

**National Institute of Technology Calicut**  
**Department of Computer Science and Engineering**  
**S8 - B. Tech. (CSE)**  
**Machine Learning (CS4044D)**  
**End Semester Exam (Winter 2019-20)**

Total Time: 2 Hrs

Max Marks: 20

**Notes:**

- All questions are compulsory.
- Total time is **Two Hours** which includes downloading the question paper, writing answers, taking snapshots of answer sheets, making a single PDF for all answers, and uploading PDF.
- Write your name and roll number in your answer sheets.
- If any plagiarism is found in answer scripts, strict action will be taken against respective group of students by concerned institute committee.

1. Suppose you want to hire a candidate to serve your organization. You take decision based on K-Nearest neighbour (K-NN) classifier. Training data set is available in the following table in which each data point (each candidate) has two features; First, grade based on qualification & experience (X1) and second, expected salary by candidate (X2). What will be the decision if one candidate has X1=4 and X2=8. Follow step-by step procedure to decide. Assume that number of nearest neighbour, K=03. **[05 Marks]**

Candidate	X1	X2	Decision
Candidate 1	8	8	No
Candidate 2	8	5	No
Candidate 3	4	5	Yes
Candidate 4	2	5	Yes


2. Mention two scenarios where Support Vector Machine (SVM) is less effective (you don't recommend to apply SVM). How you can lower the bias and variance in SVM if they occurred. Why and how do you use kernels function in SVM. **[05 Marks]**
3. "The planet Jupiter and its moons are in effect a mini solar system and Jupiter itself is often called a star that never caught fire." **[04 marks]**

Write the final output after applying text preprocessing task on the above sentence:-

- a. Sentence Tokenization

- b. Word Tokenization
- c. Stopwords removal
- d. Lemmatization
- e. Pos tagging

4. Calculate **Accuracy** and **F1 score** for the given confusion matrix: **[6 marks]**

		Predicted Output		
				
Actual Output		A	B	C
	A	100	20	12
	B	30	15	0
	C	0	22	18