# ISYE 8803 Homework 1 Problem 4

Nick DiNapoli, ndinapoli6@gatech.edu

May 29, 2022

## 1 Problem 4

In this problem, I analyze electrocardiogram (ECG) signals that are labeled as normal or abnormal. The dataset is split into training and testing sets and I use the B-splines and functional principal component analysis (FPCA) techniques to extract relevant features and reduce dimensionality when treating these signals as functional data. After applying said techniques, I classify the signals in the test set at normal or abnormal using a simple random forest classifier and I evaluate the performance of the two techniques. I aim to discover if these methods can successfully detect cardiac abnormality. The training data and testing data are shown in Figures 1 & 2 respectively.



Figure 1: ECG signals that make up the training set.

Figure 2: ECG signals that make up the test set.

## 1.1 B-splines

First, I use B-splines to make classifications on the test set after training with the training set. As mentioned, I use a random forest classifier to complete this task. To analyze the effectiveness of using B-splines I document the confusion matrix in Figure 3. Note that a 1 class label represents a normal ECG and a -1 label represents an abnormal ECG.
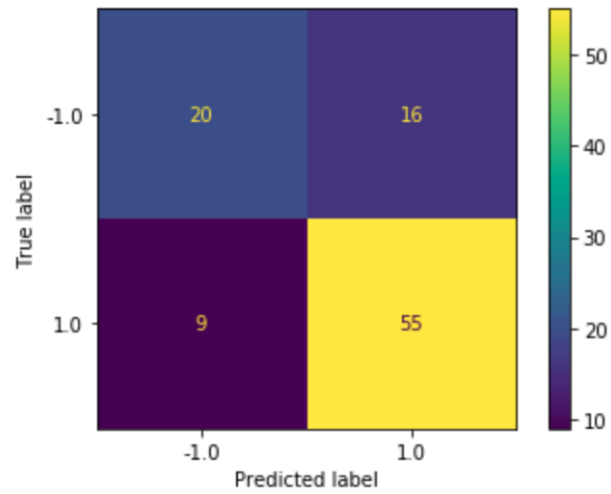


Figure 3: Confusion matrix for the test set after using B-splines and a random forest classifier.

## 1.2 FPCA

Next, to make a direct comparison to using B-splines, I complete an identical task to the previous section but use FPCA on the functional data instead. For this task I used the top 2 principal components. Once again, to evaluate and compare the performance, I visualize the confusion matrix in Figure 4. As it can be seen by the two confusion matrices, the two methods produces near identical predictions on the test set.
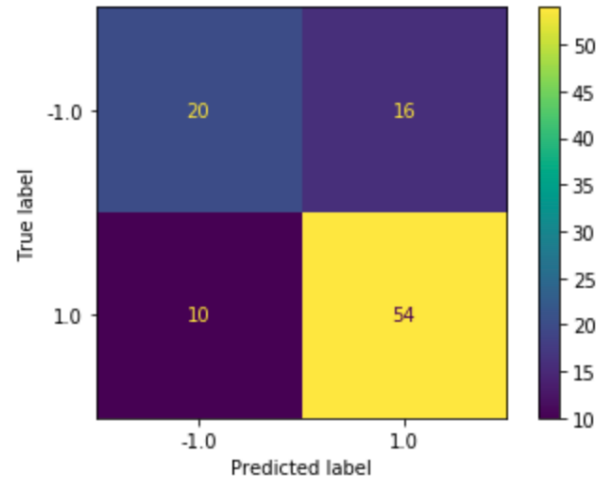


Figure 4: Confusion matrix for the test set after using B-splines and a random forest classifier.

# References

[1]