PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING

TEST-1

Even Semester: 2021-22

Course Code: ECE 299

Course Name: Computational Intelligence and Machine Learning

Programme & Sem: B.Tech & 6th Sem

Date: 27-04-2022

Time: 10.00AM to 11.00AM

Max Marks: 30

Weightage: 15%

(i) Read all the questions carefully and answer accordingly.
(ii) Answers should be visible and diagrams should be clear

Part A [Memory Recall Questions]

Answer all Questions. Each question carries 1 marks.   **(10Qx1M=10M)**

1) Curve fitting is the major task in regression concept. _____ is the scenario when the model fails to decipher or misses data point to fit on a curve and to find the underlying trend in the input data.   [C.O.No.1] [Knowledge]

    A) Over fitting

    B) Under fitting

    C) Both A and B

    D) None of the above

2) Training a generalized machine learning model means, in general, it works for all subset of unseen data. The ability to categorize correctly a new example that differ from those used for training is known as _____.   [C.O.1][Knowledge]

    A) Training data

    B) Generalization

    C) Feature extraction

    D) None of the above

3) Gaussian distribution is a bell-shaped curve, and it is assumed that during any measurement values. Gaussian distribution is also known as----------   [C.O.No.1] [Knowledge]

    A) Normal distribution

    B) Maximum likelihood

    C) Sum-of-squares

    D) Regularization

4) 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.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

5) 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

6) Machine learning is really all about using past data to either make predictions or understand general groupings in your dataset.. The expression t = y(x, w) + Є , the term Є is called as _____.   [C.O.1][Knowledge]

   A) Variance

   B) Zero mean Gaussian random variable

   C) Moore pseudo inverse matrix

   D) Non-square matrices

7) Regression is a predictive modeling technique investigating relationship between dependent and independent variables. In linear model of regression, the term Ø(x) W.R.T input variable X is   (C.O.No.1)[Knowledge]

   A) Sum function

   B) Product function

   C) Basis function

   D) Basic function

8) The bias–variance decomposition is a way of analyzing a learning algorithm's expected generalization error with respect to a particular problem. In bias variance decomposition expected loss= _____.   [C.O.1][Knowledge]

   A) Expected loss = (bias)2 + variance+ noise

   B) Expected loss = bias + variance+ (noise)2

   C) Expected loss = bias + variance+ noise

   D) Expected loss = bias +(variance)2+ noise

9) Machine learning algorithms have the ability to improve themselves through training. What are the three types of Machine Learning?   [C.O.1][Knowledge]

      A) Supervised Learning

      B) Unsupervised Learning

      C) Reinforcement Learning

      D) All of the above

10) Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine Learning is a field of AI consisting of learning algorithms that..........

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

      A) At executing some task

      B) Over time with experience

      C) Improve their performance

      D) All of the above

Part B [Thought Provoking Questions]

Answer both the Questions. Each question carries 6 marks.   **(2Qx6M=12M)**

11) 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.

12) 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.

Hint: Regularization allows complex models to be trained on data sets of limited size without severe over-fitting, essentially by limiting the effective model complexity. However, the problem of determining the optimal model complexity is then shifted from one of finding the appropriate number of basis functions to one of determining a suitable value of the regularization coefficient λ.   (C.O.No.1) [Comprehension]

Part C [Problem Solving Questions]

Answer the Question. Question carries 8 marks.   **(1Qx8M=8M)**

13) 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}. Consider a Linear model for regression with design matric $\begin{bmatrix} 1 & 0 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$

then calculate the Maximum likelihood weight vector (WML).

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

PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING

TEST-2

Even Semester: 2021-22

Course Code: ECE 299

Course Name: Computational Intelligence and Machine Learning

Programme & Sem: B.Tech & VIth Sem

Date: 2nd June 2022

Time: 10.00 AM to 11.00 AM

Max Marks: 30

Weightage: 15%

Part A [Memory Recall Questions]

Answer both the Questions. Each question carries FIVE marks. (2Qx5M=10M)

1) Hyper-planes are decision boundaries that helps to classify the data points. Data points falling on either side of the hyper-plane can be attributed to different classes. Also, the dimension of the hyper-plane depends upon the number of features. How hyper-plane will be used in classification of data? Explain in detail.
[C.O.No.2][Knowledge]

2) LDA is a supervised classification technique that is considered part of crafting competitive machine learning models. What is the goal of LDA algorithm? Explain S_W and S_B scatter matrices. [C.O.No.2]
[Knowledge]

Part B [Thought Provoking Questions]

Answer both the Questions. Each question carries SIX marks. (2Qx6M=12M)

3) 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. **(1M)**

   B) With Suitable diagram explain SVM in Detail. **(5M)**

4) 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.

[Comprehension]

Part C [Problem Solving Questions]

Answer the Question. Question carries EIGHT marks.
(1Qx8M=8M)

5) Machine Learning Algorithms whether supervised or unsupervised uses a simple matrix operations from linear algebra and statistics to calculate a projection of the original data into the same number or fewer dimensions.

A) Identify the algorithm which minimizes the complexity of large data set analysis and gives projection of data in a N-dimensional space with use of Eigen values.
(1M)

B) If you identify such an algorithm (mentioned in above), apply the same to below given data and generate the reduced dimension data.
(7M)

Hint: Use covariance matrix S= $\begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$

Feature				
X	6	10	15	9
Y	13	6	7	16

[Application]

PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING

ENDTERM EXAMINTION

Even Semester: 2021-22   **Date**: 1st July 2022

Course Code: ECE 299   **Time**: 09.30 AM to 12.30 PM

Course Name: Computational Intelligence and Machine Learning (OE)   **Max Marks**: 100

Programme & Sem: B.Tech & VI Sem   **Weightage**: 50%

Part A[Memory Recall Questions]

Answer all Questions. Each question carries TWO mark.
(15QX2M=30M)

1) 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. In what type of learning labelled training data is used   (C.O.No.1)
[Knowledge]
      A) Unsupervised learning
      B) Supervised learning
      C) Reinforcement learning
      D) Active learning

2) In statistics, maximum likelihood estimation is a method of estimating the parameters of an assumed probability distribution, given some observed data. This is achieved by maximizing a likelihood function. The main disadvantage of maximum likelihood methods is that they are _____.
(C.O.No.1)

[Knowledge]

      A) Mathematically less folded
      B) Mathematically less complex
      C) Mathematically more complex
      D) Computationally intense

3) A common job of machine learning algorithms is to recognize objects and being able to separate them into categories. This process is called classification. You are given reviews of few netflix series marked as positive, negative and neutral. Classifying reviews of a new netflix series is an example of _____.   (C.O.No.1)
[Knowledge]
      A) Reinforcement learning
      B) Unsupervised learning
      C) Semi-supervised learning
      D) Supervised learning

4) "The origin of the name "linear" comes from the fact that the set of solutions of such an equation forms a straight line in the plane." In the mathematical Equation of Linear Regression Y=mx+c, m and c refers to?   (C.O.No.1)
[Knowledge]

A) m= x intercept, c= slope

B) m= slope , c= x intercept

C) m=y- intercept , c= slope

D) m= slope , c= y-intercept

5) Oftentimes, the terms machine learning and artificial intelligence (AI) are used interchangeably; however, they are not the same. AI is basically the umbrella concept, and machine learning is a subset of artificial intelligence. Choose the correct option regarding machine learning (ML) and artificial intelligence (AI)   (C.O.No.1)

[Knowledge]

A) ML is a set of techniques that turns a dataset into a software

B) AI is a software that can emulate the human mind

C) ML is an alternate way of programming intelligent machines

D) All of the above

6) You split up the data containing known response variable values into two pieces. The training set is used to train the algorithm, and then you use the trained model on the test set to predict the response variable values that are already known._____ is the machine learning algorithms that can be used with labeled data.   (C.O.No.1)

[Knowledge]

A) Regression algorithms

B) Clustering algorithms

C) Association algorithms

D) All of the above

7) The genetic algorithm is a method for solving both constrained and unconstrained optimization problems that is based on natural selection, the process that drives biological evolution. When would the genetic algorithm terminate?   (C.O.No.1)

[Knowledge]

A) Maximum number of generations has been produced

B) Satisfactory fitness level has been reached for the population

C) Both A & B

D) None of these

8) Unsupervised Learning is a machine learning technique in which the users do not need to supervise the model. Instead, it allows the model to work on its own to discover patterns and information that was previously undetected. What is the function of Unsupervised Learning?

(C.O.No.1)

[Knowledge]

A) Find clusters of the data and find low-dimensional representations of the data

B) Find interesting directions in data and find novel observations/ database cleaning

C) Interesting coordinates and correlations

D) all of the above

9) Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. The objective of SVM algorithm is to find a hyper plane in an N-dimensional space that distinctly classifies the data points. What do you mean by a hard margin?

(C.O.No.1)

[Knowledge]

A) The svm allows high amount of error in classification

B) The svm allows very low error in classification

C) Both 1 & 2

D) None of the above

10) The method of least squares is a standard approach in regression analysis to approximate the solution of over determined systems and is the most important application in data fitting. The method of Least Squares Estimation minimizes? (C.O.No.1)

[Knowledge]

A) Summation of squares of errors

B) Summation of errors

C) Summation of absolute values of errors

D) All

11) Genetic algorithms are excellent for searching through large and complex data sets. A genetic operator is an operator used in genetic algorithms to guide the algorithm towards a solution to a given problem. Genetic operators includes (C.O.No.1)

[Knowledge]

A) Mutation

B) Crossover

C) Both A and B

D) None of the above.

12) Probability is a measure of uncertainty. Probability applies to machine learning because in the real world, we need to make decisions with incomplete information. Bayes rule can be used for? (C.O.No.1)

[Knowledge]

A) Solving queries

B) Increasing complexity

C) Answering probabilistic query

D) Decreasing complexity

13) Suppose that we have N independent variables (X1, X2… Xn) and dependent variable is Y. Now imagine that you are applying linear regression by fitting the best fit line using least square error on this data. You found that correlation coefficient for one of it's variable (Say X1) with Y is 0.95. (C.O.No.1)

[Knowledge]

A) Relation between the x1 and y is weak

B) Relation between the x1 and y is strong

C) Relation between the x1 and y is neutral

D) Correlation can't judge the relationship

14) A genetic algorithm (GA) is a heuristic search algorithm used to solve search and optimization problems. GA's are also based on the behavior of chromosomes and their genetic structure.GA techniques are inspired by _____methods. (C.O.No.1)

[Knowledge]

A) Evolutionary

B) Cytology

C) Anatomy

D) Ecology

15) This clustering algorithm terminates when mean values computed for the current iteration of the algorithm are identical to the computed mean values for the previous iteration

[Knowledge]
A) Agglomerative clustering
B) Conceptual clustering
C) Expectation maximization
D) K-means clustering

Part B [Thought Provoking Questions]

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

16) 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. (C.O.No.3) [Comprehension]

17) 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]

18) Regularization allows complex models to be trained on data sets of limited size without severe over-fitting, essentially by limiting the effective model complexity. However, the problem of determining the optimal model complexity is then shifted from one of finding the appropriate number of basis functions to one of determining a suitable value of the regularization coefficient λ.

Identify at least three possibilities with which to overcome the over-fitting problem and explain in detail. (C.O.No.3) [Comprehension]

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.

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.

$$X1= \{(4,1), (2,4), (2,3),(3,6),(4,4)\}$$
$$X2= \{(9,10),(6,8),(9,5),(8,7),(10,8)\}$$

(C.O.No.2) [Application]

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.

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	2	2	3	4	5
Y	1	1	3	2	3	5

(C.O.No.3) [Comprehension]