MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent unit of MAHE, Manipal)

End semester

SUBJECT: **Machine Learning in Chemical Engineering (CHE 4073)**

Time duration: 180 minutes Date: 21/11/2022 ***MAX.MARKS: 50***

Q. No		M	CO	PO	BL
1.	Justify the usage of Regression and Classification algorithm of Machine Learning in process systems engineering. (4)	4	1	1,4	4
2	Distinguish between supervised learning and unsupervised learning algorithms of machine learning by giving an example (4)	4	1	1,4,	4
3	Explain the following terms in reference to Machine Leaning: (i) Labels (ii) Training set (iii) Testing set (iv) Loss function (2)	2	1	1,4,	2
4	Ten data points were taken in an experiment in which the independent variable x is the mole percentage of a reactant, and the dependent variable y is the yield (in percent): x: 20, 20, 30, 40, 40, 50, 50, 60, 70, 75 y: 73, 78, 85, 90, 91, 87, 86, 75, 65, 80 Fit a quadratic model of the form $h_\theta(x) = \theta_1 x + \theta_2 x^2$ with these data and estimate the parameter vector using a closed-form linear regression solution. Determine the value of x that maximizes the yield. (4)	4	2	1,4,	5
5	Justify the use of regularization in the Machine Learning algorithm. Explain the various methods used for regularization. (4	4	3	1,4,	4
6	What's the trade-off between bias and variance in the machine learning algorithm (2)	2	3	1,4	4
7	Develop a Gradient decent algorithm to estimate the parameter of the Logistic regression model. (4)	4	2	1,2	5
8	It is decided to design a soft-sensor for a multicomponent distillation column to estimate the product composition and synthesize feedback control design. Develop a detailed step-by-step procedure to design the soft sensor using a suitable machine learning algorithm. (3)	3	4	1,2	5
9	With the help of two-dimensional space, decision boundaries and support vectors explain the classification process of a Support Vector Machine. (3)	3	4	1,2	4

10	What is the role of 'C' in the Support Vector Machine? What is the intuition of a large-margin classifier? (3)	3	4	1,2,4	4
11	Develop a K-means clustering algorithm and justify its usage in process system engineering problems. (3)	3	5	1,2,4	5
12	What are the advantages of Principal Component analysis (PCA), and What are the assumptions for PCA and its limitations. (4)	4	5	1,2,4	4
13	Obtain the Moore Penrose pseudo-inverse (left) of data matrix 'A'. (2)	2	5	1,2,4	5
14	Prove that the Eigen value and eigen vectors can be obtained through the Singular Value Decomposition (SVD) of the correlation matrix of the data set and further deduce the relation between the eigenvalue and singular values. (3)	3	5	1,2,4	4
	Formulate the principal component analysis to evaluate the first principal component (z_1), which captures the maximum variation in the data for the given data set below, and determine the percentage variance captured by each principal component. Evaluate the approximate value of original data using only first principal component (i.e., z_1) and report the actual and approximate data error. X1: 2.5, 2.3, 2.2, 2, 1, 1.5, 1.1 X2: 2.4, 2.7, 2.9, 1.6, 1.1, 1.6, 0.9 (5)	5	5	1,2,4	5