Kristianstad University   Fredrik Frisk
www.hkr.se   Computer Science Department

Content and purpose of this lab

The main focus of the lab is to more classification.

Save the code you are writing in this lab for future use. To pass the lab you need to solve/program the different bullet points and be able to explain your results. If you are not finished with the all the bullet points the remaining ones are a part of the required preparations for part 4 of the labs. The lab report in the end is an individual report, but you are allowed to work two and two with one exception all of you have to record your own sensor data.

Preparations

You need to finish part two of the lab and be able to show the result in the beginning of the lab.

Also you need to do the following recordings of data. For the below recordings you have to use three sensors:

- Accelerometer, sample frequency 100 Hz

- Rategyro, sample frequency 100 Hz

- Magnetometer/Compass, sample frequency 10 Hz

Furthermore you need to have done the following recordings:

- Running, 6 recordings, at least 10 seconds each

- Walking, 6 recordings, at least 10 seconds each

- Jumping, 6 recordings, at least 10 second each

- Push-up or other type of movement (but not the same as the above), 6 recordings, at least 10 seconds each

- 2 recordings when you do through all movements after each other. You need to do each movement for at least 5 seconds

You need to wear the phone in a similar way for all the recordings, for example in your pocket.

Take notes, or even better take a picture of how you wear the phone and take notes or record videos of your movements. But it is important that you know what you did for each recording.

If you do not have the necessary mobile for recording this data, you can borrow one from Haydar.

NOTE: before preprocessing, correlation calculation actually before you start. Separate the testdata! The test data should note be used for anything else than testing the classifier in the end! See the last paragraph regarding train, validate, test.

A first classification

Try to classify the data as is. To get a starting point make a pairplot and also a "standard" plot or time series plot. That is you plot the different features as a function of time/samples.

- Which features seems most useful?

- Can you explain the results in your confusion matrix?

Use one classifier of your choice.

Preprocess the data

During this step you should decide

- Sampling frequency. Should you keep 100 Hz? How do solve this with the magnetometer? What sampling frequency is needed for the classification?

- Calculate the standard statistical values for each recording and each feature. Use (`.decide()`) for example. Try to explain the results you get.

- Decide if you need to scale the data or not. Which scaler do you choose? What do you expect from the choice? Better or worse result? Motivate your answer.

- Transform the the acceleration to magnitude and angle. Keep this as a separate dataframe for future evaluation, i.e. classification

- Transform the magnetometer data in the same way.

Note he two transformation you have already done before in a previous part of the lab.

When we move the values from the sensors varies over time.

- Investigate if you can use some statistical measure for each class for classification purposes.

- Examine the frequency the pattern of your movements repeats itself for the different movements. Could this be helpful for the classification?

Correlation

One can use correlation to investigate the dependence between variables/features.

- Calculate the correlation matrix between all features. From this result, discuss if all features are useful for the classification

In pandas there is an autocorrelation plotfunction that can maybe of use.

- Calculate the autocorrelation for all classes. Do you see a pattern? Explain what you see.

Classification

You should use at least 2 different classifiers. You have worked with KNN, decision trees and Softmax Regression before. One of the classifiers should be the Softmax regression.

- You need to maximize the performance for each classifier by changing values of relevant hyperparameters.

- Change the number of features for your classification. How does the performance change with the choice of the features? Use your results from preprocessing and correlation.

- Performance is measured by accuracy and confusion matrix.

Train, Validate and Test

You should always divide the data in train/validation and test sets. Use crossvalidation for the training/validation. Also use 5 of the 6 recordings (for each class) for the train/validation and keep the last one for test. Also keep the mixed recordings for testing, compare the two test scenarios.