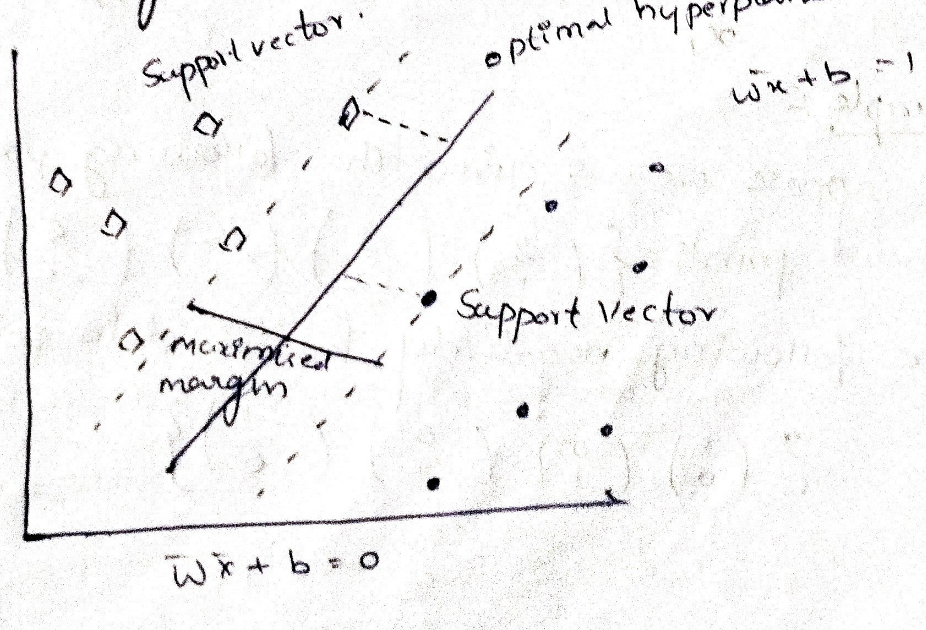
ii. Non-linear SVM :-

It is used for non-linearly Separated data which means if a dataset cannot be classified by a using a straight line. Termed as non-linear data. classifier is used as Non-linear SVM.

- linear



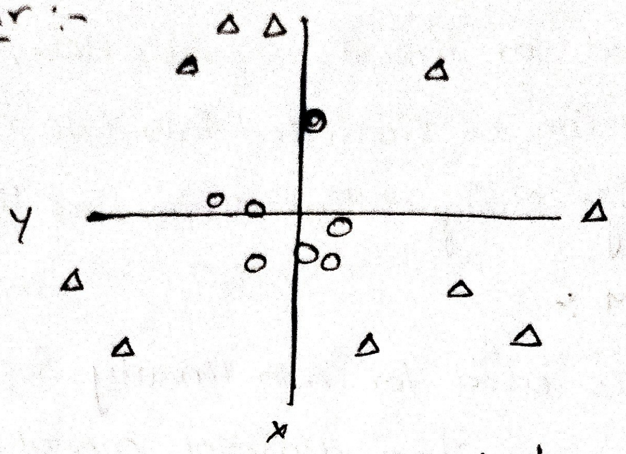
1. $f(x) = w \cdot x + b$.
2. w is the normal line, 'x' is input vector and 'b' is the bias.
3. w is the weight vector.



$(\bar{\omega}x+b) \geq +1 \ \forall$ x of class 1

$(\bar{\omega}x+b) < -1 \ \forall$ x of class 2 .

Non-linear :-

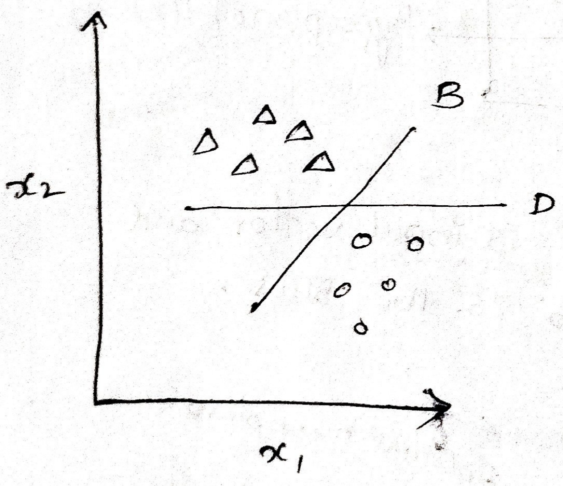


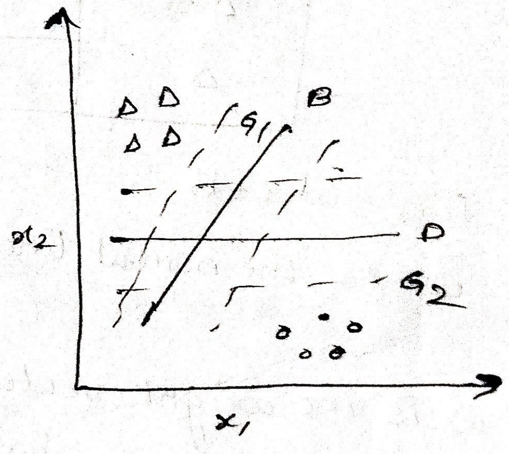
* Rules to identify the right hyperplane :-

Select the hyper-plane which Segragates the two classes better !

Maximising the distance between nearest data points (either class) and hyperplane.

The distance is called as Margin .

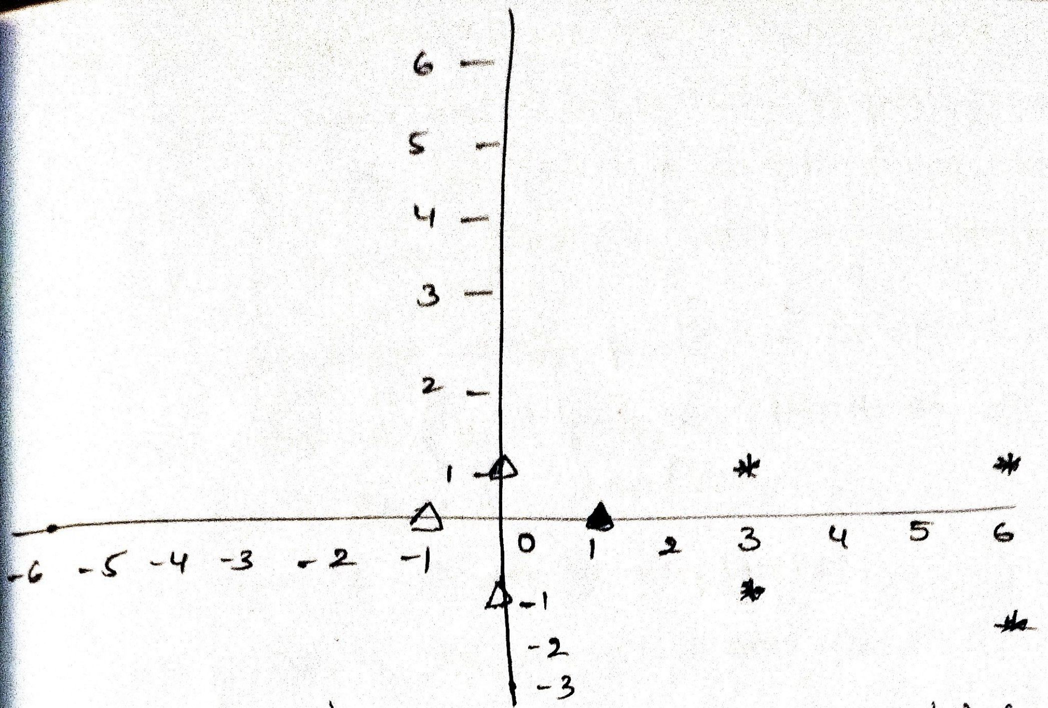




Example :-

Suppose we are given the following positively labeled points $\left\{ \begin{pmatrix} 3 \\ 1 \end{pmatrix} \begin{pmatrix} 3 \\ -1 \end{pmatrix} \begin{pmatrix} 6 \\ 1 \end{pmatrix} \begin{pmatrix} 6 \\ -1 \end{pmatrix} \right\}$ & the following negatively based data points.

$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \end{pmatrix} \right\}$



Each vector argumented with a 1 as a bias input.

By inspection, it should be obuious that there are 3-support vectors $\{ s_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, s_2 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, s_3 = \begin{pmatrix} 3 \\ -1 \end{pmatrix} \}$.

Each vector is argumented with a 1 as a bias input i.e., $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ $\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}$ $\begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}$

so, $s_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ then $\tilde{s}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

$s_2 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ then $\tilde{s}_2 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}$

$s_3 = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$ then $\tilde{s}_3 = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}$

* Ensembles of Decision Tree :-

Ensemble methods which combines several decision trees to produce better predictive performance than utilizing a single decision tree.

The main principle behind the ensemble model is that a group of weak learners come together to foam a strong learner.

Two techniques of ensemble of decision tree.

i, Bagging

ii, Boosting.

Bagging :- Bootstrap Aggregation is used when our goal is to reduce the variance of a decision tree. Average of all the predictions from different trees, are used which is more robust than a single decision tree.

Random Forest is extension over bagging. when you have many random trees. It's called Random Forest.

steps for Random Forest :-

1. suppose there are N observations and M features in training data set. First a sample from training data set is taken randomly with replacement

2. A subset of M features are selected randomly and whichever feature gives the best split is used to split the node iteratively.

3. The tree is grown to the largest.

4. Above steps are repeated and prediction is given based on the aggregation of predictions from n no.of trees.

Advantages :-

- Handles higher dimensionality data very well.
- Handles missing values and maintain accuracy for missing data.

Disadvantages :

- Since final prediction is based on the mean predictions from subset trees, it won't give precise values for the regression model.

Boosting :-

* Is another ensemble technique to create a collection of predictors. By combining the whole set at the end converts weak learners into better programming model.

Gradient Boosting is an extension over boosting method.

~~Gea~~ Gradient Boosting : Gradient Descent + Boosting

Advantages :-

- Supports different loss function
- works well with interactions.

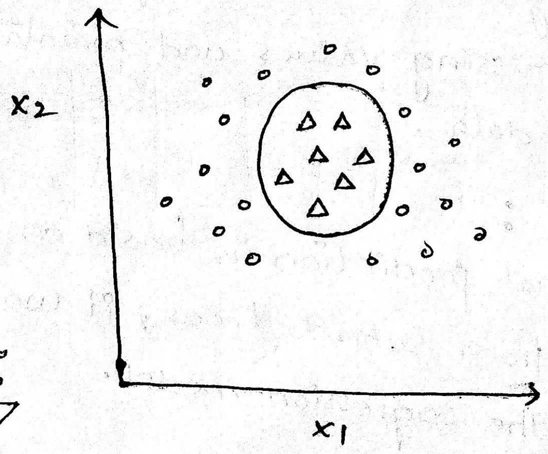
Disadvantages :

- prone to over-fitting
- Requires Careful tuning of different hyper-parameters.

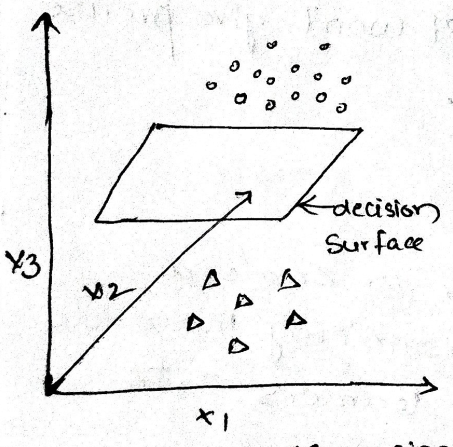
3(b)

* Kernlized Support Vector Machine :-

Kernel SVM :- Kernal function generally transforms the training set of data so that non-linear discussion Surface can be transformed to a linear equation in a higher no-of dimension Spaces.



SVM in 2-dimensions



SVM in 3-dimensions.

Standard Kernel Function Equation.

$$K(\bar{x}) = 1, \text{ if } ||\bar{x}|| <= 1.$$

* Major Ker

For

install s

- Gau

-ma

dat

- Gaus

basis

K

K

K

- Sigmoid

Th

perceptro

used o

neuron

K(x,

- polyno

in the

over p

used f

- Linear Ke

One of the mos

* Major Kernel Functions /types :

For implementing Kernel function, first we have to install scikit-learn.

- Gaussian Kernel :- It is used to perform transformation when there is no prior knowledge about data.

$$K(x, y) = e^{-}\left(\frac{\|x' - y\|^2}{2\sigma^2}\right).$$

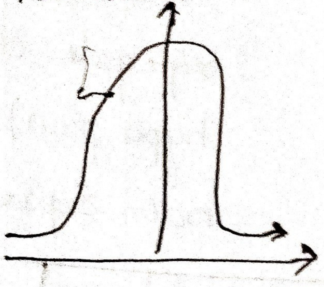
- Gaussian Kernel Radial Basis Function (RBF) :-

Same as above Kernel function, adding radial basis method to improve the transformation.

$K(x, x_1) + K(x, x_2)$ (Simplified - Formula).

$K(x, x_1) + K(x, x_2) > 0$ (Green)

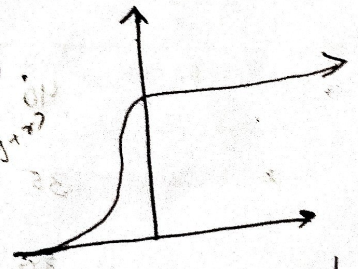
$K(x, x_1) + K(x, x_2) = 0$ (Red).



- Sigmoid Kernel :-

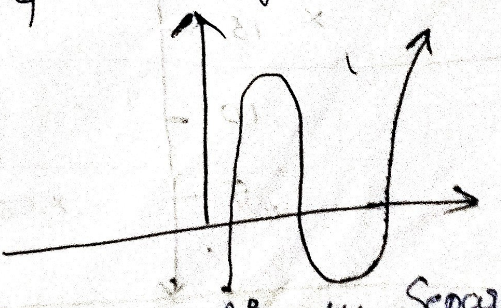
This function is equivalent to a two-layer perceptron model of the neural network, which is used as an activation function for artificial neurons.

$K(x, y) = \tanh(\gamma . x^T y + r)$.



- polynomial Kernel :-

It represents the similarity of vectors in the training set of data in a feature space over polynomials of the original variables used in Kernel.



- Linear Kernel :-

Used when data is linearly Separable: Separated using on single line.

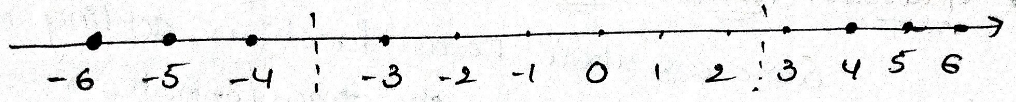
One of the most Common Kernels to be used

* understanding Kernel :-

Let us take.

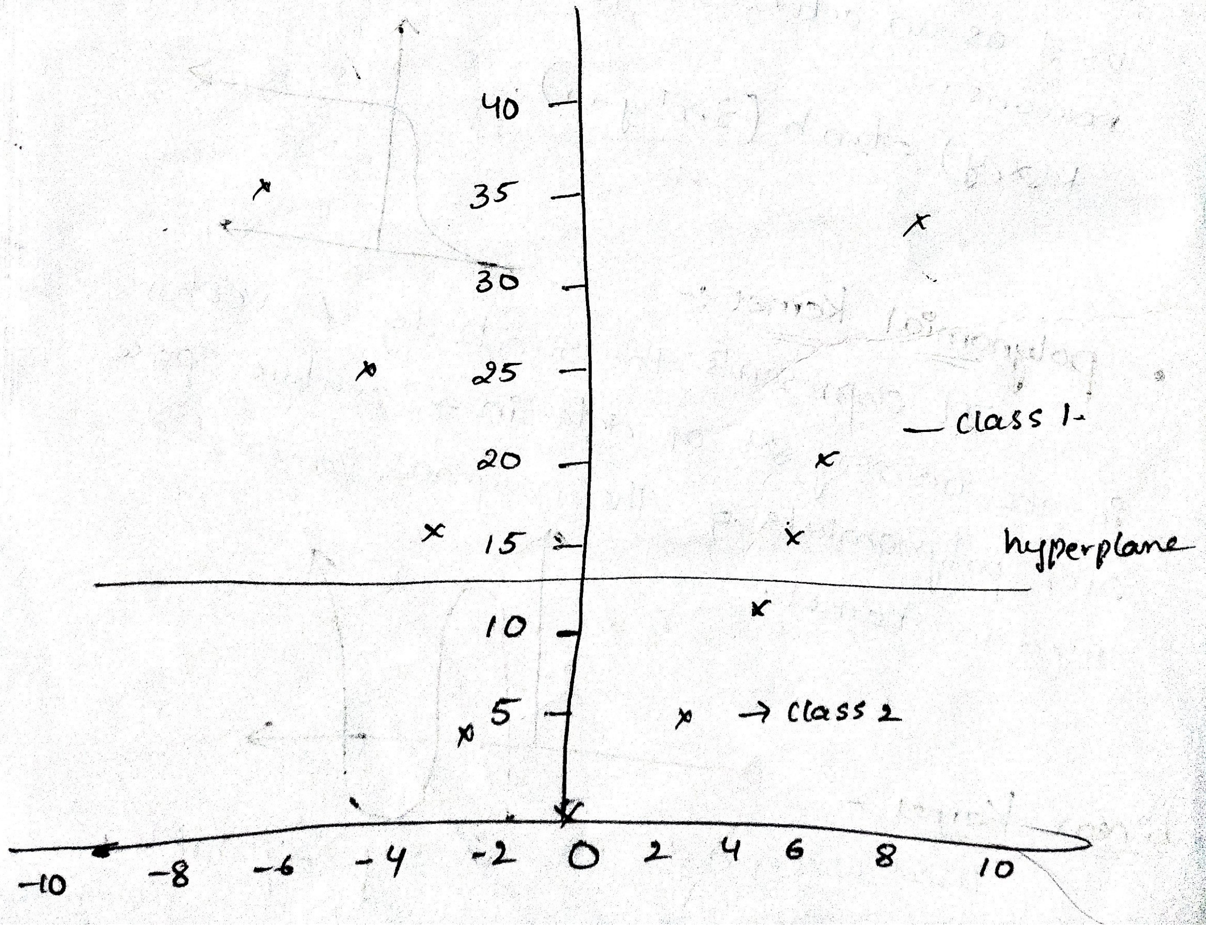
Feature(x)	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6

→ class 1
→ class 2). 3(b)



In the above graph, we cannot Separate feature(x)| Segregate classes by using a Single - line (ie., hyper plan) because sav of data points are not Separate Correctly.

feature(x)	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
x^2	86	25	16	9	4	1	0	1	4	9	16	25	36



* Ensembles

* Bias

Bias

trainig

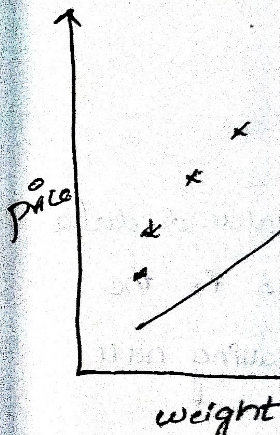
we trai

variance

and o

the da

model



bias = h

variance =

* Ensembles of Decision tree :-

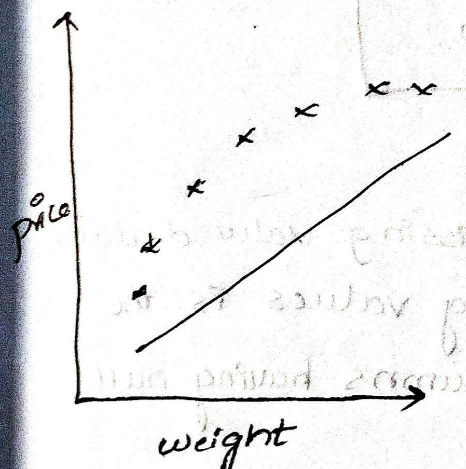
* Bias_ variance, overfitting, underfitting.

Bias — Errors of Training data.

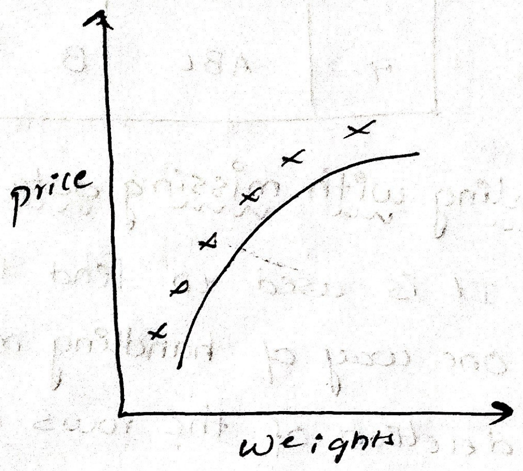
Our model will not be trained well with the training data they will be high training error, when we train your model with the data.

Variance :- Errors of Testing data.

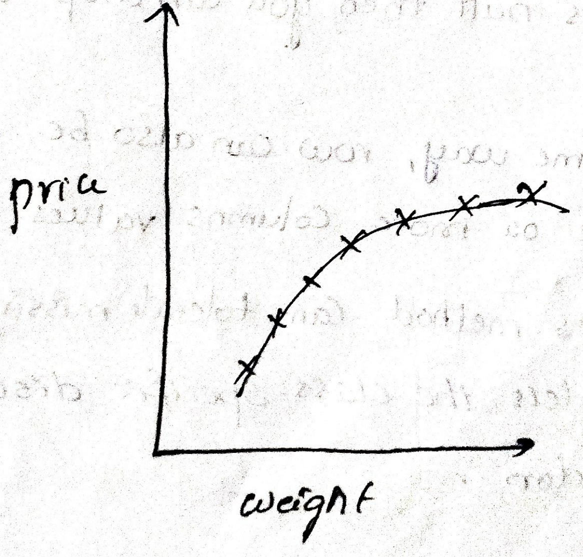
If you train your data on training data and obtain a very low error, upon changing the data and them training the same previous model the error is high. is known as variance.



bias = high
variance = high



bias = low
Variance = low



bias = low
Variance = High