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

**Presidency University**

**Bengaluru**

 **SCHOOL OF ENGINEERING**

 **MAKE UP EXAMINATION – SEPTEMBER 2023**

**Course Code**: ECE 299

**Course Name**: Computational Intelligence and Machine Learning (OE)

**Programme** : B.Tech

**Date**: 01.10.2023

**Time**: 1.00PM to 4.00PM

**Max Marks**: 100

**Weightage**: 50%

**Part A[Memory Recall Questions]**

Answer all Questions. (15QX2M=30M)

1) Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. A good clustering method will produce high quality clusters with (C.O.No.1) [Knowledge] A) High inter class similarity

B) Low intra class similarity

C) High intra class similarity

D) No inter class similarity

2) Dimensionality reduction refers to techniques for reducing the number of input variables in training data. Dimensionality Reduction Algorithms are one of the possible ways to reduce the computation time required to build a model (C.O.No.1) [Knowledge]

A) False.

B) True

C) Maybe.

D) Cannot be determined.

## 3) The real-world dataset often has a lot of missing values. The cause of the presence of missing values in the dataset can be loss of information, disagreement in uploading the data, and many more. How do you handle missing or corrupted data in a dataset? (C.O.No.1) [Knowledge]

A) Drop missing rows or columns

B) Replace missing values with mean/median/mode

C) Assign a unique category to missing values

D) all of the above

4) A machine learning model is an expression of an algorithm that combs through mountains of data to find patterns or make predictions. Father of Machine Learning is (C.O.No.1) [Knowledge]

A) Geoffrey Hill

B) Geoffrey Chaucer

C) Geoffrey Everest Hinton

D) None of the above

5) Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Which of the following sentence is correct? (C.O.No.1) [Knowledge]

A) Machine learning relates with the study, design and development of the algorithms that give computers the capability to learn without being explicitly programmed.

B) Data mining can be defined as the process in which the unstructured data tries to extract knowledge or unknown interesting patterns.

C) Both a & b

D) None of the above

6) Reduction of dimensionality is the method of reducing with consideration the dimensionality of the function space by obtaining a collection of principal features. The selection of features tries to pick a subset of the original features to be used in the machine learning model. It is not necessary to have a target variable for applying dimensionality reduction algorithms. (C.O.No.1) [Knowledge]

A) True**.**

B) False

C) Maybe.

D) Cannot be determined.

7) Machine learning can be intimidating, with its reliance on math and algorithms that most programmers don't encounter in their regular work. Which language is best for machine learning? A) C (C.O.No.1) [Knowledge]

B) java

C) Python

D) HTML

8) Everything in life depends on time and therefore, represents a sequence. To perform machine learning we can use sequential data. Sequential learning is based on considered data points in \_\_\_\_\_\_\_\_ for training. (C.O.No.1) [Knowledge]

A) One a time for training

B) All in a batch form for training

C) Set of data for training

D) None of the above

9) Essentially, the Bayes' theorem describes the probability or Total Probability Rule. The Total Probability Rule is a fundamental rule in statistics relating to conditional and marginal of an event based on prior knowledge. Bayes’ Theorem plays a central role in pattern recognition and machine learning. The term P (X|Y) in baye’s theorem is read as \_\_\_\_\_\_\_\_\_\_.

 (C.O.No.1) [Knowledge]

A) Probability of X and Y

B) Probability of Y and X

C) Probability of X and Y

D) Probability of X given Y.

10) In machine learning, multiclass or multinomial classification is the problem of classifying instances into one of three or more classes (classifying instances into one of two classes is called binary classification).Which of the following method is used for multiclass classification? (C.O.No.1)[Knowledge]

A) One vs Rest

B) One

C) All vs One

D) One vs Another

11) Gradient descent (GD) is an iterative first-order optimization algorithm used to find a local minimum/maximum of a given function. This method is commonly used in machine learning (ML) and deep learning(DL) to minimize a cost/loss function To find the minimum or the maximum of a function, we set the gradient to zero because which of the following condition (C.O.No.1) [Knowledge]

A) Depends on the type of problem

B) The value of the gradient at extreme of a function is always zero

C) Both (A) and (B)

D) None of these

12) A machine learning algorithm is the method by which the AI system conducts its task, generally predicting output values from given input data. How many types are available in machine learning? (C.O.No.1) [Knowledge]

A) 1

B) 2

C) 3

D) 4

13) Computational complexity, a measure of the amount of computing resources (time and space) that a particular algorithm consumes when it runs. Computational complexity of classes of learning problems depends on which of the following? (C.O.No.1) [Knowledge]

A) The size or complexity of the hypothesis space considered by learner

B) The accuracy to which the target concept must be approximated

C) The probability that the learner will output a successful hypothesis

D) All of the above

14) A statistical model or a machine learning algorithm is said to have under fitting when it cannot capture the underlying trend of the data. Under fitting destroys the accuracy of our machine learning model. Suppose, you got a situation where you find that your linear regression model is under fitting the data. In such situation which of the following options would you consider? (C.O.No.1) [Knowledge]

A) You will add more features

B) You will remove some features

C) You will discard data points

D) None of the above

15) Artificial neural networks, usually simply called neural networks, are computing systems inspired by the biological neural networks that constitute animal brains. Feature of ANN in which ANN creates its own organization or representation of information it receives during learning time is\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. (C.O.No.1) [Knowledge]

A) Adaptive Learning

B) Self-Organization

C) What-If Analysis

D) Supervised Learning

**Part B [Thought Provoking Questions]**

Answer all Questions. Each question carries **TEN** marks. (3Qx10M=30M)

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. 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 |

 i) Identify the name of matrix which will be used in ACO and Design the Same matrix from above said data. (2M)

 ii) Mention the optimization steps (as a flowchart) in ACO. (8M)

17) Particle swarm optimization (PSO) is a population-based stochastic optimization algorithm motivated by intelligent collective behavior of some animals such as flocks of birds or schools of fish. Since presented in 1995, it has experienced a multitude of enhancements. As researchers have learned about the technique, they derived new versions aiming to different demands, developed new applications in a host of areas, published theoretical studies of the effects of the various parameters and proposed many variants of the algorithm.

 i) 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. (5M)

 ii) How these swarms (Say P1, P2, P3, P4 and P5) modify their velocity vectors in the form of their position and reaches their destination (Say Z). Form an algorithm with suitable equations (5M)(C.O.No.3) [Comprehension]

18) Consider a data set of inputs X={X1,X2,X3,……..XN}.The target variable ’t’ is given by a deterministic function Y(X,W) with additive Gaussian noise (ɛ-zero mean Gaussian random variable with precision β). [C.O.1][Comprehension]

 i) How Maximum likelihood and least squares are related to Y(X,W). (4M)

 ii) Give the mathematical modeling behind maximum likelihood with respect to weights (w) and precision (β). (6M)

 **Part C [Problem Solving Questions]**

 Answer all Questions. Each question carries **TWENTY** marks. (2Qx20M=40M)

19) 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)}

20) 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 |