Q.No.1
The k-means algorithm is a
(A) Supervised learning algorithm
(B) Unsupervised learning algorithm
(C) Semi-supervised learning algorithm
(D) Weakly supervised learning algorithm

Q.No.2
For unsupervised learning we have ____ model.
(A) interactive
(B) predictive
(C) descriptive
(D) prescriptive

Q. No. 3
Which of the following is true about SVM?
(A) It is useful only in high-dimensional spaces
(B) It requires less memory
(C) SVM does not perform well when we have a large data set
(D) SVM performs well when we have a large data set

Q. No. 4
When you find many noises in data, which of the following options would you consider in KNN?
(A) Increase the value of k
(B) Decrease the value of k
(C) Noise does not depend on k
(D) k = 0

Q. No. 5
A multiple regression model has:
(A) Only one independent variable
(B) More than one independent variable
(C) More than one dependent variable
(D) None of the above

Q. No. 6
The average squared difference between classifier predicted output and actual output
(A) mean squared error
(B) root mean squared error
(C) mean absolute error
(D) mean relative error

Q. No. 7
Which of the following is not a benefit of using Grid search?
(A) It can be applied to non-differentiable functions.
(B) It can be applied to non-continuous functions
(C) It is easy to implement
(D) It runs reasonably fast for multiple linear regression

Q. No. 8
Data used to build a data mining model
(A) validation data
(B) training data
(C) test data
(D) hidden data

Q.No.9
Introducing a non-essential variable into a linear regression model may result in: (1).Increase in R-square, (2).Decrease in R-square
(A) Only 1 is correct
(B) Only 2 is correct
(C) Either 1 or 2
(D) None of these

Q.No.10
Which data is used to tune the parameters of supervised learning model
(A) training
(B) test
(C) verification
(D) validation

Q. No. 11
Which of the following is not TRUE for regression?
(A) It relates inputs to outputs
(B) It is used for prediction
(C) It may be used for interpretation
(D) It discovers causal relationships

Q. No. 12
Which of the following statements regarding outliers is correct?
(A) Outliers should be identified and removed from a dataset
(B) Outliers should be part of the training dataset but should not be present in the test data
(C) Outliers should be part of the test dataset but should not be present in the training data
(D) The nature of the problem determines how outliers are used

Q. No. 13
Each data instance is assigned a conditional probability value using this method.
(A) linear regression
(B) logistic regression
(C) simple regression
(D) multiple linear regression

Q. No. 14
Neural networks
(A) optimize a convex cost function
(B) always output values between 0 and 1
(C) can be used for regression as well as classification
(D) can be used in an ensemble

Q. No. 15
Which of the following is an example of a deterministic algorithm?
(A) PCA
(B) K-Means
(C) KNN
(D) None of the above

Q. No. 16
Which of the following is a performance measure for regression?
(A) Accuracy
(B) Recall
(C) Error rate
(D) RMSE

Q.No.17
Conversion of a text corpus to a numerical representation is done using ___ process.
(A) Tokenization
(B) Normalization
(C) Vectorization
(D) None of the above

Q.No.18
PCA is a technique for
(A) Feature extraction
(B) Feature construction
(C) Feature selection
(D) None of the above

Q. No. 19
Predicting whether a tumour is malignant or benign is an example of?
(A) Unsupervised Learning
(B) Supervised Regression Problem
(C) Supervised Classification Problem
(D) Categorical Attribute

Q. No. 20
For categorical data, ____ cannot be used as a measure of central tendency.
(A) Median
(B) Mean
(C) Quartile
(D) None of the above