

KATHMANDU UNIVERSITY
End Semester Examination
February/March, 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.:

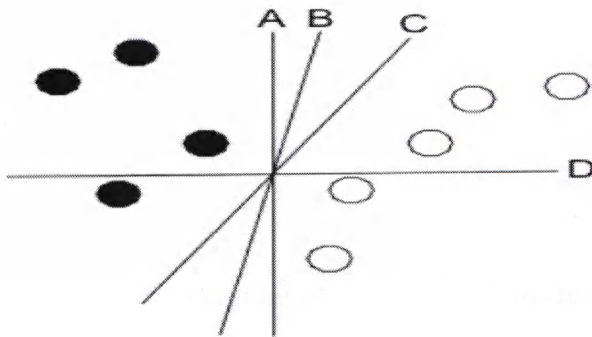
Date 0:3 MAR 2019

SECTION "A"

[20 Q. \times 0.5 = 10 marks]

Circle the most appropriate answer.

- Which of the following is FALSE regarding regression?
 - It relates inputs to outputs
 - It is used for prediction
 - It may be used for interpretation
 - It discovers causal relationship
- K-fold cross validation is:
 - Linear in K
 - Quadratic in K
 - Cubic in K
 - Exponential in K
- You observe the following while fitting a linear regression to the data:
As you increase the amount of training data, the test error decreases and the training error increases. The train error is quite low (almost what you expect it to), while the test error is much higher than the train error. What do you think is the main reason behind this behavior? Choose the most probable option:
 - High variance
 - High model bias
 - High estimation bias
 - None of the above
- Adding more bias functions in a linear model...(pick the most probable option)
 - Decreases model bias
 - Decreases estimation bias
 - Decreases variance
 - Does not affect bias and variance
- Evolution strategies have:
 - Random parents selection
 - Uniform mutation
 - Recombination by partially mapped crossover
 - Fitness proportional survivor selection
- Which separation line would SVM most likely choose?

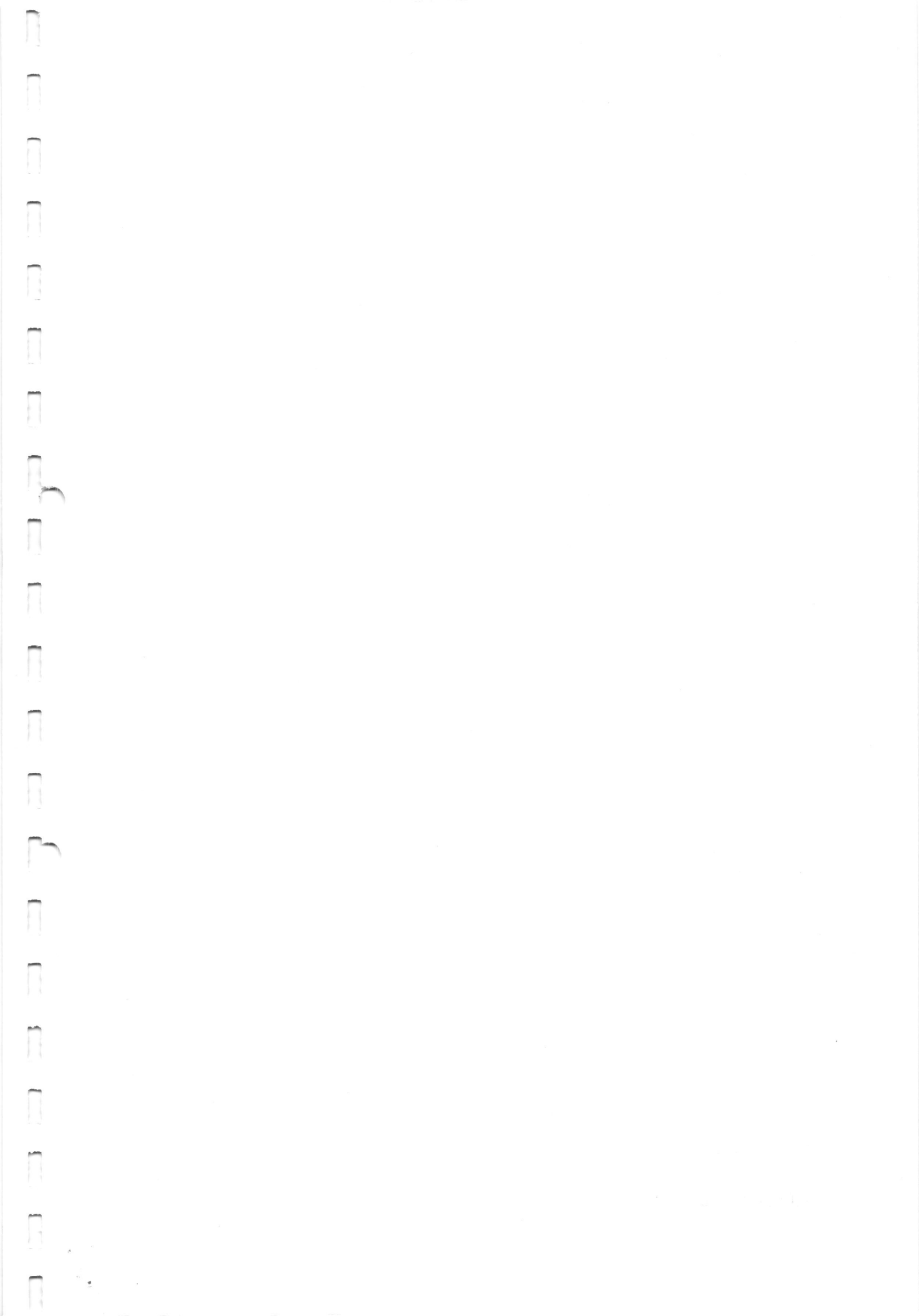


- Line A
- Line B
- Line C
- Line D

7. Epochs represent the total number of:
 - a. Input layer nodes
 - b. Passes of the training data through the network
 - c. Network nodes
 - d. Passes of the test data through the network
8. During backward propagation training, the purpose of the delta rule is to make weight adjustments so as to:
 - a. Minimize the number of times the training data must pass through the network
 - b. Minimize the number of times the test data must pass through the network
 - c. Minimize the sum of absolute differences between computed and actual outputs
 - d. Minimize the sum of squared error differences between computed and actual output
9. Neural network training is accomplished by repeatedly passing the training data through the network while
 - a. Individual network weights are modified
 - b. Training instance attribute values are modified
 - c. The ordering of the training instance is modified
 - d. Individual network nodes have the coefficients on their corresponding functional parameters modified
10. The values input into a feed-forward neural network:
 - a. May be categorical or numeric
 - b. May be either all categorical or all numeric but not both
 - c. Must be numeric
 - d. Must be categorical
11. 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
12. Selecting data so as to assure that each class is properly represented in both the training and test set:
 - a. Cross validation
 - b. Stratification
 - c. Verification
 - d. Bootstrapping
13. The hypothesis of no significant difference:
 - a. Nil
 - b. Invalid
 - c. Null
 - d. Void
14. Data used to optimize the parameter settings of a supervised learner model:
 - a. Training
 - b. Test
 - c. Verification
 - d. Validation
15. Bootstrapping allows us to:
 - a. Choose the same training instance several times
 - b. Choose the same test set instance several times
 - c. Build models with alternative subsets of the training data several times
 - d. Test a model with alternative subsets of the test data several times

03 MAR 2019

16. A single best representative of a class:
a. Mean b. Centroid c. Signature d. Prototype
17. A genetic learning operation that creates new population elements by combining parts of two or more existing elements:
a. Selection b. Crossover c. Mutation d. Absorption
18. The computational complexity as well as the explanation offered by a genetic algorithm is largely determined by the:
a. Fitness function
b. Techniques used for crossover and mutation
c. Training data
d. Population of elements
19. Supervised learning and unsupervised clustering both require at least one :
a. Hidden attribute b. Output attribute c. Input attribute d. Categorical attribute
20. Which of the following does not deal with training the model with labeled examples?
a. Decision Trees b. Concept Learning
c. Instance-based Learning d. None of the above



KATHMANDU UNIVERSITY
End Semester Examination
February/March, 2019

03 MAR 2019

Level : B.E. /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. Define Concept learning. Distinguish between the most general hypothesis and the most specific hypothesis. [2+2]
2. Given the following attributes and their respective values count:

Attributes	Sky	AirTemp	Humidity	Wind	Water	Forecast
# Values	3	2	2	2	2	2

Explain why the size of the hypothesis space in this *Enjoy Sport* learning task is 973. [4]
3. Define the Decision Tree Learning method. Construct a decision tree for the following expression:
(Outlook = Sunny AND Humidity = Normal) OR (Outlook = Overcast) OR (Outlook = Rain AND Wind = Weak) [2+2]
4. What is the learning objective in the Backpropagation algorithm? Does the backpropagation learning algorithm guarantee to find the global optimum solution? [4]
5. Explain the terms "bias" and "variance" in the context of Machine Learning with suitable examples. [4]
6. Distinguish between "lazy" and "eager" learning. Provide examples of each learning algorithms. [4]
7. Explain the terms "reward", "policy", "state value" and "action value function" in the context of reinforcement learning. What is an optimal policy and how is it computed from the state value function? [4]

SECTION "C"

[2 Q. × 8 = 16 marks]

Attempt ANY TWO questions.

8. State Bayes theorem. What is a maximum a posteriori (MAP) hypothesis? How is MAP determined? Consider a medical diagnosis problem in which there are two alternative hypotheses: 1) that the patient has a particular form of cancer, and 2) that the patient does not. The available data is from a particular laboratory test with two possible outcomes: + (positive) and - (negative). We have the following information for the given problem from the population:

$$\begin{aligned} P(\text{cancer}) &= 0.08, P(\text{not cancer}) = 0.992 \\ P(+|\text{cancer}) &= .98, P(-|\text{cancer}) = .02 \\ P(+|\text{not cancer}) &= .03, P(-|\text{not cancer}) = .97 \end{aligned}$$

Suppose, we now observe a new patient for whom the lab test returns a positive result. The maximum a posteriori hypothesis results for the new patient are as given below:

$$P(+ | \text{cancer}) P(\text{cancer}) = .0078$$

$$P(+ | \text{not cancer}) = .0298$$

Since $h_{\text{MAP}} = \text{not cancer}$, we can say that the patient does not have cancer.

Now, suppose the doctor decides to order a second laboratory test for the same patient and suppose the second test returns a positive test as well. What are the posterior probabilities of cancer, $P(\text{cancer} | ++)$ and not cancer, $P(\text{not cancer} | ++)$ following these two tests? Assume that the two tests are independent. [2+ 1+1 + 4]

9. What is meant by the term inductive bias of a learner? What are the inductive biases of the Find-S algorithm, candidate elimination algorithm and decision tree learning? What is the difference between restriction bias and preference bias? [2 + 3 + 3]
10. Explain what is meant by PAC-learnable. What is the definition of PAC-learnability for a concept class? What is the sample complexity of a learning problem and how does it relate to the notion of PAC-learnability. [3 + 2 + 3]