NAME – PRIYANSHI ASATI

SAP ID – 500108525

ROLL NO – R2142220982

BATCH- 4 (AIML-NON HONS.)

PREDICTIVE ANALYTICS LAB- 5

1.Definition:

- **Linear Regression**: Predicts a continuous outcome based on the relationship between the dependent variable and one or more independent variables.

    o **Example**: Predicting house prices based on factors like size, number of rooms, etc.

- **Logistic Regression**: Predicts a binary or categorical outcome by estimating the probability of a class label.

    o **Example**: Predicting whether a student will pass or fail an exam based on study hours, attendance, etc.

2. Dataset Compatibility:

- **Linear Regression**: Works with continuous numeric data where the target variable is continuous.

    o **Scenario**: Predicting a car's price based on its features like mileage, horsepower, and age.

- **Logistic Regression**: Works with categorical target variables, typically for classification tasks (binary or multiclass).

    o **Scenario**: Classifying emails as spam or not spam based on features like word frequency, sender, etc.

3. Model:

- **Linear Regression**: The model finds the best-fitting straight line (y = mx + c) that minimizes the sum of squared differences between actual and predicted values (least squares method).

- **Logistic Regression**: The model uses the logistic (sigmoid) function to model probabilities. Instead of a straight line, it maps values to a probability range between 0 and 1.

Key Equation:

- Linear Regression: $y = \beta_0 + \beta_1 X$

- Logistic Regression: $P(y=1|X) = \frac{1}{1+e^{-(\beta_0 + \beta_1 X)}}$

4. Validation Metrics:

- **Linear Regression**:

    o Common metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2.

    o **Scenario Example**: Predicting a company's quarterly revenue based on past data.

- **Logistic Regression**:

    o Common metrics: Accuracy, Precision, Recall, F1-score, and Area Under the ROC Curve (AUC-ROC).

    o **Scenario Example**: Predicting customer churn.

5. Visualization:

- **Linear Regression**: Visualization involves plotting the independent variable on the x-axis and the dependent continuous variable on the y-axis. The best-fit straight line (regression line) represents the model's prediction.

    o **Graph Example**: House price prediction where house size is plotted against house price, showing a line representing the predicted prices.

- **Logistic Regression**: Visualization often involves plotting the sigmoid curve, with probabilities on the y-axis and the predictor variable on the x-axis. ROC curves are also common to assess model performance.

    o **Graph Example**: Predicting whether a customer will churn based on the number of support calls, where a sigmoid curve maps values to probabilities, and a threshold defines the class boundary.