ST. JOSEPH'S COLLEGE (AUTONOMOUS), BENGALURU-27

M.Sc.(BIG DATA ANALYTICS) —III SEMESTER

SEMESTER EXAMINATION: OCTOBER 2021

(Examination conducted in January —March 2022)

BDA 3321: MACHINE LEARNING II

TIME: 2.5 Hrs   **MAXIMUM MARKS: 70**

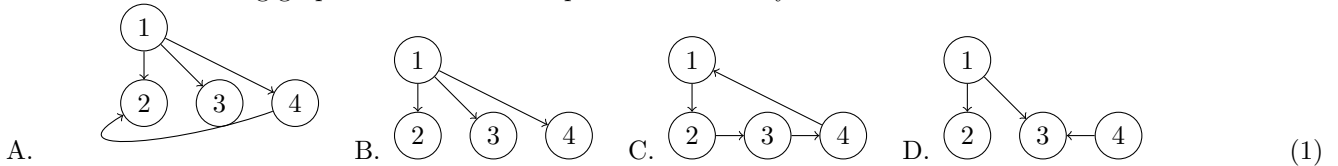**This paper has 3 printed pages and 3 parts.**

PART 1

Answer all questions. More than one options may be correct.   **($1 \times 20 = 20$)**

1. A decission tree must be binary.
   A. True   B. False   (1)

2. A decision tree may not be rooted.
   A. True   B. False   (1)

3. A decision tree can be used for:
   A. Only classificatin   B. Only Regression   C. Both classificatin and Regression   (1)

4. Naive Bayes and logistic regression can both be used for same data, always.
   A. True   B. False   (1)

5. A neural network is also known as a Percptron.
   A. True   B. False   (1)

6. Clustering algorithms use:
   A. Supervised Machine Learning Methods   B. Unsupervised Machine Learning algorithm
   C. Clustering cannot be done via machine learning.   (1)

7. Gradient decent:
   A. Always finds the global mimimum.   B. Sometimes finds the global mimimum.
   C. Aims to find local minimum.   D. Always finds the local mimimum.   (1)

8. K means clustering is:
   A. Supervised ML algorithm   B. Unsupervised ML algortihtm.   C. Same as KNN.   (1)

9. EM and MLE can be used interchangeably.
   A. True   B. False   (1)

10. Logistic regression and Naive Bayes are used to estimate the same parameters.
    A. True   B. False   (1)

11. Genetic algorithms are popular because:
    A. They are easy to build   B. They perform well for biological systems.
    C. Computations can be easily paallelized.   (1)

12. In genetic algorithm, crossover is:
    A. Genetic operator   B. A way of representing a hypothesis.
    C. Used to generate a fitness function.   (1)

13. Consider the variable outlook that can take 3 values ∈{Sunny, Overcast, Rainy}. If we use a 3 bit string to represent a hypothes. Then, the hypothesis "Sunny or Rainy" is best represented by:
    A. (1,1,1)   B. (1,0,1)   C. (1,1,0)   D. (1,1,0)   (1)

14. Consider a single point crossover genetic operator operating on the 11101 and 10010. Let the randomly chosen crossover point be 3. Then the offsprings we get are:
    A. 11011   B. 10001   C. 10101   D. 11110   (1)

15. Single point crossover is the only kind of crossover genetic operator used in genetic algortihtms.
    A. True   B. False   (1)

16. Assume a single point mutation applied to the hypothesis 1001010. Let the randomly chosen single point be 3. Then the resulting offspring is:
    A. 1001011   B. 1011010   C. 0001010   D. 1101010   (1)

17. Genetic algorithms aim to replicate the bioloigical evolution process in a machine learning setting.
    A. True   B. False   (1)

18. Which of the following graphs can be used to represent a naive bayes classifier?

    A.   B.   C.   D.   (1)

19. Can a topological ordering be found for the following graph:

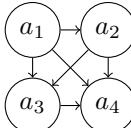   A. Yes   B. No   (1)

20. A topological ordering can be found using DFS.
    A. True   B. False   (1)

PART B

Answer ANY SIX questions.   (**6 × 5 = 30**)

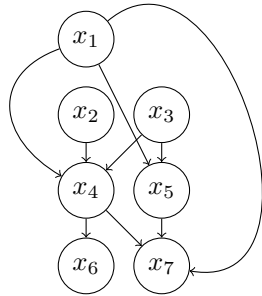1. Briefly explain random forests.   (5)

2. Briefly explain boosting.   (5)

3. Briefly explain SVMs.   (5)

4. Explain the steps involved in EM algorithm. Is there a property of EM that can be used for debugging purposes? Explain.   (5)

5. State and explain with examples the different kinds of genetic operators.   (5)

6. Discuss the general structure of genetic algorithms..   (5)

7. Discuss markov models and hidden markov models. Explain when it is appropriate to use them.   (5)

8. Explain briefly a Directed Graphical Models. For the following DGMs, give the joint probability distribution that the network represents.

1. Write p(a_1, a_2, a_3, a_4) for this DGM.

2. Write p($x_1, x_2, x_3, x_4, x_5, x_6, x_7$) for this DGM.

(5)

PART C

Answer ANY TWO questions. $(\mathbf{2 \times 10 = 20})$

1. Explain GABIL as an example of genetic algorithms. (10)

2. How can we model languages using makov models? Explain. (10)

3. Explain two different clustering algorithms. When should one use clustering? For your chosen algorithms, explain when one is used over the other. (10)