Lab-04

Regression Models

Build Linear and Logistic regression models and evaluate their performance using appropriate metrics.

Experiment-1 : Consider a dataset ‘sales’ based on 'TV' marketing budget. Build a linear regression model to predict 'Sales' using 'TV' as the predictor variable.

Dataset Link : https://github.com/Yashappin/Machine-Learning/blob/master/TvMarketing.csv

a) Importing data and understanding its structure.

b) Visualising Data Using Plot.

c) Splitting Data into Training and Testing Sets (80:20). Also display the shape of Training and testing data shape.

d) Train a Simple Linear regression Model ($\boldsymbol{y = b_0 + b_1 \times X}$) and calculate the value of optimal model parameters, thereafter, visualised the best fit line.

e) Display actual vs predicted sales.

f) Computing RMSE and R^2 Values

Experiment-2: Multiple regression is a statistical method used to model the relationship between multiple independent variables and a dependent variable. Perform multiple linear regression using the following dataset.

Dataset Link : https://www.kaggle.com/datasets/midhundasl/co2-emission-of-cars-dataset

Description: This dataset contains the co2 emission of different types of cars. The independent feature variables are VOLUME and WEIGHT. By Using Multiple Linear regression models we can predict the amount of CO2 emitted by cars per unit volume and weight.

a) Importing data and understanding its structure.

b) Calculate the co-relation co-efficient of the features volume, weight and CO2 and display a heatmap.

c) Check and display the possible outliers in volume, weight and CO2 features. (Use box plot).

d) Visualize the relation of CO2 with other two features.

e) Train a multiple linear regression model (take train-test split as 80:20)

f) Calculate weight and intercept of the best fit regression line. Visualized the true vs predicted outcome using a line chart.

g) Calculate and display Mean Absolute Error, Mean Square Error, and Root Mean Square Error.

Experiment-3: Given below the link of a dataset to predict whether a user will click on an online ad based on their demographics, browsing behaviour, the context of the ad's display, and the time of day using Logistic regression.

[Logistic Regression : Logistic Regression is commonly used to estimate the probability that an instance belongs to a particular class. If the estimated probability that an instance is greater than 50%, then the model predicts that the instance belongs to that class 1, or else it predicts that it does not. This makes it a binary classifier.]

Dataset Link : https://www.kaggle.com/datasets/gabrielsantello/advertisement-click-on-ad

This dataset provides insights into user behaviour and online advertising, specifically focusing on predicting whether a user will click on an online advertisement. It contains user demographic information, browsing habits, and details related to the display of the advertisement. This dataset is ideal for building binary classification models to predict user interactions with online ads.

a) Load and display the information about the dataset.
b) Apply Exploratory data analysis of the dataset (Missing value treatment. Null value treatment, Data transformation and normalization)
c) Find co-relation among the features and display the heatmap.
d) Train a logistic regression model and calculate the model parameters. (Train-Test Split should be 80:20).
e) Also apply K-fold cross validation split and record all the results.
f) Display the classification report like (Precision, recall and F1-score).
g) Display the confusion matrix and ROC curve, Precision Recall curve.
h) Visualised the predicted vs actual outcome.