Data Science AI/ML

Previous Activity

Started on	Thursday, 1 August 2019, 7:09 PM
State	Finished
Completed on	Thursday, 1 August 2019, 7:51 PM
Time taken	41 mins 46 secs
Marks	24.00/45.00
Grade	**16.00** out of 30.00 (**53**%)

Complete

Mark 1.00 out of 1.00

A and B are two events. If P (A, B) decreases while P(A) increases, what must be true:

Select one:

- a. All of above
- b. P(B | A) decreases
- c. P(B) decreases
- d. P(A | B) decreases

Question **2**

Complete

Mark 1.00 out of 1.00

A table with all possible value of a random variable and its corresponding probabilities is called

Select one:

- a. Cumulative distribution function
- b. Probability Density Function
- c. Probability Mass Function
- d. Probability Distribution

Complete

Mark 1.00 out of 1.00

A term used to describe the case when the independent variables in a multiple regression model are correlated is

Select one:

- a. Multicollinearity
- b. Regression
- c. None of the above
- d. Correlation

Question **4**

Complete

Mark 1.00 out of 1.00

According to Central Limit theorem, the standard deviation of the sampling distribution of the sample means is equal to the population standard deviation divided by the square root of the sample size .

Select one:

- a. False
- b. True

Complete

Mark 0.00 out of 1.00

Adding more basis functions in a linear model... (pick the most probable option)

Select one:

- a. Decreases model bias
- b. Doesn't affect bias and variance
- c. Decreases variance
- d. Decreases estimation bias

Question **6**

Complete

Mark 0.00 out of 1.00

All of the following increase the width of a confidence interval except:

Select one:

- a. Increased sample size
- b. Decreased sample size
- c. Increased variability
- d. Increased confidence level

Complete

Mark 1.00 out of 1.00

Another name for an output attribute.

Select one:

- a. dependent variable
- b. predictive variable
- c. independent variable
- d. estimated variable

Question **8**

Complete

Mark 1.00 out of 1.00

Confidence intervals are useful when trying to estimate

Select one:

- a. known parameters
- b. known statistics
- c. unknown parameters
- d. unknown statistics

Complete

Mark 0.00 out of 1.00

Consider a binary classification problem. Suppose I have trained a model on a linearly separable training set . and now I get new labelled data point which is correctly classified by the model, and far away from the decision boundary. If now add this new point to my earlier training set and re-train, in which cases is the learnt decision boundary likely to change?

Select one:

- a. When my model is a perceptron
- b. When my model is logistic Regression and Gaussian discriminant Analysis.
- c. When my model is an SVM.
- d. When my model is perceptron and logistic Regression

Question **10**

Complete

Mark 0.00 out of 1.00

Data used to optimize the parameter settings of a supervised learner model

Select one:

- a. Training
- b. Validation
- c. Testing
- d. Verification

Complete

Mark 0.00 out of 1.00

Determine the algorithm which is the best for this problem - Determine the characteristics of successful used car sales person?

Select one:

- a. Un-supervised
- b. Supervised
- c. Data Query/Retrieval

Question **12**

Complete

Mark 1.00 out of 1.00

Determine the algorithm which is the best for this problem - Determine whether a credit card transactions are fraudlulent or valid.

Select one:

- a. Data Query/Retrieval
- b. Un-supervised
- c. Supervised

Complete

Mark 0.00 out of 1.00

Determine the algorithm which is the best for this problem - Develop a profile for credit card customers likely to carry an average monthly balance of mre than $1000.00

Select one:

- a. Supervised
- b. Un-supervised
- c. Data Query/Retrieval

Question **14**

Complete

Mark 0.00 out of 1.00

Determine the algorithm which is the best for this problem - Do single men play golf more than married men ?

Select one:

- a. Data Query/Retrieval
- b. Supervised
- c. Un-supervised

Complete

Mark 0.00 out of 1.00

Determine the algorithm which is the best for this problem - What attribute simialrities group customers holding one or several insurance policies?

Select one:

- a. Supervised
- b. Un-supervised
- c. Data Query/Retrieval

Question **16**

Complete

Mark 0.00 out of 1.00

Determine the algorithm which is the best for this problem - What is the average weekly salary of all female employees under 40 years of age?

Select one:

- a. Data Query/Retrieval
- b. Un-supervised
- c. Supervised

Complete

Mark 1.00 out of 1.00

Exploratory data analysis uses data on some object to predict values for other object?

Select one:

- a. False
- b. True

Question **18**

Complete

Mark 0.00 out of 1.00

Formula of calculating standard deviation in binomial distribution

Select one:

- a. Square root of pq
- b. Square root of np
- c. Square root of p
- d. Square root of npq

Complete

Mark 1.00 out of 1.00

If value of interval a is 8.5 and value of interval b is 9.5 then value of mean for uniform distribution is

Select one:

- a. 8
- b. 7.5
- c. 10
- d. 9

Question **20**

Complete

Mark 0.00 out of 1.00

If value of μ is 75, value of x is 120 with unknown standard deviation of distribution then value of z- statistic considering probability distribution as standard normal is

Select one:

- a. = 1
- b. + ve
- c. - ve
- d. = 0

Complete

Mark 1.00 out of 1.00

In the discriminative approach to solving classification problems, we model the conditional probability of the labels given the observations.

Select one:

- a. True
- b. False

Question **22**

Complete

Mark 1.00 out of 1.00

One of the main challenge/s of NLP Is

Select one:

- a. Handling Ambiguity of Sentences
- b. Handling POS-Tagging
- c. Handling Tokenization

Complete

Mark 1.00 out of 1.00

Regarding bias and variance ,which of the following is true ?(High and low are relative to ideal model).

Select one:

- a. Models which overfit have a low bias and underfit have low variance.
- b. Models which overfit have a high bias and underfit have low variance.
- c. Models which overfit have a low bias and underfit have high variance.
- d. Models which overfit have a high bias and underfit have high variance.

Question **24**

Complete

Mark 1.00 out of 1.00

Selecting data so as to assure that each class is properly represented in both the training and test set.

Select one:

- a. verification
- b. stratification
- c. bootstrapping
- d. cross validation

Complete

Mark 1.00 out of 1.00

Simple linear regression is equipped to handle more than on predictor.

Select one:

- a. True
- b. False

Question **26**

Complete

Mark 0.00 out of 1.00

Suppose your model is overfitting. Which of the following is NOT a valid way to try and reduce the overfitting?

Select one:

- a. Increase the amount of training data.
- b. Decrease the model complexity.
- c. Reduce the noise in the training data.
- d. Improve the optimisation algorithm being used for error minimisation.

Complete

Mark 1.00 out of 1.00

The amount of risk that an analyst will accept when making a decision can be viewed as the level of significance.

Select one:

- a. False
- b. True

Question **28**

Complete

Mark 1.00 out of 1.00

The correlation coefficient for two real valued attributes is 0.85. What does this value tell you?

Select one:

- a. As the value of one attribute increases the value of the second attribute also increases.
- b. The attributes show a curvilinear relationship.
- c. The attributes are not linearly related.
- d. As the value of one attribute decreases the value of the second attribute increases.

Complete

Mark 1.00 out of 1.00

The major tasks of NLP include

Select one:

- a. Machine Translation
- b. All of the above
- c. Automatic Summarization
- d. Discourse Analysis

Question **30**

Complete

Mark 1.00 out of 1.00

The single best representative of a class.

Select one:

- a. centroid
- b. signature
- c. mean
- d. prototype

Complete

Mark 1.00 out of 1.00

This approach is best when we are interested in finding all possible interactions among a set of attributes.

Select one:

- a. K-Means algorithm
- b. association rules
- c. genetic learning
- d. decision tree

Question **32**

Complete

Mark 0.00 out of 1.00

This supervised learning algorithm technique can process both categorical and numeric input attributes.

Select one:

- a. Back propagation learning
- b. Bayes classifier
- c. Linear Regression
- d. Logistic Regression

Complete

Mark 0.00 out of 1.00

This technique associates a conditional probablity value with each data instance

Select one:

- a. Multiple linear Regression
- b. Logistic Regression
- c. Linear Regression
- d. Simple regression

Question **34**

Complete

Mark 1.00 out of 1.00

This technique uses mean and standard deviation scores to transform real-valued attributes.

Select one:

- a. decimal scaling
- b. min-max normalization
- c. z-score normalization
- d. logarithmic normalization

Complete

Mark 1.00 out of 1.00

Traditional database management tools can handle the size and complexity of big data?

Select one:

- a. True
- b. False

Question **36**

Complete

Mark 1.00 out of 1.00

Training a k-nearest-neighbours classifier takes more computational time than applying it.

Select one:

- a. True
- b. False

Complete

Mark 0.00 out of 1.00

$V(X) = E[X]^2 - E[X^2]$

Select one:

- a. True
- b. False

Question **38**

Complete

Mark 0.00 out of 1.00

What is used in determining the nature of the learning problem?

Select one:

- a. Environment
- b. Problem
- c. Feedback
- d. All of the mentioned

Complete

Mark 0.00 out of 1.00

Which among these techniques is associated with conditional probability value with each instance

Select one:

- a. Simple Regression
- b. Linear Regression
- c. Multiple Regression
- d. Logistic Regression

Question **40**

Complete

Mark 0.00 out of 1.00

Which of the following sentence is FALSE regarding regression?

Select one:

- a. It relates inputs to outputs.
- b. It discovers causal relationships.
- c. It may be used for interpretation.
- d. It is used for prediction.

Complete

Mark 1.00 out of 1.00

Which statement about outliers is true?

Select one:

- a. Outliers should be part of the test dataset but should not be present in the training data.
- b. Outliers should be part of the training dataset but should not be present in the test data.
- c. Outliers should be identified and removed from a dataset.
- d. The nature of the problem determines how outliers are used.

Question **42**

Complete

Mark 0.00 out of 1.00

Which statement is true about the decision tree attribute selection process?

Select one:

- a. Numeric and categorical attributes may appear in at most one tree node
- b. Both numeric and categorical attributes may appear in several tree nodes.
- c. A categorical attribute may appear in a tree node several times but a numeric attribute may appear at most once.
- d. A numeric attribute may appear in several tree nodes but a categorical attribute may appear at most once.

Complete

Mark 0.00 out of 1.00

Which statement is true about the K-Means algorithm?

Select one:

- a. All attributes must be numeric
- b. All attribute values must be categorical.
- c. The output attribute must be categorical.
- d. Attribute values may be either categorical or numeric.

Complete

Mark 0.00 out of 1.00

You are reviewing papers for the World's Fanciest Machine Learning Conference, and you see submissions with the following claims. Which ones would you consider accepting?

Select one:

- a. My method achieves a test error lower than all previous methods! (Footnote: When regularisation parameter λ is chosen so as to minimise cross-validation error.)
- b. My method achieves a training error lower than all previous methods!
- c. My method achieves a cross-validation error lower than all previous methods! (Footnote: When regularisation parameter λ is chosen so as to minimise cross-validation error.)
- d. My method achieves a test error lower than all previous methods! (Footnote: When regularisation parameter λ is chosen so as to minimise test error.)

Complete

Mark 1.00 out of 1.00

You observe the following while fitting a linear regression to the data: As you increase the amount of training data, the test error decreases and the training error increases. The train error is quite low (almost what you expect it to), while the test error is much higher than the train error. What do you think is the main reason behind this behaviour? Choose the most probable option.

Select one:

- a. High model bias
- b. None of the above
- c. High estimation bias
- d. High variance