Answer - 3 (a)

An association rule is in the form of implication $X \rightarrow Y$ such that $X, Y \in I$ and $X \cap Y = \phi$. Here, I is a set of items involved in the transactions.

4

Support

Support of an itemset X in a transaction database T is the number of transactions that contain X divided by the total no. of transactions in T.

$$\text{support}(X) = \frac{X.\text{count}}{n}$$

where, $X.\text{count}$ = no. of transactions that contain X

, n = Total no. of transactions in T

Confidence

Confidence of an association rule $X \rightarrow Y$ in transaction database T is defined as the

$$\text{conf}(X \rightarrow Y) = \frac{P(X \cup Y)}{P(X)}$$

$$\boxed{\text{conf}(X \rightarrow Y) = \frac{(X \cup Y).\text{count}}{X.\text{count}}}$$

It is used to measure the importance of an association rule.

An association rule $X \rightarrow Y$ describes that whenever X occurs in T, then Y also occurs in T.

An item set is referred to as an frequent item set if support of item set is greater than or equal to given minimum support.

An association rule is reffered to as an important rule if confidence of rule is greater than or equal to given minimum confidence.

Answer-3 (b)

Given

minimum support, minsup = 30% = 0.3
minimum confidence, minConf = 80% = 0.8

Sol^n

Let items are denoted as given below:-

I_1 : Bread
I_2 : Butter
I_3 : Milk
I_4 : Jelly
I_5 : Coke

Transactions (T)	Items (I)
T1	$\{I_1, I_2\}$
T2	$\{I_1, I_3, I_2\}$
T3	$\{I_1, I_2, I_4\}$
T4	$\{I_1, I_5\}$
T5	$\{I_1, I_3\}$
T6	$\{I_3, I_5\}$

Using Apriori Algorithm

C_1: $\{I_1\}, \{I_2\}, \{I_3\}, \{I_4\}, \{I_5\}$
(candidate items with size 1)

Support $\{I_1\} = \frac{5}{6} = 0.83 \geq 0.3$ (minsup)

$\{I_2\} = \frac{3}{6} = 0.5 \geq 0.3$

$\{I_3\} = \frac{3}{6} = 0.5 \geq 0.3$

$\{I_4\} = 1/6 = 0.16 < 0.3$

$\{I_5\} = \frac{2}{6} = \frac{1}{3} = 0.33 \geq 0.3$

Since, support $(\{I_4\}) < 0.3$, it will not be included while forming F_1 (frequent item set).

F_1: $\{I_1\}, \{I_2\}, \{I_3\}, \{I_5\}$

C_2: $\{I_1, I_2\}, \{I_1, I_3\}, \{I_1, I_5\}$,
$\{I_2, I_3\}, \{I_2, I_5\}, \{I_3, I_5\}$

(candidate item sets with size 2)

Support $\{I_1, I_2\} = \frac{3}{6} = 0.5$

$\{I_1, I_3\} = \frac{2}{6} = 0.33$

$\{I_1, I_5\} = \frac{1}{6} = 0.16 < 0.3$ (minsup)

$\{I_2, I_3\} = \frac{1}{6} = 0.16 < 0.3$

$\{I_2, I_5\} = \frac{0}{6} = 0 < 0.3$

$\{I_3, I_5\} = \frac{1}{6} = 0.16 < 0.3$

F_2 : $\{I_1, I_2\}$, $\{I_1, I_3\}$

C_3 : $\{I_1, I_2, I_3\}$
(candidate item set with size 3)

~~F_3~~ Support of $\{I_1, I_2, I_3\} = \frac{1}{6} = 0.16 < 0.3$

F_3 : It becomes null. So stop generating item sets with bigger size.

Frequent Item Set = $\{I_1, I_2, I_3\}$

Possible association rules are :-

X	Y	Confidence
{I_1}	{I_2, I_3}	$1/5 = 0.2 < 0.8$ (minConf)
{I_2}	{I_1, I_3}	$1/3 = 0.33 < 0.8$
{I_3}	{I_1, I_2}	$1/3 = 0.33 < 0.8$
~~{I_1, I_2}~~	{I_3}	$1/3 = 0.33 < 0.8$
{I_1, I_3}	{I_2}	$1/2 = 0.5 < 0.8$
{I_2, I_3}	{I_1}	$1/1 = 1 \geq 0.8$

Since, for association rule {I_2, I_3} $\rightarrow$ {I_1} confidence is greater than minConf (0.8), hence it is an important association rule.

Important association rule {I_2, I_3} $\rightarrow$ {I_1}

Answer - 3 (C)

Apriori algorithm assumes that all items in the data are of the same nature and have similar frequencies which is not true practically.

For Example :- Oven and Cooker are rarely bought while food items like butter, bread etc. are frequently bought from a store.

Answer - 4 (a)

To identify the root of the decision tree, we have to find information gain for each attribute in the dataset.

Info of a dataset D with n class variables is defined as,

$$\text{Info}(D) = -\sum_{i=1}^{n} p_i \log(p_i)$$

where p_i is the probability of a tuple (example) being in i^{th} class

If an attribute A divides dataset D into j subsets $\{D_1, D_2, \ldots, D_j\}$, then

$$\text{Info}_A(D) = \sum_{j=1}^{j} \frac{|D_i|}{|D|} \text{Info}(D_i)$$

$$\text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D)$$

In the given dataset, there are 3 attributes, Outlook, Humidity, Wind and 2 class variables Yes and No.

Total examples = 10
examples with class 'Yes' = 6
examples with class 'No' = 4

$$\text{Info}(D) = I(6,4) = -\frac{6}{10}\log\left(\frac{6}{10}\right) - \frac{4}{10}\log\left(\frac{4}{10}\right)$$

$$= -0.6 \times (-0.7) - 0.4 \times (-1.3)$$

$$= 0.42 + 0.52$$

$$= 0.94$$

$$\text{Info}_{outlook}(D) = \underset{(\text{Sunny})}{\frac{4}{10} I(2,2)} + \underset{(\text{Overcast})}{\frac{2}{10} I(2,0)} + \underset{(\text{Rain})}{\frac{4}{10} I(2,2)}$$

$$\text{Info}_{Humidity}(D) = \underset{(\text{Normal})}{\frac{5}{10} I(4,1)} + \underset{(\text{High})}{\frac{5}{10} I(2,3)}$$

$$\text{Info}_{wind}(D) = \underset{(\text{Strong})}{\frac{5}{10} \cdot I(2,3)} + \underset{(\text{Weak})}{\frac{5}{10} I(4,1)}$$

Now,

$$\begin{aligned} I(2,2) &= -\frac{2}{4}\log\left(\frac{2}{4}\right) - \frac{2}{4}\log\left(\frac{2}{4}\right) \\ &= -0.5 \times (-1) - 0.5 \times (-1) \\ &= 0.5 + 0.5 \\ &= 1 \end{aligned}$$

$$\begin{aligned} I(2,0) &= -\frac{2}{2}\log\left(\frac{2}{2}\right) - \frac{0}{2}\log\left(\frac{0}{2}\right) \\ &= -1 \times 0 - 0 \\ &= 0 \end{aligned}$$

$$\begin{aligned} I(4,1) &= -\frac{4}{5}\log\left(\frac{4}{5}\right) - \frac{1}{5}\log\left(\frac{1}{5}\right) \\ &= -0.8 \times (-0.3) - 0.2 \times (-2.3) \\ &= 0.24 + 0.46 \\ &= 0.70 \end{aligned}$$

$$\begin{aligned} I(2,3) &= -\frac{2}{5}\log\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\frac{3}{5}\right) \\ &= -0.4 \times (-1.3) - 0.6 \times (-0.7) \\ &= 0.52 + 0.42 \\ &= 0.94 \end{aligned}$$

∴

$$\text{Info}_{outlook}(D) = 0.4 \times 1 + 0.2 \times 0 + 0.4 \times 1$$
$$= 0.4 + 0.4$$
$$= 0.8$$

$$\text{Gain}(\text{Outlook}) = \text{Info}(D) - \text{Info}_{outlook}(D)$$
$$= 0.94 - 0.8$$
$$= 0.14 \quad \text{---} \quad (i)$$

$$\text{Info}_{Humidity}(D) = 0.5 \times 0.7 + 0.5 \times 0.94$$
$$= 0.35 + 0.470$$
$$= 0.35 + 0.47$$
$$= 0.82$$

$5\frac{1}{2}$

$$\text{Gain}(\text{Humidity}) = 0.94 - 0.82$$
$$= 0.12 \quad \text{---} \quad (ii)$$

$$\text{Info}_{Wind}(D) = 0.5 \times 0.94 + 0.5 \times 0.7$$
$$= 0.82$$

$$\cancel{\text{Info}}\ \text{Gain}(\text{Wind}) = 0.94 - 0.82$$
$$= 0.12 \quad \text{---} \quad (iii)$$

Since, we find that information gained by attribute 'Outlook' is maximum, so selected attribute for the root of the tree will be 'Outlook'.

So, root of the decision tree is 'Outlook'.

Answer - 4 (b)

Terminate the decision tree algorithm, if

① No attributes are left for further partitioning of nodes.

② All training examples have been exhausted.

③ A subset of examples are being classified by the same leaf node.

2½

Answer - 4 (C)

A model is overfitting if it performs well on the training dataset but its accuracy falls while classifying examples from test dataset.

Causes of model overfitting :-

① Insufficient generalization of the training data due to absence of data from all possible domains of the problem

② Noisy data present in the data causes overfitting.

③ Hyperparameters are not correct.

2½

Solving problem of overfitting in Decision Tree:-

① By early stopping the construction of tree

② By using pruning techniques

(i) Do not partition a node if it causes accuracy to fall below given threshold.

(ii) Allow the tree to grow fully and then prune branches that are not relevant

Answer - 5 (a)

Naive Bayes classifier is a statistical classifier which is based on the Bayes Theorem.

In this classifier it is assumed that the attributes are conditionally independent.

Given ~~class lables C_i and an~~ example X, we have to predict the class lable C_i for example X.

Posteriori probability using Bayes theorem is given by,

$$P(C_i / X) = \frac{P(X / C_i) \cdot P(C_i)}{P(X)}$$

In naive Bayes classifier, it is reduced to

$$P(C_i | X) = \underline{P(X | C_i) \cdot P(C_i)}$$

, since it is assumed that attributes are independent.

C_i Play Tennis = 'Yes'
Play Tennis = 'No'

$P(\text{PlayTennis} = \text{'Yes'}) = \frac{6}{10} = 0.6$

$P(\text{PlayTennis} = \text{'No'}) = \frac{4}{10} = 0.4$

$P(X_K | C_i)$ = Probability of kth attribute given class lable

$P(\text{Outlook} = \text{'overcast'} / \text{PlayTennis} = \text{'Yes'}) = \frac{2}{6} = 0.33$

$P(\text{Outlook} = \text{'overcast'} | \text{PlayTennis} = \text{'No'}) = \frac{0}{4} = 0$

$P(\text{Humidity} = \text{'High'} \mid \text{PlayTennis} = \text{'Yes'}) = \frac{2}{6} = 0.33$

$P(\text{Humidity} = \text{'High'} \mid \text{PlayTennis} = \text{'No'}) = \frac{3}{4} = 0.75$

$P(\text{Wind} = \text{'Weak'} \mid \text{PlayTennis} = \text{'Yes'}) = \frac{4}{6} = \frac{2}{3} = 0.66$

$P(\text{Wind} = \text{'Weak'} \mid \text{PlayTennis} = \text{'No'}) = \frac{1}{4} = 0.25$

Let X = <u>Given Sample</u>

5 $P(X \mid \text{PlayTennis} = \text{'Yes'}) = 0.33 \times 0.33 \times 0.66$

$= 0.072$

$P(X \mid \text{PlayTennis} = \text{'No'}) = 0 \times 0.75 \times 0.25$

$= 0$

Now,

$P(\text{PlayTennis} = \text{'Yes'} \mid X) = 0.072 \times 0.6$

$= 0.0432$

$P(\text{PlayTennis} = \text{'No'} \mid X) = 0 \times 0.4$

$= 0$

Since $P(\text{PlayTennis} = \text{'Yes'} \mid X) > P(\text{PlayTennis} = \text{'No'} \mid X)$

So, <u>class label = 'Yes'</u>

Answer - 5 (b)

Yes, there is an error in such prediction. There is no PlayTennis = 'No' label for attribute 'Overcast' which causes overall probability to be zero for class PlayTennis = 'No'.

2

This is a drawback of naive Bayes classification algorithm.

We can overcome this error by introducing additional examples in the dataset such that there is at least one class lable for each type for each attribute in the dataset.

Answer - C (5)

If any feature has the continuous values, such feature's values can be descretized using any descretization algorithm like Binning and then apply algorithm to predict the class lable.

0

Answer - 6 (a)

Outlook	Humidity	Wind	Actual Class Lable (C)	Predicted label by model
Sunny	Normal	Strong	Yes	Yes
Overcast	Normal	Strong	No	Yes
Rain	High	Strong	Yes	No
Sunny	High	Weak	No	No
Rain	High	Strong	No	No

4

↓ Actual label / Predicted label →	C	¬C
C	True Positive (TP) = 1	False Negative (FN) = 1
¬C	False Positive (FP) = 1	True Negative (TN) = 2

Confusion Matrix for the model

(i) Precision $= \frac{TP}{TP+FP} = \frac{1}{1+1} = \frac{1}{2} = 0.5 = 50\%$

(ii) Recall $= \frac{TP}{TP+FN} = \frac{1}{1+1} = \frac{1}{2} = 0.5 = 50\%$

(iii) F-Score $= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

$= \frac{2 \times \frac{1}{2} \times \frac{1}{2}}{\frac{1}{2}+\frac{1}{2}} = \frac{\frac{1}{2}}{1} = \frac{1}{2} = 0.5 = 50\%$

Answer- 6(b)

Holdout method

In this method, ~~training~~ available data is partitioned into two mutually exclusive subsets randomly. One set is called training dataset which usually comprises of $\frac{2}{3}$ of total data. Another set is called testing dataset which comprises of $\frac{1}{3}$ of total data.

Cross Validation method

4

In this method, dataset 'D' is randomly split into k mutually exclusive subsets $(D_1, D_2, \ldots, D_k)$. Now, training is done iteratively. In i^{th} iteration D_i is used as testing data and all other subsets are used as training data.

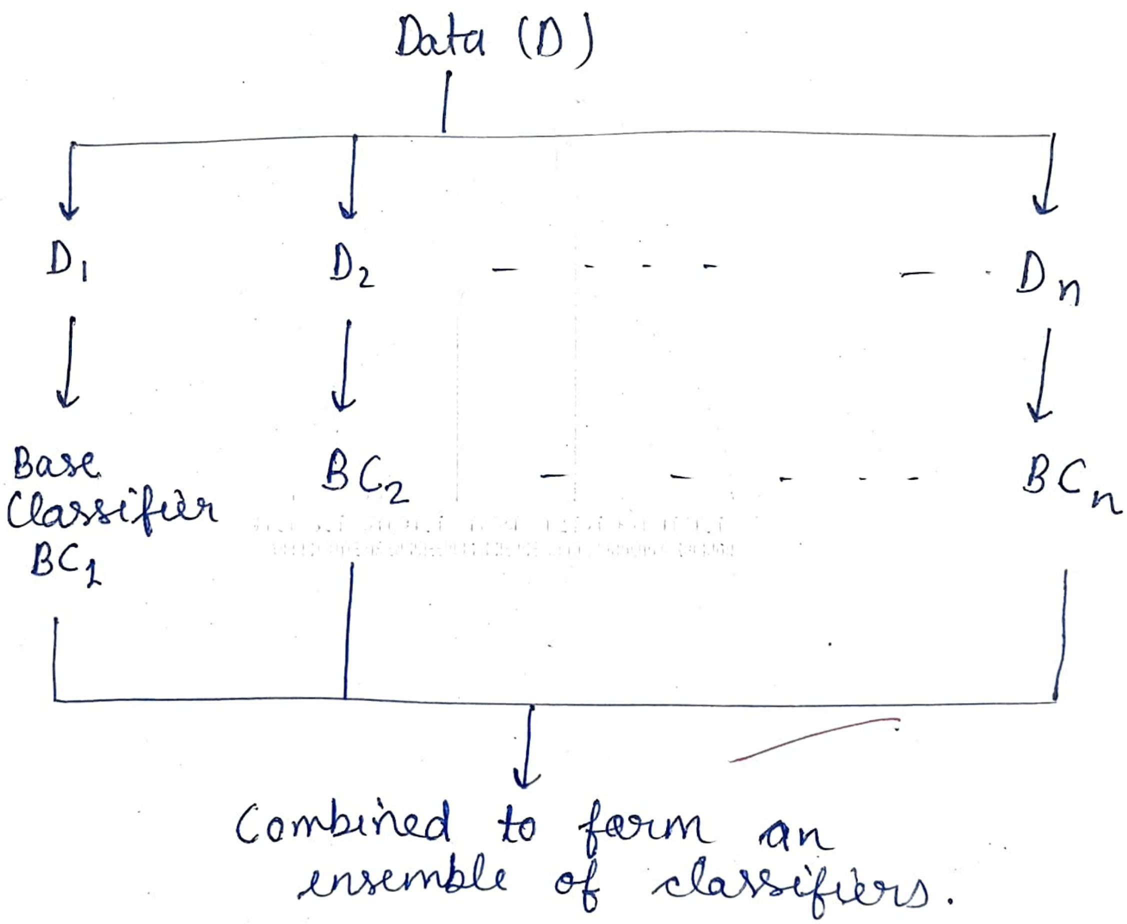
Bootstrap

In this method, a set of 'd' tuples is sampled randomly 'd' times to form a training instance of size 'd'. It is done with replacement.
It is seen that about 0.632 part of original data goes in training set and remaining data is used for testing purpose.

Answer- 6 (C)

An ensemble of classifiers is a combination of many base classifiers learned with different algorithm.

The purpose of ensemble method is to increase the overall accuracy of classification model.



Aim is to learn base classifiers such that they misclassify different types of examples instead of increasing accuracy of individual base classifier.