



BRAINWARE UNIVERSITY
Term End Examination 2020 - 21
Programme – Master of Computer Applications

Course Name – Data Science

Course Code - MCA502A

Semester / Year - Semester V

Time allotted : 85 Minutes

Full Marks : 70

[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 70=70

1. *(Answer any Seventy)*

(i) Computers are best at learning

- | | |
|------------|---------------|
| a) Facts | b) Figures |
| c) Numbers | d) Procedures |

(ii) 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 |

(iii) 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 | d) None of these |

performed by humans

(iv) 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

(v) 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

(vi) Classification is

- 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

(vii) Binary attribute are

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

(viii) 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

(ix) 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

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

- a) Complete
- b) Consistent
- c) Constant
- d) None of these

(xi) Data mining is

- 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

(xii) R is an _____ programming language.

- a) Closed source
- b) GPL
- c) Open source
- d) None of these

(xiii) A tensor is similar to

- a) Data Array
- b) ANN Model
- c) SQL query
- d) Pythoncode

(xiv) How calculations work in TensorFlow

- a) Through vector multiplications
- b) Through RDDs
- c) Through Computational Graphs
- d) Through map reduce tasks

(xv) Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging?

- a) Decision Tree
- b) Regression
- c) Classification
- d) Random Forest

(xvi) 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) Both The value of the gradient at extrema of a function is always zero and Depends on the type of problem
- d) None of these

(xvii) The most widely used metrics and tools to assess a classification model are:

- a) Confusion matrix
- b) Cost-sensitive accuracy
- c) Area under the ROC curve
- d) All of these

(xviii) Which of the following is a good test data set characteristic?

- a) Large enough to yield meaningful results
- b) Is representative of the dataset as a whole
- c) Both Large enough to yield meaningful results and Is representative of the dataset as a whole
- d) None of these

(xix) 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

(xx) How do you handle missing or corrupted data in a dataset?

- a) Drop missing rows or columns
- b) Replace missing values with mean/median/mode
- c) Assign a unique category to missing values
- d) All of these

(xxi) 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 To assess the predictive performance of the models and To judge how the trained model performs outside the sample on test data
- d) None of these

(xxii) 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 these

(xxiii) Which of the following is an example of feature extraction?

- a) Constructing bag of words vector from an email
- b) Applying PCA projects to a large high-dimensional data
- c) Removing stopwords in a sentence
- d) All of these

(xxiv) Which of the following is true about Naive Bayes ?

- a) Assumes that all the features in a dataset are equally important
- b) Assumes that all the features in a dataset are independent

c) Both Assumes that all the features in a dataset are equally important and Assumes that all the features in a dataset are independent d) None of these

(xxv) Which of the following statements about regularization is not correct?

a) Using too large a value of lambda can cause your hypothesis to under fit the data. b) Using too large a value of lambda can cause your hypothesis to over fit the data.
c) Using a very large value of lambda cannot hurt the performance of your hypothesis. d) None of these

(xxvi) How can you prevent a clustering algorithm from getting stuck in bad local optima?

a) Set the same seed value for each run b) Filtering
c) Both Set the same seed value for each run and Filtering d) None of these

(xxvii) Which of the following techniques can be used for normalization in text mining?

a) Stemming b) Lemmatization
c) Stop Word Removal d) Both Stemming and Lemmatization

(xxviii) In which of the following cases will K-means clustering fail to give good results? 1) Data points with outliers 2) Data points with different densities 3) Data points with non convex shapes

a) 1 and 2 b) 2 and 3
c) 1, 2, and 3 d) 1 and 3

(xxix) Sentence parser is typically used for

a) It is used to parse sentences to check if they are utf-8 compliant. b) It is used to parse sentences to derive their most likely syntax tree structures.
c) It is used to parse sentences to assign d) It is used to check if sentences can be

POS tags to all tokens.

parsed into meaningful tokens.

(xxx) Suppose you have trained a logistic regression classifier and it outputs a new example x with a prediction $h_0(x) = 0.2$. This means

- 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

(xxxii) Point out the correct statement?

- a) The value NaN represents undefined value
- b) Number Inf represents infinity in R
- c) NaN can also be thought of as a missing value
- d) "raw" objects are commonly used directly in data analysis

(xxxiii) Attributes of an object (if any) can be accessed using the _____ function

- a) objects()
- b) attrib()
- c) attributes()
- d) obj()

(xxxiv) R objects can have attributes, which are like _____ for the object.

- a) metadata
- b) features
- c) expression
- d) dimensions

(xxxv) Which of the following can be considered as object attribute?

- a) dimensions
- b) class
- c) length
- d) all of the mentioned

(xxxvi) Which of these measures are used to analyze the central tendency of data?

- a) Mean and Normal Distribution
- b) Mean, Median and Mode
- c) Mode, Alpha & Range
- d) Standard Deviation, Range and Mean

(xxxvi) A test is administered annually. The test has a mean score of 150 and a standard deviation of 20. If Ravi's z-score is 1.50, what was his score on the test?

- a) 180
- b) 130
- c) 30
- d) 150

(xxxvii) Which of the following measures of central tendency will always change if a single value in the data changes?

- a) Mean
- b) Median
- c) Mode
- d) All of these

(xxxviii) If the variance of a dataset is correctly computed with the formula using $(n - 1)$ in the denominator, which of the following option is true?

- a) Dataset is a sample
- b) Dataset is a population
- c) Dataset could be either a sample or a population
- d) Dataset is from a census

(xxxix) Which instruments are used for perceiving and acting upon the environment?

- a) Sensors and Actuators
- b) Sensors
- c) Perceiver
- d) None of the mentioned

(xl) What is meant by the agent's percept sequence?

- a) Used to perceive the environment
- b) Complete history of actuator
- c) Complete history of perceived things
- d) None of the mentioned

(xli) What is the rule of simple reflex agent?

- a) Simple-action rule
- b) Condition-action rule
- c) Simple & Condition-action rule
- d) None of the mentioned

(xlii) What are the compositions for agents in artificial intelligence?

- a) Program
- b) Architecture

- c) Both Program & Architecture
- d) None of the mentioned

(xlvi) Which is used to improve the agent's performance?

- a) Perceiving
- b) Learning
- c) Observing
- d) None of the mentioned

(xlvii) Which action sequences are used to achieve the agent's goal?

- a) Search
- b) Plan
- c) Retrieve
- d) Both Search & Plan

(xlviii) Which element in agent is used for selecting actions which deals with outsiders?

- a) Perceive
- b) Performance
- c) Learning
- d) Actuator

(xlix) An 'agent' is anything that,

- a) Perceives its environment through sensors and acting upon that environment through actuators
- b) Takes input from the surroundings and uses its intelligence and performs the desired operations
- c) A embedded program controlling line following robot
- d) All of the mentioned

(l) Agents behaviour can be best described by

- a) Perception sequence
- b) Agent function
- c) Sensors and Actuators
- d) Environment in which agent is performing

(li) Categorize Crossword puzzle in Fully Observable / Partially Observable.

- a) Fully Observable
- b) partially Observable
- c) All of the mentioned
- d) None of the mentioned

(xlix) An expert system differs from a database program in that only an expert system

- a) contains declarative knowledge
- b) contains procedural knowledge
- c) features the retrieval of stored information
- d) expects users to draw their own conclusions

(l) What will take place as the agent observes its interactions with the world?

- a) Learning
- b) Hearing
- c) Perceiving
- d) Speech

(li) Which modifies the performance element so that it makes better decision?

- a) Performance element
- b) Changing element
- c) Learning element
- d) None of the mentioned

(lii) What is used in determining the nature of the learning problem?

- a) Environment
- b) Feedback
- c) Problem
- d) None of the mentioned

(liii) How many types are available in machine learning?

- a) 1
- b) 2
- c) 3
- d) None of the mentioned

(liv) Which is used to choose among multiple consistent hypotheses?

- a) Razor
- b) Ockham razor
- c) Learning element
- d) None of the mentioned

(lv) What will happen if the hypothesis space contains the true function?

- a) Realizable
- b) Unrealizable
- c) Both Realizable & Unrealizable
- d) None of the mentioned

(lvi) What takes input as an object described by a set of attributes?

- a) Decision tree
- b) Graph
- c) Decision graph
- d) None of the mentioned

(lvii) How the decision tree reaches its decision?

- a) Single test
- b) Two test
- c) Sequence of test
- d) None of the mentioned

(lviii) Factors which affect the performance of learner system does not include?

- a) Representation scheme used
- b) Good data structures
- c) Training scenario
- d) None of the mentioned

(lix) Which of the following does not include different learning methods?

- a) Memo rization
- b) Analogy
- c) Introduction
- d) None of the mentioned

(lx) Which of the following is the model used for learning?

- a) Decision trees
- b) Neural networks
- c) Propositional and FOL rules
- d) All of the mentioned

(lxi) Which of the following is an example of active learning?

- a) News Recommender system
- b) Dust cleaning machine
- c) Automated vehicle
- d) All of the mentioned

(lxii) In which of the following learning the teacher returns reward and punishment to learner?

- a) Supervised learning
- b) Unsupervised learning
- c) Active learning
- d) Reinforcement

(lxiii) Decision trees are appropriate for the problems where

- a) Attributes are both numeric and nominal
- b) Target function takes on a discrete number of values.

- c) Data may have errors
- d) All of the mentioned

(lxiv) Which of the following is not an application of learning?

- a) Data mining
- b) WWW
- c) Speech recognition
- d) None of the mentioned

(lxv) Which of the following is the component of learning system?

- a) Goal
- b) Model
- c) Learning rules
- d) All of the mentioned

(lxvi) Which of the following is also called as exploratory learning?

- a) Supervised learning
- b) Unsupervised learning
- c) Active learning
- d) All of the mentioned

(lxvii) Which is not a desirable property of a logical rule-based system?

- a) Locality
- b) Attachment
- c) Detachment
- d) All of the mentioned

(lxviii) How is Fuzzy Logic different from conventional control methods?

- a) IF and THEN Approach
- b) FOR Approach
- c) WHILE Approach
- d) All of the mentioned

(lxix) In an Unsupervised learning

- a) Specific output values are given
- b) Specific output values are not given
- c) No specific Inputs are given
- d) All of the mentioned

(lxx) Inductive learning involves finding a

- a) Consistent Hypothesis
- b) Inconsistent Hypothesis
- c) Regular Hypothesis
- d) None of the mentioned