

2019W-T3 BDM 3014 - Introduction to Artificial Intelligence 01

Lab 01 – Logistic Regression

Lucicarla Silva de Santana C0724974

Valeria Ferreira de Almeida Nobrega C0724858

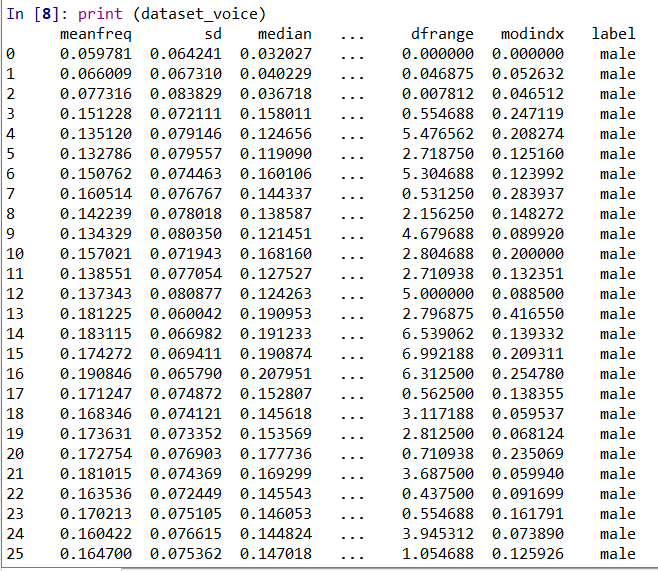
Rafael Andrade Da Conceicao C0725132

Tulio Fernandes C0722179

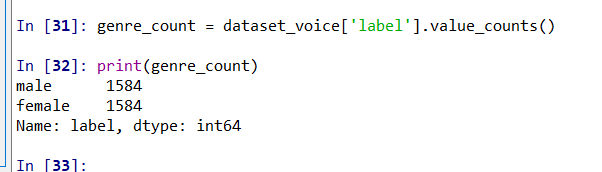
Javier Navarro Gonzalez C0725418

**Logistic regression - Python output**

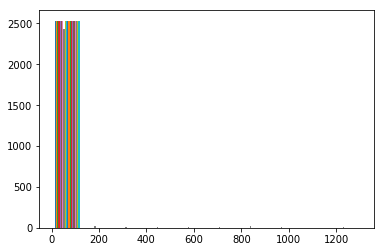
1. We will use acoustic features to distinguish a male voice from female. Load the dataset from “voice.csv”, identify the target variable and do a one-hot encoding for the same. Split the dataset in train-test with 20% of the data kept aside for testing.



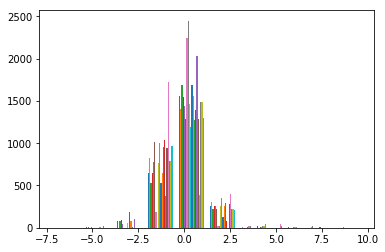
*Voice data set loaded*



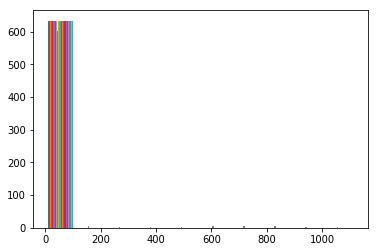
*Count of male and female*



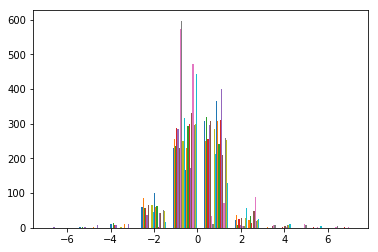
*X\_train before scaling*



*X\_train after scaling*

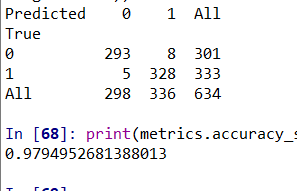


*X\_test before scaling*



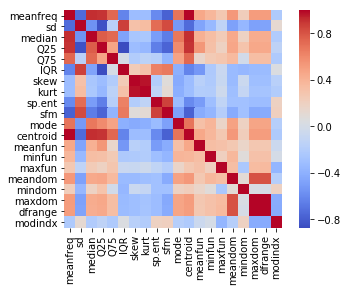
*X\_test after scaling*

1. Fit a logistic regression model and measure the accuracy on the test set.



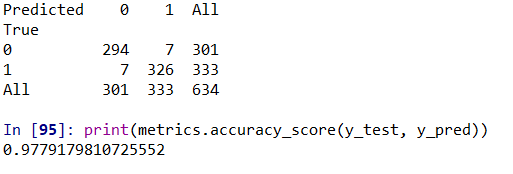
*Results of prediction using all the columns and accuracy*

1. Compute the correlation matrix that describes the dependence between all predictors and identify the predictors that are highly correlated. Plot the correlation matrix using seaborn heatmap.



*Heatmap of correlation*

1. Based on correlation remove those predictors that are correlated and fit a logistic regression model again and compare the accuracy with that of previous model.



*Confusion matrix and accuracy using 5 main features determined by RFE*