



BRAINWARE UNIVERSITY
Term End Examination 2020 - 21
Programme – Master of Science in Computer Science
Course Name – Machine Learning
Course Code - MCS303B

Semester / Year - Semester III

Time allotted : 75 Minutes

Full Marks : 60

[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 60=60

1. *(Answer any Sixty)*

(i) Computers are best at learning

- | | |
|---------------|---------------|
| a) facts | b) concepts |
| c) Procedures | d) principles |

(ii) Which of the factors affect the performance of learner system does not include?

- | | |
|-------------------------------|-------------------------|
| a) Representation scheme used | b) Training scenario |
| c) Type of feedback | d) Good data structures |

(iii) Different learning methods does not include

- | | |
|-----------------|-----------------|
| a) Memorization | b) Analogy |
| c) Deduction | d) Introduction |

(iv) In language understanding, the levels of knowledge that does not include?

- | | |
|-----------------|--------------|
| a) Phonological | b) Syntactic |
| c) Empirical | d) Logical |

(v) A model of language consists of the categories which does not include

- | | |
|-----------------------|----------------------------|
| a) Language units | b) Role structure of units |
| c) System constraints | d) Structural units |

- (vi) The action 'STACK(A, B)' of a robot arm specify to _____
- a) Place block B on Block A
 - b) Place blocks A, B on the table in that order
 - c) Place blocks B, A on the table in that order
 - d) Place block A on block B
- (vii) Which of the following library is used for boosting generalized additive models
- a) gamBoost
 - b) gbm
 - c) ada
 - d) All of the Mentioned
- (viii) Which of the following is characteristic of best machine learning method?
- a) Fast
 - b) Accuracy
 - c) Scalable
 - d) All of the Mentioned
- (ix) Data used to optimize the parameter settings of a supervised learner model.
- a) training
 - b) testing
 - c) validation
 - d) verification
- (x) Machine learning is to
- a) Plot data
 - b) Analysis data
 - c) Finding pattern in data
 - d) None of these
- (xi) Supervised machine learning is divided into two groups
- a) Supervised and Unsupervised ML
 - b) Supervised and Nonsupervised ML
 - c) Classification and Regression
 - d) None of these
- (xii) Regression analysis is to
- a) Predict numeric answers
 - b) Predict nonnumeric answers
 - c) Predict string answers
 - d) None of these

(xiii) Classification Analysis is to

- a) Predict numeric answers
- b) Predict nonnumeric answers
- c) Predict string answers
- d) None of these

(xiv) Random Forest is a

- a) Clustering algorithm
- b) Feature Selection Algorithm
- c) Classification algorithm
- d) None of these

(xv) Support Vector machine is a

- a) Clustering algorithm
- b) Feature Selection Algorithm
- c) Classification algorithm
- d) None of these

(xvi) $Y=mx+c$. Here m is

- a) Y intercept
- b) X intercept
- c) Slope of line
- d) None of these

(xvii) Naive Bayes theorem is based on

- a) Probability
- b) Calculation
- c) Conditional Probability
- d) Conditional Randomness

(xviii) A multiple regression model has

- a) Only one independent variable
- b) More than one dependent variable
- c) More than one independent variable
- d) None of these

(xix) Logistics Regression is a

- a) Ternary Classifier
- b) Binary Classifier
- c) MultiValued Classified
- d) None of these

(xx) 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
- b) Depends on the type of problem

function is always zero

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

(xxi) 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 & Assumes that all the features in a dataset are independent
- d) None of these

(xxii) Which of the following sentence is FALSE regarding 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

(xxiii) Grid search is,

- a) Linear in D.
- b) Polynomial in D.
- c) Exponential in N.
- d) Linear in N.

(xxiv) Computational complexity of Gradient descent is,

- a) linear in D
- b) linear in N
- c) polynomial in D
- d) dependent on the number of iterations

(xxv) Adding more basis functions in a linear model... (pick the most probably option)

- a) Decreases model bias
- b) Decreases estimation bias
- c) Decreases variance
- d) Doesn't affect bias and variance

(xxvi) What do you mean by a hard margin?

- a) The SVM allows very low error in
- b) The SVM allows high amount of error in

classification

c) All of the above

classification

d) None of the above

(xxvii) The minimum time complexity for training an SVM is $O(n^2)$.

According to this fact, what sizes of datasets are not best suited for SVM's?

a) Large datasets

b) Small datasets

c) Medium sized datasets

d) Size does not matter

(xxviii) 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

(xxix) 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

(xxx) To calculate the Median

a) Middle value of samples

b) Arrange the samples in ascending order

c) Calculate middle position

d) All of these

(xxxi) The measures of dispersion include

a) Range

b) Standard Deviation

c) Variance

d) None of the above

(xxxii) The range is

a) Highest value-Lowest Value

b) Lowest Value- Highest value

c) Mean Value- Highest value

d) None of these

(xxxiii) Variance is

- a) Sample mean of the squared deviations from the arithmetic mean
- b) Arithmetic mean of the squared deviations from the sample mean
- c) Sample mean of the squared deviations from the sample mean
- d) None of the above

(xxxiv) Coefficient of the correlation ranges from

- a) -1 to +1
- b) 0 to +1
- c) -1 to 0
- d) None of the above

(xxxv) Chebysheff's theorem deals with

- a) Range
- b) Variance
- c) Standard deviation
- d) None of these

(xxxvi) Conditional probability is related with

- a) Naïve Bayes theorem
- b) Chebysheff's Theorem
- c) Pythagorean theorem
- d) None of the above

(xxxvii) What would you do in PCA to get the same projection as SVD?

- a) Transform data to zero mean
- b) Transform data to zero median
- c) Not possible
- d) None of these

(xxxviii) How does generalization performance change with increasing size of training set?

- a) Improves
- b) Deteriorates
- c) No Change
- d) None

(xxxix) Which of the following are recommended applications of PCA?

- a) To get more features to feed into a learning algorithm.
- b) Data compression: Reduce the dimension of your data, so that it takes up less memory / disk space

- c) Preventing overfitting: Reduce the number of features (in a supervised learning problem), so that there are fewer parameters to learn.
- d) As a replacement for (or alternative to) linear regression: For most learning applications, PCA and linear regression give substantially similar results.

(xl) K-fold cross-validation is

- a) linear in K
- b) quadratic in K
- c) cubic in K
- d) exponential in K

(xli) 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) Underfitting
- b) Nothing, the model is perfect
- c) Overfitting
- d) None of these

(xlii) 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

(xliii) Confusion matrix is used for

- a) Predicting values
- b) Assesses a model
- c) Summarizing of prediction results on a classification problem.
- d) None of these

(xliv) Accuracy formula for Confusion Matrix

- a) $(TN)/(TP+FN)$
- b) $(TP+TN)/(TP+TN+FP+FN)$
- c) $(TP+TN)/(TP+FN)$
- d) None of these

(xlv) Which is method of cross validation?

- a) K Fold
- b) Precision
- c) Recall
- d) None of these

(xlvi) LOOCV is

- a) Leave out one cross-validation
- b) Leave out one crypto-validation
- c) Leave one out crisis-validation
- d) None of these

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

- a) Regression
- b) Naïve Bayes
- c) Random Forest
- d) None of these

(xlvi) A two-layered neural network used for unsupervised clustering

- a) back propagation network
- b) Kohonen network
- c) perceptron network
- d) agglomerative network

(xlix) 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 & Is representative of the dataset as a whole
- d) None of the above

(l) 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

(li) 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 the above

(lii) Which of the following is a reasonable way to select the number of principal components "k"?

- a) Choose k to be the smallest value so that at least 99% of the variance is retained. b) Choose k to be 99% of m ($k = 0.99 * m$, rounded to the nearest integer).
c) Choose k to be the largest value so that 99% of the variance is retained. d) Use the elbow method

(liii) The time complexity of hierarchical clustering is

- a) linear b) quadratic
c) cubic d) none of these

(liv) Which metrics are used for deciding the closeness of two clusters?

- a) Euclidean distance b) Manhattan distance
c) Maximum distance d) All of these

(lv) Components of machine learning are

- a) Representation b) Evaluation
c) Optimization d) All of these

(lvi) PAC is

- a) Probably Approximately Correct Learning b) Probably Approximately Looping
c) Priority Approximately Correct Learning d) Priority Approximately Looping

(lvii) I have trained a classifier, and to evaluate its performance I perform a 10-fold validation. I have obtained the following accuracies on the validation set in each of the runs - 0.90, 0.98, 0.95, 0.98, 0.97, 0.96, 0.94, 0.99, 0.96, 0.96.

What is the mean accuracy?

- a) 0.93 b) 0.959
c) 0.98 d) 0.97

(lviii) Suppose that for some linear regression problem (say, predicting housing prices as in the lecture), we have some training set, and for our training set we managed to find some θ_0, θ_1 such that $J(\theta_0, \theta_1) = 0$. Which of the statements below must then be true?

- a) For this to be true, we must have $\theta_0 = 0$ and $\theta_1 = 0$ so that $h_{\theta}(x) = 0$
- b) Our training set can be fit perfectly by a straight line, i.e., all of our training examples lie perfectly on some straight line
- c) For this to be true, we must have $y(i) = 0$ for every value of $i = 1, 2, \dots, m$.
- d) We can perfectly predict the value of y even for new examples that we have not yet seen.

(lix) You run gradient descent for 15 iterations with $\eta = 0.3$ and compute $J(\eta)$ after each iteration. You find that the value of $J(\eta)$ decreases slowly and is still decreasing after 15 iterations. Based on this, which of the following conclusions seems most plausible?

- a) Rather than use the current value of η , it'd be more promising to try a larger value of η (say $\eta = 1.0$).
- b) Rather than use the current value of η , it'd be more promising to try a smaller value of η (say $\eta = 0.1$).
- c) $\eta = 0.3$ is an effective choice of learning rate.
- d) None of these

(lx) Which of the following statements are true?

- a) Gradient checking is useful if we are using one of the advanced optimization methods (such as in `fminunc`) as our optimization algorithm. However, it serves little purpose if we are using gradient descent.
- b) Computing the gradient of the cost function in a neural network has the same efficiency when we use backpropagation or when we numerically compute it using the method of gradient checking.
- c) Using gradient checking can help verify if one's implementation of backpropagation is bug-free.
- d) None of these