INDIAN INSTITUTE OF TECHNOLOGY ROPAR
CS503 - MACHINE LEARNING
Second Semester of Academic Year 2022- 2023
End Semester Examination

Duration:3 Hours | Max.Marks: 50 | Date: 29-04-2023

Instructions:

- No clarifications will be entertained during the examination.
- Make appropriate assumptions wherever necessary and explicitly mention the same.
- All questions are compulsory.
- Be precise and concise in your answers. Partial marks can be awarded for steps/ explanations.
- There are a total of 7 questions for 50 marks. Question 0 can fetch you bonus marks.
- Calculators are allowed during the exam.
- Go through all the questions before you start answering as there might be some questions present later that you may attempt easily.
- Write legibly so that it could be understood what you want to convey in your answers.
- All the Best!

Q.0. Write the full form of DBSCAN. [**1 bonus mark**]

Q.1. Mention whether the following statements are *True/False*. Provide your reasoning in one line. Answers without reasoning will not be awarded any marks. [**10X2 = 20 marks**]

(i) For training an ANN model, it is a usual practice to initialize the weights with zero while the biases must always be initialized with one.

(ii) Given a test instance, the class label assigned by a MAP hypothesis and a Bayes Optimal Classifier will always be the same.

(iii) In K-means clustering, if we increase the value of 'K' then the Error/Loss over the clusters will also increase.

(iv) In DBSCAN, if there are two border points belonging to two different ϵ-neighbourhoods with no common point in between them, then they can't be density connected.

(v) In agglomerative clustering, the optimal number of clusters can be found by running the algorithm for different values of K.

(vi) In K-medoid clustering using the partitioning around medoids method, we cannot use Euclidean distance as the dissimilarity measure.

(vii) K-medoid is robust to the outliers in the dataset while K-means is susceptible to outliers.

(viii) A trained GMM from a dataset can be used to generate new data points to augment the dataset.

(ix) In an HMM, if we multiply the forward and backward probabilities for a particular time **t** and a particular state **j**, it will result in a joint probability of the whole observation sequence **O** along with being in state **j** at time **t**.

(x) In the Viterbi algorithm, the final trellis values $v_T(j), \forall j \in \{1,2,..,N\}$ represent the probability of finding all the ways to reach the j^{th} state at time T from the beginning of the series.

Q.2. Answer the following questions within 4-5 lines (short answers only). **[3X3 = 9 marks]**

I. In DBSCAN, what makes a cluster? Under what conditions could the DBSCAN method result in a lot of noise points?

II. Why a perceptron training rule may not always converge? How does gradient descent solve this problem?

III. Minimizing the errors over the training examples can come at the cost of decreasing the generalization accuracy. Explain briefly.

Q.3. Write the likelihood function for the GMM model. Derive the expressions for GMM parameters μ, Σ, and π. Show your work.

[4 marks]

Q.4. Assume that you have been recruited as a machine learning expert in a multinational taxi aggregator company. The first task assigned to you is to find the pattern of high (H) and low (L) taxi demand at each hour in the city of Chandigarh. Another observation shared with you is that the number of trips in an hour given the demand follows Gaussian distribution with parameters $\mathcal{N}(50,1)$ for high demand and $\mathcal{N}(10,1)$ for low demand. Further, the following likelihood of transitions are available P(H|H) = 0.2, & P(H|L) = 0.25. Given the number of trips in the last three hours as:10, 40 and 25, find the pattern of taxi demand in these three hours. **[4 marks]**

Q.5. Consider the following distance table across data points A to M. Perform DBSCAN clustering on this dataset, using ϵ = 1.5 and minpoints = 3. Assume that the order of processing the points is from A to M in alphabetic order. **[1+1+2 = 4 marks]**

a) For each data point, list its neighbour points within the ϵ neighbourhood.

b) For each point, determine whether it's a core point or not.

c) Construct the final clustering result, including the core points, border points, and noise points.

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	0												
B	1.4	0											
C	3.6	2.2	0										
D	4.5	3.2	1	0									
E	4.7	3.6	1.4	1	0								
F	5.1	3.6	3	2.2	1.4	0							
G	6.3	5	4.5	3.6	2.8	1.4	0						
H	7.8	6.6	5.7	4.7	4	2.8	1.4	0					
I	9	7.7	6.9	5.8	5.1	3.9	2.5	1.1	0				
J	9.2	8.6	7.5	6.3	5.7	4.5	3	1.4	1.4	0			
K	8.2	7.1	6.3	5.1	6.0	5.0	4.0	3.0	2.0	1.0	0.0		
L	7.1	6.0	5.1	4.0	5.1	4.1	3.2	2.2	1.4	1.0	1.4	0.0	
M	8.0	7.1	6.0	5.1	6.3	5.4	4.5	3.6	2.8	2.2	2.0	1.4	0.0

Q.6. Consider a binary classification problem with two classes, A and B, and two features, x and y representing the instances. The class-conditional distributions of the features for each class are given as:

$$P(x, y \mid A) = \frac{1}{2\pi} \exp(-(x^2 + y^2)/2)$$

$$P(x, y \mid B) = \frac{(1 - x)}{(2\pi)^{1.5}} \exp(-(x^2 + y^2)/2)$$

Assume that the prior probabilities of class A and B are equal, i.e. P(A) = P(B) = 0.5.

(i) What is the Bayes optimal classifier for this problem?

(ii) Suppose a test point has features x = 0.5 and y = 0.5. Which class would the Bayes optimal classifier predict for this point? **[2+2=4 marks]**

Q.7. Consider a mixture of Bernoulli distributions given as

$$p(X \mid \mu, \pi) = \sum_{k=1}^{K} \pi_k p(X \mid \mu_k) \qquad (7.1)$$

where $\pi = \{\pi_1, \ldots, \pi_K\}$ are the mixture coefficients and $\mu = \{\mu_1, \ldots, \mu_k\}$ are the mean of K distributions. As with the mixture of Gaussians, in order to use EM algorithm for estimating the parameters of a mixture of Bernoullis, we associate a latent variable z for each instance of X where $z = (z_1, \ldots, z_K)^T$ is a binary K-dimensional variable having a single component equal to 1, with all other components equal to 0. The conditional distribution of X, given the latent variable z, can be given as:

$$p(X \mid z, \mu) = \prod_{k=1}^{K} p(X \mid \mu_k)^{z_k} \qquad (7.2)$$

and the prior of the latent variable z is given as,

$$p(z|\pi) = \prod_{k=1}^{K} \pi_k^{z_k} \tag{7.3}$$

Write the expression for the joint distribution of latent and observed variables for the mixture of Bernoulli distributions. Further, show that by marginalizing the joint distribution with respect to z, Equation (7.1) can be obtained. **[5 marks]**

*********** The End ******************