**Predictive Analytics**

**NAME - MANAS ARORA**

**B.TECH CSE AIML BATCH 5**

**SAP ID – 500109627**

**ROLL.NO - R2142221204**

**Lab 7:**

**Differentiate between logistic regression and linear regression through two real-world scenarios.**

**Hint: Differentiate in terms of i) Definition, ii) Datasets compatibility, iii) Model, iv) Validation Metrics, v) Visualization ( through graphs)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| |  | | --- | | Criteria | | |  | | --- | | Logistic Regression |  |  | | --- | |  | | Linear Regression |
| Definition | Used for binary classification tasks. Predicts probabilities for an outcome that can be 0 or 1. Uses a sigmoid function to map predictions between 0 and 1. | Used for predicting continuous values. Assumes a linear relationship between input features and the output. |
| Dataset Compatibility | Works with categorical target variables, especially binary (0 or 1). Input features can be continuous or categorical. | Works with continuous target variables. Input features can be continuous or categorical (requires encoding). |
| Model | The model fits a log-odds linear equation and uses a sigmoid function to transform predictions into probabilities. | The model fits a straight line by minimizing the sum of squared differences between actual and predicted values. |
| Validation Metrics | Common metrics include accuracy, precision, recall, F1-score, AUC-ROC. Suitable for evaluating classification tasks. | Common metrics include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared. Used for continuous value prediction. |

|  |  |
| --- | --- |
| Visualization (Graphs) | |
| Logistic Regression | Linear Regression |
| The relationship between the independent variables (inputs) and the probability of the outcome (binary: 0 or 1) is represented by a **sigmoid curve**. This curve is **S-shaped**, where the probability sharply increases from 0 to 1 as the input values cross a threshold.  **Example:** In a graph showing the probability of customer churn, the x-axis could represent customer usage data, and the y-axis would show the probability of churn, with the curve smoothly transitioning from low probability (0) to high probability (1). | The relationship between the independent variables and the dependent variable (continuous) is represented by a straight line on a scatterplot of data points. The line of best fit is drawn such that the sum of the squared distances from the actual data points to the line is minimized.  **Example**: In a graph predicting house prices based on square footage, the x-axis could represent square footage, and the y-axis would show house prices. The line would show the predicted price as a linear function of square footage. |

