|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No |  |  |  |  |  |  |  |  |  |  |  |  |

 ****

**Presidency University**

**Bengaluru**

 **SCHOOL OF INFORMATION SCIENCE**

**Make up Examinations July 2024**

**Date**: 10/July/2024

**Time**: 1:30Pm – 4:30pm

**Max Marks**: 100

**Weightage**: 50%

**Odd Semester**: IV

**Course Code**: ECE3020

**Course Name**: Computational Intelligence and Machine Learning

**Department:** ECE

 **Instructions:**

1. *Read the all questions carefully and answer accordingly.*
2. *Do not write any matter on the question paper other than roll number.*

**PART A**

**Answer any SIX Questions. Each question carries 10 marks. (6Qx 10M= 60M)**

1. The phenomenon of over-fitting is really an unfortunate property of maximum likelihood and does not arise when we marginalize over parameters in a Bayesian setting. With help of graphical and mathematical evidence, prove that the total expected loss depends on variance, square of Bias and Noise.

Hint: If we model the g(x) using a parametric function y(x,w) governed by a parameter vector ‘w’, then from a Bayesian perspective the uncertainty in our model is expressed through a posterior distribution over w.

1. Linear Discriminant Analysis (LDA) used for supervised classification problems. It is used for modeling differences in groups i.e. separating two or more classes. It is used to project the features in higher dimension space into a lower dimension space. [C.O.No.2] [Comprehension]

Other than dimensionality reduction, Can we apply LDA for any other application? Explain in detail such an application with hierarchal steps.

1. Probabilistic discriminative models are majorly divided into two types a) generative model b) discriminative models. Consider a data set C={(0,1),(0,1),(1,0),(1,1)},find the relation between generative and discriminant models for the given data set.
2. SVM( support vector machine) works by mapping data to a high-dimensional feature space so that data points can be categorized. A separator between the categories is found then the data is transformed in such a way that the separator could be drawn as a hyper-plane. [C.O.No.2] [Comprehension]

A) Justify the term “Support vector” with respect to perpendicular distance

B) With Suitable diagram explain SVM in Detail.

1. Probabilistic discriminative models are majorly divided into two types a) generative model b) discriminative models. Consider a data set C={(1,1),(0,1),(0,1),(1,0)}, and how you find the relation between generative and discriminant models for the given data set. [C.O.No.2] [Comprehension]
2. Fuzzy K-Means is exactly the same algorithm as K-means, which is a popular simple clustering technique. The only difference is, instead of assigning a point exclusively to only one cluster, it can have some sort of fuzziness or overlap between two or more clusters.

Assume that you have three clusters C1, C2 and C3 with membership values M1, M2, and M3. How convergence will happens to these membership values and how these membership values will update, Explain in detail with suitable algorithm steps.

1. In computational science, particle swarm optimization (PSO) is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. The algorithm was simplified and it was observed to be performing optimization.

Consider a scenario that, five particles (Say A, B, C, D and E) are moving around the solution space (Say P). Each particle moves around the solution space randomly but at the same time attracted by other poles, its past best position (solution) and the best position (solution) of the whole swarm (collection of particles). These poles modify the velocity vector of the particles at each iteration.

How these swarms modify their velocity vectors in the form of their position. Form an algorithm with suitable equations. (C.O.No.3) [Comprehension]

1. Ant colony optimization (ACO) is an optimization algorithm which employs the probabilistic technique and is used for solving computational problems and finding the optimal path with the help of graphs. An ACO is a population-based Algorithm that can be used to find approximate solutions to difficult optimization problems. In ACO, a set of software agents called artificial ants search for good solutions to a given optimization problem. (C.O.No.2) [Application]

Consider the problem of finding the optimum order in which the numbers from 1 to 9 are arranged so that the cost of order is maximum. Assume that six ants have the cost functions (C1, C2, C3, C4, C5, and C6). Consider the following are the orders selected by the six ants along with the corresponding Cost as given below.

|  |  |  |
| --- | --- | --- |
| ANT Number | ORDER | COST |
| ANT1 | 3 | 2 | 4 | 1 | 7 | 5 | 8 | 6 | 9 | C1 |
| ANT2 | 5 | 8 | 9 | 2 | 7 | 3 | 6 | 4 | 1 | C2 |
| ANT3 | 5 | 6 | 8 | 2 | 3 | 9 | 7 | 1 | 4 | C3 |
| ANT4 | 8 | 4 | 9 | 6 | 3 | 1 | 2 | 7 | 5 | C4 |
| ANT5 | 5 | 6 | 7 | 3 | 4 | 1 | 9 | 2 | 8 | C5 |
| ANT6 | 4 | 6 | 7 | 8 | 1 | 2 | 5 | 9 | 3 | C6 |

Identify the name of matrix which will be used in ACO and Design the Same matrix from above said data and also Mention the optimization steps (as a flowchart) in ACO.

**PART B**

**Answer any TWO Questions. Each question carries 20 marks. (2Qx 20M= 40M)**

1. Principal Component Analysis (PCA) is a statistical procedure that uses an orthogonal transformation that converts a set of correlated variables to a set of uncorrelated variables. PCA is the most widely used tool in exploratory data analysis and in machine learning for predictive models. Moreover, PCA is an unsupervised statistical technique used to examine the interrelations among a set of variables.

A) Many times, analyzing the higher dimension data will be typical than lower dimension, what algorithm/method you will choose to overcome this problem.

B) If you identify such an algorithm (mentioned in above), apply the same to below given data and generate the reduced dimension data.

Hint: Use covariance matrix S= $\left[\begin{matrix}14&-11\\-11& 23\end{matrix}\right]$

|  |
| --- |
| Feature |
| X | 4 | 8 | 13 | 7 |
| Y | 11 | 4 | 5 | 14 |

1. Logistic Regression is one of the most popular linear classification models that perform well for binary classification but falls short in the case of multiple classification problems with well-separated classes. While Linear Discriminant Analysis (LDA) handles these quite efficiently. LDA can also be used in data preprocessing to reduce the number of features just as Principle component analysis which reduces the computing cost significantly. (C.O.No.2) [Application]

Consider two data sets as mentioned below, what will be suitable weight vector which will be used to perform classification as well as dimensionality reduction.

X1= {(9,10),(6,8),(9,5),(8,7),(10,8)}

X2= {(4,1), (2,4), (2,3),(3,6),(4,4)}

1. K-means algorithm is an iterative algorithm that tries to partition the dataset into K-pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster’s centroid is at the minimum. (C.O.No.3) [Comprehension]

Consider data sets X and Y as given below. Show the steps of calculation for data points until final clustering is done where no data points are changing clusters.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 1 | 3 | 2 | 3 | 5 |
| Y | 1 | 2 | 2 | 3 | 4 | 5 |