KATHMANDU UNIVERSITY
End Semester Examination [C]
May/June, 2019

Marks Scored:

Level : B.E. /B.Sc.
Year : IV

Course : COMP 484
Semester: I

Exam Roll No. :

Time: 30 mins.

F. M. : 10

Registration No.:

Date : 31 MAY 2019

SECTION "A"
[20Q. × 0.5 = 10 marks]

Choose the most appropriate answer among the given options and ***encircle*** *the letter of your choice.*

1. In neural networks, nonlinear activation functions such as sigmoid, tanh, and ReLU
   a. speed up the gradient calculation in backpropagation, as compared to linear units
   b. are applied only to the output units
   c. help to learn nonlinear decision boundaries
   d. always output values between 0 and 1

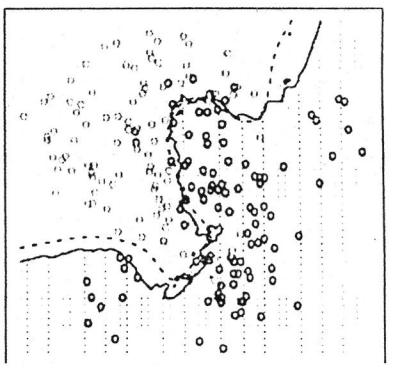
2. Consider k-fold cross validation. With a higher number of folds, the estimated error will be on average:
   a. Higher b. Lower c. Same d. Can't tell

3. Machine learning techniques differ from statistical techniques in that machine learning methods :
   a. typically assume an underlying distribution for the data.
   b. are better able to deal with missing and noisy data.
   c. are not able to explain their behavior.
   d. have trouble with large-sized datasets.

4. The values input into a feed-forward neural network
   a. may be categorical or numeric.
   b. must be either all categorical or all numeric but not both.
   c. must be numeric.
   d. must be categorical.

5. The standard error is defined as the square root of this computation:
   a. The sample variance divided by the total number of sample instances.
   b. The population variance divided by the total number of sample instances.
   c. The sample variance divided by the sample mean.
   d. The population variance divided by the sample mean.

6. Which value of k in the k-nearest neighbors algorithm generates the solid decision boundary depicted here? There are only 2 classes. (Ignore the dashed line, which is the Bayes decision boundary.)
   a. *k=1*
   b. *k=10*
   c. *k=2*
   d. *k=100*



7. Classification problems are distinguished from estimation problems in that:
   a. classification problems require the output attribute to be numeric
   b. classification problems require the output attribute to be categorical.
   c. classification problems do not allow an output attribute.
   d. classification problems are designed to predict future outcome.

8. The test set accuracy of a backpropagation neural network can often be improved by:
   a. increasing the number of epochs used to train the network.
   b. decreasing the number of hidden layer nodes.
   c. increasing the learning rate.
   d. decreasing the number of hidden layers.

9. The average positive difference between computed and desired outcome values.
   a. root mean squared error
   b. mean squared error
   c. mean absolute error
   d. mean positive error

10. Which statement is true about neural network and linear regression models?
   a. Both models require input attributes to be numeric.
   b. Both models require numeric attributes to range between 0 and 1.
   c. The output of both models is a categorical attribute value.
   d. Both techniques build models whose output is determined by a linear sum of weighted input attribute values.

11. A nearest neighbor approach is best used
   a. with large-sized datasets.
   b. when irrelevant attributes have been removed from the data.
   c. when a generalized model of the data is desireable.
   d. when an explanation of what has been found is of primary importance.

12. A statement to be tested is:
   a. Theory b. Procedure c. Principle d. Hypothesis

13. With Bayes theorem the probability of hypothesis H, specified by P(H)is referred to as :
   a. an a priori probability
   b. a conditional probability
   c. a posterior probability
   d. a bidirectional probability

14. The process of forming general concept definitions from examples of concepts to be learned.
   a. Deduction
   b. Abduction
   c. Induction
   d. Conjunction

15. Computers are best at learning
   a. facts
   b. concepts
   c. procedures
   d. principles

16. With Bayes classifier, missing data items are :
   a. treated as equal compares
   b. treated as unequal compares
   c. replaced with a default value
   d. ignored

17. For an image recognition problem (recognizing a cat in a photo), which architecture of neural network would be better suited to solve the problem?
    a. Convolution Neural Network
    b. Recurrent Neural Network
    c. Multi Layer Perceptron
    d. Perceptron

18. Which of the following holds true for the "validation" set?
    a. It is interchangeably referred to as the "test set"
    b. It is a "pretend test set"
    c. It is a subset of the "training set"
    d. Both "b" and "c"

19. Which of the following is not true about "variance in the estimate"?
    a. The smaller the set of test examples, the greater the expected variance.
    b. The measured accuracy can still vary from the true accuracy depending on the makeup of the particular set of examples.
    c. This problem of variance in the estimate arises when we have a limited set of data
    d. Variance increases as the size of the data sample increases

20. Overfitting suffers from:
    a. Poor generalization
    b. Poor specialization
    c. Both of the above
    d. None of the above

KATHMANDU UNIVERSITY
End Semester Examination [C]
May/June, 2019

31 MAY 2019

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

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

SECTION "B"
[6Q. × 4 = 24 marks]

Attempt *ANY SIX* questions.

1. Discuss supervised and unsupervised learning methods with suitable examples.

2. Consider the following set of training examples:

Instance	Classification	a_1	a_2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

What is the information gain of a_2 relative to the training examples?

3. Given the following dataset with prior and conditional probabilities:

P(weather=sunny|class=go-out) = 0.8
P(weather=rainy|class=go-out) = 0.2
P(weather=sunny|class=stay-home) = 0.4
P(weather=rainy|class=stay-home) = 0.6
P(car=working|class=go-out) = 0.8
P(car=broken|class=go-out) = 0.2
P(car=working|class=stay-home) = 0.2
P(car=broken|class=stay-home) = 0.8
P(class=go-out) = 0.5
P(class =stay-home)=0.5

Predict using Naïve Bayes theorem which of the classes ("go-out" or "stay-home") is more probable for the instance "weather=sunny, car=working".

4. What is the trade-off between bias and variance? What measures do you make in your model to reduce bias and/or variance?

5. Explain the difference between the perceptron learning rule and the delta rule for a single layer network. In particular, from what principles are the rules derived and under what conditions do the algorithms converge and find a solution that classifies all training examples correctly.

6. How does the k-Nearest Neighbor algorithm work? What is a Voronoi diagram?

7. Explain your understanding of the Probably Approximately Correct (PAC) learning model. How does it contrast with the agnostic learning model?

SECTION "C"

[2Q. × 8 = 16 marks]

Attempt *ANY TWO* questions.

8. What is the key idea in the CANDIDATE-ELIMINATION algorithm? Define the term "version space". Explain the most specific and general hypotheses. [3+3+2]

9. Discuss Decision trees as a classification technique. Shed light on the characteristics of problems to which Decision trees are more suitable. [4+4]

10. State and define Bayes theorem. What is a maximum a posteriori (MAP) hypothesis (h_{MAP})? How is h_{MAP} determined? How do you derive maximum likelihood hypothesis (h_{ML}) from h_{MAP}? [2+2+4]