Part I

1. a) What does the learning rate do in Back-Propagation training? [2]
   b) Describe what is likely to happen when a learning rate is used that is too large, and when one is used that is too small. How can one optimize the learning rate? [3]
   c) Explain the main reasons why a Back-Propagation training algorithm might not find a set of weights which minimizes the training error for a given feed-forward neural network. [3]
   d) Explain the purpose of the momentum term that is often included in the Back-Propagation learning algorithm. [2]

2. a) The Pants Pizza Parlour sells pizzas with optional toppings: pepperoni, pineapple and pickled onion. Every day this week you have tried a pizza (A to E) and kept a record of which you liked: [6]

	Pepperoni	Pineapple	Pickled onion	Liked
A	True	True	True	False
B	True	False	False	True
C	False	True	True	False
D	False	True	False	True
E	True	False	False	True

   Using Hamming distance throughout, show how the 3-NN classifier with majority voting would classify {pepperoni = false, pineapple=true, pickledOnion=true}.
   b) What are the drawbacks of using k-NN algorithm to classify objects? Also, provide the remedies to overcome the drawbacks. [4]

3. Use Naïve Bayes Classifier for the dataset given in the following table to determine the Class for an instance with GATE qualified = No, Publications = No, Written Test qualified = No, Interview performance = Bad. [10]

Id	Gate Qualified	Publications	Written Test Qualified	Interview Performance	Decision
1	Yes	Yes	No	Bad	Reject
2	No	Yes	Yes	Bad	Reject
3	Yes	No	No	Good	Accept
4	No	No	Yes	Bad	Reject
5	No	Yes	No	Bad	Reject
6	Yes	No	Yes	Good	Accept
7	No	Yes	Yes	Good	Accept
8	Yes	Yes	No	Good	Accept
9	No	No	Yes	Good	Reject
10	No	Yes	No	Bad	Reject
11	No	No	No	Good	Reject
12	No	Yes	No	Good	Accept
13	Yes	Yes	Yes	Bad	Accept
14	Yes	No	No	Bad	Reject
15	Yes	No	Yes	Bad	Accept

4. Describe the principal component analysis method: what is it used for, what is the algorithm, what are limitations? [10]

5. a) Write short notes on any two of the followings: [6]
      I. Data space and feature space [4]
      II. Support vector
      III. Margin in Support Vector Machine
   b) Suppose a support vector machine for separating pluses from minuses finds a plus support vector at the point x1= (1, 0), a minus support vector at x2= (0, 1). You are to determine values for the classification vector w and the threshold value b.

6. a) Perform K-means clustering on all the points in the following table, where K=2. Randomly select the initial seeds and perform the algorithm for two iterations. [7+3]

Points	X co-ordinate	Y co-ordinate
p1	1	9
p2	2	10
p3	7	4
p4	10	3
p5	5	9
p6	7	2
p7	3	8
p8	4	10
p9	8	1
p10	9	3

	b) Describe the major drawbacks of K-means algorithm for clustering.	
	Part II	
7.	a) Derive the Lagrangian for the optimization problem as defined by linear SVM – separable case. b) A linearly separable dataset is given in the following Table. Predict the class of (0.6, 0.8) using a support vector machine classifier.	[8] [12]

X_1	X_2	Y	Lagrange Multiplier
0.3858	0.4687	+1	65.5261
0.4871	0.611	-1	65.5261
0.9218	0.4103	-1	0
0.7382	0.8936	-1	0
0.1763	0.0579	+1	0
0.4057	0.3529	+1	0
0.9355	0.8132	-1	0
0.2146	0.0099	+1	0

Linearly separable dataset

8.	a) Define, with example, Core point, Border Point and Noise point in the perspective of DBSCAN clustering algorithm. b) Describe the DBSCAN Algorithm c) Describe the process of selecting the parameters Eps (radius that defines the neighborhood of a point) and MinPts (minimum number of points in the neighborhood of the core point) in DBSCAN. d) Explain why DBSCAN does not work well for the data having varying density.	[6] [6] [5] [3]
9.	e) Define Information gain, Gain Ratio and Gini Index. f) Draw a decision tree to predict whether a student will be accepted in the post-graduate program using the data provided in the Table of Question 3.	[5] [15]

Name & Signature of the moderator

Name & Signature of the paper setter