ML_QP2_Format_ESE_50 marks_R-2022

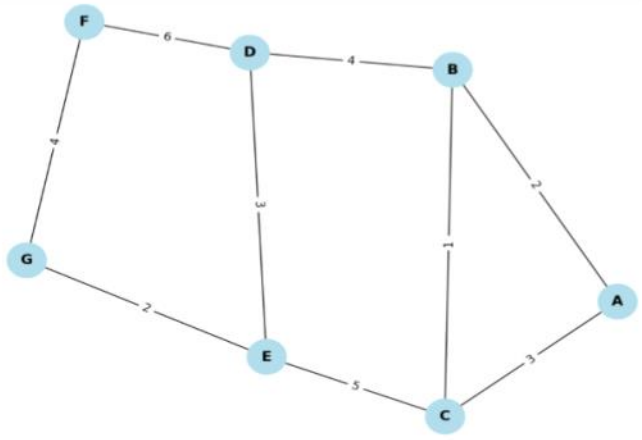
VIT Vidyalankar Institute of Technology Accredited A+ by NAAC (Autonomous College Affiliated to University of Mumbai)	End Semester Examination (R-2022 Scheme) -(2024-25)	
Date:	Branch: CMPN	Time: 2 Hrs.
Semester: VII	Subject: Machine Learning	Marks: 50
N B. :- All Questions are Compulsory		CO

Q. 1)	Attempt any two (5 Marks Each)	
a)	Compare and contrast supervised, unsupervised, and reinforcement learning. Provide an example for each type.	CO1
b)	Elaborate on Karl Pearsons coefficient of Correlation.	CO2
c)	Elaborate on following terms with respect to DBSCAN: (i) Directly Density Reachable (ii) Density Reachable	CO5
Q. 2)	Attempt any two. (5 Marks Each)	
a)	Justify the need of cross validation. Explain k fold cross validation in details.	CO3
b)		CO3

	INDEPENDENT	DEPENDENT
SR.NO	X	Y
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

Tasks:
Fit a linear regression model to the data and determine the equation of the regression line.

c)	How does the AdaBoost algorithm iteratively adjust weights and combine weak classifiers to form a strong classifier, and what are the main steps involved in this process?	CO4
Q 3)	Attempt any two. (5 Marks Each)	
a)	Justify the need of pruning in XGBOOST Algorithm.	CO4
b)	How does a polynomial kernel work, and how does it help in modeling non-linear relationships in Support Vector Machines (SVMs)?	CO5
c)	How does graph-based clustering using a Minimum Spanning Tree (MST) work to identify clusters in a dataset, and how would you apply this method to a given graph example to demonstrate the process of separating clusters?	CO6



Q 4) Attempt any One (10 Marks Each)

a) CO4

Weekend	Weather	Parents	Money	Decision
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay In
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

For the dataset given below, construct a decision tree using Gini Index, and determine which attribute is a root attribute.

b) CO4

An insurance company wants to predict the likelihood of claims based on customer demographics and policy details. Suggest how a Random Forest model can be used for this task and explain how bagging improves the model's robustness and accuracy. Additionally, compare the performance of the Random Forest model with that of a single decision tree.

Q 5) Attempt any One (10 Marks Each)

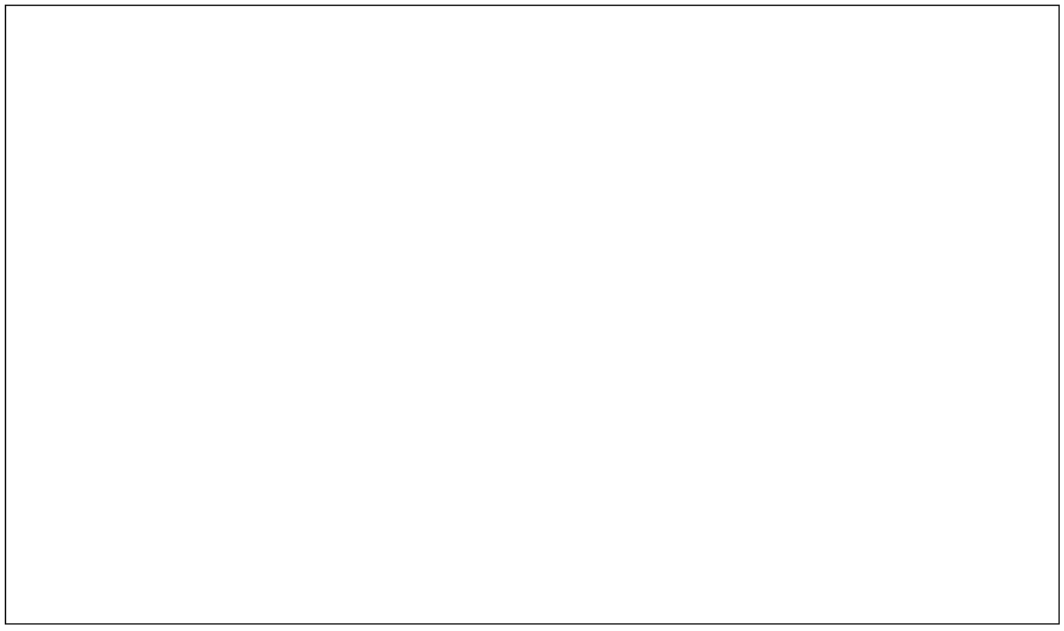
a) CO5

How do we derive the loss function in a Support Vector Machine (SVM) that penalizes misclassification errors, and how does this loss influence the overall cost function for the classifier?

b) CO6

Explain SVD as dimensionality reduction technique. Consider the matrix A as given below and find the Eigen values in SVD for A^TA.

$$A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$



Q1.A Compare and contrast supervised, unsupervised, and reinforcement learning. Provide an example for each type.

Solution

(1) Supervised Learning

→ Model is trained on a labelled dataset

→ goal is to learn the mapping between input and outputs, enabling it to make predictions on Unseen data.

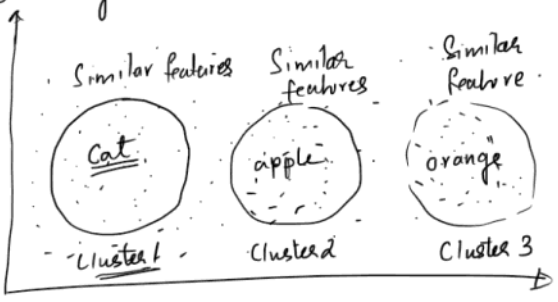
→ Model learns by comparing its predictions with the Correct output and adjust its internal parameters to minimize the prediction error.

Ex → Image classification → Given labelled images the model learns to classify new images.

Spam Detection: Training on emails labelled as Spam or not Spam, the model can filter new emails.

Unsupervised Learning

Given 10000 images and no label are specified.



* Here the input is only data and no labels are associated with data.
* Not Sure about type of output
* Unsupervised Algo will work on 10000 images and will create cluster of images based on the similarities.
* It is "SELF LEARNING"

Semi Supervised

Text Document Classifier

Ex 1000000 articles and need to classify them into News, Literature, Research Paper, medical Report

Label is not provided.

Manually labelling the 1000000 articles not possible.

→ We will label (10000) articles [Supervised Learning]

→ My model will be trained on these 10000 articles and will use the pattern identified to classify the remaining 990000 articles [Unsupervised]

* Uses small amount of labelled data
* and large amount of unlabelled data.
* Benefits of Both Labelled and unlabelled dates.
* Overcome the challenge of finding large amount of Labelled data.

Ex ① Speech Analysis → we will label audio files which will need lot of human resource and thus we will use some technique that will help to improve traditional speech analysis tool.

② Web Content Classification → Organizing the knowledge available on billions of web pages will need advance processing, but the task needs human intervention to classify

Q1.B Elaborate on Karl Pearsons coefficient of Correlation.

Solution

① Karl Pearsons Coefficient of Correlation (γ)

→ To calculate Relationship bet two variable

$$\gamma = \frac{N\Sigma xy - \Sigma x \Sigma y}{\sqrt{\left(N\Sigma x^2 - (\Sigma x)^2\right)\left(N\Sigma y^2 - (\Sigma y)^2\right)}}$$

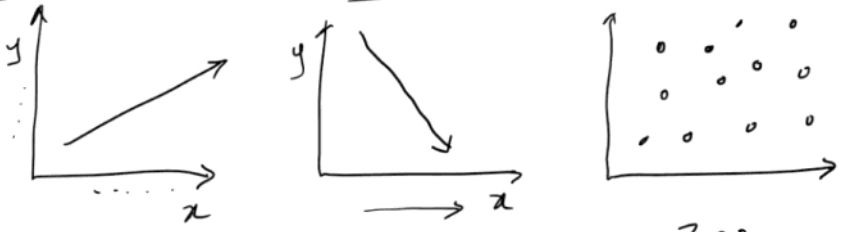
γ = quantifies the strength of relationship bet^n two variables.

The value of γ be between $+1$ and -1

If $\gamma = 1$ ⇒ Total +ve correlation ⇒ If $x \uparrow$ then $y \uparrow$

If $\gamma = -1$ ⇒ Total -ve correlation ⇒ If $x \downarrow$ then $y \uparrow$

→ If gives strength (degree) & direction of correlation. or $x \uparrow$ then $y \downarrow$



+ve Correlation

- ve Correlation

Zero Correlation.

Q1.C Elaborate on following terms with respect to DBSCAN:
(i) Directly Density Reachable
(ii) Density Reachable

Solution :-

Directly Density Reachable → A point X is directly density reachable from point Y wrt epsilon & Minpt if

1. X belong to neighbourhood of Y (dist(X,Y) <= ∈)
2. Y is core Point.

Minpt = 4

① as Y a core point →
no of point in neighbourhood of Y = 5 > Minpt
∴ Y is core point

② X belong to neighbourhood of Y.

∴ X is Directly Density Reachable from Y.

② Density Reachable →

A point X is density reachable from Point Y wrt ∈ & Minpt if there is a chain of points $P_1, P_2 \dots P_n$ and $P_1 = X$ & $P_n = Y$ such that P_{i+1} is directly density Reachable from P_i

Minpt = 4 Here X & Y are Core points

Here P_2 is directly Density Reachable from X

P_3 is directly Density Reachable from P_2

Y is directly Density Reachable from P_3

∴ X is Density Reachable from Y.

Q2.A Justify the need of cross validation. Explain k fold cross validation in details.

Solution

Cross Validation :→

→ In Supervised M.L

→ Train a model on a Dataset

→ Trained model is used to predict the target given new sample.

→ How to know if model we have trained will produce effective and accurate result on new input

Cross Validation → It is process that ensures the model will perform well on new Data.

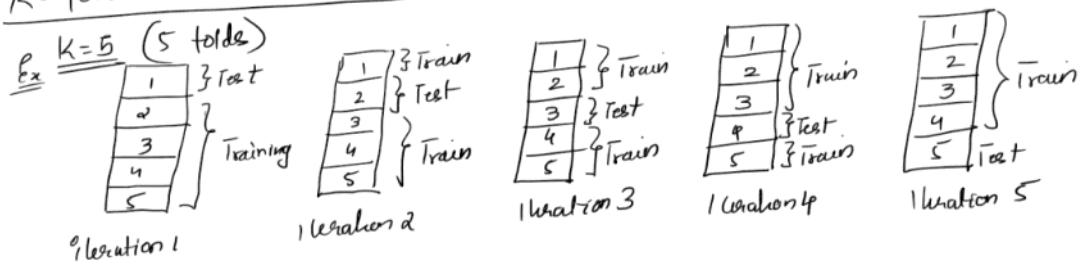
Split the Dataset
- Training Set (60%)
- Validation Set (Hold-out set) (20%)
- Test Set (20%)

Training Set = part of data on which model is trained.
(This dataset will help to * build model)

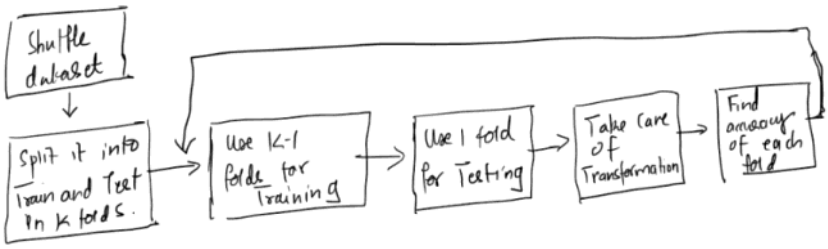
Training Set => part of data on which model is trained.
(This dataset will help to * build model)

* Validation Set => * Evaluate the Model
→ will help to chk if model overfits or Underfits.
→ Update the parameters and again train the model.
→ Repeat this until the model performs best on Validation set

K-fold Cross Validation →



K fold => → K fold helps us to build the model in generalized form.
→ To achieve this K fold Cross Validation splits the dataset into Training Testing & Validation.
→ Here Test and Train data will support building the model.
→ Life cycle of K-fold Cross Validation.



* The No of iterations ideally is K time
* Finding mean of accuracy score of each iteration will give the consistency of the Trained model.

Rules

① $K \geq 2$

if $K == 2$ → just 2 iterations.

if $K = n >> 2$ ⇒ n-1 for Training
1 for Testing

② most commonly used value of $K = 10$

③ If K is very large then the running time of process will increase.

④ The value of K is inversely proportional to size of data i.e if dataset size is small then number of folds can increase.

Q2.B INDEPENDENT DEPENDENT

SR.NO	X	Y
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

Tasks:

Fit a linear regression model to the data and determine the equation of the regression line.

Solution →

The linear regression line is represented as: $Y = a + bX$

where $$b = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$a = \frac{\Sigma y - b\Sigma x}{n}$$

Here $\Sigma x = 247$

$\Sigma y = 486$

$\Sigma x^2 = 11409$

$\Sigma y^2 = 40022$

$\Sigma xy = 20485$

$\therefore \quad y = 65.14 + 0.39x$

Q2.C How does the AdaBoost algorithm iteratively adjust weights and combine weak classifiers to form a strong classifier, and what are the main steps involved in this process?

Solution:-

Summary of Adaboost

1. Assign Sample Weight (Initial Equal)
2. Create Stump for Each feature.
3. Use Gini Index to Identify first stump.
4. For first Stump.
   1. Calculate Total Error
   2. Amt of Say
5. Update the Sample Weights for each Sample

   For Incorrectly Classified = Updated weight = old weight $\times e^{+\text{Amt of Say}}$

   For Correctly Classified = Updated weight = old weight $\times e^{-\text{Amt of Say}}$

6. Now Specify updated weight for Each Sample & Normalize the Weight to generate New Sample weight.
7. Identify New stump based on New Sample Weight & Repeat step 3 to 7
8. For Classification (Testing).

   8.1 ⇒ Run the test Sample through all the Stumps.

   8.2 ⇒ Calculate Amt of Say for Sample Classifying Yes & No

   8.3 ⇒ Classify the test Sample based on Largest Sum of Amount of Say

Idea Behind ADABOOST. (Boosting Technique → reduce Bias)

1. Adaboost combines lot of weak learners to make classification
2. The weak learners are almost always stump.
3. Some stump will have more say in classification than other stump.
4. Each stump is made by taking previous stump mistakes into account

Q3.A Justify the need of pruning in XGBOOST Algorithm.

Solution:-

Pruning : Kyon chahiye

1. Mujhe restriction of Tree
2. Ignoring Unnecessary branches
3. Decrease Computation
4. Less Time
5. Avoid Overfitting

Pruning of Trees in XGBoost is purely done on gain

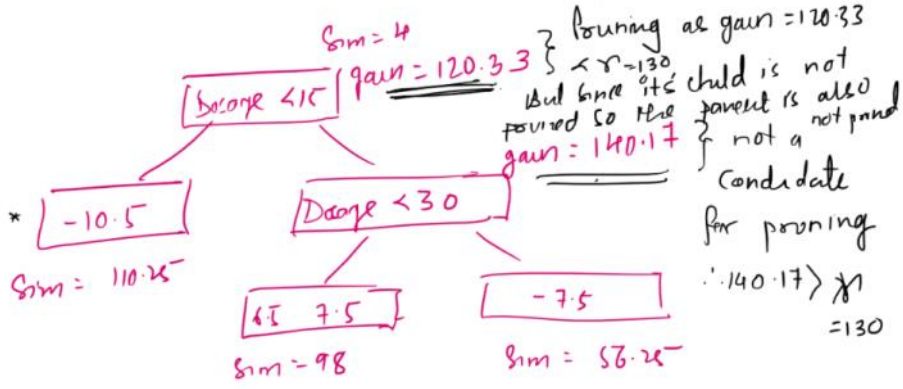
Let us Assume a Threshold gain = γ [gamma].

Rule for Pruning * If at branch gain - γ < 0 then Prune and go above

* If gain - $\gamma \geq 0$ then do not Prune
* If child is not pruned then parent also will not be pruned]

∴ Tree $\lambda = 0$

Let $\gamma = 130$



This is candidate for pruning as gain = 120.33 < γ = 130

But since its child is not pruned so the parent is also not pruned

gain = 140.17 } not a candidate for pruning

∴ 140.17 > γ = 130

* Note
If α is sufficiently large then it may result into Complete Tree pruning till root → Extreme Pruning

* Regardless of value of $\underline{\underline{\lambda}}$ and $\underline{\underline{\gamma}}$ but we assume the Tree as :→

Q3.B How does a polynomial kernel work, and how does it help in modeling non-linear relationships in Support Vector Machines (SVMs)?

Solution:-

Polynomial Kernel

* It calculates Higher Dimensional Relationship Between observations (data points).

* The Kernel that transforms 1D to 2D is Polynomial Kernel.

* It may look like $(a \times b + \gamma)^d$

where a and b are two different observations in dataset (any two data points).

γ = coefficient of Polynomial.

d = degree of Polynomial.

* Here we use SVM with Polynomial Kernel to Compute Relationship bet^n observations in Higher dimension and then find Good Classifier.

* Let a and b be two observations.
* Let $\gamma = 1/2$ & $d = 2$ } Note ⇒ value of γ and d is determined by Cross Validation.

$$\Rightarrow (a \times b + \gamma)^d = \left(a \times b + 1/2\right)^2$$

$$= \left(a \times b + 1/2\right) \cdot \left(a \times b + 1/2\right)$$

$$= a^2 b^2 + \frac{1}{2} ab + \frac{1}{2} ab + 1/4$$

$$= ab + a^2 b^2 + \frac{1}{4} \Rightarrow \text{Polynomial.}$$

= can be written as dot product of

$= \left(a, a^2, \frac{1}{2}\right) \cdot \left(b, b^2, \frac{1}{2}\right)$

For data point a

original value a

Higher dimension value of a

original value b

value of b in higher dimension

Dot product is sum of 1st term multiplied, and term multiplied and so on

Note: $\frac{1}{2}$ is third axis but value same so ignore.

So $(a + b \times \gamma)^d$ is used to get Higher dimension Rel^n between two data points a & b.

Q3.C How does graph-based clustering using a Minimum Spanning Tree (MST) work to identify clusters in a dataset, and how would you apply this method to a given graph example to demonstrate the process of separating clusters?

Solution :-

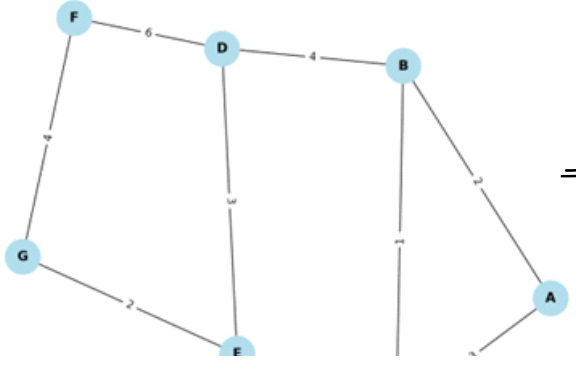
* Graph Based Clustering Using Minimum Spanning Tree →

Spanning Tree
→ Subgraph of given graph
→ Same No of vertex
→ and no cycle
→ Subgraph is Connected (All vertex)

Algo
1. Determine MST of given graph.
2. Delete branch iteratively.
3. Each Connected Component is a cluster

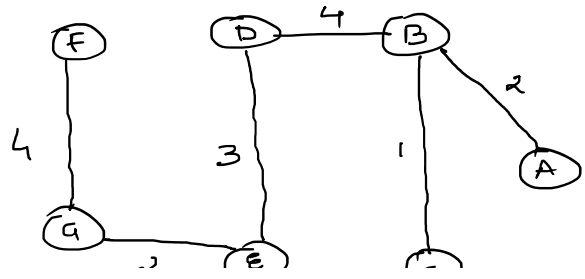
Different Stralgies to Delete Branches.
(1) Delete Branch with max weight →.
(2) Delete Inconsistent Branch →
(3) Delete by Analysis of weight.

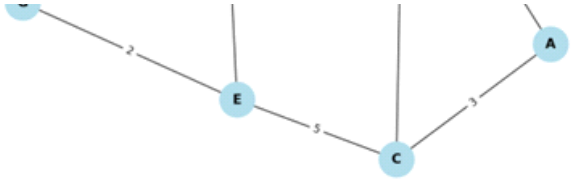
Delete Inconsistent Branches.
* A branch is inconsistent if the corresponding weight (de) is much larger than the reference value $\bar{de}$
* The reference value $\bar{de}$ can be defined by the average weight of all the branches adjacent to edge e

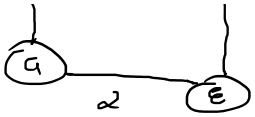


⇒

MST for the graph

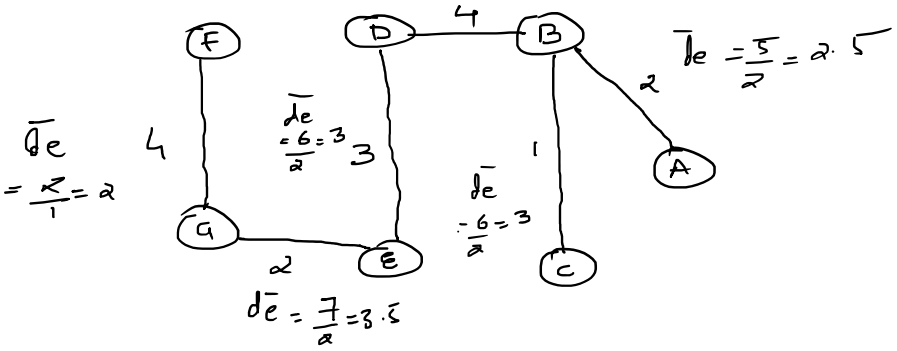




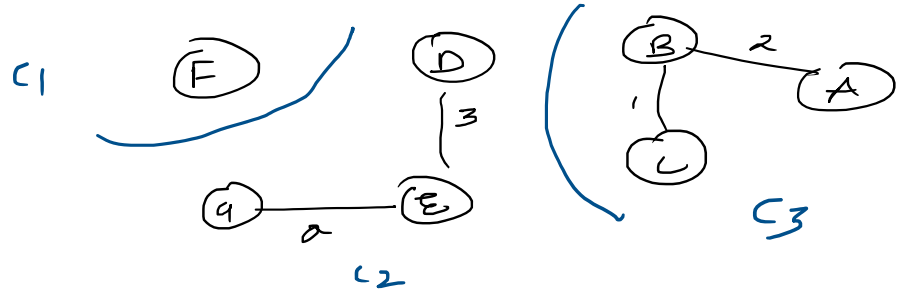


Calculating $\bar{d}_e$ for every edge

$\bar{d}_e = \frac{6}{3} = 2$



can delete edge with cost 4 i.e FG and DB



we have 3 clusters.

Q4.A a)

	Weekend	Weather	Parents	Money	Decision
W1	Sunny	Yes	Rich		Cinema
W2	Sunny	No	Rich		Tennis
W3	Windy	Yes	Rich		Cinema
W4	Rainy	Yes	Poor		Cinema
W5	Rainy	No	Rich		Stay In
W6	Rainy	Yes	Poor		Cinema
W7	Windy	No	Poor		Cinema
W8	Windy	No	Rich		Shopping
W9	Windy	Yes	Rich		Cinema
W10	Sunny	No	Rich		Tennis

For the dataset given below, construct a decision tree using Gini Index, and determine which attribute is a root attribute.

Solution → Independent Features: Weather, Parent, Money

Decision/Outcome: Cinema, Tennis, StayIn, Shopping.

Step 1

We will calculate Gini Index for Overall collection of Outcomes of Training Examples.

There are 4 possible outcomes for decision: Cinema — 6 instances; Tennis — 2 instances; StayIn — 1 instances; Shopping — 1 instances.

$$Gini_{(decision)} = 1 - \left(\left(\frac{6}{10}\right)^2 + \left(\frac{2}{10}\right)^2 + \left(\frac{1}{10}\right)^2 + \left(\frac{1}{10}\right)^2\right)$$

$$= 1 - \left(\frac{42}{100}\right) = 0.58$$

Note → In machine learning, Gini index/coefficient is utilized as an impurity measures in decision tree for classification.

$$Gini = 1 - \sum_{i=1}^{n} (P_i)^2$$ where P_i probability of outcome of specific class

Step 2: To find Gini Index for money

Gini (money) → possible value → Rich, poor

① Gini (money = Rich) → 7 instance → possible Decision → Cinema → 3 time, Tennis → 2 time, Stay In → 1 time, Shopping → 1 time.

$$\text{Gini}_{(\text{money}=\text{Rich})} = 1 - \left(\left(\frac{3}{7}\right)^2 + \left(\frac{2}{7}\right)^2 + \left(\frac{1}{7}\right)^2 + \left(\frac{1}{7}\right)^2 \right)$$

$$= 0.694$$

② Gini (money = poor) → 3 instance → Decision → Cinema.

$$= 1 - \left(\frac{3}{3}\right)^2 = 0 //$$

$$\text{Weighted Average Gini}_{(\text{money})} = \left(\text{Gini}_{(\text{money}=\text{Rich})} * \text{Proportion of rich}\right) + \left(\text{Gini}_{(\text{money}=\text{poor})} * \text{Proportion of poor}\right)$$

$$= (0.694 * 7/10) + (0 * 3/10)$$

$$\text{Gini}_{(\text{money})} = 0.485$$

Step 3 Gini Index on Parent

For Parent feature → Possible values → Yes, NO

Gini (parent = Yes) = 5 instances → Possible Decision → Cinema

$= 1 - \left((5/5)^2\right) = 0$ //

Gini (parent = NO) = 5 instance → possible decision → Tennis → 2 times, Stay/In → 1 times, Shopping → 1 time, Cinema → 1 time

$= 1 - \left((2/5)^2 + (1/5)^2 + (1/5)^2 + (1/5)^2\right)$

$= 1 - (9/25) = 0.72$

Weighted Average of $Gini_{(parent)} = (0 * 5/10) + (0.72 * 5/10) = 0.36$

$Gini_{(parent)} = 0.36$

Step 4) Gini Index for Weather.

Weather → Possible values → Sunny = 3 instances, Windy = 4 instances, Rainy ⇒ 3 instances.

Gini (weather = Sunny) ⇒ 3 instances → possible outcomes: Cinema → 1 time, Tennis → 2 time

$= 1 - \left((1/3)^2 + (2/3)^2 \right) =$ 0.444

Gini (weather = Windy) → 4 instances → possible outcome: 3 time Cinema, 1 time Shopping

$= 1 - \left((3/4)^2 + (1/4)^2 \right) =$ 0.375

Gini (weather = Rainy) → 3 instance → possible outcomes: 2 cinema, 1 Stay In

$= 1 - \left((2/3)^2 + (1/3)^2 \right) =$ 0.444

Weighted Average Gini (weather) $= (0.44 * 3/10) + (0.375 * 4/10) + (0.44 * 3/10)$

$=$ 0.419

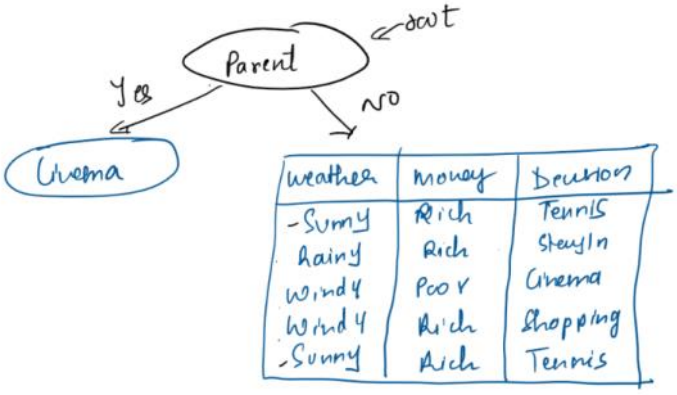
∴

Gini (money)	= 0.486
Gini (parent)	= 0.36
Gini (weather)	= 0.416

Minimum Gini ⟶ Minimum Impurity in Decision.

Here Minimum Gini Value = Gini (parent) = 0.36

So the root of Decision is Parent



Weather	Money	Decision
Sunny	Rich	Tennis
Rainy	Rich	Stay In
Windy	Poor	Cinema
Windy	Rich	Shopping
Sunny	Rich	Tennis

Q4.B An insurance company wants to predict the likelihood of claims based on customer demographics and policy details. Suggest how a Random Forest model can be used for this task and explain how bagging improves the model's robustness and accuracy. Additionally, compare the performance of the Random Forest model with that of a single decision tree.

Solution =>

Using a Random Forest Model in Claim Prediction.

→ A Random Forest model is an ensemble method that combines multiple decision tree to make predictions.

Feature Selection

⇒ Customer demographics (ex age, gender, location) and policy details (ex policy type, coverage amount, previous claims) are treated as input features for the model

⇒ Label → the likelihood of a claim (ex claim or no claim, probability of claim) can be target label.

How Bagging Improves Robustness and Accuracy

- Process → In Bagging multiple samples are drawn with replacement from the original dataset to create multiple bootstrapped datasets. Each decision tree in the Random Forest is trained on a different bootstrapped dataset.

- Diversity → Since each tree is trained on a unique sample and may use a random subset of features, the trees becomes diverse, capturing different patterns in data.

→ Reduced Overfitting → By averaging the results of multiple trees, bagging reduced the likelihood of overfitting which is common problem with individual decision trees.

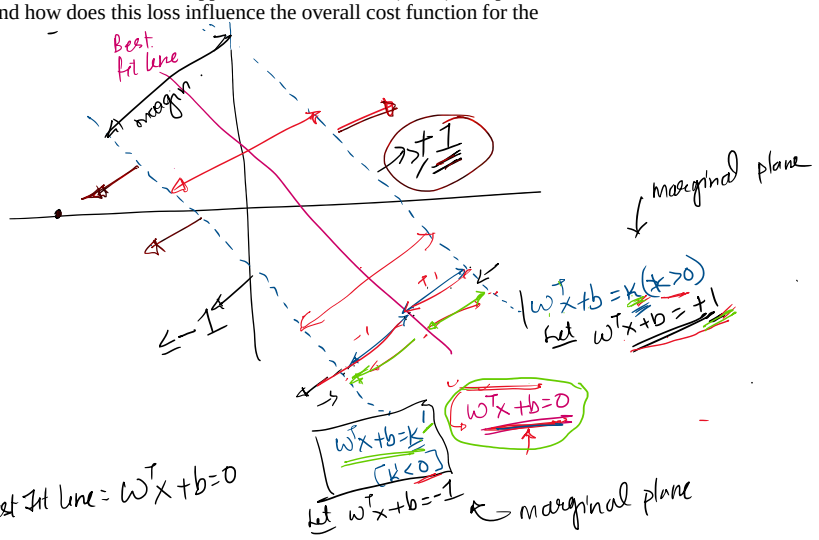
⇒ Increased Robustness ⇒ The Random Forest model is more robust to outliers and noise, as the effect of errors in individual trees are minimized by combining results across many trees.

Comparison :

Feature	Random Forest	Single Decision Tree
Accuracy	Higher, due to ensemble effect	Lower, prone to overfitting
Variance	low	High
Interpretability	lower	Higher
Overfitting	less likely	more likely.
Prediction Stability	High (due to averaging results)	low (depends on only one model)

Q5.A How do we derive the loss function in a Support Vector Machine (SVM) that penalizes misclassification errors, and how does this loss influence the overall cost function for the classifier?

Solution



The Hyperplane / Best Fit line : $w^T x+b=0$

Let the Marginal plane be $w^T x+b=+1$ [on +ve side]

And $w^T x+b=-1$ [on -ve side]

Our Aim is to draw two marginal planes (+ve & -ve side) and need to ensure the distance (Margin) is maximum.

* we want to find distance b/n the marginal planes.

Lets find difference :

$$w^T x_1 + b = +1$$
$$w^T x_2 + b = -1$$
$$w^T(x_1 - x_2) = 2$$

$\therefore\ w^T(x_1-x_2)=2$

Here w = slope (coefficient) < magnitude, direction } Vector.

To Convert w^T into Vector ie $\vec{w}^T$

divide by $\|w\|$

$\therefore\ \frac{w^T}{\|w\|}(x_1-x_2) = \frac{2}{\|w\|}$ = $\vec{w}(x_1-x_2)=\frac{2}{\|w\|}$

$\frac{w^T}{\|w\|}(x_1-x_2)$ ↑ difference (margin)

$\frac{2}{\|w\|}$ ↑ value

$\therefore$ To Maximize the Margin →

$$\underset{(w,b)}{\text{Maximize}} = \frac{2}{\|w\|}$$

Constraints → , $(+1$ when $w^T x+b \geq$:

(w, b)

Constraints ⇒ y_i $\begin{cases} +1 & \text{when } w^T x + b \geq 1 \\ & (\text{point lies outside of } w^T x + b = +1) \\ -1 & \text{when } w^T x + b \leq -1 \\ & (\text{point lies below of } w^T x + b = -1) \end{cases}$

The Constraints are for Correctly Classified Data point

∴ Final Constraints for S.V Classifier.

$$y_i * (w^T x + b) \geq 1$$

For Correctly Classified Data point

y_i ↑ observed (actual) ; $w^T x + b$ ↑ Predicted value

∴ To Maximize $= \frac{2}{\|w\|}$ Subjected to $y_i * (w^T x + b) \geq 1$

Also can be written as

To Minimize $= \frac{\|w\|}{2}$ Subjected to $y_i * (w^T x + b) \geq 1$

only for Correctly Classified Data points.

Here we have not Considered for Misclassification

But in real world there will be always Misclassification.

∴ To Consider for Misclassification we have to use

"Hyper parameters"

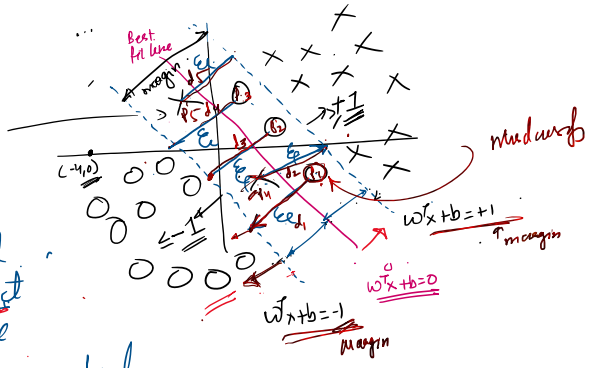
We allow Misclassification

Now we will use

two Hyperparameter

ξ_i ⇒ ξ_i = distance bet^n the Misclassified point and correct Marginal plane

C_i ⇒ No of Allowed Misclassified points.



Now Final Cast F^n (For All data points) [Correctly & Incorrectly Classified]

↓ Mathematical

Now Final Cast Fn (For All data points) (Incorrectly classified)

* Objective's

To Minimize $\frac{\|w\|}{2} + C_i \sum_{i=1}^{n} \xi_i$ Constraint $y_i^o * (w^T x + b) \geq 1$

C_i → for Hyperparameters.

Correctly identified — Misclassification.

The total Cost Fn for SVC used for classification [For Allowed Misclassification].

= Sum of Eta values of all the Misclassified points.

Q5.B Explain SVD as dimensionality reduction technique. Consider the matrix A as given below and find the Eigen values in SVD for ATA.

$$A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

Solution → To find U := $AA^T = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} * \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}$

$\Rightarrow \quad AA^T - \lambda I = 0$

for Eigen Values $\Rightarrow \begin{vmatrix} 11-\lambda & 1 \\ 1 & 11-\lambda \end{vmatrix} = 0 \longrightarrow (I)$

$\lambda^2 - 22\lambda + 120 = 0$

$\lambda_1 = 10 \quad \lambda_2 = 12$

For Eigen Vector → $\begin{bmatrix} 11-\lambda & 1 \\ 1 & 11-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Case 1) Substitute $\lambda = \lambda_1 = 10$

$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\frac{x_1}{1} = \frac{-x_2}{1} = 1$

$x_1 = 1 \quad x_2 = -1$

Eigen Vector $X_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

(cont'd) Substitute $\lambda = \lambda_2 = 12$

$$\begin{bmatrix} 11-\lambda & 1 \\ 1 & 11-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Eigen Vector $X_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

* In U the Eigen Vector generated by largest Eigen value will be the first column.

$$\therefore U = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

↑ Eigen vector of $\lambda = 12$ ↑ Eigen Vector of $\lambda = 10$

Now we need to Normalize the matrix ⇒ divide by length of respective Vector.

$$U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

↑ length of Vector $\lambda = 12$ $\sqrt{1^2+1^2} = \sqrt{2}$

↑ length of Vector $\lambda = 10$ $\sqrt{(1)^2+(-1)^2} = \frac{1}{\sqrt{2}}$

Step 2 To find V

$$A^TA = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}_{3\times3}$$

Here for Eigen values $A^TA - \lambda I = 0$

$$\lambda^3 - 22\lambda^2 + 132\lambda = 0$$

$$\therefore \boxed{\begin{aligned} \lambda_1 &= 0 \\ \lambda_2 &= 10 \\ \lambda_3 &= 12 \end{aligned}}$$

We need to find Eigen Vectors for the 3 Eigen values.

Case 1) Eigen Vector for $\lambda_1 = 0$

$$\Rightarrow \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

By Cramers Rule

$$\frac{x_1}{\begin{vmatrix} 0 & 2 \\ 10 & 4 \end{vmatrix}} = \frac{x_2}{\begin{vmatrix} 10 & 2 \\ 0 & 4 \end{vmatrix}} = \frac{x_3}{\begin{vmatrix} 10 & 0 \\ 0 & 10 \end{vmatrix}}$$

$$\frac{x_1}{-20} = \frac{-x_2}{40} = \frac{x_3}{100} = \frac{-1}{20}$$

$$\boxed{x_1 = 1 \qquad x_2 = 2 \qquad x_3 = -5}$$

Eigen Vector $X_1 = \begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix}$

Case 2) Eigen Vector for $\lambda = 10$

$$\begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 4 \\ 2 & 4 & -8 \end{bmatrix} \begin{bmatrix} x1 \\ x2 \\ x3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

By Cramers Rule

$$\frac{x1}{\begin{vmatrix} 0 & 4 \\ 4 & -8 \end{vmatrix}} = \frac{-x2}{\begin{vmatrix} 0 & 4 \\ 2 & -8 \end{vmatrix}} = \frac{x3}{\begin{vmatrix} 0 & 0 \\ 2 & 4 \end{vmatrix}} \Rightarrow \frac{X_1}{-16} = \frac{-x_2}{-8} = \frac{x_3}{0} = \frac{-1}{8}$$

$\therefore x_1 = 2 \quad x_2 = -1 \quad x_3 = 0$

Eigen Vector $= X_2 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$

Case 3) Eigen Vector for $\lambda_3 = 12$

$$\begin{bmatrix} -2 & 0 & 2 \\ 0 & -2 & 4 \\ 2 & 4 & -10 \end{bmatrix} \begin{bmatrix} x1 \\ x2 \\ x3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

By Cramers Rule

$$\frac{x1}{\begin{vmatrix} 0 & 2 \\ -2 & 4 \end{vmatrix}} = \frac{-x2}{\begin{vmatrix} -2 & 2 \\ 0 & 4 \end{vmatrix}} = \frac{x3}{\begin{vmatrix} -2 & 0 \\ 0 & -2 \end{vmatrix}} \Rightarrow \frac{X_1}{4} = \frac{-x_2}{-8} = \frac{x_3}{4} = \frac{1}{4}$$

$\therefore x_1 = 1 \quad x_2 = 2 \quad x_3 = 1$

Eigen Vector $X_3 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

Now $V = \begin{bmatrix} 1 & 2 & 1 \\ 2 & -1 & 2 \\ 1 & 0 & 5 \end{bmatrix}$

Eigen Vector for $\lambda = 12$ | Eigen Vector for $\lambda = 10$ | Eigen Vector for $\lambda = 0$.

Now Normalizing the matrix → Divide by length of Vector.

$$V = \begin{bmatrix} 1/\sqrt{6} & 2/\sqrt{5} & 1/\sqrt{30} \\ 2/\sqrt{6} & -1/\sqrt{5} & 2/\sqrt{30} \\ 1/\sqrt{6} & 0 & -5/\sqrt{30} \end{bmatrix}$$

Length of Vector $\sqrt{6}$ | Length of Vector $\sqrt{5}$ | Length of Vector $\sqrt{30}$

$$V^T = \begin{bmatrix} 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \\ 2/\sqrt{5} & -1/\sqrt{5} & 0 \\ 1/\sqrt{30} & 2/\sqrt{30} & -5/\sqrt{30} \end{bmatrix}$$

Step3 To find Σ (or D)

$$\Sigma = \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \\ 0 & 0 & \sqrt{0} \end{bmatrix}_{3\times 3}$$

diag matrix ⇒ the diag elements are Square root of Eigen values in Decreasing order.

$$\therefore A = U \Sigma V^T$$

$$\begin{bmatrix} 7 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \\ 2/\sqrt{5} & -1/\sqrt{5} & 0 \\ 1/\sqrt{30} & 2/\sqrt{30} & -5/\sqrt{30} \end{bmatrix}$$

U Σ V^T