Quiz

Due Dec 6 at 10am **Points** 30 **Questions** 24 **Available** Dec 6 at 9:30am - Dec 6 at 10am 30 minutes
Time Limit 30 Minutes

Instructions

This is a short in-class quiz consisting of 24 questions with either 1 or 1.5 point for each question. Time duration to submit the quiz is 30 mins.

This quiz was locked Dec 6 at 10am.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	30 minutes	17.75 out of 30

Score for this quiz: **17.75** out of 30
Submitted Dec 6 at 10am
This attempt took 30 minutes.

Question 1 — 0 / 1.5 pts

Compute Gini Gain for the below diagram:

Quiz q5.JPG

- 0.0033
- **You Answered** — 0.25
- **Correct Answer** — 0.0075
- 0.1013

Question 2 — 0 / 1.5 pts

In the Below Tree Diagram, T1 Represent the No. of Responders for the Marketing Campaigns. If you are Business Analyst for executing the Campaigns going forward, which node will you suggest for targeting

Quiz q6.JPG

- Root Node
- **You Answered** — Node where Holding Period is greater than or equal to 10
- **Correct Answer** — Node where Holding Period is less than 10

All the Nodes

Question 3

1.5 / 1.5 pts

We have data of a Pilot Personal Loans Cross-Sell Campaign executed by MyBank. The bank reached out to 20000 customers. 1733 of those customers responded to the campaign. The customers who responded are tagged as Target = 1 (i.e. Responders) and the rest of the customers are marked as Target = 0 (i.e. Non-Responders).

We are now splitting this data set into Development Sample and Hold-out Sample. The Development Sample has 14000 observations and 1235 responders. The rest of the data is in Hold-Out sample. What is the Response Rate for Hold-out Sample data?

8.66%

8.82%

Correct! 8.30%

91.7%

Question 4

0 / 1 pts

Which of the following node(s) is held by the leaf node?

Branch Node

Root Node

You Answered Branch and Root nodes

Correct Answer None of these

Question 5

0.75 / 1.5 pts

In regression problem, the best decision tree has the minimum_____ of a target variable at leaf level (select ALL the applicable options)

Correct! Standard deviation

Correct! Variance

You Answered

- [x] Standard error
- [ ] Average

Question 6

1 / 1 pts

k-NN algorithm does more computation on test time rather than train time

Correct!

- (•) True
- ( ) False

Question 7

1.5 / 1.5 pts

In the image below, which would be the best value for k assuming that the algorithm you are using is k-Nearest Neighbour.

- ( ) 3

Correct!

- (•) 10
- ( ) 20
- ( ) 50

Question 8

1.5 / 1.5 pts

Which of the following is/are true about k-NN algorithm?

Correct!

- [x] It can be used in both classification and regression

Correct!

- [x] k-NN makes no assumptions about the functional form of the problem being solved
- [ ] k-NN struggles with a small number of input variables (p), but works well when the number of inputs is very large

All of the above

Question 9

1.5 / 1.5 pts

Following are the two statements given for k-NN algorthm, which of the statement(s) is/are true?

1. We can choose optimal value of k with the help of cross validation
2. Euclidean distance treats each feature as equally important

1

2

Correct!

1 and 2

None of the above

Question 10

1 / 1 pts

Which of the following will be true about k in k-NN in terms of Bias?

Correct!

When you increase the k the bias will be increased

When you decrease the k the bias will be increased

Can't say

None of these

Question 11

0 / 1 pts

Which of the following is true about k in k-NN in terms of variance?

You Answered

Variance increases with increase in k

Correct Answer

Variance increases with decrease in k

Can't say

None of these

Unanswered

Question 12

0 / 1.5 pts

1. k-NN is a memory-based approach is that the classifier immediately adapts as we collect new training data.
2. The computational complexity for classifying new samples grows linearly with the number of samples in the training dataset in the worst-case scenario.

Which of the above statements is/are true?

- 1
- 2

Correct Answer

- Both 1 and 2
- None

Question 13

1 / 1 pts

The Naïve Bayes algorithm assumes the independent variable/features are independent of each other.

Correct!

- True
- False

Question 14

0 / 1.5 pts

You are given a data set of 10,000 students with their sex, height, and hair color. You are trying to build a classifier to predict the sex of a student, so you randomly split the data into a training set and a testing set. Here are the specifications of the data set: (2 pts)

- sex ∈ {male,female}
- height ∈ [0,300] centimeters
- hair ∈ {brown, black, blond, red, green}
- 3240 men in the data set
- 6760 women in the data set

Under the assumptions necessary for Naive Bayes (not the distributional assumptions you might naturally or intuitively make about the dataset) answer the below question with T or F :

As height is a continuous valued variable, Naive Bayes is not appropriate since it cannot handle continuous valued variables.

You Answered

- True

Correct Answer

- False

Question 15

1 / 1 pts

Naïve Bayes classifier is a lazy learner.

- True

Correct!

- False

Question 16

0 / 1 pts

You are given a data set of 10,000 students with their sex, height, and hair color. You are trying to build a classifier to predict the sex of a student, so you randomly split the data into a training set and a testing set. Here are the specifications of the data set: (2 pts)

- sex ∈ {male,female}
- height ∈ [0,300] centimeters
- hair ∈ {brown, black, blond, red, green}
- 3240 men in the data set
- 6760 women in the data set

Under the assumptions necessary for Naive Bayes (not the distributional assumptions you might naturally or intuitively make about the dataset) answer below question with T or F :

Is P(height, hair|sex) = P(height|sex)P(hair|sex)

Correct Answer

- True

You Answered

- False

Question 17

0 / 1.5 pts

Which of the following statements about Naive Bayes is incorrect?

- Simple, fast in processing and effective.
- No idea for data sets with large number of numerical attributes

You Answered

- Easy to obtain estimated probability for a prediction

Correct Answer

- All the above

Question 18 1.5 / 1.5 pts

Which of the following statement(s) about Naive Bayes is incorrect?

- [ ] Attributes are equally important.

Correct!

- [x] Attributes are statistically dependent of one another given the class value
- [ ] Attributes are statistically independent of one another given the class value
- [ ] Attributes can be nominal or numeric

Question 19 1 / 1 pts

The word Naïve in Naïve refers to the assumptions and features in the data set are mutually not independent.

- ( ) True

Correct!

- (•) False

Question 20 1 / 1 pts

In real world, the independence assumptions if often violated, but naïve Bayes classifiers still tend to perform poor.

- ( ) True

Correct!

- (•) False

Question 21 1 / 1 pts

Which of the following machine learning algorithm can be used for imputing missing values of both categorical and continuous variables?

Correct!

- (•) k-NN

Linear regression

Logistic Regression

None of the above

Question 22

1 / 1 pts

_____ probability is assigned after a query point is passed into decision tree model

Conditional Probability

Probability of odds

Correct! Posterior Probability

Prior Probability

Question 23

1.5 / 1.5 pts

CART divides the dataset into —— homogenous nodes based on the best split

Correct! Two

Three

Four

Multiple

Question 24

0 / 1 pts

Decision trees splits the dataset vertically

You Answered True

Correct Answer False

Quiz Score: **17.75** out of 30