Quiz2

Due Mar 30 at 11:59pm **Points** 50 **Questions** 24
Available Mar 16 at 12am - Apr 6 at 11:59pm 22 days **Time Limit** None
Allowed Attempts Unlimited

Instructions

Dear students,

Answer all questions. You can consult other students. You can take your time until the due date. You have unlimited attempts.

All questions in this quiz are open book. You may work or discuss in groups.

Cheers,

:)

Jagan

Take the Quiz Again

Attempt History

	Attempt	Time	Score
KEPT	Attempt 2	27 minutes	38 out of 50 *
LATEST	Attempt 2	27 minutes	38 out of 50 *
	Attempt 1	71 minutes	29 out of 50 *

* Some questions not yet graded

ⓘ Correct answers will be available on Apr 9 at 12am.

Score for this attempt: **38** out of 50 *
Submitted Mar 24 at 7:13pm
This attempt took 27 minutes.

Question 1 1 / 1 pts

For non-linear relationships, correlations can give correct results.

- True
- False

Question 2

1 / 1 pts

Correlations are never distorted if the data is not standardized.

- True
- False

Question 3

1 / 1 pts

Linear Regression cannot not be applied on every dataset.

- True
- False

Question 4

1 / 1 pts

Higher level aggregations may have more variations than lower level aggregations.

- True
- False

Question 5

1 / 1 pts

Dimensionality reduction helps to eliminate irrelevant attributes or reduce possible noise.

- True
- False

Question 6

1 / 1 pts

Discretized values in a decision tree may be combined into a single branch if order is not preserved.

- True
- False

Question 7

1 / 1 pts

If a branch separates all records into a single class, then the purity is very low.

- True

- False

Question 8

1 / 1 pts

XOR function mappings can easily be classified by decision trees.

- True
- False

Question 9

1 / 1 pts

Jaccard coefficient ignores 00 combinations since it is meant to eliminate skewness when 00 combinations are common and irrelevant.

- True
- False

Question 10

1 / 1 pts

Bias toward selecting an attribute at a node of the decision tree may happen if the attribute has many branches.

- True
- False

Question 11

5 / 5 pts

Gini Index	1- (∑[P(j | t)]^2 for all j)
Interactions	Cannot classify properly
Dividing Gain by SplitINFO	Can overcome disadvant
Misclassification Error	1- (max(P(i | t)) for all i)
Underfitting	Model too simple

Question 12

2 / 2 pts

Decision trees use a _____ approach which often is unable to find the best tree.

greedy

Question 13

2 / 2 pts

A continuous attribute range may be split at the point where the GINI index values is _____.

least

Question 14

2 / 2 pts

The loss function for linear regression is the square of the difference between the original Y value and the _____ Y value.

predicted

Question 15

2 / 2 pts

_____ = GINI measure before splitting - GINI measure after splitting.

Gain

Incorrect

Question 16

0 / 2 pts

The process of _____ data before calculating correlations is the best way to get good correlations.

transforming

Question 17

2 / 2 pts

BoxPlots are centric to median

Answer 1:

median

Question 18

2 / 2 pts

Standardization transformation is centric to Mean

Answer 1:

Mean

Question 19

2 / 2 pts

The Mean of the transformed data after standardization becomes 0 :

Answer 1:

0

Question 20

2 / 2 pts

The standard deviation of the new transformed data after standardization is 1 :

Answer 1:

1

Question 21

2 / 2 pts

Outliers are values outside the range between Q1 - 1.5 * IQR and this Q3 + 1.5 * IQR :

Answer 1:

Q3 + 1.5 * IQR

Question 22

2.5 / 2.5 pts

Is this statement true? When outliers are important then it is important not to change the current minimum and maximum for normalization.

- (●) True
- ( ) False

Question 23

2.5 / 2.5 pts

Is this statement true? When outliers are not significant then it is important to change the maximum and minimum by subtracting outlier end points from minimum and maximum to get the new minimum and maximum.

- ( ) False

- True

Question 24

Not yet graded / 10 pts

Read this article and provide your summary of the article:

https://statisticsbyjim.com/regression/interpret-r-squared-regression/

Also discuss your understanding of the equation:

R^2 = SSR/SST = 1 - SSE/SST

(Note: All of you will get full points for this question for answering. Do bit worry about quality. The purpose is: the paper gives you a new perspective of how to look at things.)

Your Answer:

R^2 describes how good of a fit your data has with the ideal best fit curve. An R^2 value close to 1 indicates a strong correlation, while an R^2 value close to 0 indicates a low correlation. We can sometimes draw statistical conclusions from either high or low R^2 value.

Quiz Score: **38** out of 50