FYS-STK3155/4155 Applied Data Analysis and Machine Learning - Project 3

Lotsberg, Bernhard Nornes
Nguyen, Anh-Nguyet Lise

https://github.com/liseanh/FYS-STK4155-project3

November - December 2019

Abstract

Whole page: 6.24123in Column: 3.01682in

1 Introduction

2 Data

The data set we will analyse in this project is the MAGIC Gamma Telescope data set retrieved from the UCI Machine Learning Repository. The set consists of ten explanatory variables and a binary response variable. The data set is generated through a Monte Carlo simulation to **unfinished**

For technical reasons, the number of h events is underestimated. In the real data, the h class represents the majority of the events.

1. `fLength`: continuous # major axis of ellipse [mm]
2. `fWidth`: continuous # minor axis of ellipse [mm]
3. `fSize`: continuous # 10-log of sum of content of all pixels [in #phot]
4. `fConc`: continuous # ratio of sum of two highest pixels over fSize [ratio]
5. `fConc1`: continuous # ratio of highest pixel over fSize [ratio]
6. `fAsym`: continuous # distance from highest pixel to center, projected onto major axis [mm]
7. `fM3Long`: continuous # 3rd root of third moment along major axis [mm]
8. `fM3Trans`: continuous # 3rd root of third moment along minor axis [mm]
9. `fAlpha`: continuous # angle of major axis with vector to origin [deg]
10. `fDist`: continuous # distance from origin to center of ellipse [mm]
11. `class`: g,h # gamma (signal), hadron (background)

3 Method

4 Results

5 Discussion

6 Conclusion

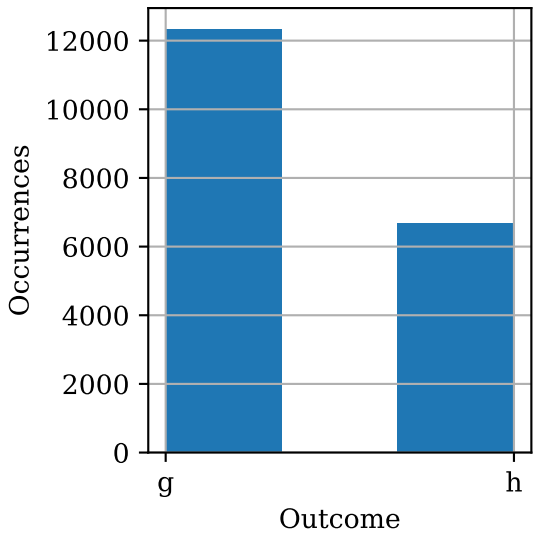

Figure 1: Frequencies of the binary outcomes in the data set.

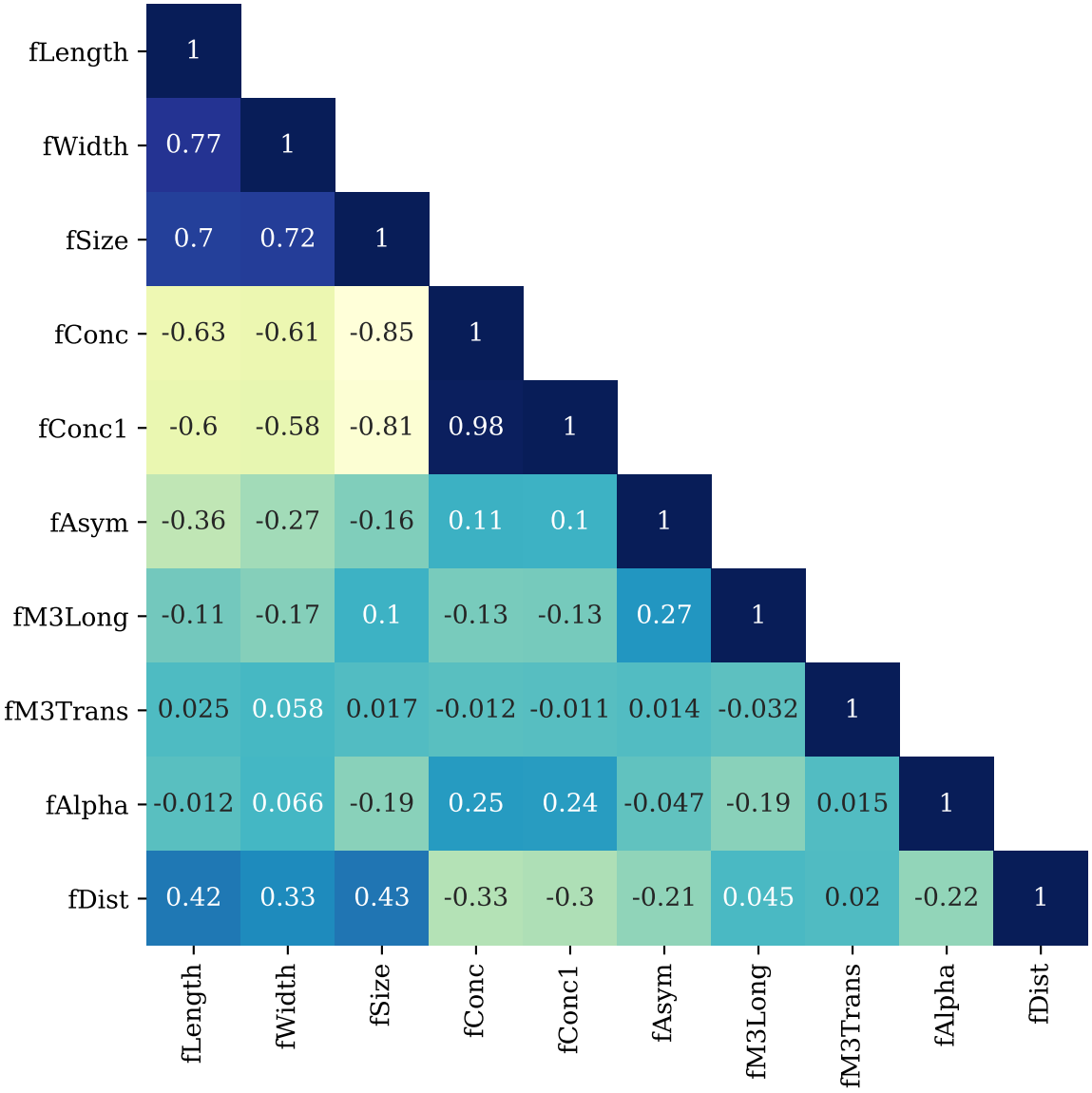

Figure 2: Correlation matrix of the features in the train set. Upper triangle excluded for readability.

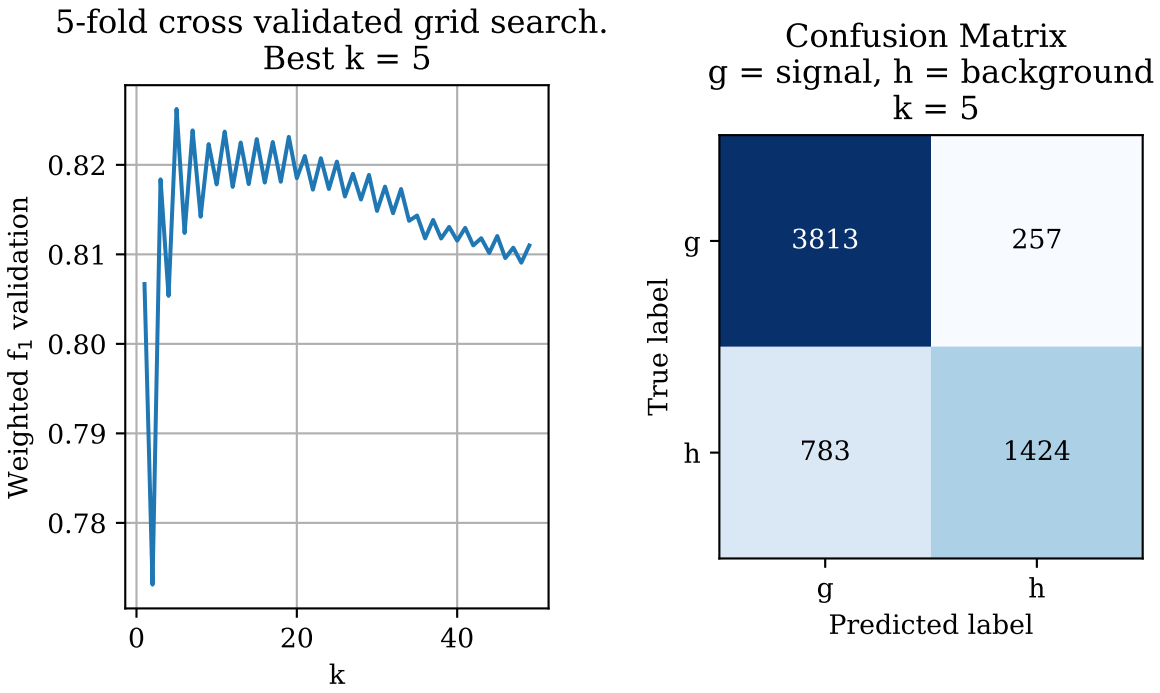

Figure 3: Results from tuned kNN using cross validation. The confusion matrix was found using the test set.

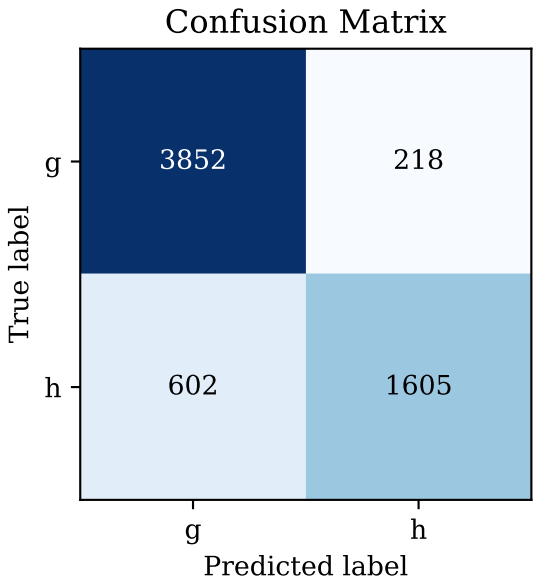

Figure 4: Confusion matrix of the neural network model applied to the test set.

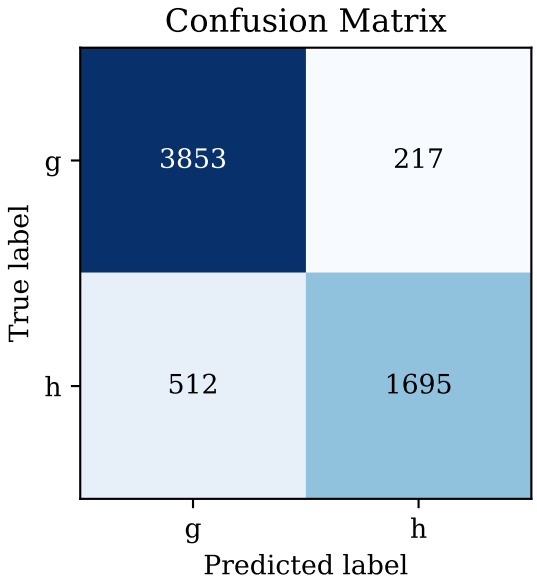

Figure 5: Confusion matrix of the gradient boosted model applied to the test set.