# GANDHI INSTITUTE OF ENGINEERING AND TECHNOLOGY UNIVERSITY, ODISHA, GUNUPUR
# (GIET UNIVERSITY)

B. Tech(Fourth Semester - Regular) Examinations, April – 2025

## 23BCMP24011 – Introduction to Machine Learning

(CSE-AIML)

Time: 3 hrs Maximum: 60 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. | CO1 | K1 |
| b. | What is residual errors? Give a suitable example. | CO2 | K1 |
| c. | Write down the steps for calculating the ROC curve | CO3 | K2 |
| d. | Write down the algorithm for Gradient descent with linear regression | CO4 | K2 |
| e. | In K-NN, what is K, and what is the justification of K? | CO5 | K2 |

**PART – B** **(10 x 5=50 Marks)**

Answer ***ALL*** the questions

| | | Marks | CO # | Blooms Level |
|---|---|---|---|---|
| 2. a. | Consider a dataset containing the following observations:<br>I. Compute the slope and intercept of the linear regression line for the given dataset.<br>II. Use the equation of the regression line to predict the value of Y when X is 6.<br>III. Calculate the coefficient of determination ($R^2$) for this regression model and interpret its meaning in the context of the dataset | 5 | CO1 | K3 |

| X | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Y | 3 | 4 | 5 | 6 | 8 |

| | | | | | |
|---|---|---|---|---|---|
| b. | Differentiate between the training and testing phases in machine learning. Discuss how training-testing splits, cross-validation, and performance metrics help ensure model robustness and generalization to unseen data. | 5 | CO1 | K2 |
| | (OR) | | | |
| c. | Discuss the components of a machine learning system and explain how machine learning differs from traditional design-based approaches. Include a diagrammatic representation of the end-to-end ML workflow. | 5 | CO2 | K2 |
| d. | A dataset with the following values:<br>X = [1, 2, 3, 4, 5]<br>Y = [2, 3, 4, 5, 6]<br>Perform the following:<br>1. Fit a linear regression model to the data using the least squares method.<br>2. Find the slope and intercept of the regression line.<br>3. Predict the value of Y when X = 6.<br>4. Calculate the following performance metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) | 5 | CO2 | K3 |
| 3.a. | The logistic regression model for binary classification is: **logit(p) = -1 + 0.5·$X_1$ + 1.2·$X_2$**<br>Use this model to compute the predicted probability **p** that Y = 1 for an input where: $X_1$ = 2 and $X_2$ = 3<br>1. Calculate the log-odds (logit).<br>2. Convert the logit into a probability.<br>3. Classify the output if threshold = 0.5. | 5 | CO2 | K3 |

| | | | | |
|---|---|---|---|---|
| b. | Discuss the various types of regularization techniques in regression. Explain L1 (Lasso), L2 (Ridge), and Elastic Net regularization with appropriate mathematical formulations. | 5 | CO2 | K2 |
| | (OR) | | | |
| c. | Explain the working of Logistic Regression in classification problems. Discuss the sigmoid function, log-odds, probability estimation, decision boundary, and the loss function used in logistic regression. | 5 | CO2 | K2 |
| d. | Write an algorithm to demonstrate the K-NN classifier to classify the test point (X1=2, X2=2) with K=3 | 5 | CO2 | K3 |

| X1 | X2 | Class |
|---|---|---|
| 1 | 2 | 1 |
| 2 | 3 | 1 |
| 3 | 3 | -1 |
| 2 | 1 | -1 |

| | | | | |
|---|---|---|---|---|
| 4.a. | Explain truncation and pruning in decision trees. Given a decision tree with 5 levels and 95% accuracy on training data but 70% on test data, should pruning be applied? Justify your answer. | 5 | CO3 | K4 |
| b. | Explain the Decision Tree algorithm. How do ID3, C4.5, and CART differ from one another? Discuss how each algorithm chooses the splitting attribute, handles missing values, and deals with continuous attributes. | 5 | CO3 | K2 |
| | (OR) | | | |
| c. | Explain the concept and importance of cross-validation in machine learning. Discuss K-fold cross-validation. Explain how cross-validation helps in reducing overfitting and selecting hyper parameters. | 5 | CO3 | K2 |
| d. | Write a short notes on Truncation and Pruning. | 5 | CO3 | K1 |
| 5.a. | What is the Silhouette Score? Explain how it is computed for each data point and its significance in cluster validation. | 5 | CO4 | K2 |
| b. | Describe the working of Principal Component Analysis (PCA). Explain how eigenvectors and eigenvalues are used in reducing dimensions. | 5 | CO4 | K2 |
| | (OR) | | | |
| c. | Consider the following 2D dataset and apply the PCA of the followings: -<br>1. Calculate the covariance matrix.<br>2. Find the eigenvalues and eigenvectors.<br>3. Reduce the data to 1 principal component (1D).<br>4. Reconstruct the data from the reduced form. | 5 | CO4 | K2 |

| X1 | 2 | 0 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| X2 | 0 | 2 | 1 | 3 | 5 |

| | | | | |
|---|---|---|---|---|
| d. | Describe the structure and use of K-d Trees. How are they constructed and used for nearest neighbor searches in clustering and classification? | 5 | CO4 | K2 |
| 6.a. | Differentiate between Bagging and Boosting. Provide the workflow, advantages, and limitations of each technique as well as write the algorithms in each | 5 | CO5 | K4 |
| b. | What is Random Forest? Describe its working with the help of Decision Trees and explain how it handles overfitting and feature importance. | 5 | CO5 | K2 |
| | (OR) | | | |
| c. | Explain the Gradient Boosting and AdaBoost algorithms. Highlight their differences in terms of model construction, error handling, and performance. | 5 | CO5 | K4 |
| d. | What is Stacking (Stacked Generalization)? Explain the concept of meta-learners with a use case example. | 5 | CO5 | K3 |

--- End of Paper ---