Avinashilingam Institute for Home Science and Higher Education for Women

(Deemed to be University under Category 'A' by MHRD, Estd. u/s 3 of UGC Act 1956)
Re-accredited with 'A+' Grade by NAAC. Recognised by UGC Under Section 12B
Coimbatore - 641 043, Tamil Nadu, India

Bachelor of Vocation Degree Examination – January 2021
III Semester

Class : III B. Voc **Time: 3 Hours**
Major : Artificial Intelligence and Machine Learning **Max. Marks: 100**

19VAIC10 Fundamentals of Machine Learning

Part A **10 x 1 = 10**

Choose the Correct Answer

1. The problem of finding hidden structure in unlabeled data is called
   a. Supervised learning b. Unsupervised learning
   c. Reinforcement learning d. Enforcement

2. You are given data about seismic activity in Japan, and you want to predict the magnitude of the next earthquake, this is in an example of
   a. Supervised Learning b. Unsupervised Learning
   c. Serration d. Dimensionality reduction

3. Suppose we train a hard-margin linear SVM on n > 100 data points in R 2, yielding a hyperplane with exactly 2 support vectors. If we add one more data point and retrain the classifier, what is the maximum possible number of support vectors for the new hyperplane (assuming the n+1 points are linearly separable)?
   a. 2 b. 3 c. n d. n+ 1

4. _____ algorithm looks for the minimum value of the error function in weight space using a technique called the delta rule or gradient descent.
   a. Propagation b. Back propagation c. Neural Network d. none of the above

5. What strategies can help reduce over fitting in decision trees?
   i) Pruning.
   ii) Make sure each leaf node is one pure class.
   iii) Enforce a minimum number of sample in leaf nodes.
   iv) Enforce a maximum depth for the tree
   a. i , iii and iv b. i and ii c. iii and iv d. i and iv

6. In which of the following cases will K-means clustering fail to give good result?
   i) Data points with outliers
   ii) Data points with different densities
   iii) Data points with non convex shapes
   a. i and ii b. ii and iii c. i and iii d. all the above

7. When performing regression or classification, which of the following is the correct way to pre-process 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

8. A _____ is a search heuristic that is inspired by Charles Darwin's theory of natural evolution.
   a. Genetic algorithm b. Back propagation c. Neural Networks d. none of the above

9. Which of the following statement about Naive Bayes is incorrect?
   a. Attributes are equally important
   b. Attributes are statistically dependent of one another given the class value
   c. Attributes are statistically independent of one another given the class value
   d. Attributes can be nominal or numeric

10. _____ methods comprise a class of algorithm for sampling from a probability distribution.
    a. Markovechain Monte Carlo b. Boosting
    c. Bagging d. Statistics

Part B **5 x 6 = 30**

Answer ALL questions

Each answer should not exceed 400 words or two pages

11.a. i) Distinguish supervised learning from unsupervised learning.
ii) Give an account on parameters in a perception network and its significance.

(or)

11.b. Describe the version spaces and the candidate elimination algorithm.

12.a. Write short notes on Multi layer perception.

(or)

12.b. What are the steps involved in back propagation algorithm? Discuss it.

13.a. Provide outline of the ID3 algorithm used for inducing decision tree from the training tuples. Also list down the different attribute selection measures used in the process of decision tree construction.

(or)

13.b. i) List the advantages of bagging over boosting.
ii) State the principle in self organizing feature map. Where is it applicable?

14.a. Explain the Evolutionary Learning techniques.

(or)

14.b Write short notes on Markov Decision Process.

15.a Give an account on Proposal Distribution.

(or)

15.b. Explain the Markov Random Fields.

Part C **5 x 12 = 60**

Answer ALL questions

Each answer should not exceed 800 words or four pages

16.a. Elaborate on the types of machine learning and discuss the components in the design of a learning system.

(or)

16.b .Explain the version spaces and the candidate elimination algorithm.

17.a. Describe the working behavior of support vector machine with diagrams.

(or)

17.b. What are the steps involved in back propagation algorithm? Discuss it.

18.a. Explain the steps in k-means algorithm. Cluster the following set of 4 objects into clusters using k-means A(3,5),B(4,5), C(1,3),D(2,4). Consider the objects A and C as the initial cluster centers.

(or)

18.b. Write short notes on:
i) Ensemble Learning.
ii) Gaussian Mixture Models.

19.a. State the algorithm and the operators of genetic algorithm.

(or)

19.b. i) Explain all the parameters of a hidden Markov model.
ii) Determine the Principal Components for the given Two-Dimentional dataset. (1,2), (2,4), (3,6)

20.a. Explain the Bayesian networks with an example.

(or)

20.b. Describe the Markov chain Monte Carlo Methods.

****