Module 7: Supervised Learning-1

Case Study – 1

edureka!

Case Study – 1

Objectives:

- Fit a model using binary classification using logistic regression.
- Identify correlated variables and form a less complex model.

Questions:

1. We will use acoustic features to distinguish a male voice from a female. Load the dataset from "voice.csv", identify the target variable and do a one-hot encoding for the same. Split the dataset into train-test with 20% of the data kept aside for testing.
   [Hint: Refer to LabelEncoder documentation in scikit-learn]

2. Fit a logistic regression model and measure the accuracy of the test set.
   [Hint: Refer to Linear Models section in scikit-learn]

3. Compute the correlation matrix that describes the dependence between all predictors and identifies the highly correlated predictors. Plot the correlation matrix using a seaborn heatmap.
   [Hint: Explore dataframe methods to identify appropriate methods]

4. Based on correlation remove those predictors that are correlated and fit a logistic regression model again and compare the accuracy with that of the previous model.
   [Hint: Identify correlated variable pairs and remove one among them]