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) Give the Likelihood Expression for the situation of throwing of one dice with faces marked 1-6. There are N throws and the number of times the ith face appears is N_i is for the face marked 'i' (i=1-6). DERIVE the probability of obtaining N_i 'i's.

6

(b) Now assume there are 2 dice. Do as above with the observations coming out as in (a).

10

(c) Give a machine learning situation which is like the throwing of 2 dice. What are the parameters and what are the hidden variables?

4

2. Two dimensional data is given below,

X1	X2
1.4	1.65
1.6	1.95
-1.4	-1.77
-2.0	-2.52
-3.0	-3.95
2.4	3.07
1.5	2.02
2.3	2.75
-3.2	-4.05
-4.1	-4.85

Obtain first and second Principle component. Calculate and show obtained values in different steps of PCA. Obtain coordinates of first two datapoints in the direction of Principle components.

3. a)   10

If Epsilon is 2 and minpoint is 2, what are the clusters that DBScan would discover with the following 8 examples: A1=(3,3), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9). Draw the 10 by 10 space and illustrate the discovered clusters. What if Epsilon is increased to 10?

b) How to automatically determine the values of epsilon and minpoint from the given data set?

c) Use single and complete link agglomerative clustering to group the below data set. Show the dendrograms.

	A	B	C	D
A	0	1	4	5
B		0	2	6
C			0	3
D				0

10+4+14=28

4. (a) Show the steps after application of K-medoid clustering on the following data.

There are 8 examples and number of clusters=3: A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9). Suppose that the initial seeds (centroids of each cluster) are A1, A4 and A7. Run the K-medoid algorithm for 1 epoch only. At the end of this epoch shows:
i) The new clusters (i.e. the examples belonging to each cluster)
ii) The medoids of the new clusters

(b) What are cluster validity indices? Differentiate between internal and external validity indices. Define Silhouttee index. Explain the different components.

5+ 2+3+4=14

5. a) What is semi-supervised clustering? Differentiate between seeded K-means and constrained K-means approaches for semi-supervised clustering.
b) Write the procedure of performing non-dominated sorting in case of NSGA-II. What is the time complexity of this process? Explain.

2+6+5+3=16

6. What is weighted-vote based classifier ensemble problem.
Consider there are 3 classifiers (A1, A2, A3) and 3 classes (C1, C2, C3). Assume the following is a representation of a weighted-vote based ensemble:
0.12 0.70 0.45 0.95 0.51 0.15 0.12 0.07 0.81
Let the F-measure values of these classifiers be 0.5, 0.7, and 0.75, respectively. Find out the final accuracy of the weighted ensemble classifier for the following test data:

Sample	Predicted class by A1	Predicted class by A2	Predicted class by A3	ACTUAL O/P
S1	C1	C1	C2	C2
S2	C2	C2	C2	C2
S3	C1	C3	C3	C3
S4	C2	C1	C1	C1
S5	C1	C1	C1	C1
S6	C3	C2	C3	C3

2+ 10 =12