Akansh Grover

500102060

B4

Lab Experiment 7

Difference Between Logistic Regression and Linear Regression

1. **Definition**
   Linear Regression: This is a statistical technique that is known to predict a continuous numerical output. It puts forward a linear relationship among one or more independent variables and the dependent variable by determining the difference between the predicted and actual values.
   For example, using the features of a house such as size, number of rooms, and location to predict its price.

   Logistic Regression: This is a form of classification algorithm that predicts the probability of a binary outcome-0 or 1. Instead of modeling a continuous output, it models a dependent variable that represents categories or classes, which makes it well suited for tasks like spam detection and medical diagnoses.
   Example: Based on a customer's browsing behavior and interaction with the web site, predicting whether that customer will buy a product (yes/no).

2. **Dataset Compatibility**
   Linear Regression:
   Linear Regression supports datasets where the dependent variable is continuous.
   Example: This will include predicting a house price or temperature where the outcome takes numeric values.

   Logistic Regression:
   This is applicable for datasets with a categorical-dependent variable, usually binary, such as a yes/no, success/failure.
   Example: predicting whether a customer is going to buy a product using his or her browsing behavior is a typical use case.

3. **MODEL**
   Linear Regression:

In the case of Linear Regression, it fits a straight line to the data points. This can be mathematically represented as y=mx+c where y is the dependent variable, m is the slope, and c is the intercept. Therefore, this line represents the relation between the input variables and output.
Example: A linear model that will predict the prices of houses. The house price increases with the size of the house in direct proportion.

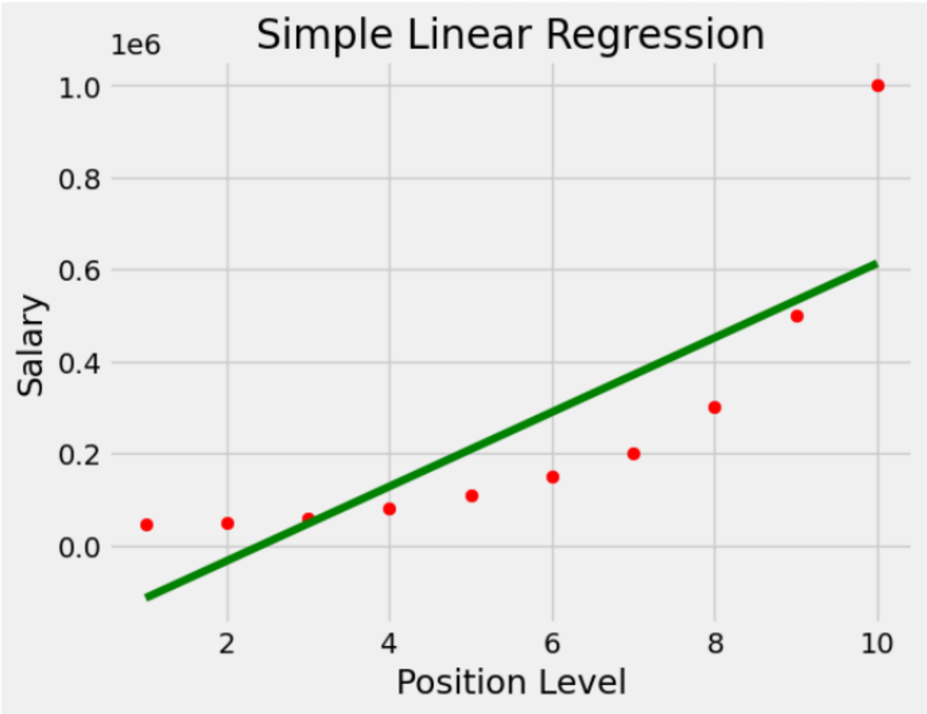
Logistic Regression:
The model of Logistic Regression maps input features to probabilities using a sigmoid function. It is an S-shaped curve that confines the predicted values between 0 and 1 and thus is interpretable as probabilities themselves. A decision threshold is then applied-e.g., 0.5-to classify data.
Example: A logistic model that predicts whether a user will make a purchase. A browsing time of 5 minutes might correspond to a 70% probability of purchasing, which is classified as "yes."

4. **Validation Metrics**
Linear Regression:
Linear Regression uses validation metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared. These metrics assess how close the predicted values are to the actual values, providing insights into the model's accuracy and explanatory power.
Example: If the model predicts house prices, the MSE can calculate the average squared difference between the predicted prices and the actual prices.

Logistic Regression:
Validation metrics for Logistic Regression include Accuracy (percentage of correct predictions), Precision (positive predictive value), Recall (true positive rate), F1 Score (harmonic mean of Precision and Recall), and the Receiver Operating Characteristic - Area Under Curve (ROC-AUC), which measures the model's ability to distinguish between classes.
Example: Evaluating the model's performance in predicting whether users make a purchase, using metrics like accuracy or ROC-AUC to assess classification quality.

5. **Visualization**

Linear Regression:



Logistic Regression: