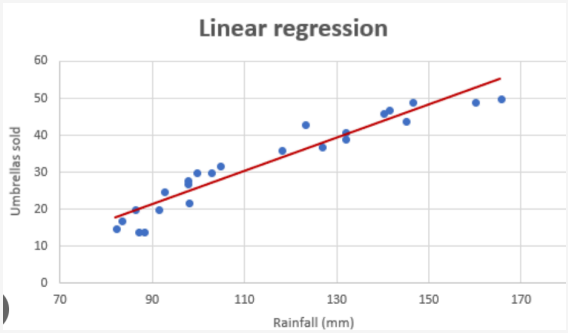
Predictive Analytics Lab
Lab-7

Name: Rishi raj Jain
SAP: 500105417
Batch: 5

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)

Parameter	Linear Regression	Logistic Regression
Definition	This is a type of regression analysis that models the relationship between a dependent (target) variable and one or more independent (predictor) variables by fitting a linear equation to observed data. It predicts a continuous outcome.	A type of classification algorithm used when the dependent variable is categorical (often binary). It predicts the probability of a given class by applying a logistic function to the output of a linear equation.
Datasets compatibility	Works best with continuous numerical datasets. For example, it's used when the target variable is a number (e.g., predicting house prices based on square footage, number of rooms, etc.).	Suitable for binary or categorical datasets. It's used when the target variable is categorical (e.g., whether a customer will churn or not based on age, income, etc.).
Model	Model is a straight line (in case of single feature) described by the equation: $$y = \beta_0 + \beta_1 x + \epsilon$$	Model is based on the sigmoid function to constrain the output between 0 and 1 (for binary classification). The equation is: $$P(y = 1\|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$
Validation Metrics	**R-squared (R^2)**: Measures the proportion of variance in the dependent variable explained by the independent variables. **Mean Squared Error (MSE)**: Measures the average squared difference between the observed actual outcomes and predicted outcomes.	**accuracy:** Measures the percentage of correct predictions. **Precision, Recall, F1-score**: Evaluate classification performance, especially in imbalanced datasets. **AUC-ROC**: Assesses the model's ability to differentiate between classes.

Visualization (through graphs)	The relationship between variables can be visualized as a **straight line** in a scatter plot, where the x-axis represents the independent variable, and the y-axis represents the dependent variable. 	The S-form curve is called the Sigmoid function or the logistic function. In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.