

Indian Institute Of Technology Patna  
Department of Computer Science and Engineering

Mid-Sem Exam : CS-6123 Advance Pattern Recognition

27th September 2024

Full Marks: 60

Duration: 2 hours

**Instruction:** Make reasonable assumptions as and whenever necessary. Markings will be based on the correctness and soundness of the answers. Proper indentation and appropriate comments (if required) are mandatory. You can answer the questions in any sequence. However, answers to all the parts of any particular question should appear together.

1. Explain the process of building a decision tree model. How does the model split the data at each node? Discuss one of the criteria used for splitting and how overfitting can be handled in decision trees. (10 marks)

2. Discuss the assumptions of logistic regression. How does logistic regression handle binary classification problems, and what is the role of the sigmoid function in this context? Also, explain how the model's coefficients are estimated. (10 marks)

3. Explain the concept of Support Vector Machine (SVM). How does SVM find the optimal hyperplane for classification? Also, discuss the significance of the margin and support vectors in this context. (10 marks)

4. Perform the principal component analysis of the following data- (10 marks)

CLASS 1 :  $X = 2, 3, 4$     $Y = 1, 5, 3$

CLASS 2 :  $X = 5, 6, 7$     $Y = 6, 7, 8$

5. Given the following two classes of data points: (10 marks)

Class 1:  $X_1 = \{(2,3), (3,4), (4,5)\}$    Class 2:  $X_2 = \{(6,8), (7,9), (8,10)\}$

a. Calculate the mean vectors for each class.

b. Compute the within-class scatter matrix  $S_W$ .

c. Find the between-class scatter matrix  $S_B$ .

d. Compute the LDA projection vector  $w$  using  $S_W^{-1}S_B$ .

e. Project the points from both classes onto the new LDA vector  $w$ .

6. Explain the concepts of bagging and boosting in the context of ensemble methods. Compare and contrast the two techniques. Write the steps of AdaBoost. (10 marks)