Machine Learning 09012

Group Assignment

Machine Learning

Submitted by: **Maksim Gusev, David Mc Gowan, Senan Stanley, Rabin Sapkota**

Submitted to: **Donny Hurley**

Thursday 17 November 2022

1. Introduction

This report is a group based assignment that has been part of the Machine Learning module taken at ATU Sligo, to presents supervised analysis of the cars dataset (Peugeot and Opel-Corsa provided) based on their features using Linear aggression and Support Vector Machine (SVM) model.

Within the data provided we have browsed it, identified the feature functions, formatted it to remove inconsistency, selected the two possible model (Linear Regression and Support Machine Learning MVC) to train it with expectation of best possible result to compare the two algorithms through best scores and visualisations.

The finalised version of the project was drafted by individual coding efforts that has been put together in GitHub[1] Repository and with weekly based meeting, discussion and improvisation through the Microsoft Team application to agree upon this version of the report.

2. Methodology

2.1. Sample Variance and Standard Deviation

[1] https://github.com/DavinMcGowan/Machine-Learning-Group-Assignment-.git

3. Data Prepration

On the real world information obtained can be very vague and uneven. To get a certain output it should be cleansed and check for any errors to begin with, that assist the accuracy for the end-result. For this context, the original data we have cleared commas used in floating point numbers and replaced it with the decimal points, similarly with the NaNs being replaced by the most frequent value in column using SimpleImputer. Standardisation of the columns by fitting and transforming it to standard scalar to scale the variance makes our task more efficient for analysis. The first 14rows of the data i.e. ~20% of the data has been used as a training dataset to test it with the rest of the data i.e. ~80%. The last three columns of the data set we are interested in, the traffic, drivingStyle and roadSurface has non numeral values and has been handled using OrdinalEncoder that encodes its value as a categorical features with an integer array.

4. Results

5. Conclusion

6. References