Department: Computer Science and Engineering
Course Name: Machine Learning **Code: CS 564**
Full Marks-60 **Time: 2 hours**

Make reasonable assumptions as and whenever necessary. Answer the questions to any sequence but answers to all the parts of any question should appear together. Marks will be deducted if this is not followed properly.

1. (a) Given the following 3D input data, project the data along two principle components.

1 1 9
2 4 6
3 7 4
4 11 4
5 9 2

(b) Differentiate between LDA and PCA.

10 +3

2. Write the algorithm for sequential forward feature selection? Given an objective function J(X) below, perform a sequential forward feature selection algorithm to select the feature.
$J(X)=2x_1x_2x_4 + 7x_2x_1x_3 - 10x_2 - 5x_1 + 3x_3 - 4x_1x_2x_3x_4$

5

3. Write the algorithm of Plus L Minus R feature selection? What is the drawback of this approach? Write the approach which can be used to remove the drawback.

3+2+3=8

4. Given are the points A = (1,2), B = (2,2), C = (2, 1), D = (-1, 4), E = (-2, -1), F = (-1,-1)
Starting from initial cluster, Cluster1 = {A} which contains only the point A and Cluster2 = {D} which contains only the point D.
Show each step of K-means clustering algorithm for two iterations and report the final clusters.
Show a 2D plot of the final clusters.

6

5. (a) If Epsilon is 3 and minpoint is 2, what are the clusters that DBScan would discover with the following 8 examples: A1=(4,3), A2=(2,6), A3=(9,4), A4=(5,9), A5=(7,6), A6=(6,5), A7=(2,2), A8=(3,9). What if Epsilon is increased to 8?

8

b) Use single link agglomerative clustering to group the below data set. Show the dendrograms.

	A	B	C	D
A	0	2	6	10
B		0	5	8
C			0	6
D				0

6. What is linear regression? Write the cost function of linear regression for a single feature. What is regularization? Write down the cost function of regularized linear regression. What is the cost function of logistic regression? Show the steps of gradient descent to optimize the cost function. 1+2+2+2+2+6=15