

KATHMANDU UNIVERSITY
End Semester Examination [C]
April, 2022

Marks Scored:

Level : B.Sc.

Course : COMP 484

Year : IV

Semester : I

Exam Roll No. :

Time: 30 mins.

F. M. : 10

Registration no.:

Date :

SECTION "A"

[20Q. × 0.5 = 10 marks]

Encircle the most appropriate option.

1. Higher value of entropy would mean which of the following?
 - a. Attributes of instances cannot be related to any class of values
 - b. Attribute of instances belong to just one class of values
 - c. Attributes of instances belong to a mixed class of values
 - d. None of these
2. The average positive difference between computed and desired outcome values is:
 - a. Mean absolute error
 - b. Mean squared error
 - c. Mean positive error
 - d. Root mean squared error
3. Averaging the output of multiple decision trees helps:
 - a. Decrease variance
 - b. Increase bias
 - c. Increase variance
 - d. Decrease bias
4. Overfitting suffers from:
 - a. Poor generalization
 - b. Poor specialization
 - c. Both a and b
 - d. Neither a nor b
5. FIND-S algorithm starts from the most specific hypothesis and generalize it by considering only _____ examples.
 - a. Negative
 - b. Positive
 - c. Negative or Positive
 - d. None of the above
6. The general limitations of backpropagation rule is/are:
 - a. Scaling
 - b. Slow convergence
 - c. Local minima problem
 - d. All of the above
7. Which of the following is assumed by the FIND-S algorithm?
 - a. H indeed contains the target concept c
 - b. Training examples are error-free
 - c. At each step, the hypothesis is generalized just as much to cover the current observed positive example
 - d. All of the above
8. What does an entropy value of 1 mean?
 - a. All the members of S belong to one class either positive or negative
 - b. S has equal number of positive and negative members
 - c. S has a mix of non-uniform number of positive and negative members
 - d. None of the above

9. Which of the following systems is devoid of any inductive bias?
 - a. Rote-learner
 - b. Candidate Elimination
 - c. FIND-S
 - d. None of the above
10. Parameter estimation is all about:
 - a. Identifying input parameters
 - b. Identifying output parameters
 - c. Identifying model parameters
 - d. All of the above
11. Which of the following is true about Naïve Bayes?
 - a. Assumes that all features in a dataset are equally important
 - b. Assumes that all features in a dataset are independent
 - c. Both of a and b
 - d. Neither a nor b
12. Which of the following is provided by the Bayesian Network?
 - a. Complete description of the problem
 - b. Partial description of the network
 - c. Complete description of the domain
 - d. All of the above
13. The difference between the expected sample value and the estimated parameter value is called
 - a. Bias
 - b. Error
 - c. Contradiction
 - d. Difference
14. Which of the following is a subset of population?
 - a. Distribution
 - b. Sample
 - c. Data
 - d. Set
15. The sampling error is defined as:
 - a. Difference between population and parameter
 - b. Difference between sample and parameter
 - c. Difference between population and sample
 - d. Difference between parameter and sample
16. A nearest-neighbor approach is best used:
 - a. with large-sized database
 - b. when irrelevant attributes have been removed from the data
 - c. when a generalized model of the data is desirable
 - d. when an explanation of what has been found is of primary importance
17. What do you expect will happen with bias and variance as you increase the size of the training data?
 - a. Bias increases and Variance increases
 - b. Bias decreases and Variance increases
 - c. Bias decreases and Variance decreases
 - d. Bias increases and Variance decreases
18. What does Computational Learning Theory (CLT) comprise of?
 - a. Sample complexity
 - b. Computational complexity
 - c. Mistake Bound
 - d. All of the above
19. When you find noise in the data, which of the following option would you consider in kNN?
 - a. Increase the value of k
 - b. Decrease the value of k
 - c. Noise is not dependent on the value of k
 - d. None of the above
20. Which of the following techniques would perform better for reducing dimensions of a dataset?
 - a. Removing columns which has too many missing values
 - b. Removing columns which have high variance in data
 - c. Removing columns with dissimilar data trends
 - d. None of the above

KATHMANDU UNIVERSITY
End Semester Examination [C]
April, 2022

Level : B.Sc.
Year : IV
Time : 2 hrs. 30 mins.

Course : COMP 484
Semester: I
F. M. : 40

SECTION "B"

[6 Q. × 4 = 24 marks]

Attempt *ANY SIX* questions.

1. What is the major shortcoming of Concept Learning as a Machine Learning Approach? Differentiate between the hypothesis space and version space.
2. Define the terms – “Instances”, “Hypotheses”, “Target concept” and “Training Examples”.
3. Distinguish between “Preference bias” and “Restriction bias”. Which of the bias do the Decision trees have and how does it impact in the overall classification?
4. Explain the architecture and working mechanism of Artificial Neural Networks.
5. Explain the two difficulties with limited set of data – “Bias in the estimate” and “Variance in the estimate”.
6. State “Bayes theorem” and explain the parameters involved in the theorem. Derive the maximum posterior hypothesis (h_{MAP}) and maximum likelihood hypothesis (h_{ML}) from the theorem.
7. Distinguish between “lazy” and “eager learning” with suitable examples. What is the key factor that differentiates “instance-based learning” and “general supervised learning models”?

SECTION "C"

[2 Q. × 8 = 16 marks]

Attempt *ANY TWO* questions.

8. Explain the working mechanism of “Decision trees”. How are “Entropy” and “Information Gain” used for classification by Decision trees? Shed light on some of the issues of Decision trees classification and the possible solutions to the problems.
9. Discuss the stochastic gradient descent version of the Backpropagation algorithm for feedforward networks containing two layers of sigmoid units. What is the “local minima” problem and what are some of the common heuristics to attempt to alleviate the problem of the local minima?
10. How does the k-nearest neighbor algorithm work? Discuss the curse of dimensionality problem in the context of the k-nearest neighbor algorithm.

