BRAINWARE UNIVERSITY

Term End Examination 2023
Programme – B.Tech.(CSE)-2019/B.Tech.(CSE)-2020
Course Name – Data Mining
Course Code - PEC-602C
(Semester VI)

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) The dual optimization problem in SVM design is classifying using
   a) Linear programming
   b) Quadratic programming
   c) Dynamic programming
   d) Integer programming

(ii) Compute the main point of difference between the adaline & perceptron model
   a) weights are compared with output
   b) sensory units result is compared with output
   c) analog activation value is compared with output
   d) all of the mentioned

(iii) 'Name of a person', can be located as an attribute of type.
   a) Nominal
   b) Ordinal
   c) Interval
   d) Ratio

(iv) A frequent itemset is one which quote the
   a) Support criteria
   b) Confidence criteria
   c) Both support and confidence criteria
   d) None of the above

(v) Show the fundamental unit of network
   a) brain
   b) nucleus
   c) neuron
   d) axon

(vi) The amount of rain that falls in a day is usually measured in either millimeter (mm) or inches. Suppose you use a learning algorithm to predict how much rain will fall tomorrow. Determine this as a classification or a regression problem
   a) Regression
   b) Classification
   c) All of the above
   d) None of these

(vii) Leaf nodes of a decision tree refer to:
   a) Attributes
   b) Classes
   c) Data instances
   d) None of the above

(viii) In principal component analysis, the projected lower dimensional space report to

a) subset of the original co-ordinate axis
b) eigenvectors of the data covariance matrix
c) eigenvectors of the data distance matrix
d) orthogonal vectors to the original co-ordinate axis

(ix) If dimensionality reduction is performed on a record data matrix, the transformed data matrix is identified

a) has reduced number of rows
b) has reduced number of columns
c) has reduced number of both rows and columns
d) has same number of rows and columns

(x) Bayes classifier is also describe as

a) Maximum aposteriori classifier
b) Maximum apriori classifier
c) Most likely classifier
d) Maximum margin classifier

(xi) The classification boundary realised by the perceptron is discover as

a) Circle
b) Parabola
c) Straight line
d) Ellipse

(xii) Choose the following is not true about K-means clustering algorithm

a) It is a partitional clustering algorithm
b) The final cluster obtained depends on the choice of initial cluster centres
c) Number of clusters need to be specified in advance
d) It can generate non-convex cluster shapes

(xiii) Explain the internal state of neuron, is the function of the inputs the neurons receive

a) Weight
b) activation or activity level of neuron
c) Bias
d) None of these

(xiv) Report the meaning of Perceptron

a) General class of approaches to a problem.
b) Performing several computations simultaneously
c) Structures in a database those are statistically relevant
d) Simple forerunner of modern neural networks, without hidden layers

(xv) Consider three item sets: I1={milk, sugar, bread}; I2={milk, sugar}; I3={milk} Identify the correct statement

a) support(I1) > support(I2)
b) support(I2) > support(I3)
c) both statements A and B
d) none of the statements A and B

Group-B
(Short Answer Type Questions) 3 x 5=15

2. Define "Text data" with proper example. (3)
3. Interpret the data pre-processing techniques used in KDD. (3)
4. Explain the overfitting and underfitting problem with proper example. (3)
5. Illustrate the Single linkage and complete linkage clustering algorithm (3)
6. Illustrate the Apriori probability for Bayes Classifier. (3)

OR

Point out the Mahalanobis distance to identify the class boundary in Bayes Classifier. (3)

Group-C
(Long Answer Type Questions) 5 x 6=30

7. Explain the various types of OLAP cubes along with its advantage and disadvantage. (5)
8. Explain about the Fuzzification approaches. (5)
9. Judge the Hessian Matrix to optimize the dual problem for identifying good decision boundaries in support of vector machine-based classifier. (5)
10. Define Data Mining technique with the proper example and recall the various characteristic of Data Mining approaches. (5)
11. Explain the various Data Mining functionalities with proper example. (5)

12. Explain the confusion Matrix along with formulation of various Cost-Sensitive Measure. (5)

OR

Calculate to identify the proper Margin in Support Vector Machine with the help of primal and dual problem. (5)

****