Mid Term (23rd Sep 2021)
Foundation of Machine Learning (CS-564)

Instructions: - **Total Marks: - 120**

- **Answer all the questions over the A4 sheet and scan in a single file with the name “yourRoll_Name.pdf”**
- **Upload the soft copy of the answer using the link :** https://www.dropbox.com/request/81V5gi10RFIPWYO5PQMq
- **Submit your answer before (24th Sep 2021: 9:00 AM)**
- **Plagiarism will be checked strictly and zero marks will be given in case of copied answers.**

1. Consider the dataset given in table 1. It is desired to transform the data to one-dimensional data using Fisher LDA. Find the one-dimensional transformed data. What is the significance of Fisher linear discriminants for classification problems? Use PCA for dimensionality reduction from two dimensions to one dimension. Confirm that classification performance on the reduced dataset is better using Fisher’s Linear Discriminant analysis compared to PCA. Explain why so? **(Marks: 14 = 5+2+5+2)**

Table 1

$s^{(i)}$	x_1	x_2	Class
$s^{(1)}$	1	2	1
$s^{(2)}$	2	3	1
$s^{(3)}$	3	3	1
$s^{(4)}$	4	5	1
$s^{(5)}$	5	5	1
$s^{(6)}$	1	0	2
$s^{(7)}$	2	1	2
$s^{(8)}$	3	1	2
$s^{(9)}$	3	2	2
$s^{(10)}$	5	3	2
$s^{(11)}$	6	5	2

2. How can we decide the optimal values for epsilon (Eps) and Minpoints (Minpts) for running DBSCAN? Does DBSCAN find the Clusters with variable density and overlapping regions? Justify your answer with a proper explanation. **(Marks: 6 = 4+2)**

3. Consider the 10 companies listed in Table 2. For each company, we have information on whether or not charges were filed against it, whether it is a small or large company, and whether(after investigation) it turned out to be fraudulent or truthful in financial reporting. A "small" company has just been "charged" with fraudulent financial reporting. Using naive Bayes classification technique, compute the probability that the company is "fraudulent". **(Marks: 5)**

Table 2

$s^{(i)}$	X_1: charges filed	X_2: company size	Y: status
1	yes	small	truthful
2	no	small	truthful
3	no	large	truthful
4	no	large	truthful
5	no	small	truthful
6	no	small	truthful
7	yes	small	fraudulent
8	yes	large	fraudulent
9	no	large	fraudulent
10	yes	large	fraudulent

4. Using Genetic Algorithm maximize $f(x) = x^2$ over {0,1,2,..,31} with initial x values of {13,24,8,16}. Show one crossover and mutation operation. **(Marks: 6)**

5. What is the need for feature selection? What are the popular feature selection strategies used in practice? Suppose, we are given a data sample containing 2^{20} features. According to you, which of the said feature selection methods would perform the best and why? **(Marks: 10 = 3+5+2)**

6. K-medoid is more robust than k-means in the presence of noise and outliers because a medoid is less influenced by outliers or other extreme values than a k-means. But, K-medoid is relatively more costly, complexity is O(ik(n-k)^2), where i is the total number of iterations, k is the total number of clusters, and n is the number of objects. To overcome this problem, one researcher proposed an efficient method for finding the k medoids.

    i. Fix the number of clusters *k* and the number of the nearest neighbors *p.*

    ii. Randomly select *k* medoids (without replacement) from the data set X. These Objects represent initial *k* medoids.

    iii. Assign each object in X to the cluster C_j with the closest medoid under the Euclidean distance metric.

    iv. Update *k* medoids. For j=1 to the number of clusters *k* do

        a) Within the cluster C_j. choose a subset C subset that corresponds to m_j and its *p* nearest neighbors (which have not been evaluated before current iteration) of m_j

        b) Calculate the new medoid $q = \underset{x_k \in C_{subset}}{\operatorname{argmin}} \sum_{x_i \in C_j} d(x_k, x_i)$

        after that, the old medoid m_j is replaced by *q* if it is different from m_j.

        c) Repeat steps (a) and (b) until the medoid does not change anymore.

    v. Repeat steps iii and iv until k medoids do not change anymore.

The cost function used in this method is Sum of Euclidean Distance (SED) instead of mean squared error.

$$SED = \sum_{i=1}^{n} \sum_{j=1, x_i \in C_j}^{k} d(x_i, m_j).$$

What will be the advantage of using this new cost function over the traditional cost function? Explain with proper intuition. Also, derive the time-complexity of the above-mentioned method for finding optimal medoids. Find the best possible 3 - medoids using the above-mentioned method and 3 centroids obtained from traditional k-means over the following dataset. Compare if both are the same or different.

A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)

(Marks: 18 = 2+2+7+7)

7. Compute the Silhouette for the following clustering that consists of 2 clusters: {(0,0), (0.1), (2,3)}, {(3,3), (3,4)}; use Manhattan distance for distance computations. Compute each point's silhouette; interpret the results (what do they say about the clustering of the 5 points; the overall clustering?)! **(Marks: 5)**

8. Derive the time complexity of single linkage, complete linkage, and average linkage-based hierarchical clustering algorithm. **(Marks: 9 = 3+3+3)**

9. Let $Y = \beta 0 + \beta 1X + \epsilon$ represent a simple linear regression model which predicts the value of Y as $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$

   Derive the formula for finding the values of intercept ($\beta 0$) and slope ($\beta 1$) for the linear regression model and convert the given linear regression model to logistic regression model by deriving the probability of finding the occurrence Y over the given value of X. Find linear regression equation for the following two sets of data:
   (Marks: 15= 5+5+5)

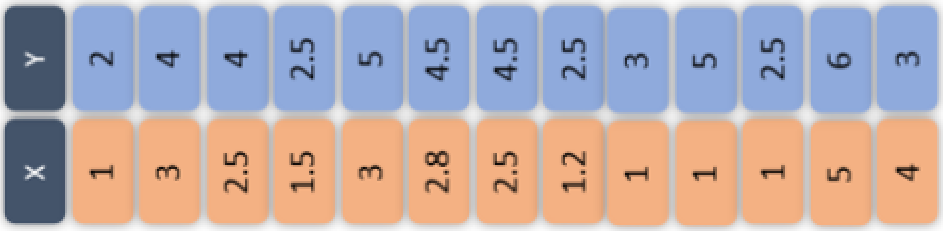
x	2	4	6	8
y	3	7	5	10

10. Assume that the database D is given by the table below. Follow the single link technique to find clusters in D. Use Euclidean distance measure. Also, draw the dendogram for this problem. **(Marks: 12= 10+2)**

D

	x	y
p1	0.40	0.53
p2	0.22	0.38
p3	0.35	0.32
p4	0.26	0.19
p5	0.08	0.41
p6	0.45	0.30

11. Let's take a dataset of 13 points as shown in the image below. Let us choose eps = 0.6 and MinPts =4 and apply the DBSCAN algorithm to find out clusters. Each step and terminologies related to DBSCAN algorithms must be explained clearly. **(Marks: 10)**



Y	2	4	4	2.5	5	4.5	4.5	2.5	3	5	2.5	6	3
X	1	3	2.5	1.5	3	2.8	2.5	1.2	1	1	1	5	4

12. Explain the Expectation-Maximization clustering algorithm step by step and implement the same on the given dataset for k=2 clusters.
X ={X1,X2,X3,X4,X5,X6}= {1,2,3,10,11,12}. Initial selection of $\mu 1$ and $\mu 2$ are random while **σ**1 = 0.82, **σ**2 = 0.82.
Note: Use the normal distribution function for solving this problem. **(Marks: 10)**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$