



BRAINWARE UNIVERSITY

Term End Examination 2021 - 22

Programme – Bachelor of Science (Honours) in Computer Science

Course Name – Introduction to Data Sciences

Course Code - BCSD601B

(Semester VI)

Time allotted : 1 Hrs.15 Min.

Full Marks : 60

[The figure in the margin indicates full marks.]

Group-A

(Multiple Choice Type Question)

1 x 60=60

Choose the correct alternative from the following :

(1) Adaptive system management is

- | | |
|---|--|
| a) It uses machine -learning techniques. Here program can learn from past experience and adapt themselves to new situations | b) Computational procedure that takes some value as input and produces some value as output. |
| c) Science of making machines performs tasks that would require intelligence when performed by humans | d) None of these |

(2) Bayesian classifiers is

- | | |
|---|--|
| a) A class of learning algorithm that tries to find an optimum classification of a set of examples using the probabilistic theory. | b) Any mechanism employed by a learning system to constrain the search space of a hypothesis |
| c) An approach to the design of learning algorithms that is inspired by the fact that when people encounter new situations, they often explain them by reference to familiar experiences, adapting the explanations to fit the new situation. | d) None of these |

(3) Algorithm is

- | | |
|--|---|
| a) It uses machine-learning techniques. Here program can learn from past experience and adapt themselves to new situations | b) Computational procedure that takes some value as input and produces some value as output |
| c) Science of making machines performs tasks that would require intelligence when performed by humans | d) None of these |

(4) Bias is

- a) A class of learning algorithm that tries to find an optimum classification of a set of examples using the probabilistic theory
- b) Any mechanism employed by a learning system to constrain the search space of a hypothesis
- c) An approach to the design of learning algorithms that is inspired by the fact that when people encounter new situations, they often explain them by reference to familiar experiences, adapting the explanations to fit the new situation.
- d) None of these

(5) Background knowledge referred to

- a) Additional acquaintance used by a learning algorithm to facilitate the learning process
- b) A neural network that makes use of a hidden layer
- c) It is a form of automatic learning.
- d) None of these

(6) Case-based learning is

- a) A class of learning algorithm that tries to find an optimum classification of a set of examples using the probabilistic theory.
- b) Any mechanism employed by a learning system to constrain the search space of a hypothesis
- c) An approach to the design of learning algorithms that is inspired by the fact that when people encounter new situations, they often explain them by reference to familiar experiences, adapting the explanations to fit the new situation.
- d) None of these

(7) Classification is

- a) A measure of the accuracy of the classification of a concept that is given by a certain theory
- b) A subdivision of a set of examples into a number of classes.
- c) The task of assigning a classification to a set of examples.
- d) None of these

(8) Binary attribute are

- a) The natural environment of a certain species
- b) Systems that can be used without knowledge of internal operations
- c) This takes only two values. In general, these values will be 0 and 1 and they can be coded as one bit
- d) None of these

(9) Classification accuracy is

- a) A subdivision of a set of examples into a number of classes
- b) 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

(10) Cluster is

- a) Group of similar objects that differ significantly from other objects
- b) Operations on a database to transform or simplify data in order to prepare it for a machine-learning algorithm
- c) Symbolic representation of facts or ideas from which information can potentially be extracted
- d) None of these

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

- a) Complete
c) Constant
- b) Consistent
d) None of these
- (12) R is an _____ programming language?
a) Closed source
c) Open source
- b) GPL
d) None of these
- (13) R was named partly after the first names of _____ R authors?
a) 2
c) 3
- b) 1
d) 4
- (14) R is an interpreted language so it can access through _____ ?
a) Disk operating system
c) Operating system
- b) User interface operating system
d) Command line interpreter
- (15) Predictive analysis is the branch of _____ analysis?
a) Advanced
c) Both advanced and core
- b) Core
d) None of these
- (16) How many steps does the predictive analysis process contain?
a) 8
c) 5
- b) 7
d) 6
- (17) How many data types are present in R?
a) 5
c) 6
- b) 4
d) 7
- (18) Classification task referred to
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) A measure of the accuracy, of the classification of a concept that is given by a certain theory
d) None of these
- (19) Discovery is
a) It is hidden within a database and can only be recovered if one is given certain clues (an example IS encrypted information).
c) An extremely complex molecule that occurs in human chromosomes and that carries genetic information in the form of genes.
- b) The process of executing implicit previously unknown and potentially useful information from data
d) None of these
- (20) Hidden knowledge referred to
a) A set of databases from different vendors, possibly using different database paradigms.
c) Information that is hidden in a database and that can't be recovered by a simple SQL query.
- b) An approach to a problem that isn't guaranteed to work but performs well in most cases
d) None of these
- (21) Heterogeneous databases referred to
a) A set of databases from different vendors, possibly using different database paradigms
c) Information that is hidden in a database and that cannot be recovered by a simple SQL query
- b) An approach to a problem that is not guaranteed to work but performs well in most cases.
d) None of these

(22) Heuristic is

- a) A set of databases from different vendors, possibly using different database paradigms
- b) An approach to a problem that is not guaranteed to work but performs well in most cases
- c) Information that is hidden in a database and that cannot be recovered by a simple SQL query.
- d) None of these

(23) KDD (Knowledge Discovery in Databases) is referred to

- a) Non-trivial extraction of implicit previously unknown and potentially useful information from data
- b) Set of columns in a database table that can be used to identify each record within this table uniquely.
- c) collection of interesting and useful patterns in a database
- d) None of these

(24) Inductive logic programming is

- a) A class of learning algorithms that try to derive a Prolog program from examples
- b) A table with n independent attributes can be seen as an n-dimensional space
- c) Prediction made using an extremely simple method, such as always predicting the same output
- d) None of these

(25) Prediction is

- a) The result of the application of a theory or a rule in a specific case
- b) One of several possible enters within a database table that is chosen by the designer as the primary means of accessing the data in the table.
- c) Discipline in statistics that studies ways to find the most interesting projections of multi-dimensional spaces
- d) None of these

(26) Node is

- a) A component of a network
- b) In the context of KDD and data mining, this refers to random errors in a database table.
- c) One of the defining aspects of a data warehouse
- d) None of these

(27) Point out the wrong statement

- a) The trapezoidal rule is used to compute the area under the ROC curve
- b) For regression, the relationship between each predictor and the outcome is evaluated
- c) An argument, para, is used to pick the model fitting technique
- d) All of the mentioned

(28) Which of the following functions tracks the changes in model statistics?

- a) varImp
- b) varImpTrack
- c) findTrack
- d) None of the mentioned

(29) Which of the following models include a backwards elimination feature selection routine?

- a) MCV
- b) . MARS
- c) MCRS
- d) All of the mentioned

(30) Which tool is best suited for solving Deep Learning problems

- a) R
- b) Sk-learn
- c) Excel
- d) TensorFlow

individual

- (44) Which of the following is commonly referred to as 'data fishing'?
- a) Data bagging
 - b) Data booting
 - c) Data merging
 - d) None of the mentioned
- (45) Which of the following is the probability calculus of beliefs, given that beliefs follow certain rules?
- a) Bayesian probability
 - b) Frequency probability
 - c) Frequency inference
 - d) Bayesian inference
- (46) Which of the following can be considered as a random variable?
- a) The outcome from the roll of a die
 - b) The outcome of flip of a coin
 - c) The outcome of exam
 - d) All of the mentioned
- (47) Which of the following is also referred to as a random variable?
- a) stochast
 - b) aleatory
 - c) eliette
 - d) All of the mentioned
- (48) Which of the following functions is associated with a continuous random variable?
- a) pdf
 - b) pmv
 - c) pmf
 - d) All of the mentioned
- (49) Point out the wrong combination.
- a) True negative=correctly rejected
 - b) False negative=correctly rejected
 - c) False positive=correctly identified
 - d) All of the mentioned
- (50) Which of the following is not a machine learning algorithm?
- a) SVM
 - b) Random forest are easy to interpret but often very accurate
 - c) SVG
 - d) None of the mentioned
- (51) Which of the following is a categorical outcome?
- a) RMSE
 - b) RSquared
 - c) Accuracy
 - d) All of the mentioned
- (52) Which of the following can be used to create the most common graph types?
- a) qplot
 - b) quickplot
 - c) plot
 - d) All of the mentioned
- (53) Which of the following is required by K-means clustering?
- a) defined distance metric
 - b) number of clusters
 - c) initial guess as to cluster centroids
 - d) all of the mentioned
- (54) Which of the following clustering requires a merging approach?
- a) Partitional
 - b) Hierarchical
 - c) Naive Bayes
 - d) None of the mentioned
- (55) How calculations work in TensorFlow
- a) Through vector multiplications
 - b) Through RDDs
 - c) Through Computational Graphs
 - d) Through map reduce tasks
- (56) To find the minimum or the maximum of a function, we set the gradient to zero because
- a) The value of the gradient at extrema of a function is always zero
 - b) Depends on the type of problem

- c) All of the mentioned
d) None of the mentioned
- (57) Which of the following is a disadvantage of decision trees?
a) Factor analysis
b) Decision trees are robust to outliers
c) Decision trees are prone to be over fit
d) None of these
- (58) What is the purpose of performing cross-validation?
a) To assess the predictive performance of the models
b) To judge how the trained model performs outside the sample on test data
c) Both of the mentioned
d) None of the mentioned
- (59) When performing regression or classification, which of the following is the correct way to preprocess the data?
a) Normalize the data → PCA → training
b) PCA → normalize PCA output → training
c) Normalize the data → PCA → normalize PCA output → training
d) None of the mentioned
- (60) What is `pca.components_` in Sklearn?
a) Set of all eigen vectors for the projection space
b) Matrix of principal components
c) Result of the multiplication matrix
d) None of the mentioned