DHARMSINH DESAI UNIVERSITY, NADIAD
FACULTY OF TECHNOLOGY
SECOND SESSIONAL – CE
SUBJECT: Machine Learning (CE622)

Examination	**: B.Tech Semester VI**	**Seat No.**	**:**
Date	**: 03/02/2025**	**Day**	**: Monday**
Time	**: 2:30 PM to 3:45 PM**	**Max. Marks**	**: 36**

INSTRUCTIONS:
1. Figures to the right indicate maximum marks for that question.
2. The symbols used carry their usual meanings.
3. Assume suitable data, if required & mention them clearly.
4. Draw neat sketches wherever necessary.

[12]

Q.1 Do as directed.

CO2 U (a) Define sigmoid activation function. What are the benefits of using sigmoid activation function? [2]

CO2 R (b) Which are the major limitations of K – Means clustering algorithm? [2]

CO1 N (c) State true or false and justify your answer, "The running time of Hierarchical Top Down clustering is more than the K-Means clustering". [2]

CO2 E (d) Consider a binary classification problem with linearly separable data points. **X1= [-1 0]** is a support vector from **positive class** and **X2= [0 1]** is a support vector from **negative class**. Find the equation of the decision boundary for Linear SVM. [3]

CO1 U (e) Explain the curse of dimensionality and its impact on machine learning models. [2]

CO2 R (f) State whether the following statements are true or false. [1]

a) Lasso and Ridge can be used for dimensionality reduction or feature subset selection.

b) Lasso computes linear transformations of the input features.

[12]

Q.2 Attempt ***Any TWO*** from the following questions.

CO2 A (a) Demonstrate the working of K-Means Clustering for the dataset given in the **Table 1**, [6]

	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12
X1	1	2	3	4	5	6	7	5	9	7	8	10
X2	1	3	4	2	6	5	8	5	8	9	10	10

Table 1

For K means clustering, Number of Clusters=3 and Initial clusters p1, p4, p10. Measure distance using the Euclidean distance metric

CO2 A (b) Demonstrate the working of Hierarchical Agglomerative clustering for the dataset given in the **Table 1**. Measure distance using the Euclidean distance metric. [6]

CO3 A (c) Describe the perceptron architecture with the learning algorithm. Demonstrate why a single layer perceptron cannot solve XOR problem. Show a multilayer perceptron model to solve XOR problem. [6]

[12]

Q.3 Attempt the following questions.

CO2 N (a) i) What is Regularization? Explain the working of Linear Regression model with Regularization. [6]

ii) We train a regularized (Ridge) Linear Regression model using **Stochastic Gradient Descent** with value of λ=1000 and the learning rate=0.01. The hypothesis function of the linear regression model is defined as: $h(X) = \theta 0 + \theta 1 x1 + \theta 2\, x2 + \theta 3 x3$. In one of the training iteration the value of the parameters is $\boldsymbol{\theta 0 = 0.2, \theta 1 = 0.5,}$ $\boldsymbol{\theta 2 = -1.1}$ ***and*** $\boldsymbol{\theta 3 = 2.7}$. How much change will be there in the value of the parameter $\theta 1$ for the **example** x1= -1, x2= 2, x3= -2 and Y (output value) =1.

CO2 C (b) Apply PCA Algorithm on the given scaled dataset and find the Principal components. [6]

No.	X1	X2
1	0.29	0.09
2	-1.11	-1.04
3	0.83	0.95

Determine the percentage of variance explained by each principal component. Which principal component would you choose to retain for a reduced dimensionality and why?

OR

CO2 **N** (a) i) What is the log-odds interpretation in logistic regression? Why do we use cross-entropy (log loss) instead of mean squared error (MSE) in logistic regression? [6]

ii) Consider a labelled set of examples for three **classes (y= (Y1, Y2, Y3))** with an input space **x= (x1, x2, x3).** We train the model using multi class logistic regression model using gradient descent. In one of the iteration of the gradient descent, the parameter matrix is as follows (columns correspond to the parameters for the individual classes):

	Y1	Y2	Y3
$\theta 0$	3	3	3
$\theta 1$	2	0	-2
$\theta 2$	3	-4	6
$\theta 3$	-3	0	2

Consider one example in the training set x= (-4,-1,-3) with the label y= (1, 0, 0). What is the amount of cross-entropy loss incurred by this example in this iteration?

CO2 **C** (b) Consider the following set of data points. A=(2,10), B=(2,5), C=(8,4), D=(5,8), E=(7,5), F=(6,4),G=(1,2), H=(4,9) **i)** Take **minpts=3** and **eps(epsilon) = 2** . Apply the DBSCAN algorithm on the given data points and identify the clusters and outliers. **ii)** If the value of **epsilon** is changes to **10**, what is the change in the number of clusters? Find and list the new clusters formed. [6]