Tallinn University of Technology

School of Information Technologies

Department of Health Technologies

Classification Validation Methods and Decision Tree Classification

Laboratory Guide for Lab Assignments 3 and 4, v1.0

Tallinn 2020

Aim: The aim of these lab assignments is to give the student a practical experience on how to deploy a multi-class classification tree based algorithm to classify between different physical activities. This is often done in multiple human activity monitoring and recognition studies. Laboratory guide is based on MATLAB software.

Tasks of the lab assignment: 1) Open file *AccSignal.mat*. This file contains 3D acceleration signals (vectors *accX*, *accY*, *accZ*) measured with an accelerometer attached on human body during which the human conducted various physical activities. Sampling rate of the accelerometer was 25 Hz. Plot these vectors together in one graph and observe the signal. Are you able to differentiate between different activities based on the graph?

2) In human activity recognition studies we want to classify the activity of the user during a certain time window. For example – we want to recognise the activity after every 5 seconds. In this case this time window is called "classification window length". The previously observed acceleration signal is already divided into smaller segments with as window length of 5 seconds. These segments are stored in matrices (*accSegX*, *accSegY*, *accSegZ*, for each different axis). Open one of the segments and compare it to the original acceleration vector. Understand how the data is segmented.

(Observe that the length of the matrix (rows) is the window length times the sampling frequency of the signal. The length of the matrix (columns) is the number of segments. In total, the length of x axis multiplied by the length of y axis should give the length of the original acceleration vector.)

The activity the user conducted during each segment is stored in a vector (*accLabel*). There are four different types of activities during which the signals were measured (and which need to be classified):

1. Static (either lying, standing or sitting)
2. Walking
3. Running
4. Other (this includes various upper-body movement)

3) For classification, we need to extract features based on the signals. The feature set must capture specific and diverse information of posture and human motion to allow precise activity classification. In this lab assignment, for each segment, calculate the following features (separately for X, Y and Z axis) and store them in a matrix:

1-3. Mean of the absolute values of the segment (function *mean*)
4-6. Standard deviation of the segment (function *std*)

4) Train a decision tree based classifier based on the training set. For this, use the function *fitctree*.

5) It is possible to graphically represent the classification tree. For this, use function *view* with: view(classifierTree,'mode','graph'), where classifierTree is the name of the classifier variable created in the last step. Observe the tree: is it possible to understand how the classifier works based on the image?

If everything is done correctly, then we should now have a working classifier that is able to classify between four different physical activities.

6) To see how well our classifier performs, we need evaluate the classifier. In order to evaluate our classifier, we need to separate data for the training set and evaluation set. As the names say, the training set is used to train the classifier and evaluation set is used to evaluate the classifier.

In the lab assignment we are going to evaluate how the classifier works for signals measured in the same way but from another individual. This validation set can be found in *accSignalVal.mat*, which contains same segments as our original file.

1. Find the same features based on the validation set.
2. Classify the evaluation set features using function *predict*.

What would be other solutions for having separate data for the validation set?

7) There are different characteristics that can be used for classifier evaluation. To validate the results of the classifier created during this lab assignment, create/find the following elements:

- Confusion Matrix of the classification results. Possible to use the function *confusionmat*.
- Find the True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) of each activity type.
- For each different activity type, provide the following characteristics:
    1. True Positive Rate (sensitivity/recall)
    2. True Negative Rate (specifity/selectivity)
    3. False Positive Rate (false alarm rate/type I error)
    4. False Negative Rate (miss rate/type II error)
    5. Accuracy

How well did the classifier perform? For which activity type did the classifier have the best performance with (and for which type the lowest performance)?

8) Present your findings and code to the supervisor. Be ready to explain your code and the steps taken to achieve the results.