# GIET UNIVERSITY, GUNUPUR – 765022

B. Tech (Fourth Semester Regular) Examinations, May – 2024

## 22BCMPE24001 - Introduction to Machine Learning

(CSE - AIML )

Time: 3 hrs                                                      Maximum: 70 Marks

**(The figures in the right hand margin indicate marks)**

**PART – A**                                                      **(2 x 5 = 10 Marks)**

Q.1. Answer *ALL* questions

|  |  | CO # | Blooms Level |
|---|---|---|---|
| a. | Explain what is linearly ,non linearly and almost linearly separable with a suitable example | CO3 | K2 |
| b. | What is residual errors? Give a suitable example . | CO1 | K2 |
| c. | Write down the steps for calculating the ROC curve. | CO1 | K2 |
| d. | Write down the algorithm for Gradient descent with linear regression. | CO2 | K2 |
| e. | The regression coefficient of the regression equation of X on Y is 2.4 and Y on X is 0.8. Are the regression coefficients consistent? | CO1 | K3 |

**PART – B**                                                      **(15 x 4 = 60 Marks)**

Answer *ALL* questions                                        Marks     CO #     Blooms Level

2. a. Consider a dataset containing the following observations:               5     CO1     K3

| X | Y |
|---|---|
| 2 | 5 |
| 3 | 7 |
| 4 | 9 |
| 5 | 11 |
| 6 | 13 |

    (i) Compute the slope and intercept of the linear regression line for the given dataset.

    (ii) Use the equation of the regression line to predict the value of Y when X is 10.

    (iii) Calculate the coefficient of determination ($R^2$ ) for this regression model and interpret its meaning in the context of the dataset

  b. Explain briefly logistic regression and find the loss function.                 10     CO1     K2

(OR)

  c. 

| X1 | X2 | X3 | Y |
|---|---|---|---|
| 1 | 2 | 3 | 10 |
| 4 | 5 | 6 | 15 |
| 2 | 3 | 5 | 20 |

5     CO1     K3

Compute the coefficients (intercept and slopes) of the multiple linear regression model and generate the regression line.

  d. Explain briefly Lasso and Ridge regularization and find the loss function on it. How elastic net regression is working           10     CO2     K2

3.a. Explain briefly how K-NN is used in the mentioned dataset: -            8     CO2     K3
Point A: (2, 3), Class: 0     Point B: (4, 4), Class: 1     Point C: (5, 5), Class: 1
Point D: (7, 6), Class: 0     Point E: (6, 3), Class: 1
Classify a new point, F (5, 4), using the K-Nearest Neighbors algorithm with $k$=3 and the Euclidean distance metric. Determine the class label for point F.

  b. Explain briefly about linear regression with gradient descent approaches as well as write down the detailed steps and perform at least two iterations .      7     CO2     K2

(OR)

| c. | Explain briefly how supervised, unsupervised learning, and reinforcement learning works | 5 | CO2 | K2 |
|---|---|---|---|---|

d.

| X1 | X2 | Y |
|---|---|---|
| 2 | 3 | Class A |
| 2 | 4 | Class A |
| 1 | 3 | Class B |
| 8 | 4 | Class B |
| 6 | 7 | Class B |

| | | 10 | CO2 | K2 |
|---|---|---|---|---|

Explain briefly how SVM works on the modified dataset with two features, $X1$ and $X2$, and a target variable $Y$. Build a model with a Support Vector Machine (SVM) classifier to classify the data points into two classes.

| 4.a. | Given dataset: X = [1,2,3,1,] and Y = [1,3,2,4] | 12 | CO2 | K2 |
|---|---|---|---|---|

   (i) Compute the regression coefficients $B0$ (intercept) and $B1$ (slope) for the linear regression model.

   (ii) Calculate the standard errors for the regression coefficients $SEB0$ and $SEB1$.

   (iii) Use the t-test to determine whether each regression coefficient is statistically significantly different from zero. Perform hypothesis testing with a significance level of $\alpha=0.05$.

   (iv) Interpret the results of the hypothesis testing and discuss the implications for the linear regression model.

| b. | Explain the Wilcoxoin Rank-sum test. | 3 | CO2 | K2 |
|---|---|---|---|---|

(OR)

| c. | The data below shows the salaries in randomly selected adv. in two different occupations (Education and Marketing) | 10 | CO2 | K3 |
|---|---|---|---|---|

Education: 22,40,18,25,15,23,16,19,21,30
Marketing:28,24,20,45,50,39,26,55,48,41,42
**Use mann-whiteney test** at 1%(-2.33) level of significance to test the median salary in the field of education is lower than the median salary in the field of marketing

| d. | Given the following data, Calculate the spearman's correlation coefficient and interpret R. | 5 | 3 | K3 |
|---|---|---|---|---|

| X | Y |
|---|---|
| 10 | 7 |
| 5 | 8 |
| 4 | 9 |
| 15 | 7 |
| 8 | 9 |
| 8 | 7 |
| 15 | 4 |

| 5.a. | Suppose we have four different teaching methods (A, B, C, and D) and we want to determine if there is a significant difference in the exam scores of students taught using these methods. The exam scores for each group are as follows: | 10 | CO3 | K3 |
|---|---|---|---|---|

Teaching Method A: [8, 8, 9, 9, 10]   Teaching Method B: [7, 8, 8, 9 9]
Teaching Method C: [7, 7, 8, 8, 9]    Teaching Method D: [8, 9, 9, 10, 10]
Perform a one-way ANOVA to test the null hypothesis that there is no significant difference in the mean exam scores among the four teaching methods. Use a significance level of $\alpha=0.05$.

| b. | What do you mean by ANN? Explain briefly about it. | 5 | CO4 | K2 |
|---|---|---|---|---|

(OR)

c. Implement the XOR function using McCullah-pitts neuron (Consider the binary data)   8   CO4   K2

| X1 | X2 | Y |
|----|----|---|
| 0  | 0  | 0 |
| 0  | 1  | 1 |
| 1  | 0  | 1 |
| 1  | 1  | 0 |

7   CO4   K3

d. Consider the dataset:

| Actual value | Predicted value |
|--------------|-----------------|
| 100          | 130             |
| 150          | 170             |
| 200          | 220             |
| 250          | 260             |
| 300          | 325             |

Explain the significance of each error metric (MSE, RMSE, RMSLE, MAE) in evaluating the performance of a regression model. Provide examples of scenarios where each metric would be preferred over others.

--- End of Paper ---