Computer Science Department

COMP4388

Machine Learning

Dr: Radi Jarrar

Project one

Name: Amir Mohammad Eleyan

ID: 1191076

Date: 26/11/2021

- Part 3:

A correlation coefficient measures the strength of the linear association between two variables x and y. Therefore, the closer the value is to 1, the relationship will be strong. (1: Positive relation, 0: No relation, -1: Negative relation). i.e., ISI Vs. FWI, the relation is positive; because the correlation coefficient =0.92 which is close to 1.

- Part 4:

Based on the result that appeared, we note that there is no strong positive relationship between the others features and Classes, but there are some semi-strong negative relationships such as FFMC.

- Part 5.a:

The best feature that is best used to predict the "Fire Weather Index (FWI)" is ISI because the correlation between them is strong, around 0.93.

- Part 5.e:

Measures type	First Linear Model	First Linear Model	Third Linear Model
MAE	1.68	0.89	0.63
MSE	6.35	1.29	0.89
MSE	2.52	1.13	0.95
R-Squared	0.85	0.97	0.98

From these results we noticed the best model is third model because it has l ess value than MAE -values, that's mean the model has lower percentage of errors (That is, the lower its value, the lower the error rate in the model). When we compar e between these models we depend on the MAE because is the best choice to comp are between models. In contrast, to detect the accuracy of the specific model usuall y RME is the best option.

We can also compare models based on the value of R-Squared. Because the higher its value, the higher the accuracy of the model Also, the third model has the highest R-Squared value.

- Part 8:

Recall of logistic regression model = 0.9362
Precision of logistic regression model = 0.9362
Accuracy of logistic regression model = 0.9362
Error rate of logistic regression model = 0.0638

Recall of KNN model = 0.8936
Precision of KNN model = 0.8936
Accuracy of KNN model = 0.8936
Error rate of KNN model = 0.1064

The strange thing, based on these results, the measurements are the same for each model, and I think this is due to the nature of the training and test data. So, the performance of the logistics model is better than KNN model because he has highest accuracy.