



15812



BRAINWARE UNIVERSITY

Library
Brainware University
398, Ramkrishnapur Road, Barasat
Kolkata, West Bengal-700125

Term End Examination 2024-2025
Programme – MCA-2022/MCA-2023
Course Name – Applied Data Science
Course Code - MCA401B
(Semester IV)

Full Marks : 60

Time : 2:30 Hours

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

Group-A

(Multiple Choice Type Question)

1 x 15=15

1. Choose the correct alternative from the following :

(i) Binary attributes are defined as

- a) This takes only two values . In general, these values will be 0 and 1 and they can be coded as one bit
- c) Systems that can be used without knowledge of internal operations

- b) The natural environment of a certain species
- d) None of these

(ii) Classification accuracy is identified as

- a) A subdivision of a set of examples into a number of classes
- c) The task of assigning a classification to a set of examples

- b) Measure of the accuracy, of the classification of a concept that is given by a certain theory
- d) None of these

(iii) Cluster is identified as

- a) Group of similar objects that differ significantly from other objects
- c) Symbolic representation of facts or ideas from which information can potentially be extracted

- b) Operations on a database to transform or simplify data in order to prepare it for a machine-learning algorithm
- d) None of these

(iv) A definition of a concept is defined as -----if it recognizes all the instances of that concept

- a) Complete
- c) Constant

- b) Consistent
- d) None of these

(v) Data mining is identified as

- a) The actual discovery phase of a knowledge discovery process

- b) The stage of selecting the right data for a KDD process

- c) A subject-oriented integrated time variant non-volatile collection of data in support of management
- d) None of these
- (vi) Who developed R?
- a) Dennis Ritchie
b) John Chambersii
c) Bjarne Stroustrup
d) None of these
- (vii) R is described as _____ programming language.
- a) Closed source
b) Open source
c) GPL
d) None of these
- (viii) Data selection is identified as
- a) The actual discovery phase of a knowledge discovery process
b) The stage of selecting the right data for a KDD process
c) A subject-oriented integrated time variant non-volatile collection of data in support of management
d) None of these
- (ix) Classification task refers to
- a) A subdivision of a set of examples into a number of classes
b) A measure of the accuracy, of the classification of a concept that is given by a certain theory
c) The task of assigning a classification to a set of examples
d) None of these
- (x) Hybrid is identified as
- a) Combining different types of method or information
b) Approach to the design of learning algorithms that is structured along the lines of the theory of evolution
c) Decision support systems that contain an information base filled with the knowledge of an expert formulated in terms of if-then rules.
d) None of these
- (xi) How calculations work in TensorFlow is estimated ?
- a) Through vector multiplications
b) Through RDDs
c) Through Computational Graph
d) Through map reduce tasks
- (xii) Suppose you have trained a logistic regression classifier and it outputs a new example x with a prediction $h_0(x) = 0.2$. This concludes
- a) Our estimate for $P(y=1 | x)$
b) Our estimate for $P(y=0 | x)$
c) Our estimate for $P(y=1 | x)$
d) None of these
- (xiii) _____ infer initiation of an infinite loop right from the start.
- a) Never
b) Repeat
c) Break
d) Set
- (xiv) Which of the following methods are distinguished to be present in caret for regularized regression?
- a) Ridge
b) Lasso
c) Relaxo
d) All of these
- (xv) Which of the following commands will infer a list?
- a) `list1 = list()`
b) `list1=[]`
c) `list1 = list([1, 2, 3])`
d) All of these

Group-B

(Short Answer Type Questions)

3 x 5=15

2. Compare data science and confusion matrix. (3)
3. Summarize the reason why Python is used for data cleaning in Data Science (3)
4. Describe Variance in statistics with example (3)

5. Explain Selection bias with example.
6. Evaluate Normal distribution

Compare bias and variance

OR

Library
Brainware University
398, Ramkrishnapur Road, Barasat
Kolkata, West Bengal-700125

(3)

(3)

(3)

Group-C

(Long Answer Type Questions)

5 x 6=30

7. When creating a statistical model, how do we determine overfitting? (5)
8. State Bayes' theorem in statistics. Outline the Naive Bayes algorithm to build classification models. (5)
9. Compare quantitative data and qualitative data. (5)
10. Evaluate Mean with simplification using a sample dataset. (5)
11. Explain correlation with diagram. (5)
12. Imagine that Jeremy took part in an examination. The test is having a mean score of 160, and it has a standard deviation of 15. If Jeremy's z-score is 1.20, Evaluate his score on test. (5)

OR

A regression analysis between apples (y) and oranges (x) resulted in the following least-squares line: $y = 100 + 2x$. Predict the implication if oranges are increased by 1 (5)
