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

PRESIDENCY UNIVERSITY BENGALURU

SCHOOL OF ENGINEERING

MAKE-UP EXAMINATION - JULY 2024

|  |  |
| --- | --- |
| **Semester :** IV | **Date :**10-07-2024 |
| **Course Code :** ECE3020 | **Time :**9:30AM-12:30PM |
| **Course Name :** Computational Intelligence and Machine Learning | **Max Marks :** 100 |
| **Program :** B.Tech | **Weightage :** 50% |

**Instructions:**

1. *Read all questions carefully and answer accordingly.*
2. *Question paper consists of 3 parts.*
3. *Scientific and non-programmable calculator are permitted.*
4. *Do not write any information on the question paper other than Roll Number.*

|  |  |  |  |
| --- | --- | --- | --- |
| **PART A** | | | |
| **ANSWER ANY 4 QUESTIONS 4Q X 5M=20M** | | | |
| 1 | Statement: “Batch techniques are involved in the process of the entire training set in one go which can be computationally costly for large data sets”.  How to overcome the above (statement) drawback in case of large data sets and explain the same in detail. | (CO 1) | [Knowledge] |
|  | | | |
| 2 | There are few possibilities to come out from the over-fitting problem. Among them increasing dataset size, using Bayes functions and adding regularization term to an error function are the most efficient methods.  How regularized least squares are useful to control the error while we are training a system. | (CO 1) | [Knowledge] |
|  | | | |
| 3 | In the field of machine learning, sequence models are those that take in or produce data in the form of a sequence. Sequential data includes but is not limited to text streams, audio snippets, video clips, time series data, and so on. With suitable equation, explain sequential learning. | (CO 1) | [Knowledge] |
|  | | | |
| 4 | The goal in classification is to take an input vector ‘x’ and to assign it to k – discrete classes Ck where k=1,2,3,…..k. In the most common scenario, the classes are taken to be disjoint so that each input is assigned to one and only one class. In case of multiclass, with diagrams explain one-versus-one and one-versus- rest classification. | (CO 2) | [Knowledge] |
|  | | | |
| 5 | LDA is a dimensionality reduction technique. It is used as a pre-processing step in Machine Learning and applications of pattern classification.  A) What are the terms SB, Sw in the linear discriminant analysis?  B) What is PCA, explain briefly. | (CO 2) | [Knowledge] |
|  | | | |
| 6 | Covariance Matrix is a measure of how much two random variables gets change together.  Assuming two 2 features x and y for a data set, write the number of ordered pairs which can be formed and also mention the pairs used to find the covariance matrix in Principal component analysis. | (CO 3) | [Knowledge] |
|  | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **PART B** | | | |
| **ANSWER ANY 5 QUESTIONS 5Q X 10M=50M** | | | |
| 7 | 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. PSO has a main advantage of having fewer parameters to tune. PSO obtains the best solution from particles' interaction, but through high-dimensional search space, it converges at a very slow speed towards the global optimum.  Consider a scenario that, five particles (Say P1, P2, P3, P4 and P5) are moving around the solution space (Say Z). 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. | (CO 3) | [Comprehension] |
|  | | | |
| 8 | Probabilistic discriminative models are majorly divided into two types a) generative model b) discriminative models. Consider a data set C={(0,1),(1,1),(1,0),(0,1),(1,0),(1,1),(0,1)},find the relation between generative and discriminant models for the given data set | (CO 3) | [Comprehension] |
|  | | | |
| 9 | 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 happen to these membership values and how these membership values will update, Explain in detail with suitable algorithm steps | (CO 2) | [Comprehension] |
|  | | | |
| 10 | Perceptron models are the simplest type of artificial neural network which carries input and their weights, the sum of all weighted input, and an activation function. Perceptron models are continuously contributing to Artificial Intelligence and Machine Learning, and these models are becoming more advanced. Perceptron enables the computer to work more efficiently on complex problems using various Machine Learning technologies. With suitable diagrams explain perceptron algorithm | (CO 3) | [Comprehension] |
|  | | | |
| 11 | Consider a data set X={X1, X2, X3……….XN} and the target vector t= {1 1 1 1}T. the weight vector W= {W1,W2, W3, W4}. Consider a Linear model for regression with design matric then Calculate the Maximum likelihood weight vector (WML) | (CO 2) | [Application] |
|  | | | |
| 12 | Statement: “Batch techniques are involves in the process of the entire training set in one go which can be computationally costly for large data sets”.  A) How to overcome the above (statement) drawback in case of large data sets and explain the same in detail.  B) Suppose if you identify any such algorithm (to overcome the above statement) then explain in detail.  C) Assume that you have to data sets and these data sets are linearly separable. Which algorithm will give you 100% guaranty to do the classification? Explain the same in detail. | (CO 2) | [Comprehension] |
|  |  |  |  |
| 13 | 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 that you will choose to overcome this problem. (2M)  B) If you identify such an algorithm (mentioned in above), apply the same to below given data and generate the reduced dimension data. (8M)  Hint: Use covariance matrix S=   |  | | --- | | Feature | | X | 4 | 8 | 13 | 7 | | Y | 11 | 4 | 5 | 14 | | (CO 2) | [Comprehension] |
|  | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **PART C** | | | |
| **ANSWER ANY 2 QUESTIONS 2Q X 15M=30M** | | | |
| 14 | 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.  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 | 4 | 3 | 2 | 5 | 3 | 7 | 1 | | Y | 6 | 4 | 1 | 3 | 5 | 8 | 2 | | (CO 3) | [Comprehension] |
|  | | | |
| 15 | Linear Discriminant Analysis, or LDA for short, is a predictive modelling algorithm for multi-class classification. It can also be used as a dimensionality reduction technique, providing a projection of a training dataset that best separates the examples by their assigned class.  .  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)} | (CO 2) | [Application] |
|  | | | |
| 16 | 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. ACO has a good performance in solving discrete problems, but it inevitably has some disadvantages. Although it has good stability, it has some shortcomings in the convergence speed and solution accuracy when dealing with a large amount of data.  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. With help of pheromone matrix explain optimization steps in ACO.   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | ANT Number | ORDER | | | | | | | | | COST | | ANT1 | 3 | 4 | 1 | 9 | 2 | 8 | 5 | 6 | 7 | C1 | | ANT2 | 8 | 4 | 9 | 6 | 3 | 1 | 2 | 7 | 5 | C2 | | ANT3 | 3 | 2 | 4 | 1 | 7 | 5 | 8 | 6 | 9 | C3 | | ANT4 | 5 | 8 | 9 | 2 | 7 | 3 | 6 | 4 | 1 | C4 | | ANT5 | 5 | 6 | 8 | 2 | 3 | 9 | 7 | 1 | 4 | C5 | | ANT6 | 4 | 6 | 7 | 8 | 1 | 2 | 5 | 9 | 3 | C6 | | (CO 2) | [Application] |