Assignment No:- 4

Problem statement :-

Create a Linear Regression model using Python/R to predict home prices using Boston housing dataset. The Boston Housing dataset contains information about various houses in Boston through different parameters. There are 506 samples and 14 Feature variable in dataset.

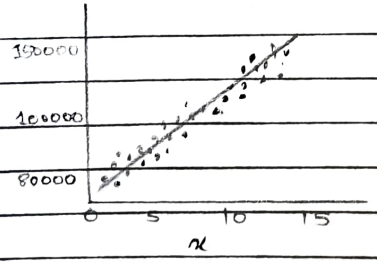
Objective :-

To predict the value of prices of the house using the given Features.

Theory :-

Linear Regression :-

- It is a machine Learning algorithm based on supervised learning.
- It performs a Regression task.
- Regression model is a target prediction value based on Independant variables.
- Linear Regression performs the task to predict a dependant variable value(y) based on a given independant variable(x)
- so, this Regression technique finds out a linear Relationship between x (input) and y (output)



Equation For a Line :-

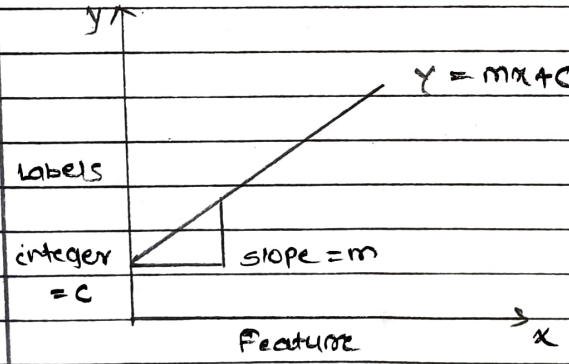
- The equation For a line,

$$y = mx + b$$

- In the equation For a line,
  - Y = The vertical Value.
  - M = slope.
  - X = The Horizontal Value.
  - B = The value of Y when X = 0 (i.e., Y intercept).

$$\therefore \text{slope} = \frac{\text{change in } Y}{\text{change in } X}$$

Regression Line Equation graph :-



Cost Function ():-

- By achieving best-fit Regression line, the model aims to predict y value such that the error difference between predicted value and true value is minimum.
- Cost Function of linear Regression is the Root Mean Squared Error (RMSE) between predicted y value (pred) and true y value (y).

Result:-

Thus we have predicted Housing price of Boston city using linear Regression.

Assignment No:- 5

Problem statement:-

1. Implement Logistic Regression using python/R to perform classification on social_Network_Ads.csv dataset.
2. Compute Confusion Matrix to Find TP, FP, TN, FN, Accuracy, Error Rate, precision, Recall on the given dataset.

Objectives:-

1. to performed classification on social_Network_Ads.csv dataset using Logistic Regression.
2. Computing the Confusion Matrix.

Theory:-

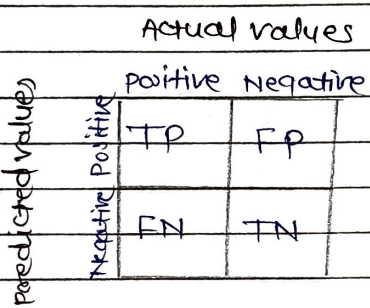
Logistic Regression:-

- Logistic Regression is one of the most popular Machine Learning Algorithm.
- It is used to predicting the categorical dependant variable using a given set of independant variables.
- It gives the probabilistic values which the lie between 0 and 1.
- Logistic Regression used for solving the classification problems.
- It is denoted by sigmoid Function.
- The sigmoid Function is a mathematical Function used to map the predicted values to probabilities.
- The dependant variable must be categorical in nature.
- The independant variable should not have multi-collinearity.

$$\therefore \sigma = \frac{1}{1+e^{-Y}}$$

Confusion Matrix :-

- A confusion matrix is an N x N matrix used for evaluating the performance of a classification model,
- where N is a number of target classes.
- The matrix compares the actual target values with those predicted by the machine learning model.



True positive (TP) :-

- The predicted value matches the actual value.
- The actual value was positive and the model predicted a positive value.

True Negative (TN) :-

- The actual value was negative and the model predicted a negative value.

False positive (FP) :-

- The predicted value was falsely predicted.
- The actual value was negative but the model predicted a positive value.
- It is also known as TYPE1 error.

False Negative (FN) :-

- The actual value was positive but the model predicted a negative value.
- It is also known as type 2 error.

Accuracy :-

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

Precision :-

- Precision tells us how many of the correctly predicted cases actually turned out to be positive.

$$\text{Precision} = \frac{TP}{TP + FP}.$$

Recall :-

- Recall tells us how many of the actual positive cases we where able to predict correctly with our model.

$$\text{Recall} = \frac{TP}{TP + FN}.$$

Result :-

Implemented Logistic Regression to perform classification on Social_Network_Ads.csv dataset using Python.