Q1a) Reinforcement learning (Experiential learning)

* Here the agent learns how to behave in an environment by performing actions and experiencing the result.

Types

Episodic learning

Here we have starting & ending state
Thus an episode is created

→ Here the further action is based on feedback of result of earlier action.

Ex → Fear of dog after being bitten is Episodic learning

Continous Task learning

* There is no terminal state here
* Agent that does Automated Stock trading goes for Continous learning.

Q1b)

Clustering

① Model will identify the pattern in data and will create clusters

Ex → Above given Market Segment"

Association

① We find dependencies of one data item on another.
② This dependency will help to map the relation and will enhance prediction
③ It's like "If - Then"

Clustering

① Model will identify the pattern in data and will create clusters

Ex -> Above given

Market Segment"

Created Clusters of
- High Sale
- Medium Sale
- Low Sale

Association

① We find dependencies of one data item on another.

② This dependency will help to map the relation and will enhance prediction

③ It's like "If - Then"

Ex Market Basket Analysis used by Amazon

If a person purchases cell phone then the person has tendency to purchase screenguard & batteries.

Q1 c)

1 Mark | Pearsons Coefficient of Correlation.

$$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{\left[N\sum x^2 - (\sum x)^2\right]\left[N\sum y^2 - (\sum y)^2\right]}}$$

$r \Rightarrow$ quantifies the strength of relationship between two variables

The value of r will be bet^n ± 1 and -1.

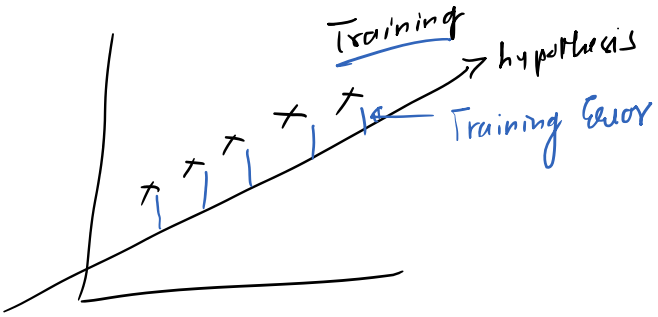
- If $1 \Rightarrow$ Total +ve correlation of $x \uparrow$ then $y \uparrow$
- If $-1 \Rightarrow$ Total -ve correlation of $x \downarrow$, then $y \downarrow$,
- If $0 \Rightarrow$ No correlation.

→ It summarizes the digree & direction of corelation.

Old>

① Training Error →

It is prediction Error that we get when we apply the model on same data from which we trained.



$$\mathcal{E}_{Train} = \frac{1}{n} \sum_{i=1}^{n} Error\left(f_D(x_i), y_i\right)$$

($f_D(x_i)$: prediction on x_i; y_i: actual value; $i=1$ (for all samples))

② Test Error → It is prediction Error we get when we apply model on altogether different data set (Test set) and not on the data it is trained.

$$\mathcal{E}_{test} = \frac{1}{n} \sum_{i=1}^{n} Error\left(f_D(x_i), y_i'\right)$$

over test set

Q1 e)

Performance Metric.

Consider a Skewed Data Set (one Sided)

Ex 100 MRI Scans → 99 No cancer (No)
→ 1 cancer (Yes)

Error → 1%

Normally allowed

But Here will not work.

In Such Cases to Evaluate Performance →

Confusion Matrix

Predicted \ Actual	True	False
True	True +ve (TP)	False +ve (FP)
False	False -ve (FN)	True -ve (TN)

For 2 classes True / False

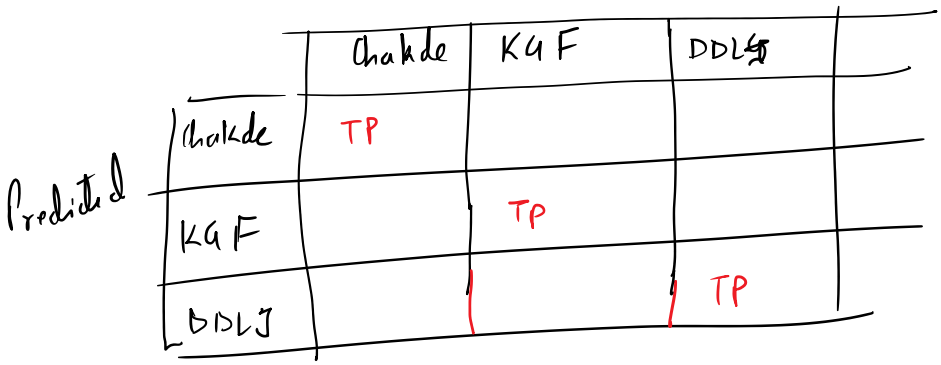
Ex

Actual

Predicted	Has Heart Disease	Do not have Heart Disease
Has Heart Disease	TP	FP
Do not have Heart Disease	FN	TN

* If we have more than Two classes.

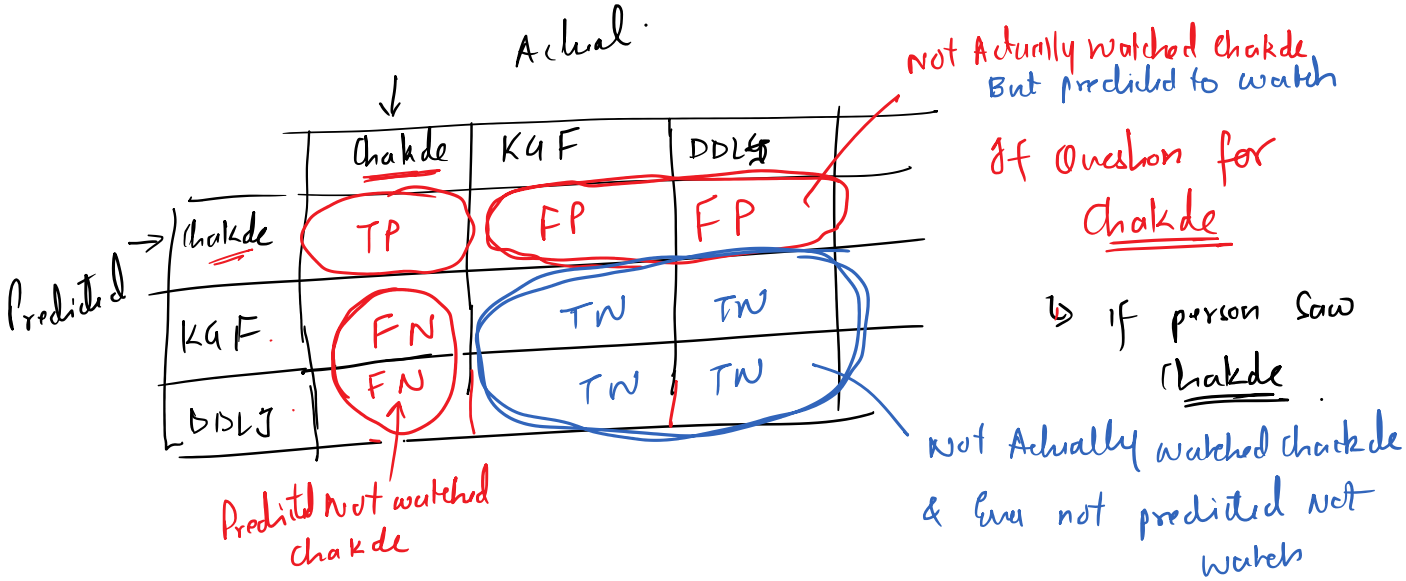
Actual



* Size of confusion matrix is determined by No of things we want to predict.
* Confusion matrix summarizes what ML Algo did right and what it did wrong

Tone true is evident easily

But for Tone -ve, FP, FN it depends on Question Asked.



Q1 f)

(1) Sensitivity ⇒ (TPR)

Ex ⇒ It tells us what percentage of patient with Heart Disease are correctly identified

$$= \frac{\text{Correctly Identified}}{\text{Total no of Patients with Heart Disease}} = \frac{TP}{\text{Actual +ve}}$$

(2) Specificity = Tells us what percentage of patients without Heart Disease were correctly identified.

$$= \frac{\text{Correctly Identified}}{\text{Total No of Patients without Heart Disease}} = \frac{TN}{\text{Actual -ve}}$$

FPR = Fake +ve Rate

$$= 1 - \text{Specificity} = 1 - \frac{TN}{TN+FP} = \frac{TN+FP-TN}{TN+FP} = \frac{FP}{TN+FP}$$ ✓

Ex

Predicted \ Actual	Chakde	KGF	DDLJ
Chakde	TP 12	102	93
KJF	FN 112	23	77
DDLJ	83	92	17

FP

TN

Calculate Sensitivity & Specificity for Chakde.

$$\text{Sensitivity} = \frac{\text{Correct +ve}}{} = \frac{TP}{TP+FN} = \frac{12}{12+112+83} = 0.06$$

Sensitivity = Correct +ve / Total +ve = TP / TP + FN = 12 / 12 + 112 + 83 =

chakde

Specificity = Correct -ve / Total -ve = TN / TN + FP = 23 + 77 + 92 + 17 / 23 + 77 + 92 + 17 + 102 + 93 = 0.52

chakde

Q2a)

Linear Regression Numerical Ex with multiple independent features (Here 2).

[For >2 features we will need Matrix Algebra.]

Consider

Sr.NO	X_1	X_2	Y.	
1	3	8	-3.7	given.
2	4	5	3.5	
3	5	7	2.5	
4	6	3	11.5	
5	2	1	5.7	
6	3	2	?	Predict

Independent features

Dependent on X_1 & X_2.

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

Linear Reg for 2 Independent Variable.

$$\theta_0 = \bar{y} - \theta_1 \bar{x_1} - \theta_2 \bar{x_2}$$

$$\theta_1 = \frac{(\Sigma x_2^2)(\Sigma x_1 y) - (\Sigma x_1 x_2)(\Sigma x_2 y)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

$$\theta_2 = \frac{(\Sigma x_1^2)(\Sigma x_2 y) - (\Sigma x_1 x_2)(\Sigma x_1 y)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

Here $\Sigma x_1^2 = \Sigma x_1 x_1 - \frac{(\Sigma x_1)(\Sigma x_1)}{N} = \Sigma (x_1)^2 - \frac{(\Sigma x_1)^2}{N}$

$$\Sigma x_2^2 = \Sigma x_2 x_2 - \frac{(\Sigma x_2)(\Sigma x_2)}{}$$

$$\Sigma x_1 y = \Sigma x_1 y - \frac{(\Sigma x_1)(\Sigma y)}{N}$$

$$\Sigma x_2 y = \Sigma x_2 y - \frac{(\Sigma x_2)(\Sigma y)}{N}$$

$$\Sigma x_1 x_2 = \Sigma x_1 x_2 - \frac{(\Sigma x_1)(\Sigma x_2)}{N}$$

S.NO	y	x_1	x_2	$(x_1)^2$	$(x_2)^2$	$x_1 y$	$x_2 y$	$x_1 x_2$
1	-3.7	3	8	9	64	-11.1	-29.6	24
2	8.5	4	5	16	25	14	17.5	20
3	2.5	5	7	25	49	12.5	17.5	35
4	11.5	6	3	36	9	69	34.5	18
5	5.7	2	1	4	1	11.4	5.7	2
Σ	19.5	20	24	90	148	95.8	45.6	99.

$$\Sigma x_1^2 = \Sigma (x_1)^2 - \frac{(\Sigma x_1)^2}{N} = 90 - \frac{(20)^2}{5} = 90 - 80 = 10$$

$$\Sigma x_2^2 = = 148 - \frac{(24)^2}{5} = 32.8$$

$$\Sigma x_1 y = 95.8 - \frac{20 * 19.5}{5} = 17.8$$

$$\Sigma x_2 y = 45.6 - \frac{24 * 19.5}{5} = -48$$

$$\Sigma x_1 x_2 = 99 - \frac{20 \times 24}{5} = 99 - 96 = 3$$

$$\beta_1 = \frac{(32.8)(17.8) - (3)(-48)}{(10)(32.8) - (3)^2} = 2.28$$

$$a_2 = \frac{(10)(-48) - (3)(17.8)}{(10)(32.8) - (3)^2} = -1.67$$

$$a_0 = \bar{y} - a_1\bar{x}_1 - a_2\bar{x}_2 = \frac{19.5}{5} - \frac{2.28 \times 20}{5} - \frac{(-1.67)(24)}{5}$$

$$= 2.796$$

$$\therefore y = 2.796 + 2.28x_1 - 1.67x_2$$

Substitute $x_1 = 3$ $x_2 = 2$ y_0 ?

Q2 b)

Create Decision Tree for following Using Gini Index (Classification & Regression Tree) → CART.

Weekend	Weather	Parent	Money	Decision (y)
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

Soln →

Calculate Gini Index for Overall collection of Outcomes of Training Example.

There are 4 possible Outcome Variable
- Cinema → 6 instances
- Tennis → 2 instances
- StayIn → 1 instance
- Shopping → 1 instance.

$$\text{Gini(Decision)} = 1 - \left[\left(\frac{6}{10}\right)^2 + \left(\frac{2}{10}\right)^2 + \left(\frac{1}{10}\right)^2 + \left(\frac{1}{10}\right)^2\right]$$

(Cinema, Tennis, StayIn, Shopping)

=

$= 0.58$.

* We will calculate Gini Index for every attribute and based on value of gini for Attribute we will select root and start constructing Tree.

(1) Let find Gini Index for Money feature.

2 possible values → Rich ⇒ 7 instances
→ Poor ⇒ 3 instances.

we need to find Gini for Both Rich & Poor

Gini (Rich) → 7 instances → 2 instances Decision = Tennis
→ 3 " " = Cinema
→ 1 " " = Stay In.
→ 1 " " = Shopping

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

$$= 0.694$$

Gini (Poor) ⇒ 3 instances → Decision is Cinema.

$$\text{Gini(Poor)} = 1 - \left(\left(\frac{3}{3}\right)^2\right)$$

$$= 1 - 1 = 0$$

Weighted Avg Gini for Money

Weighted Avg Gini for Money

$$= (\text{Gini Index} * \text{Proportion}) \text{ for each possible values.}$$

$$= \underbrace{\left(0.694 * \frac{7}{10}\right)}_{\text{Rich.}} + \underbrace{\left(0 * \frac{3}{10}\right)}_{\text{Poor}}$$

$$= 0.486$$

Gini (Money) = 0.486

(2) Gini Index for Parent.

Parent — Possible Outcomes. → Yes ⇒ 5, No = 5

Gini (Parent = Yes) = 5 instances — Possible Decision — 5 Cinema.

Gini (Parent = Yes) = $1 - [(5/5)^2] = 0$.

Gini (Parent = NO) = 5 instances → 2 times Tennis, 1 Shopping, 1 Stay In, 1 Cinema.

Gini (Parent = NO) = $1 - [(2/5)^2 + (1/5)^2 + (1/5)^2 + (1/5)^2]$

$$= 0.72$$

$= 0.72$

Weighted Avg Gini for Parent $= \left(0 * \frac{5}{10}\right) + \left(0.72 * \frac{5}{10}\right)$

$= 0.36$

$\therefore$ **Gini (Parent) = 0.36**

(3) Gini Index for Weather.

Weather — possible values:
- Sunny ⇒ 3 instances
- Windy = 4 instances
- Rainy = 3 instances.

Gini (Weather = Sunny) — 3 instances:
- 1 Cinema
- 2 Tennis

Gini (Weather = Sunny) $= 1 - \left((1/3)^2 + (2/3)^2\right) = 0.444$.

Gini (Weather = Windy) = 4 instances:
- 3 Cinema
- 1 Shopping

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

Gini (Weather = Rainy) = 3 instances:
- 2 Cinema
- 1 Stayin

$\therefore$ Gini (Weather = Rainy) $= 1 - \left((2/3)^2 + (1/3)^2\right) = 0.444$

weighted Avg Gini for weather $= \left(0.444 * \frac{3}{10}\right) + \left(0.375 \times \frac{4}{10}\right) + \left(0.444 * \frac{3}{10}\right)$

Sunny Windy Rainy

Gini (weather) = 0.416.

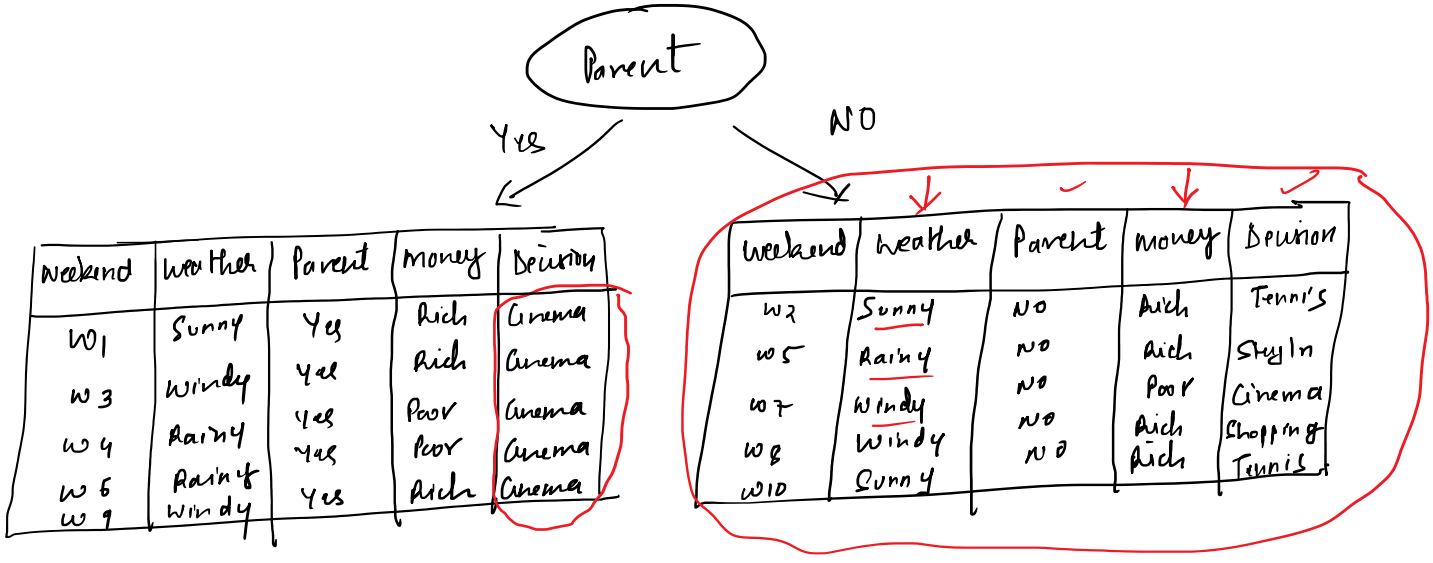
∴ Gini (weather) = 0.416

Gini (Parent) = 0.36

Gini (money) = 0.486.

Gini Index for Parent is Minimum So it is said to have maximum information.

So Select Parent as Root Node of Decision Tree



Weekend	Weather	Parent	Money	Decision
W1	Sunny	Yes	Rich	Cinema
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W6	Rainy	Yes	Poor	Cinema
W9	Windy	Yes	Rich	Cinema

Weekend	Weather	Parent	Money	Decision
W2	Sunny	NO	Rich	Tennis
W5	Rainy	NO	Rich	Stay In
W7	Windy	NO	Poor	Cinema
W8	Windy	NO	Rich	Shopping
W10	Sunny	NO	Rich	Tennis

* Computation of Gini Index for Parent = Yes Not needed. as In All cases Decision = Cinema

* For parent = NO we do not have Single Decision

So again we have to calculate Gini for.

Gini (Parent = NO & Weather) ——— possible value

Sunny - 2 instance
Windy - 2 instance
Rainy - 1 instance

Gini (Parent = NO & weather = Sunny) ⇒ 2 instance < 2 tennis

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

Gini (Parent = NO & weather = Windy) = 2 instance < 1 Cinema, 1 Shopping

$= 1 - \left[\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2\right]$

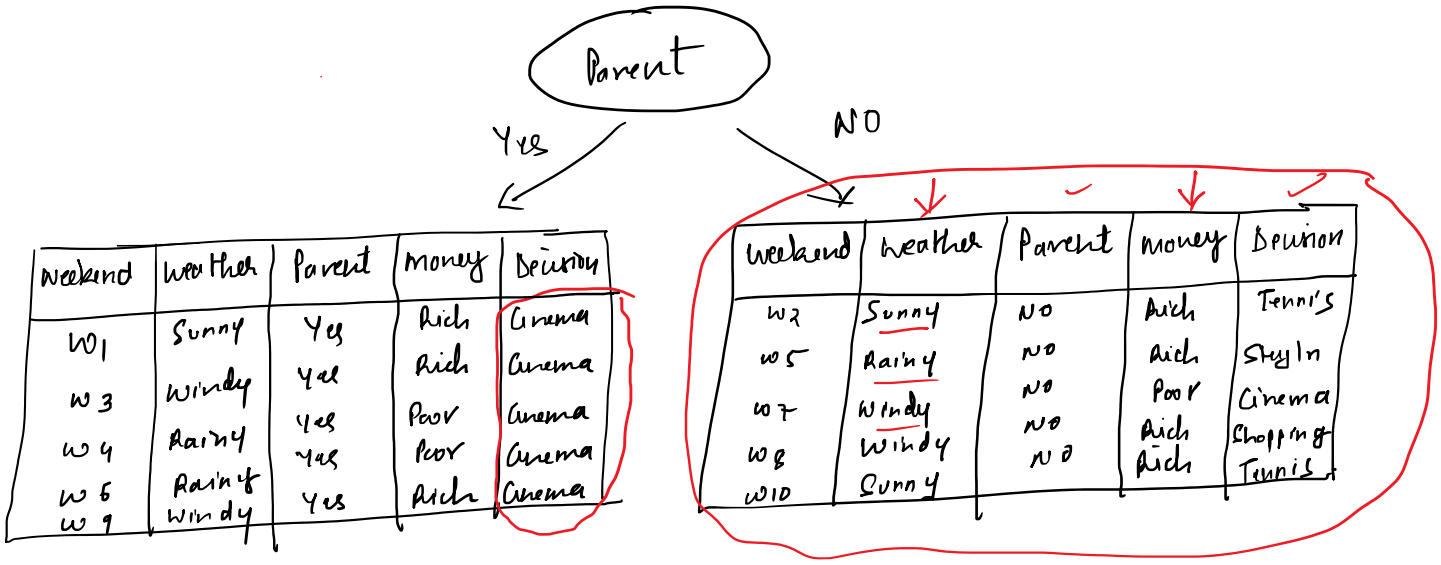
$= 0.5$

Gini (Parent = NO & weather = Rainy) = 1 instance → 1 Study In

$= 1 - \left[\left(\frac{1}{1}\right)^2\right] = 0.$

∴ Avg weighted gini (Parent = NO & weather) $= \left(0 * \frac{2}{5}\right) + \left(0.5 * \frac{2}{5}\right) + \left(0 * \frac{1}{5}\right)$

$= 0.2$



Weekend	Weather	Parent	Money	Decision
W1	Sunny	Yes	Rich	Cinema
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W6	Rainy	Yes	Poor	Cinema
W9	Windy	Yes	Rich	Cinema

Weekend	Weather	Parent	Money	Decision
W2	Sunny	NO	Rich	Tennis
W5	Rainy	NO	Rich	Stay In
W7	Windy	NO	Poor	Cinema
W8	Windy	NO	Rich	Shopping
W10	Sunny	NO	Rich	Tennis

Gini Index Parent = NO and money: < Rich → 4 instances = / Poor → 1 instances.

gini Index (Parent = NO & Money = Rich) = < 2 Tennis / 1 Stay In / 1 Shopping.

gini Index (Parent = NO & money = Rich) = $1 - \left((2/4)^2 + (1/4)^2 + (1/4)^2 \right)$

$= 0.625$

gini Index (Parent = NO & money = Poor) — 1 instance Cinema

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

Weighted Average gini Index for (Parent = NO & money) =

$= 0.625 \times \frac{4}{5} + 0 \times \frac{1}{5}$

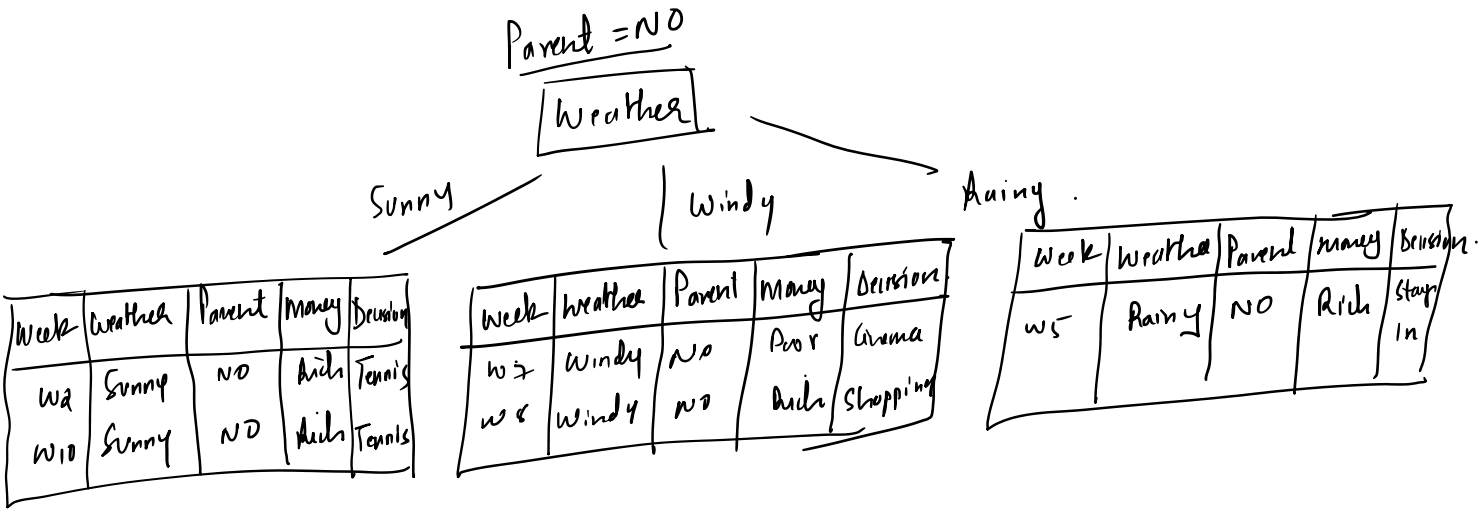
Rich

$= 0.5$

∴ Gini Index (Parent=NO & Weather) = 0.2
Gini Index (Parent=NO & money) = 0.5.

Here Weather Attribute has Smaller Gini Index

∴ Select Weather as Node.

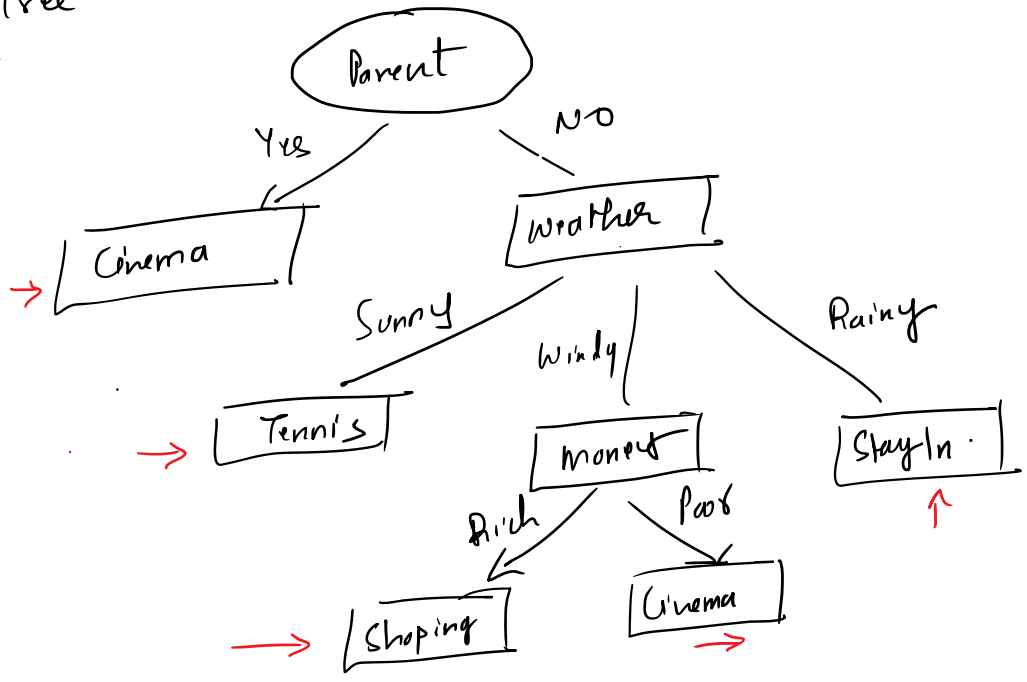


Week	Weather	Parent	Money	Decision
W2	Sunny	NO	Rich	Tennis
W10	Sunny	NO	Rich	Tennis

week	weather	Parent	Money	Decision
W3	Windy	NO	Poor	Cinema
W8	Windy	NO	Rich	Shopping

Week	Weather	Parent	money	Decision
W5	Rainy	NO	Rich	Stay In

Desired Decision Tree
For Given DataSet



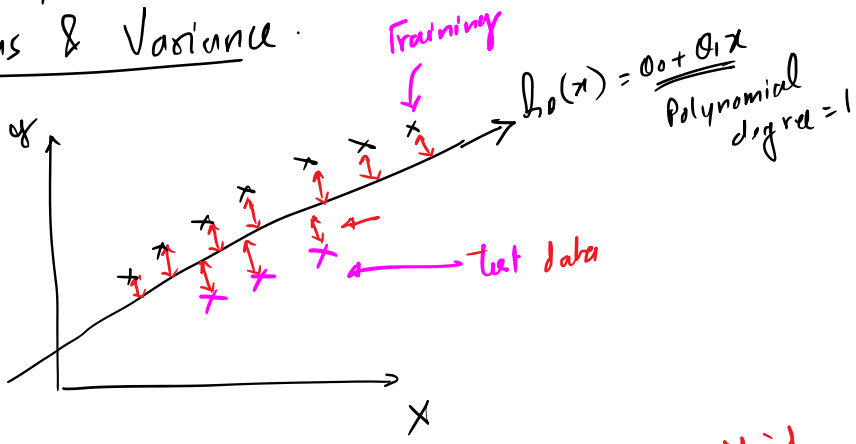
HW

Construct an optimal Decision Tree for following.

Outlook	Temperature	Humidity	Windy	Play (Decision)
Sunny	Hot	High	False	NO
Sunny	Hot	High	True	NO
Overcast	Hot	High	F	Y
Rainy	mild	High	F	Y
Rainy	Cool	Normal	F	Y
Rainy	cool	Normal	T	NO
overcast	cool	Normal	T	Y
Sunny	mild	High	F	NO
Sunny	cool	Normal	F	Y
Rainy	mild	Normal	F	Y
Sunny	mild	Normal	T	Y
Overcast	mild	High	T	Y
overcast	Hot	Normal	F	Y
Rainy	mild	High	T	NO

Q3a)

Bias & Variance.



Normally
- 70% Training
- 30% Test

Here line is of lower degree Polynomial

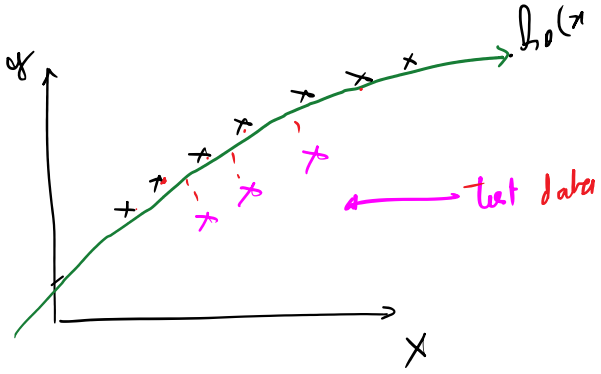
Underfitting ⇒ Whatever data we have trained the model the error is quite High.

We can have Higher Training Error ⇒ High Bias
& Higher Test Error ⇒ High Variance.

However It is Normally Found that Underfitting Model has

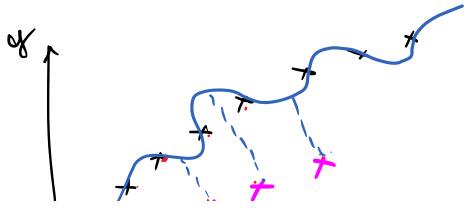
↑ Bias & ↓ Variance

In Comparison to Earlier Model degree of Polynomial 2.



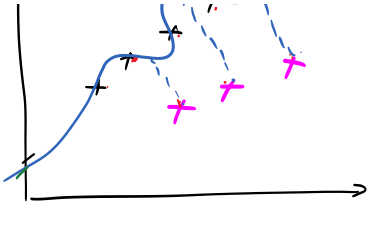
Here Training Error ↓ (Bias)

Test Error ↓ (variance) (as compared to overfit)



Higher degree Polynomial

Perfectly Fitting the Training Data



Training Data

Training Error ≈ 0 (Bias)

Test Error (Variance) Very High

Overfitting → Here for Training Each & Every point is satisfied by Regression line.

* Here training data perfectly fit the line But Test Data may show Error (High).

Note

Normally Degree of Polynomial low → Underfit → High Bias low Variance

If Degree of Polynomial is very High (Complex model) → Overfit → Low Bias High Variance.

Training Error ⇒ Bias → Represent Error when Regression line fits the Training Data.

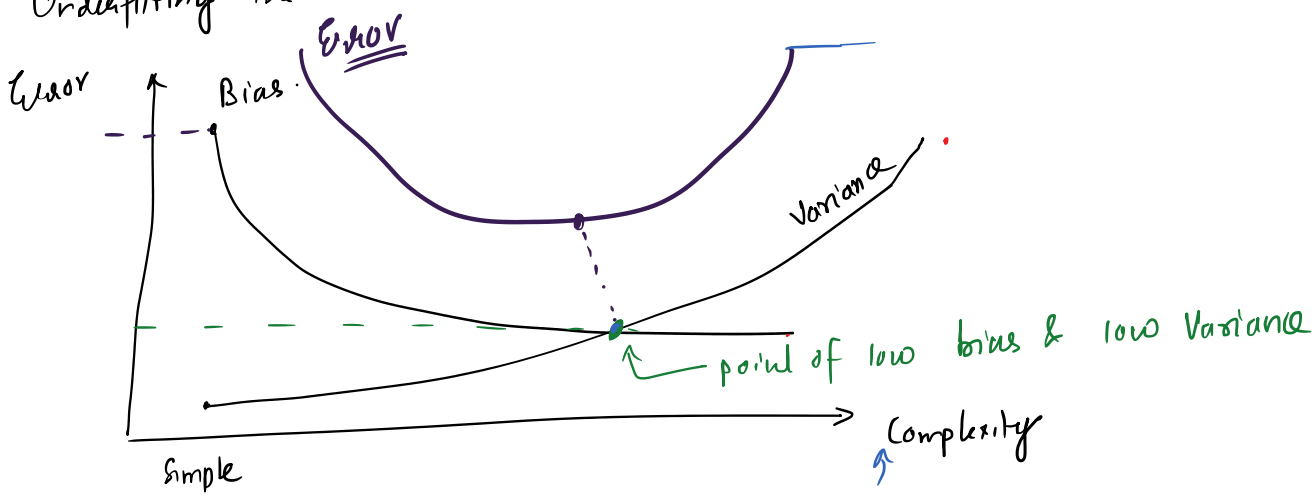
Test Error (Cross Validation) → Variance = Represents Error when Regression line fits the Test Data.

Bias - Variance Tradeoff

① If model is too simple and few parameters → Underfit → High Bias Low Variance

② If model has large no of Parameters and → Overfit
i.e complex → High Variance
→ low Bias

③ we need to find good balance without overfitting & Underfitting the data.



Bias Variance Tradeoff Says
we need a model that gives
① low Bias
② low Variance.

Ie we need to find point of low Bias & low Variance.

* The method are:
  * Regularization (Penalize ∝ (parameter)).
  * Boosting
  * Bagging

We need to Minimize the total Error = Bias² + Variance + Irreducible Error.

Note
Model 1 Training Error 1% ⇒ low Bias } Overfitting (complex model)

Now

Model 1 — Training Error 1% ⇒ low Bias; Test Error 20% ⇒ High Variance. } Overfitting (Complex model)

Model 2 — Training Error = 25% ⇒ High Bias; Test Error = 26% ⇒ High Variance } Underfitting

Model 3 — Training Error < 10%, low Bias; Test Error < 10%, low Variance } Recommended.

Q3b)

Kappa Statistic →

* Kappa statistic or cohen's kappa is statistical measure of inter-rater reliability for categorical variable.

* It is used when two raters apply a criteria based on a tool to assess whether or not some condition occur

Ex Two doctors rate whether or not each of 20 patients has diabetes based on symbons

* If two raters uses same criteria on same target to Evaluate and then their agreement is very high then we will have evidance of reliable rating

* If their agreement is not very high then
  → either Criterian tool is not useful
  → or raters are not trained enough.

* Kappa statistics corrects for chance agreement and not percent agreement.

		Raters A	
		Yes	NO
Rater B	Yes	35	20
	NO	15	40

→ 35 times they agree on Yes.
40 times they agree on NO
20 time Rater A is NO But Rater B is Yes
15 time Rater A is Yes But Rater B is NO.

Cohen Suggested following Statistics.

value ≤ 0 → No agreement

0.01 → 0.20 → as none to slight

0.21 → 0.40 → as fair

0.41 → 0.60 → as moderate

0.61 → 0.80 → as substantial

0.81 → 1.00 → perfect agreement.

* Rather than calculating the percentage of items raters agreed on, Cohens kappa attempts to account the fact that raters may happen to agree on some items purely by chance.

Ex Two Curators asked to rate 70 paintings

C1 \ C2	Yes	NO
Yes	25	10
NO	15	20

Step 1> Calculate Relative Agreement bet^n Raters

$$P_o = \frac{\text{Both said Yes + Both said NO}}{\text{Total}} = \frac{25+20}{70} = 0.6429$$

Step 2 > Calculate hypothetical probabilities of chance agreement between raters

Pe ??

$$P(\text{Yes}) = \frac{C_1(\text{Yes})}{\text{Total Ans}} * \frac{C_2(\text{Yes})}{\text{Total Ans}} = \left(\frac{25+10}{70}\right) * \left(\frac{25+15}{70}\right) = 0.285714$$

$P(No) = \frac{C_1(No)}{\text{Total Res}} * \frac{C_2(No)}{\text{Total Res}} = \left(\frac{15+20}{70}\right) * \left(\frac{20+10}{70}\right) = 0.214285$

$P_e = P(Yes) + P(No) = 0.5$

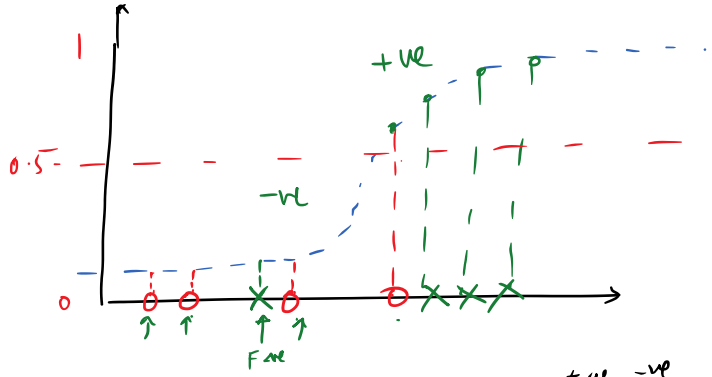
Calculate Cohen's Kappa $= K = \frac{P_o - P_e}{1 - P_e} = \frac{(0.6429 - 0.5)}{1 - 0.5} = 0.2857$

Its in range $0.21 \rightarrow 0.40$ so the Agreement bet^n two Curator is fair

Q3 c)

ROC & AUC →

ROC [Reciever Operator Characteristic]



for threshold 0.5

	+ve	-ve
+ve	3	1
-ve	1	3

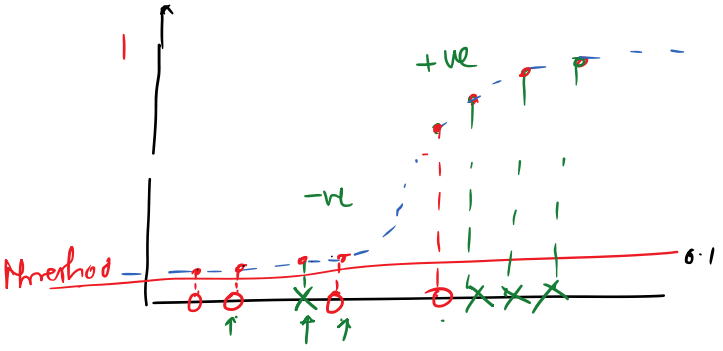
Sensitivity = $\frac{3}{4}$

Specificity = $\frac{3}{4}$

	+ve	-ve
+ve	4	4
-ve	0	0

TPR = Sensitivity = $\frac{4}{4} = 1$

Specificity = $\frac{0}{4} = 0$ FPR = 1 - 0 = 1



Here if Threshold 0.1 then the confusion matrix changes and according the Sensitivity and Specificity changes.

* Consider for logistic Regression where a identified a threshold point and we prepared confusion matrix and calculated Sensitivity & Specificity.

* If Threshold changes then the confusion matrix and accordingly the Sensitivity and Specificity changes.

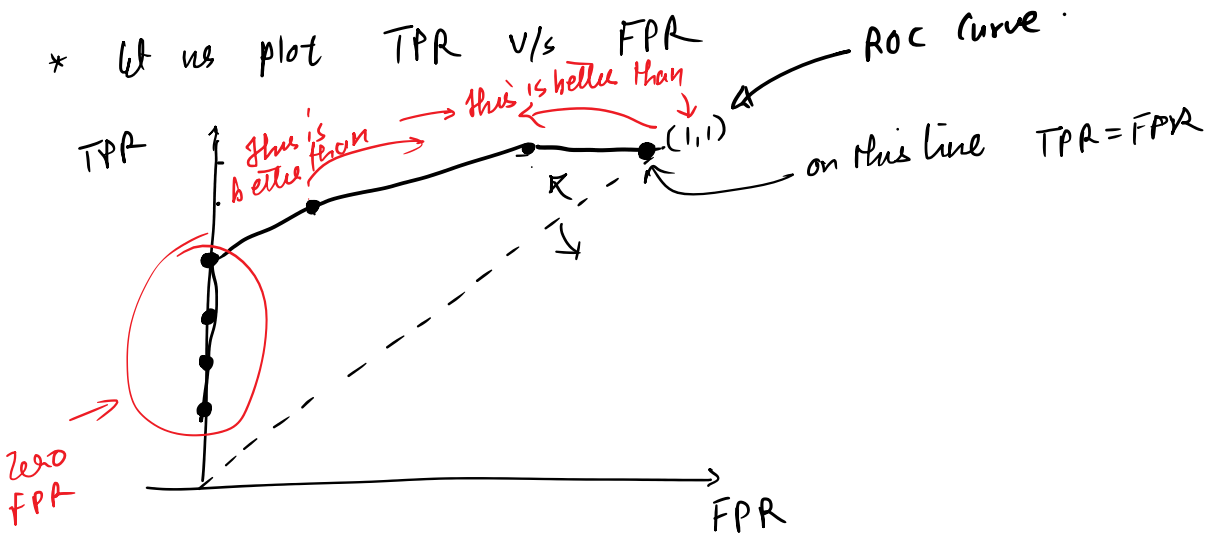
* We can have many such thresholds bet^n 0 → 1.

* We want to analyse the performance at different threshold and want to identify the best of it

* Instead of using so many confusion matrix we can use

ROC Curve

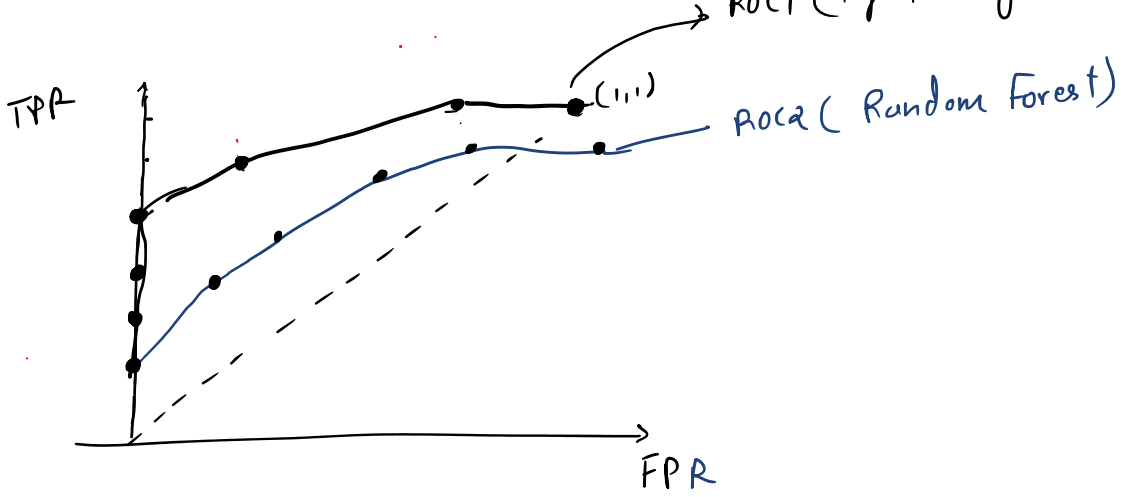
* Let us plot TPR v/s FPR



For diff threshold we calculated TPR and FPR at plotted accordingly & in form of ROC Curve.

→ From ROC Curve we can directly determine which threshold is better.

AUC → Area Under Curve → It is a method to compare ROC for more than one method and will help us to judge which one is better.



Here AUC for ROC1 > AUC for ROC2

So method for ROC1 (logistic Regression) is better than method for ROC2 (Random Forest)