**LAB 7**

**1. Real-World Scenarios**

* **Scenario for Linear Regression**: Predicting house prices based on square footage.
* **Scenario for Logistic Regression**: Predicting whether a customer will buy a product (yes/no) based on their online activity.

**i) Definition**

* **Linear Regression**: It models the relationship between a **dependent variable** (continuous) and one or more **independent variables** by fitting a linear equation to observed data. The dependent variable is **continuous** (e.g., house price).
* **Logistic Regression**: It is used to model the probability of a **binary outcome** (0/1, yes/no) based on one or more independent variables. It estimates probabilities and uses a **logistic function** to ensure outputs are between 0 and 1.

**ii) Dataset Compatibility**

* **Linear Regression**:
  + Compatible with datasets where the target variable is **continuous** (e.g., house price).
  + Independent variables can be continuous or categorical (with proper encoding).
* **Logistic Regression**:
  + Works with datasets where the target variable is **binary/categorical** (e.g., customer buys or not).
  + Independent variables can also be continuous or categorical, but the outcome variable must be binary.

**iii) Model**

* **Linear Regression**:
  + The model is a straight line represented by the equation:

**y=β0​+β1​x1​+β2​x2​+⋯+βn​xn**

where y is the predicted house price.

* **Logistic Regression**:
  + The model uses the logistic function (sigmoid) to predict probabilities:

**P(y=1)=1/1+e−(β0​+β1​x1​+⋯+βn​xn​)**

where P(y=1) is the probability of the customer buying the product.

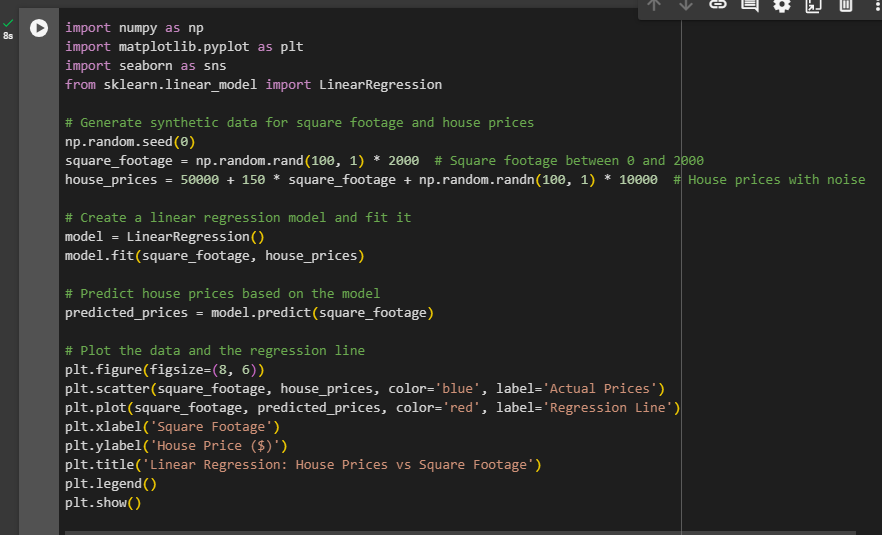
**iv) Validation Metrics**

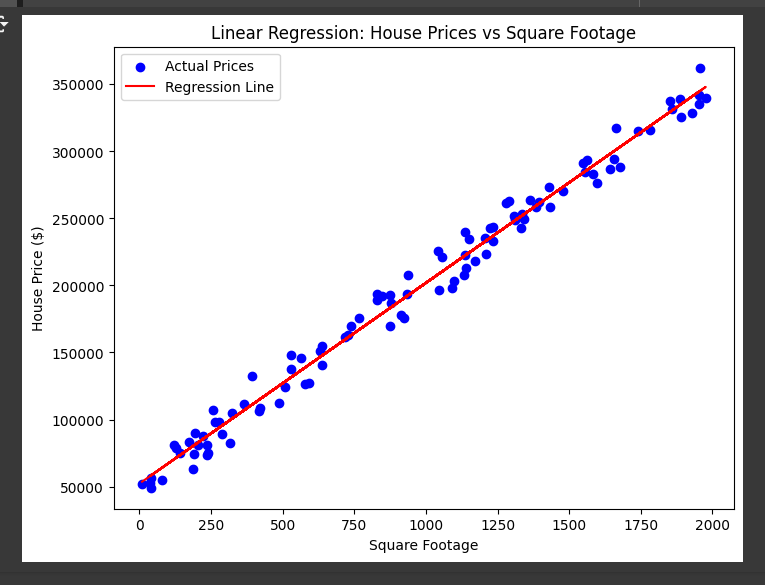
* **Linear Regression**:
  + Common metrics include:
    - **Mean Squared Error (MSE)**: Measures the average squared difference between actual and predicted values.
    - **R-squared**: Represents the proportion of variance explained by the model (higher is better).
* **Logistic Regression**:
  + Since the output is binary, different metrics are used:
    - **Accuracy**: Percentage of correct predictions.
    - **Precision and Recall**: Precision measures how many predicted positives are correct, and recall measures how many actual positives were correctly identified.
    - **F1 Score**: Harmonic mean of precision and recall.
    - **AUC-ROC**: Measures the trade-off between true positives and false positives.

**v) Visualization (Through Graphs)**

* **Linear Regression**:
  + **Graph**: A scatter plot of the data with a **straight line** fitted to show the relationship between variables.
  + **Example**: A plot of square footage (x-axis) vs. predicted house price (y-axis), where the line captures the trend.

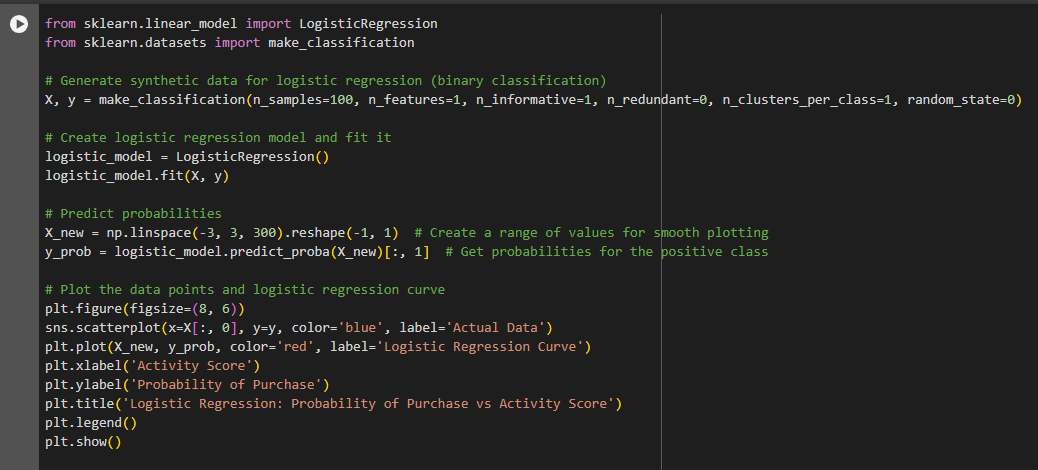
**CODE:**

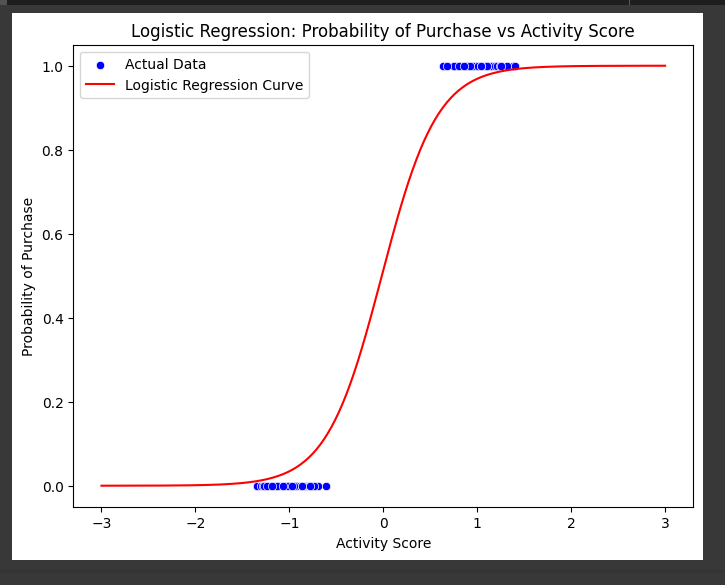




* **Logistic Regression**:
  + **Graph**: A plot of the logistic curve (S-shaped) showing the probability of the binary outcome as a function of the independent variables.
  + **Example**: A plot where the x-axis represents an online activity score, and the y-axis represents the probability of a customer purchasing the product, with the logistic curve illustrating the transition between 0 and 1.

**CODE**:





**Summary of Differentiation**

| **Sl.No.** | **Linear Regression** | **Logistic Regression** |
| --- | --- | --- |
| 1. | Linear Regression is a supervised regression model. | Logistic Regression is a supervised classification model. |
| 2. | Equation of linear regression: **y = a0 + a1x1 + a2x2 + … + aixi** Here, **y**= response variable **xi**= ith predictor variable **ai**= average effect on y as xi increases by 1 | Equation of logistic regression **y(x) = e(a0 + a1x1 + a2x2 + … + aixi) / (1 + e(a0 + a1x1 + a2x2 + … + aixi))** Here, **y**= response variable **xi**= ith predictor variable **ai**= average effect on y as xi increases by 1 |
| 3. | In Linear Regression, we predict the value by an integer number. | In Logistic Regression, we predict the value by 1 or 0. |
| 4. | Here no activation function is used. | Here activation function is used to convert a linear regression equation to the logistic regression equation |
| 5. | Here no threshold value is needed. | Here a threshold value is added. |
| 6. | Here we calculate Root Mean Square Error(RMSE) to predict the next weight value. | Here we use precision to predict the next weight value. |
| 7. | Here dependent variable should be numeric and the response variable is continuous to value. | Here the dependent variable consists of only two categories. Logistic regression estimates the odds outcome of the dependent variable given a set of quantitative or categorical independent variables. |
| 8. | It is based on the least square estimation. | It is based on maximum likelihood estimation. |
| 9. | Here when we plot the training datasets, a straight line can be drawn that touches maximum plots. | Any change in the coefficient leads to a change in both the direction and the steepness of the logistic function. It means positive slopes result in an S-shaped curve and negative slopes result in a Z-shaped curve. |
| 10. | Linear regression is used to estimate the dependent variable in case of a change in independent variables. For example, predict the price of houses. | Whereas logistic regression is used to calculate the probability of an event. For example, classify if tissue is benign or malignant. |
| 11. | Linear regression assumes the normal or gaussian distribution of the dependent variable. | Logistic regression assumes the binomial distribution of the dependent variable. |
| 12. | Applications of linear regression:   * Financial risk assessment * Business insights * Market analysis | Applications of logistic regression:   * Medicine * Credit scoring * Hotel Booking * Gaming * Text editing |