BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: MCA & Pre-PhD **SEMESTER : III**
BRANCH: COMP. SC. **SESSION : MO/2023**

SUBJECT: CA511 BASICS OF MACHINE LEARNING

TIME: 3 Hours **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
***4*. Before attempting the question paper, be sure that you have got the correct question paper.**
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO
Q.1(a)	Explain the difference between classification and regression in the context of machine learning. Provide examples of problems that can be framed as classification and regression tasks.	[5]	1,1,1
Q.1(b)	Discuss Principal Component Analysis (PCA) and its role in machine learning. How does PCA help in reducing the dimensionality of data?	[5]	1,2,1
Q.2(a)	Describe the gradient descent algorithm for optimizing the parameters in linear regression with one variable. What are the key steps involved, and how does it converge to the optimal solution? Explain with a suitable example.	[5]	2,2,2
Q.2(b)	Derive the cost function for logistic regression. How does it differ from the cost function in linear regression, and what challenges does it address in the context of classification? Extend the idea of regularization to logistic regression. Explain how regularization is incorporated into the logistic regression cost function and its impact on the model.	[5]	2,3,2
Q.3(a)	Explain the fundamental concepts of Support Vector Machines (SVM). What is the main objective of SVM, and how does it achieve the task of classification or regression?	[5]	3,2,3
Q.3(b)	Describe the role of regularization techniques in addressing overfitting. How do regularization methods, such as L1 and L2 regularization, help in achieving a better balance between bias and variance?	[5]	3,3,3
Q.4(a)	Describe the backpropagation algorithm. How does backpropagation work to minimize the cost function and adjust the weights in a neural network? Include the key steps and mathematical expressions involved.	[5]	4,4,4
Q.4(b)	Define recurrent neural networks (RNNs). What makes them suitable for handling sequential data, and how do they differ from feedforward neural networks?	[5]	4,1,4
Q.5(a)	Consider a dataset with the following points in a two-dimensional space: A(2, 3), B(5, 8), C(1, 1), D(6, 6), E(8, 5) Perform Agglomerative Clustering using the Euclidean distance metric. Initially, treat each point as a separate cluster. Perform the agglomeration steps manually following the single-linkage (minimum distance) criterion until there are only two clusters left. a) Calculate the Euclidean distance between all pairs of points: AB, AC, AD, AE, BC, BD, BE, CD, CE, DE. b) Identify the pair of points with the minimum distance and merge them into a single cluster. Update the distances between this new cluster and the remaining points. c) Repeat step until there are only two clusters left. d) Draw a dendrogram illustrating the agglomeration process.	[5]	5,5,5
Q.5(b)	Explain the steps involved in applying the EM algorithm to perform soft clustering. How does EM assign probabilities to data points belonging to different clusters?	[5]	5,3,5

:::::21/11/2023 E:::::