Quiz 10

1) a) ~~Ko~~ Predictive accuracy is a method of evaluating classification methods / techniques.

It is a simple formula of the form

$$\text{Predictive accuracy} = \frac{\text{No of correct classification}}{\text{Total no of Test data / cases}}$$

b) The key to building a decision tree is to find and select which attribute branch in the tree.

c) Evaluation method is any method used to evaluate classification techniques / methods.

d) Machine Learning is the process of training a computer using data in the place of past experiences in order to allow the machine to predict the outcome / make decision in the future without human interaction.

e) True. For computers data is the only way ~~we~~ we can give it past experiences.

f) The ~~data~~ data is a set of data records, which are used to capture past experiences/instances/cases/outcomes in a application domain. Example would be students and their past grades.

A goal is ~~to~~ to train/learn a classification model using the data in order to enable the system to predict outcomes for new data. Example grade prediction of students ~~a cours~~ in a cours.

g) Data records are labelled with a specific ~~class~~ predefined class. Test data for training are labelled with these ~~classes~~ class labels.

h) A heuristic alg. aims to find an approximate/fast solution to a problem insted of finding the most optimal solution to a problem.

i) True all most tree building algorithms are heuristic algorithms because finding the best optimal decision is a NP-hard problem and heuristics are the only way of finding the closest approximation to the best tree in a reasonable runtime.

---

j) Yes a ~~gre~~ decision tree is a greedy divide and conquer because at each step we find the best possible decision in each branch to find the optimal solution, thus making it a greedy divide and ~~con~~ conquer algorithm.

---

k) Deterministic is when we can ~~gi~~ assign a definitive class to the output like Yes/No in our classification model.

Probabilistic models output the chances ~~chances~~ of something happening for each test instance instead of something ~~positive~~ boolean in nature.

2) for ~~1~~ 4 fold cross validation we pick ~~one~~ 1 fold as our Test data and 3 folds for our Training data from the entire test data.

After one iteration, we select another 1 fold as Test data and the other 3 folds as training data and train our model again.

At the end of all iterations, cross validation is complete.

---

3) Naive Bayes classification is a classification technique based on bayes theorem of conditional probability where we assume that the predictors involved in the model are independent of each other.

The given model

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

For exampl lets consider below table in C

$$P(A=m \mid C=t) = \frac{P(C=t \mid A=m)\ 3/10}{1/2}$$

$$P(A=m \mid C=t) = 2/25$$

---

3)

C) i) $A=h$ $B=s$ $C=?$

for $C=t$ we have,

$Pr(A=h \mid C=t) = \cancel{\ }\ 1/5$

$Pr(B=s \mid C=t) = \cancel{3/5}\ 2/5$

$Pr(C=t) = 1/2$

for test instance ~~Pr(A=h~~ $Pr(C=t) = 1/5 \times 2/5 \times 1/2$

$$= \frac{1}{25}$$

for $C=f$ we have.

$Pr(A=h \mid C=f) = 2/5$

$Pr(B=s \mid C=f) = 1/5$

$Pr(C=f) = 1/2$

for test instance

$$Pr(C=f) = \frac{2}{5} \times \frac{1}{5} \times \frac{1}{2}$$

$$= \frac{1}{25}$$

Both C=t and C=f are equal.

C can be either t or f. We can choose t

ii) for $A=h$ $B=b$

for $C=t$

$Pr(A=h \mid C=t) = \frac{1}{5}$

$Pr(\cancel{A=h}\ B=b \mid C=t) = \frac{1}{5}$

$Pr(C=t) = 1/2$

for test instance, $Pr(C=t) = \frac{1}{5} \times \frac{1}{5} \times \frac{1}{2} = \frac{1}{50}$

for $C=f$

$Pr(A=h \mid C=f) = \frac{2}{5}$

$Pr(B=b \mid C=f) = \frac{1}{5}$

$Pr(C=f) = 1/2$

for test instance, $Pr(C=f) = \frac{2}{5} \times \frac{1}{5} \times \frac{1}{2}$

$= \frac{1}{25}$

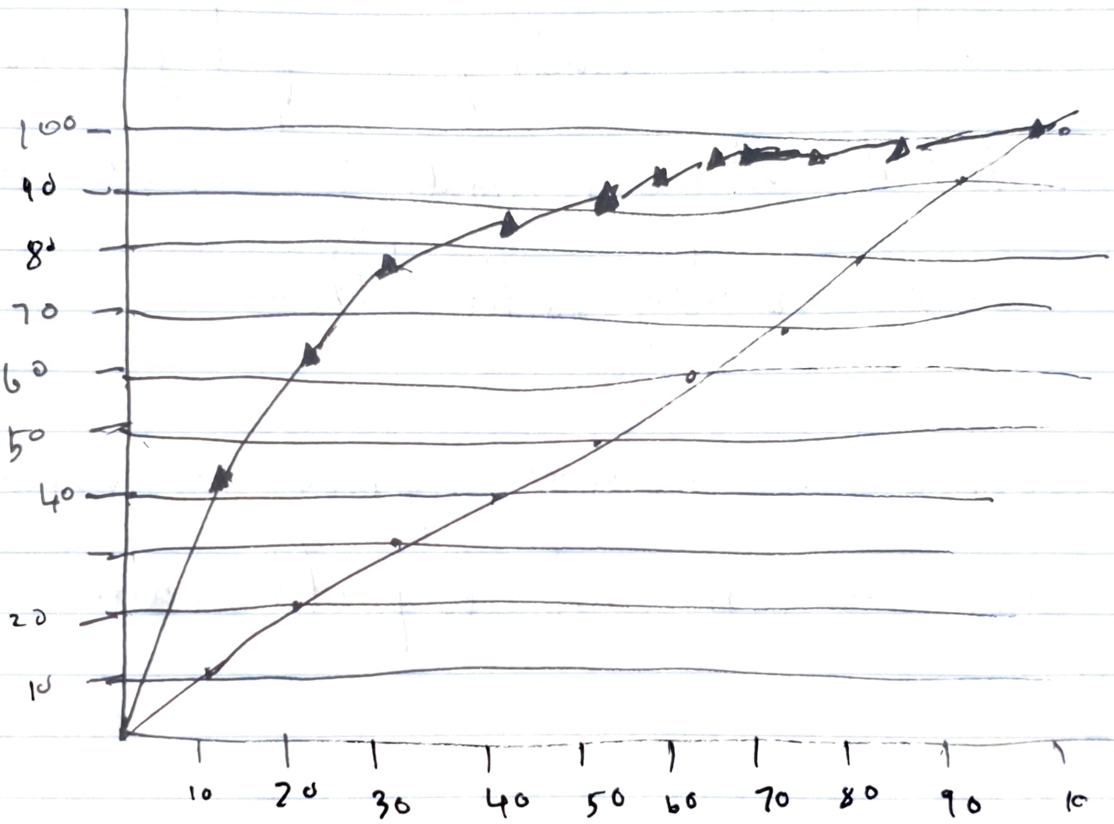
$\frac{1}{25}$ is larger. $\boxed{\text{Thus } C=f}$ is final class

5).

A)

Bin	1	2	3	4	5	1	7	8	9	10
~~% +ve sales~~ No of +ve instar	42%	24%	12%	8%	4.40%	3.6%	2.40%	1.40%	1.20%	1%
% of %ve instal	210	120	80	40	~~18~~ 22	18	12	7	6	5
Total posite %	42% ~~66%~~	67% ~~78%~~	78% ~~86%~~	86% ~~90%~~	90.4% ~~94%~~	94% ~~96%~~	96.4% ~~97.8~~	97.8%	99%	100%

Lift ~~curv~~ Curve



• - Baseline / Random

▲ - lift curve from model.

$\frac{96.40}{100} \times 1000$ 964

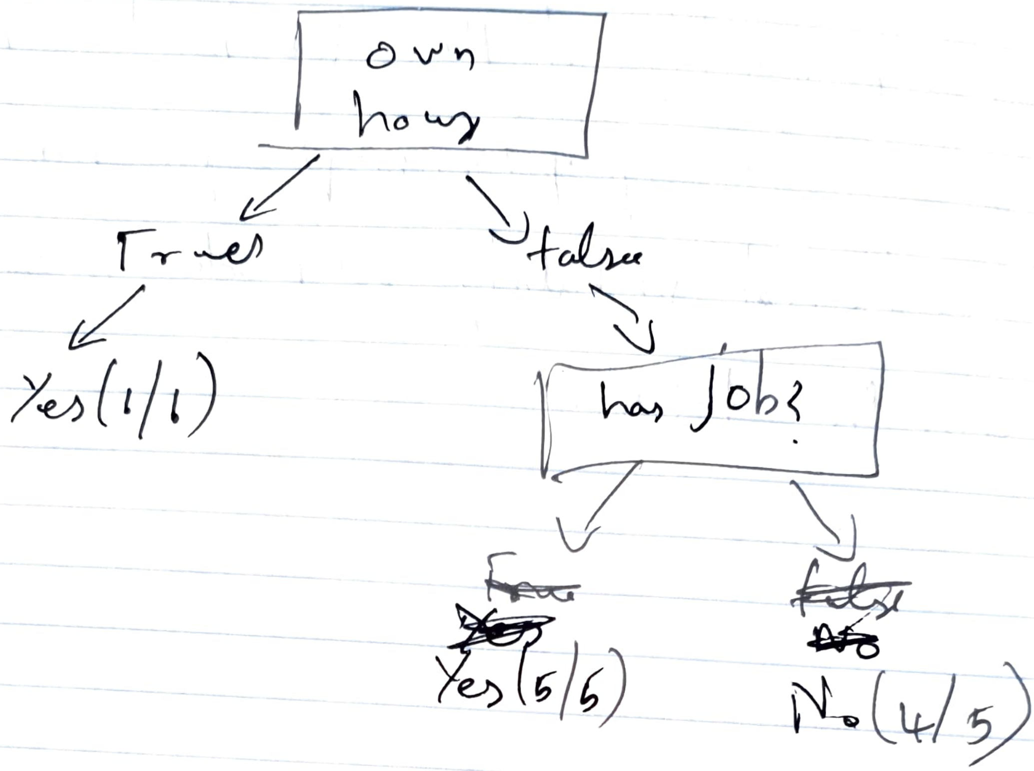
B) from lift curve analysis we can see that at bin 7 (~~60%~~ 70% of customers) ~~600~~ ~~700~~ 7000 customers, we reach 96.40% of all positive instances (sales) which is 964 out of 1000 total sales made.

Thus sending ~~around~~ packages to around ~~600~~ 7000 customers would get us ~~around~~ approx 1000 units sold. Sending 7000 packages to the top 70% of customers would get us the maximum gain according to the lift curve.

Above 7000 customers, gains are minimal and not worth the cost to profit ratio.

1)

A) No decision tree is not optimal.



This makes the tree shorter and more accurate.

c) Gini for has job node $= (0.4)\times(0.4) + (0.6)\times(0.6) = 0.376$

Gini for own house

Sub node $= (0.6\times0.6) + (0.4\times0.4) = 0.376$.

Gini for credit rating

Sub node $= (0.8\times0.8) + (0.2\times0.2) = \boxed{0.68}$

We should split on credit rating as it has highest gini rating.

B) Gini is used to ~~inde~~ identify the subnode to split the decision tree on. It calculates the probability of the two braches by using the formula $(p \wedge p) + (q \wedge q) =$ gini probability. The highest probability subnode is chosen to split the d-tree.