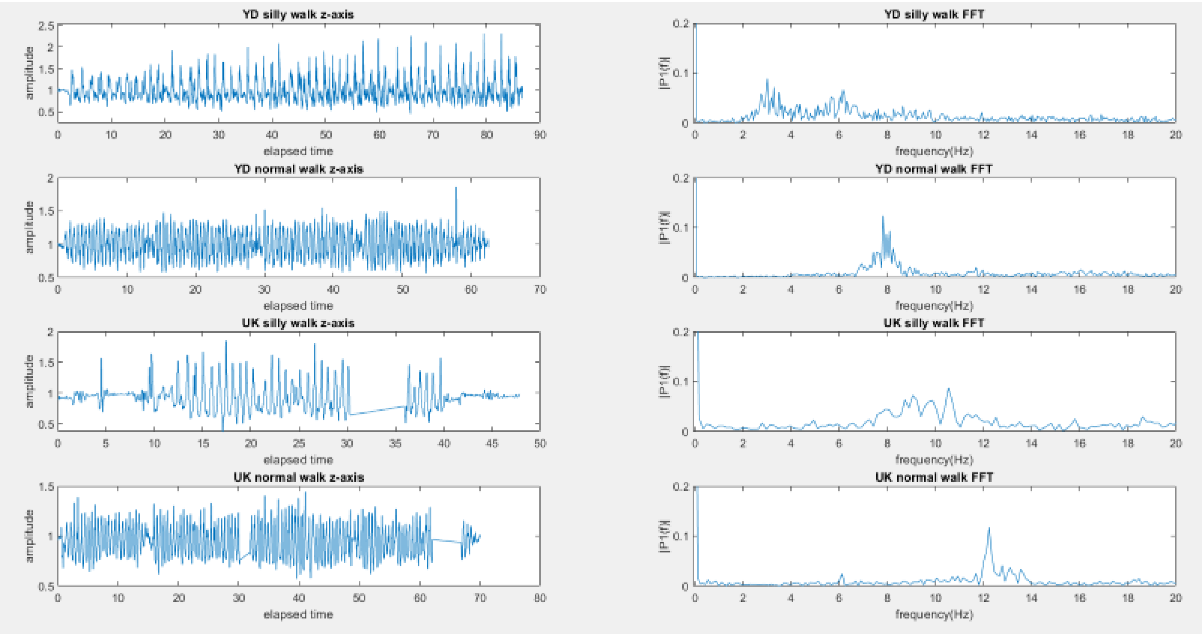
Walking acceleration signal analysis

Problem: Use FFT to create features from your examples of Silly Walks, plus examples of normal walking, and classify them using hand-labeled supervised learning

Deliverable 1: Document your approach to classification. This should be the same for each of your Silly Walks, but you should both train and run the classifier separately with your walk as the positive example. Compare the results. Describe your shared assessment of good and bad habits of data acquisition.

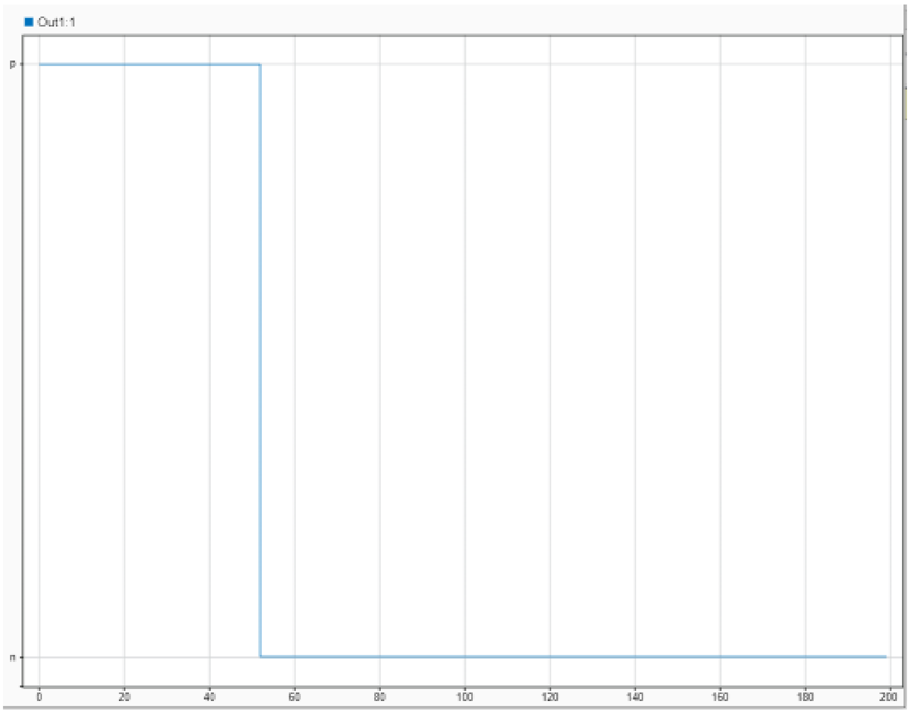
Accelerometer data from YD and UK's silly walks and normal walks:



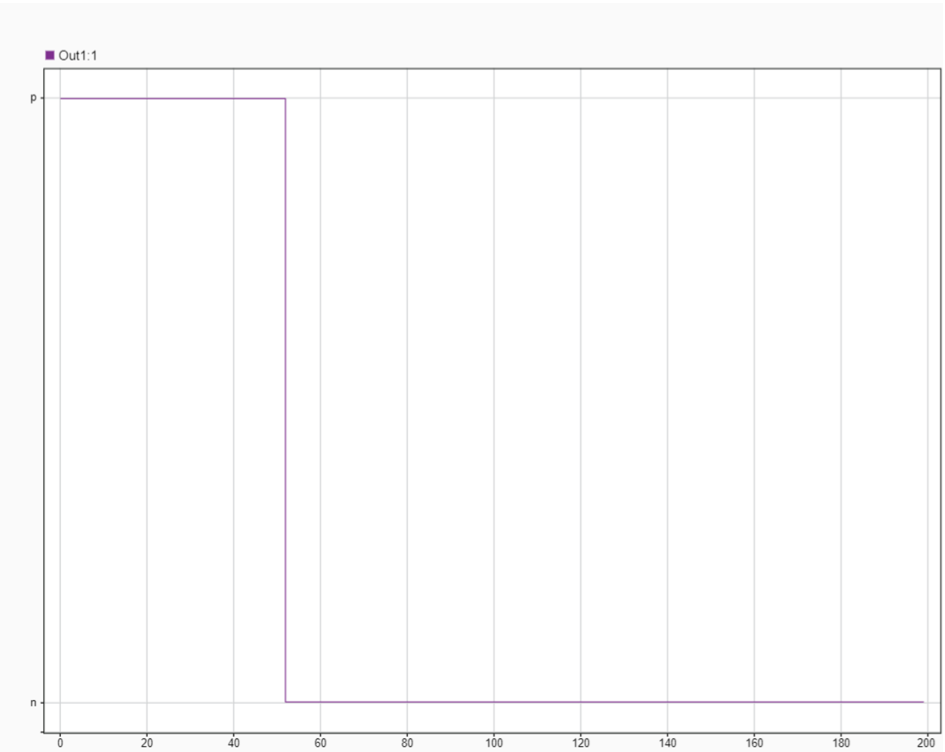
Approach for classification:

We took 150 random segments each from YD silly walk, YD normal walk, UK silly walk, and UK normal walk, then did FFT to each pf the segment and extract the first 10 largest components as the features for classification. These 600 segments with 10 components are used to create a dataset (600 rows) that contains both the positive and negative data. We hand-labeled the 600 rows of data as either 'p' for positive samples and 'n' for negative samples. Then we use 400 rows of data (prsntX) and their labels (prsntY) as the training set, and the rest of the data as the testing set. Next, we use the training set to train the SVM model then input ftrX to the SVM model to classify and get the generated label.

SVM predictions when using YD silly walk as positive examples:



SVM predictions when using UK silly walk as positive examples:



Results of classifying YD silly walk and UK silly walks both show 100% correct predictions from the SVM model. However, this model might still have the issue of overfitting.

Good and bad habits of data acquisition:

- Good habits of data acquisition include trying to hold the phone stable at a position to reduce noise when doing silly walks and normal walks for recording accelerometer data. In addition, making sure to keep the walking pattern consistent when collecting data. The length of the data sequences should include at least 2 good cycles of the entire motion for the walk. The shape of the data collected should be similar and repeated.

- Bad habits would be doing too many sharp turns or sudden stops when recoding data, as well as frequently changing the position of the phone relative to our body during data recording that can add a lot of noise to the data. Plus, if we record the data sequence too short (e.g. less than 2 cycles of the walk) or walking in a way that has too many variations such as the speed, the length of the steps etc., the data may not be good enough to train a classifier.

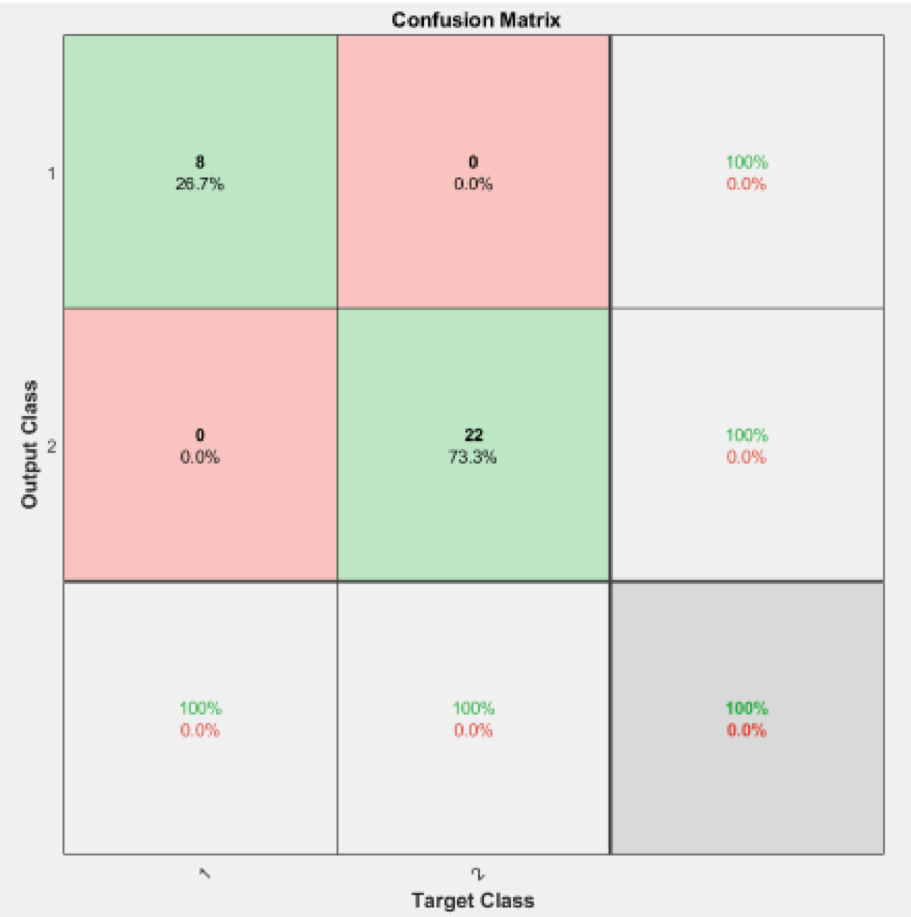
Problem: Apply Shallow Neural Networks to examples of Silly Walks, plus examples of normal walking

Deliverable 1: Include your results from the example.

Apply neural networks to your Silly Walks

Results when apply neural networks to examples (YD silly walk as positive examples):

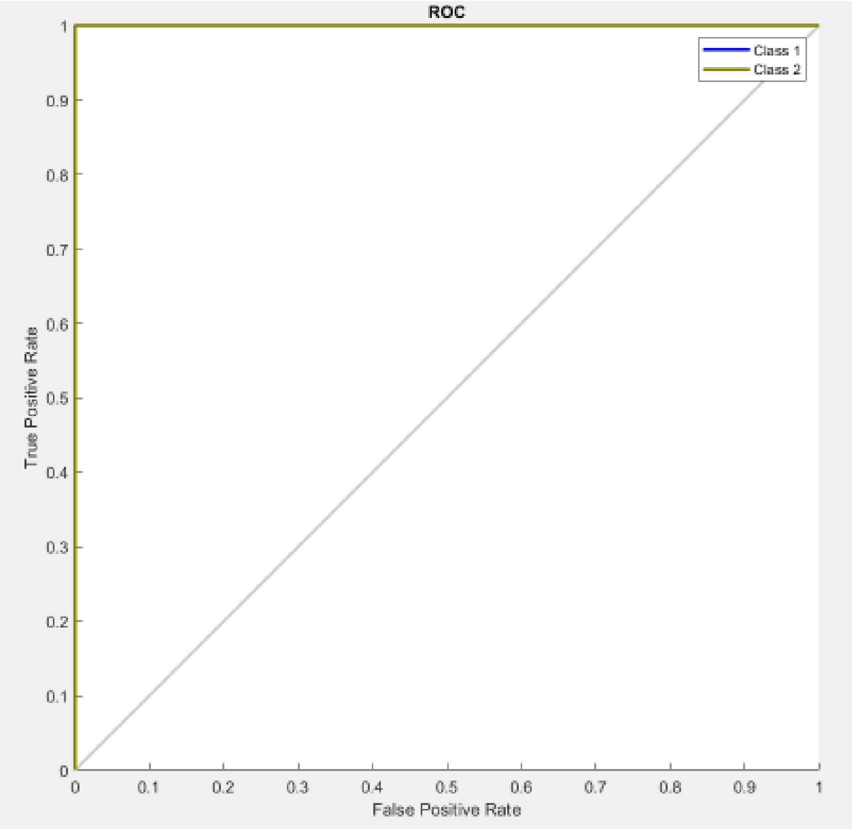
Confusion matrix:



Overall percentages of correct and incorrect classification:

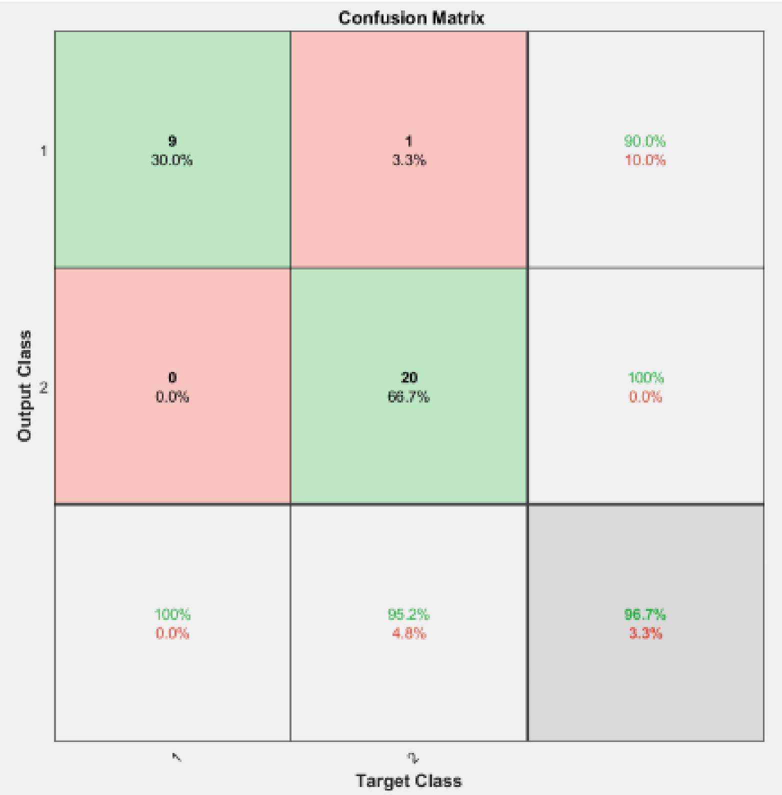
```
>> fprintf('Percentage Correct Classification   : %f%%\n', 100*(1-c));
fprintf('Percentage Incorrect Classification : %f%%\n', 100*c);
Percentage Correct Classification   : 100.000000%
Percentage Incorrect Classification : 0.000000%
>> |
```

Receiver operating characteristic plot:



Results when apply neural networks to examples (UK silly walk as positive examples):

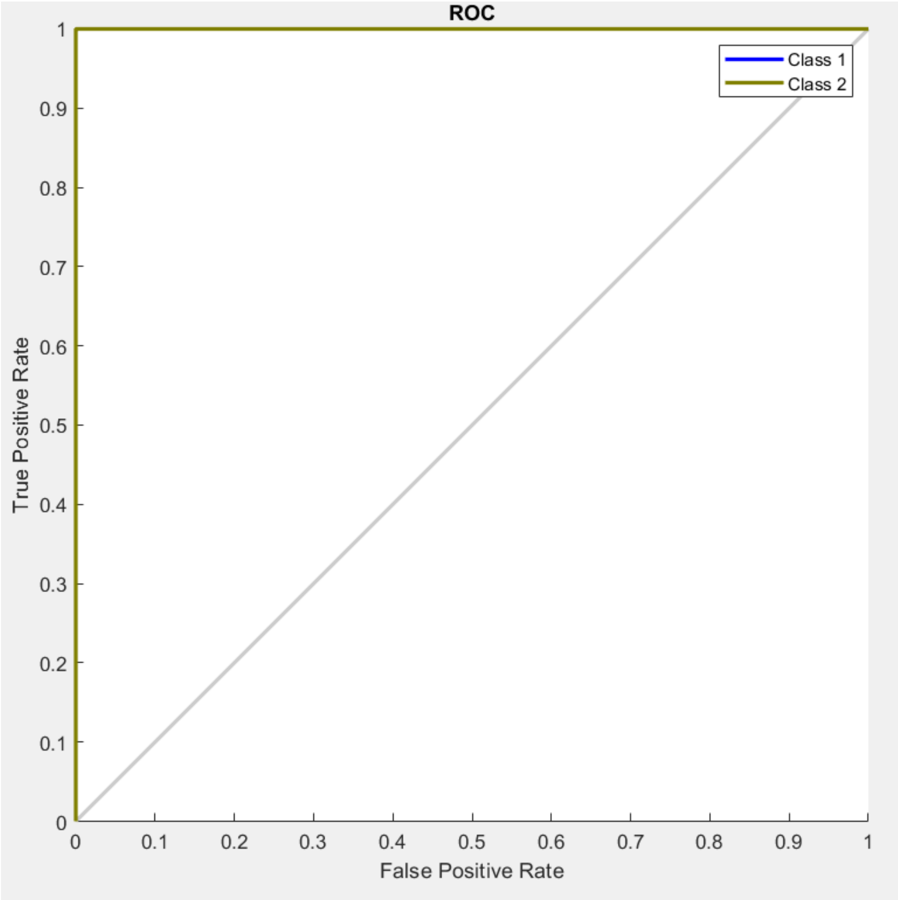
Confusion matrix:



Overall percentages of correct and incorrect classification:

```
>> fprintf('Percentage Correct Classification   : %f%%\n', 100*(1-c));
fprintf('Percentage Incorrect Classification : %f%%\n', 100*c);
Percentage Correct Classification   : 96.666667%
Percentage Incorrect Classification : 3.333333%
```

Receiver operating characteristic plot:



Deliverable 2: Describe your training and testing data sets, your classification approach, and the results. Discuss what could you have done differently, in your data acquisition, to get a better result?

We took 50 random segments each from YD silly walk, YD normal walk, UK silly walk, and UK normal walk, then did FFT to each pf the segment and extract the first 10 largest components as the features for classification. These 200 segments with 10 components are used to create a dataset (200 rows) that contains both the positive and negative data. We hand-labeled the data that has 1 in the first row and 0 in the second row for positive samples and 0 in the first row and 1 in the second row for negative samples. Then we use those data (X) and their labels (T) to train the neural networks for classification.

The results both showed high percentage of correct classification for YD silly walk as positive example or UK silly walk as positive example.

What we could have done differently is to slightly change the way we do for our silly walks and do more recordings for all those different variations of the same person's silly walk. For example, walking in a faster speed or exaggerate our movement when doing the silly walk to increase the amplitude. By doing those things, the classifier would be more adaptive to more different situations when doing classification and keep high accuracy at the same time.