3. (a) Suppose there is a rule such that,

$$X \rightarrow Y$$

Then the support of this rule is defined as the

$$\text{Support} = \frac{(X \cup Y).\text{Count}}{N}$$

where N is total number of samples i.e. it is the ratio of samples containing X ∪ Y divided by total No. of instances.

and the item set is referred to as frequent if the it's ~~min~~ support is greater than or equal to the minimum support ~~and~~.

An item is referred to as important rule if it is frequent and the confidence of item set is greater than or equal to the minimum confidence.

4

$$\text{Confidence} = \frac{(X \cup Y).\text{Count}}{X.\text{Count}}$$

b) Let I_1 = Bread

I_2 = Butter

I_3 = Milk

I_4 = Jelly

I_5 = Coke

Given:

T_1 : I_1 I_2

T_2 : I_1 I_3 I_2

T_3 : I_1 I_4 I_2

T_4 : I_1 I_5

T_5 : I_1 I_3

T_6 : I_3 I_5

Min. Confidence = 0.8

Min - support = 0.3

So, C_1 are

~~Confidence~~ support

I_1 : 5/6

I_2 : 3/6

I_3 : 3/6

I_4 : 1/6

I_5 : 2/6

So, F_1 are

I_1 I_2 I_3 I_5

So, C_2 are

$I_1\ I_2$ - 3/6

$I_1\ I_3$ - 2/6

$I_1\ I_5$ - 1/6

$I_2\ I_3$ - 1/6

$I_2\ I_5$ - 0/6

$I_3\ I_5$ - 1/6

So F_2 are

$I_1\ I_2$

$I_1\ I_3$

$5\frac{1}{2}$

So, C_3 are

$I_1\ I_2\ I_3$ - 1/6

So, F_3 is None :

Rules $x \to y$ possible are

X	Y	confidence
I_1	I_2	3/5
I_2	I_1	- 3/3
I_1	I_3	- 2/5
I_3	I_1	- 2/3

So, only Rule having confidence greater than 0.8 is

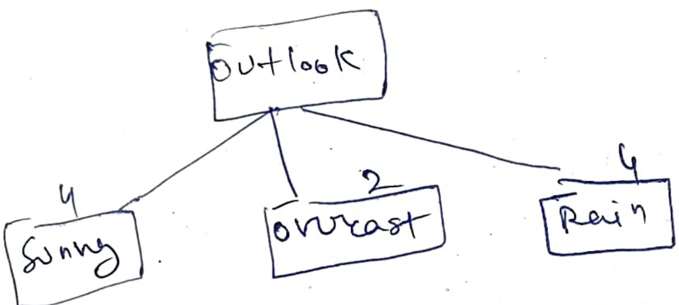
$\boxed{I_2 \to I_1}$

c) The major drawback of a-priori algorithm is also it's computational complexity. Sometimes the possible No. of cases may be very large and hence it is slower.

There is ~~always~~ some ambiguity in providing the result and hence it may lead to some errors as well. The result machine is not very good and hence may need improvement.

4.

a) If we split upon outlook



Entropy = 0.8

2 Y 2 N   2 Y 0 N   2 Y 2 N

So, entropy upon splitting on outlook is

Entropy is

For 2 Y 2 N

$-\frac{2}{4}\log\frac{2}{4} - \frac{2}{4}\log\frac{2}{4}$

$-\log\frac{2}{4} \Rightarrow -\log\frac{1}{2} = \log 2 = 1$

For 2 Y 0 N   entropy is

$-\frac{2}{2}\log\frac{2}{2} - \frac{0}{2}\log\frac{0}{2}$

$= 0$

Again for Rain Entropy is 1.

Again for Rain entropy is 1

So, weighted entropy is

$$\frac{4}{10}\times 1 + \frac{2}{10}\times 0 + \frac{4}{10}\times 1$$

$$\Rightarrow \frac{8}{10} = 0.8$$

3/2

Or we split upon Humidity

Humidity

$\boxed{\text{Normal}}$ $\boxed{\text{High}}$ = Entropy = 0.82

4 Yes 1 NO

$$-\frac{4}{5}\log\frac{4}{5} - \frac{1}{5}\log\frac{1}{5}$$

$$\Rightarrow -\frac{4}{5}(\log 4 - \log 5) + \log 5 \times \frac{1}{5}$$

$$\Rightarrow \frac{4}{5}\log(5 - \log 4) + \log 5 \times \frac{1}{5}$$

$$\frac{4}{5}(2.3 - 2) + 2.3 \times \frac{1}{5}$$

$$\frac{4}{5}\times 0.3 + \frac{2.3}{5}$$

$$\Rightarrow \frac{1.2}{5} + \frac{2.3}{5} = \frac{3.5}{5} = 0.7$$

2 Yes 3 NO

$$-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}$$

$$\Rightarrow \frac{2}{3}(\log 3 - \log 2) + \frac{1}{3}\log 3$$

$$\Rightarrow \frac{2}{3}(0.6) + \frac{1}{3}\times 1.6$$

$$\Rightarrow \frac{1.2}{3} + \frac{1.6}{3} = \frac{2.8}{3}$$

$$= 0.93$$

$$= -\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5}$$

$$\frac{2}{5}(1.3) + \frac{3}{5}(0.7)$$

$$\Rightarrow \frac{2.6}{5} + \frac{2.1}{5} \Rightarrow \frac{4.7}{5}$$

$$= 0.94$$

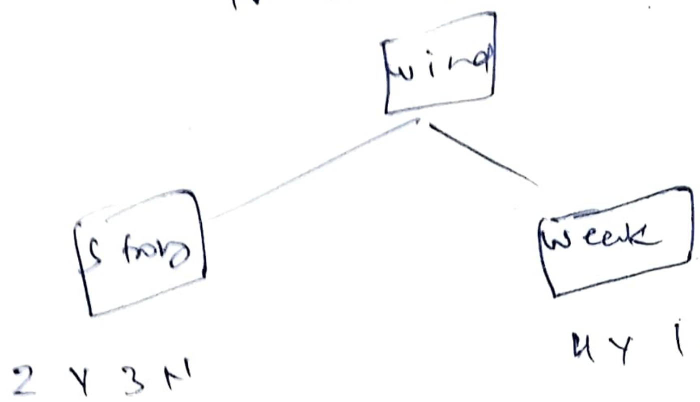
So, weighted

So, Entropy is $\frac{1}{2}\times 0.7 + \frac{1}{2}\times 0.94$

$$\Rightarrow 0.35 + 0.47$$

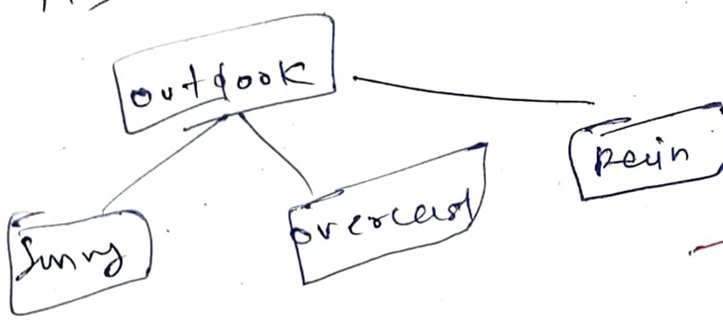
$$\Rightarrow 0.82$$

If we split on wind



It will be same as Humidity
because both are splitting equally
So, it's entropy is 0.82

So, it's root would be outlook



b) The possible terminating criterion of a decision tree algorithm is

i) If all the values in the Node have same labels then we should terminate that

ii) If the entropy gain upon splitting the node is less than a ~~threshold~~ threshold value then also no splitting

1½

iii) If upon splitting the depth of tree will become greater than a threshold depth then also we will not split.

c) The causes ~~a~~ model overfitting can be beacause the data given had too many outliers and it was not properly segmented.
overfitting corresponds to when a decison tree work very well on training data but worst on the test data.
It can be solved by mostly 2 ways.

i) pre-pruning
ii) post-pruning

3

i) pre-pruning : The pre-pruning a decision ~~tree~~ tree means is to set ~~a~~ some sort of rules so that you further do not split the node and terminate that. For Example By defining a minimum information gain or the max depth of the ~~tree~~ tree you can terminate a node.

ii) post-pruning : you First develop a decison tree and prune the splits or ~~ad~~ nodes whose split does not seem ~~a~~ useful. In this way you save the decision tree from overfitting and hence it provide good result.

~~80~~

6.

Outlook	Humidity	Wind	Play Tennis (class variable)	Play Tennis (predicted)
Sunny	Normal	Strong	Yes	Yes
Overcast	Normals	Strong	No	Yes
Rain	High	Strong	Yes	No
Sunny	High	Weak	No	No
Rain	High	Strong	No	No

4

The confussion Matrix correspond to

TP	FN
FP	TN

1	1
1	2

i) precision $= \dfrac{TP}{TP + FP} = \dfrac{1}{1+1} = \dfrac{1}{3} = 0.5$

ii) Recall $\dfrac{TP}{TP+FN} = \dfrac{1}{1+1} = \dfrac{1}{2} = 0.5$

iii) F-score is Harmonic mean of precision and Recall

$= 2 \times \dfrac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

$= \dfrac{1}{2}$

$\dfrac{2 \times \frac{1}{2} \times \frac{1}{2}}{\frac{1}{2} + \frac{1}{2}} \Rightarrow 0.5$

$\dfrac{2/6}{5/6} = 0.4$

b) i) Holdout method :-

The Hold out method used for estimating classification model is that we hold the classifier ~~the~~ model so that it gives some correct output

(ii) Cross-Validation :- In the cross-validation method the dataset is divided into k sub sets so that (k-1) datasets are used for training and 1 is used for testing

(iii) Bootstrap : In the bootstrap technique all the ~~no~~ classifiers are boostrapped so that it gives correct output upon predicting the data

1/2

c) The purpose of ensembles of classifiers is that the majority voting. It leads to avoid the overfitting of the classification model. So like instead of predicting on 1 strong classifier, we have more than one classifier so that we have many weak models and we predict ~~the~~ result on the basis of the majority voting.

The Ada boost algorithm works like this when we are training the data set on the classification algorithm, if ~~a~~ on ~~training data~~ some instance of training data if it gives correct output then, we multiply the weight of that instance by $\frac{e}{1-e}$ given e is the error rate of classifier so that it's weight decreases and ~~it~~ next time it focus more on instances where it predicts wrong. 2

on testing if some classifior predicts some class then the weight of that class is increased by $-\log\left(\frac{e}{1-e}\right)$ . .

In the end the model will predict the class with the highest weight.

$2\frac{1}{2}$

2. b) Intercluster distance: The inter cluster distance is defined as the distance between elements or data points of the different clusters. Minimum Inter cluster distance is defined as the minimum distance between two points both belonging to different clusters.

$$\text{Min- intercluster distance} = \text{Min} \sum_{i=1}^{N} \sum_{j=1}^{N} \delta(x_i x_j)$$

3

where x_i belongs to cluster 1 and x_j belongs to cluster 2.

where $\delta(x_i x_j)$ is the distance between point x_i and x_j

Also, the maximum inter cluster distance is defined as the maximum distance between two data points both belonging to different cluster so,

$$\text{Max - intercluster distance} = \text{max} \sum_{i=}^{N} \sum_{j=1}^{m} \delta(x_i x_j)$$

$\Delta (X_k)$ MAX

Intra Cluster distance corresponds to the distance between the points belonging to the same cluster.

Maximum intra- cluster distance is the distance maximum distance between points belonging to the same cluster.

$$\text{Min} \quad i \neq j = \frac{N \times (N-1)}{2} \quad \delta(x_i, x_j)$$

The avrage intra cluster distance is the avrage distance between the points belonging to the same cluster.

$$(\Delta)_{avg} \quad \frac{\sum_{i,j = \frac{N \times (N-1)}{2}} \delta(x_i, x_j)}{\frac{N \times (N-1)}{2}}$$

c) The Dunn's cluster validation index is defined as

$$\text{Min} \; \min_{i=}^{N} \; \min_{j=1}^{M} \frac{\delta(x_i x_j)}{\text{MAX}_{k=1}^{N} \Delta(x_k)}$$

2

With the help of dunn's cluster validation index we try to Maximise the inter cluster distance and Minimize the intra-cluster distance.

The maximum is the dunn's index the good the cluster result is.

In this way it helps us in evaluating the cluster.

2. a)

(i) Purity :- Purity corresponds to validate the cluster index in a different way so that the it measures the purity of clusters. The more distant the different clusters are the more greater it is

(ii) Rand index :- Rand index corresponds to the random cluster formation by the clustering algorithms so that when the clusters are evaluated we get the best out of them.