Assignment 9B: Logistic Regression and KNN

- S. Vishakan, SSN College of Engineering, ML Using Python Online Course

Batch – 1

Report:

The provided CSV files, ServiceTrainData.csv and ServiceTestData.csv were used to create and qualify the Logistic Regression and KNN model.

The data under ServiceTrainData.csv was used to create the machine learning models.

The accuracy of prediction was checked using ServiceTestData.csv.

Analysis of Data:

The given data contained only float values with no missing columns or outliers. The data was cleaned well in advance. The dependent variable, Service, was a categorical data type, (Yes / No) which made it a classical problem for classification. Pair Plots and Correlation Matrices were drawn up to analyze the data.

From the correlation matrix and the heat-map used to represent it, it was evident that all 5 columns were highly correlated with at least 0.9 correlation values.

All of the 5 columns were used to create the model, nevertheless.

Both the logistical regression model and the KNN model fared well. KNN gave 100% accurate results.

Logistical Regression:

Co-efficients:

Oil Qual: 0.33572197
EnginePerf: -0.3356375
NormMileage: -1.15976788
TyreWear: -0.42440671
HVACwear: 1.47265798

Intercept: 5.24658803

Accuracy: 91.11%

Misclassified Samples: 12

K Nearest Neighbors:

Neighbors: 5

Accuracy: 100.00%

Misclassified Samples: 0

Output:

The models' output were included in a .csv file and included in this assignment's zip folder for reference.