



BRAINWARE UNIVERSITY

Term End Examination 2021 - 22

Programme – Bachelor of Technology in Computer Science & Engineering

Course Name – Machine Learning

Course Code - BCSE605C

(Semester VI)

Time allotted : 1 Hrs.25 Min.

Full Marks : 70

[The figure in the margin indicates full marks.]

Group-A

(Multiple Choice Type Question)

1 x 70=70

Choose the correct alternative from the following :

- (1) The process of forming general concept definitions from examples of concepts to be learned is _

a) Deduction	b) Abduction
c) Induction	d) Conjunction
- (2) Supervised learning and unsupervised clustering both require at least one

a) Hidden attribute	b) Output attribute
c) Input attribute	d) Categorical attribute
- (3) Supervised learning differs from unsupervised clustering in that supervised learning requires

a) at least one input attribute	b) input attributes to be categorical
c) at least one output attribute	d) output attributes to be categorical.
- (4) Supervised Learning uses

a) labeled dataset	b) unlabeled dataset
c) both a and b	d) none of these
- (5) Classification uses which type of output variable

a) categorical	b) continuous
c) both a and b	d) none of these
- (6) This type of supervised network architecture does not contain a hidden layer

a) back propagation	b) perceptron
c) self-organizing map	d) None of these
- (7) The total delta measures the total absolute change in network connection weights for each pass o
f the training data through a neural network. This value is most often used to determine the conve
rgence of a

a) perceptron network	b) feed-forward network
c) back propagation network	d) self-organizing network
- (8) Predicting the amount of rainfall in a region based on various cues is a _____ problem.

a) Supervised learning	b) Unsupervised learning
c) Clustering	d) Reinforcement Learning

- (9) MLE estimates are often undesirable because
- a) they are biased
 - b) they are not consistent estimators
 - c) they have high variance
 - d) None of these
- (10) The model obtained by applying linear regression on the identified subset of features may differ from the model obtained at the end of the process of identifying the subset during
- a) Forward stepwise selection
 - b) Best-subset selection
 - c) Forward stage wise selection
 - d) All of these
- (11) Suppose we 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) Our estimate for $P(y=0 | x)$
- (12) Which is method of cross validation?
- a) K Fold
 - b) Precision
 - c) Recall
 - d) ROC curves
- (13) Precision is
- a) how many of the positives does the model return
 - b) how many of the returned documents are correct
 - c) both a and b
 - d) none of these
- (14) LOOCV is
- a) Leave out one cross-validation
 - b) Left one out cross-validation
 - c) Leave one out cross-validation
 - d) Leave one out cut-validation
- (15) In Bayes Theorem, $P(A | B) = \{ P(B | A) * P(A) \} / P(B)$, where $P(B | A)$ is:
- a) The probability of event A (hypothesis) occurring given that B (evidence) has occurred
 - b) The probability of the event B (evidence) occurring given that A (hypothesis) has occurred
 - c) The probability of event B (hypothesis) occurring.
 - d) The probability of event A (evidence) occurring.
- (16) Bootstrap Method is
- a) method of cross validation
 - b) resampling technique
 - c) classifier performance measure
 - d) none of these
- (17) Dependent Variable in Regression analysis is known as
- a) target variable
 - b) predictor
 - c) Outliers
 - d) Multicollinearity
- (18) Independent Variable in Regression analysis is known as
- a) target variable
 - b) predictor
 - c) Outliers
 - d) Multicollinearity
- (19) If an algorithm works well with the training dataset but not well with test dataset, then such problem is called
- a) Multicollinearity
 - b) Overfitting
 - c) under fitting
 - d) Outlier
- (20) In linear regression , the mathematical expression used is
- a) $Y = aX + b$
 - b) $F(x) = 1 / (1 + e^{-x})$
 - c) $Y = b_0 + b_1x + b_2x^2 + b_3x^3 + \dots + b_nx^n$
 - d) None of these
- (21) A regression model in which more than one independent variable is used to predict the dependent variable is called
- a) an independent mode
 - b) multiple regression models
 - c) none of these
 - d) simple linear regression model
- (22) Regression trees are often used to model _____ data
- a) Linear
 - b) Non linear
 - c) Categorical
 - d) symmetrical

- (23) If I am using all features of my dataset and I achieve 100% accuracy on my training set, but ~70% on validation set, what should I look out for?
- a) Under fitting
 - b) Overfitting
 - c) None of these
 - d) Midfitting
- (24) 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
- (25) K-fold cross-validation is
- a) linear in K
 - b) quadratic in K
 - c) cubic in K
 - d) exponential in K
- (26) As the number of training examples goes to infinity, your model trained on that data will have:
- a) Lower variance
 - b) Higher variance
 - c) Same variance
 - d) None of these
- (27) Given two Boolean random variables, A and B, where $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$, and $P(A | \neg B) = \frac{1}{4}$, what is $P(A | B)$?
- a) 1
 - b) 0.75
 - c) 0.25
 - d) 0.16666666666666667
- (28) The K-means algorithm:
- a) Requires the dimension of the feature space to be no bigger than the number of samples
 - b) Minimizes the within class variance for a given number of clusters
 - c) Has the smallest value of the objective function when $K = 1$
 - d) Converges to the global optimum if and only if the initial means are chosen as some of the samples themselves
- (29) If N is the number of instances in the training dataset, nearest neighbors has a classification run time of
- a) $O(1)$
 - b) $O(\log N)$
 - c) $O(N)$
 - d) $O(N^2)$
- (30) Computational complexity of Gradient descent is,
- a) linear in D
 - b) linear in N
 - c) dependent on the number of iterations
 - d) polynomial in D
- (31) 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) All of these
 - d) None of these
- (32) Which of the following is a reasonable way to select the number of principal components "k"?
- a) Choose k to be 99% of m ($k = 0.99 * m$, rounded to the nearest integer)
 - b) Choose k to be the largest value so that 99% of the variance is retained
 - c) Choose k to be the smallest value so that at least 99% of the variance is retained
 - d) Use the elbow method
- (33) When we run gradient descent for 15 iterations with $a=0.3$ and compute $J(\theta)$ after each iteration. You find that the value of $J(\theta)$ decreases quickly and then levels off. Based on this, which of the following conclusions seems most plausible?
- a) Rather than using the current value of a, use a larger value of a (say $a=1.0$)
 - b) Rather than using the current value of a, use a smaller value of a (say $a=0.1$)
 - c) $a=0.3$ is an effective choice of learning rate
 - d) None of these
- (34) 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
- (35) Which of the following is a good test dataset 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
- (36) High entropy means that the partitions in classification are:
 a) pure b) not pure
 c) useful d) useless
- (37) Which of the following are the spatial clustering algorithms?
 a) Partitioning based clustering b) K-means clustering
 c) Grid based clustering d) All of these
- (38) 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) Does not depend on the type of problem d) None of these
- (39) A machine learning problem involves four attributes plus a class. The attributes have 3, 2, 2, and 2 possible values each. The class has 3 possible values. How many maximum possible different examples are there?
 a) 12 b) 24
 c) 48 d) 72
- (40) Compared to the variance of the Maximum Likelihood Estimate (MLE), the variance of the Maximum A Posteriori (MAP) estimate is _____ .
 a) higher b) same
 c) lower d) none of the above
- (41) Suppose we would like to perform clustering on spatial data such as the geometrical locations of houses. We wish to produce clusters of many different sizes and shapes. Which of the following methods is the most appropriate?
 a) Decision Trees b) Density-based clustering
 c) Model-based clustering d) K-means clustering
- (42) ID3 Algorithm is used to build
 a) Regression tree b) cluster
 c) Decision tree d) None of these
- (43) The average positive difference between computed and desired outcome values.
 a) root mean squared error b) mean squared error
 c) mean absolute error d) Data used to optimize the parameter settings of a supervised learner model
- (44) 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
- (45) A feed-forward neural network is said to be fully connected when
 a) all nodes are connected to each other b) all nodes at the same layer are connected to each other
 c) all nodes at one layer are connected to all nodes in the next higher layer d) all hidden layer nodes are connected to all output layer nodes
- (46) A two-layered neural network used for unsupervised clustering
 a) back propagation network b) Kohonen network
 c) perceptron network d) agglomerative network
- (47) How the decision tree reaches its decision?
 a) Single test b) Two test
 c) Sequence of tests d) No test
- (48) Which of the following are real world applications of the SVM?

- a) Text and Hypertext Categorization
 b) Image Classification
 c) Clustering of News Articles
 d) All of these
- (49) Suppose you are using SVM with linear kernel of polynomial degree 2. Think that you increase the complexity (or degree of polynomial of this kernel). What would you think will happen?
 a) Increasing the complexity will overfit the data
 b) Increasing the complexity will underfit the data
 c) Nothing will happen since your model was already 100% accurate
 d) None of these
- (50) What do you mean by generalization error in terms of the SVM?
 a) How far the hyper plane is from the support vectors
 b) How accurately the SVM can predict outcomes for unseen data
 c) The threshold amount of error in an SVM
 d) None of these
- (51) The effectiveness of an SVM depends upon:
 a) Selection of Kernel
 b) Kernel Parameters
 c) Soft Margin Parameter C
 d) All of these
- (52) 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 above
- (53) Decision Trees are a type of
 a) Supervised machine learning
 b) Unsupervised machine learning
 c) Reinforcement machine learning
 d) Semi-supervised machine learning
- (54) What are the issues on which biological networks prove to be superior than AI networks?
 a) flexibility
 b) robustness & fault tolerance
 c) collective computation
 d) all of these
- (55) How many types of perceptron are there?
 a) 1
 b) 2
 c) 3
 d) 4
- (56) Hyperbolic Tangent i.e. Tanh Activation function is
 a) $f(x) = 1 / (1 + \exp(-x))$
 b) $f(x) = (1 - \exp(-2x)) / (1 + \exp(-2x))$
 c) $R(x) = \max(0, x)$
 d) None of these
- (57) ReLU is an activation function
 a) Rectified linear units
 b) Rotatory linear units
 c) Rectified non-linear units
 d) None of these
- (58) Perceptron is
 a) single layer neural network
 b) multi layer neural network
 c) none of these
 d) 3 layer neural network
- (59) How many input does bias vector need in ANN?
 a) 1
 b) 2
 c) 0
 d) 3
- (60) Which algorithm is used for solving temporal probabilistic reasoning?
 a) Hill-climbing search
 b) Hidden markov model
 c) Depth-first search
 d) Breadth-first search
- (61) Basic problem(s) of HMM are
 a) Evaluation
 b) Decoding
 c) Learning
 d) All of these
- (62) Which method is used for combining the predictions from different models?
 a) Evaluation
 b) Learning

- c) Decoding
- d) Bagging
- (63) What is Voting?
- a) Building multiple models (typically of the same type) from different subsamples of the training dataset.
- b) Building multiple models (typically of the same type) each of which learns to fix the prediction errors of a prior model in the chain
- c) Building multiple models (typically of differing types) and simple statistics (like calculating the mean) are used to combine predictions
- d) None of these
- (64) Stacking is used in which method?
- a) Random subspace
- b) Gradient Descent
- c) Blending
- d) All of these
- (65) Boosting aims to
- a) increase bias, not variance
- b) decrease bias and variance
- c) decrease bias, not variance
- d) increase bias and variance
- (66) Which of the following is/are true regarding an SVM ?
- a) For two dimensional data points, the separating hyperplane learnt by a linear SVM will be a straight line
- b) In theory, a Gaussian kernel SVM cannot model any complex separating hyperplane
- c) For every kernel function used in a SVM, one can obtain an equivalent closed form basis expansion
- d) Overfitting in an SVM is not a function of number of support vectors
- (67) Which among the following prevents overfitting when we perform bagging?
- a) The use of sampling with replacement as the sampling technique
- b) The use of weak classifiers
- c) The use of classification algorithms which are not prone to overfitting
- d) The practice of validation performed on every classifier trained
- (68) Averaging the output of multiple decision trees helps _____.
- a) Increase bias
- b) Increase variance
- c) Increase variance
- d) Decrease variance
- (69) In evaluation problem of HMM deals with
- a) What is probability that observations are generated by model
- b) What is most likely state sequence in model that produced the observations
- c) How to adjust model parameters to maximize
- d) None of these
- (70) K-Means clustering algorithm is example of which model?
- a) Connectivity models
- b) Centroid models
- c) Distribution models
- d) Density Models