VRIJE UNIVERSITEIT AMSTERDAM

Assignment 1
Machine Learning for the Quantified Self
Basics, removing noise, and extracting features

Ali el Hassouni and Mark Hoogendoorn

Deadline: 09/06/2019, 23:59

1 SELECTED BOOK ASSIGNMENTS

Within this assignment, you are going to explore the first aspects of machine learning for the quantified self. More concrete, you are going to work on more theoretical aspects of the algorithms and domain, and also work with some real data and do some programming. Three topics will be addressed:

- Basics of sensory data
- Removal of sensory noise
- Feature engineering

For the theoretical part, you should answer the following assignments from the book:

- Chapter 2: 1, 2, and 3
- Chapter 3: 2 and 4
- Chapter 4: 1, 6, and 7

For the practical assignments the same holds, although you can choose which assignments to do:

- Chapter 2: compulsory: 1 and select either 2 or 3
- Chapter 3: compulsory: 3 and select either 2 or 4
- Chapter 4: compulsory: 1 and select either 2 or 3

For the practical assignments we will (among other datasets) use the CrowdSignals dataset. It can be downloaded from

`http://www.cs.vu.nl/~mhoogen/ml4qs/crowdsignals.zip`

The source code you can use as a basis for the practical assignments can be downloaded from GitHub

`https://github.com/mhoogen/ML4QS/`

Note that we will use the Python code, not the R code as the R code is not complete and has not been properly tested. If you want to use the code in combination with data you generate yourself, please use the format used in the Tables posted on page 17 of the book for your data, otherwise you will need to make changes to the code (which should not be too difficult). **Warning**: when you use Excel with the time stamps in milliseconds since the start of UNIX time, Excel will round these numbers, which you do not want.

Running the algorithms can be time consuming, be aware of this and start early. You can change the step size to reduce the computation time required if you think you will run out of time. In addition, in the directory `http://www.cs.vu.nl/~mhoogen/ml4qs/` you can find intermediate data files (resulting after each of the chapters). Furthermore, you are also allowed to work with a subset of the data (e.g. the first half hour or full hour), you will need to select this subset yourself.

2 Submission requirements

You should work in groups of three students. You should write a report covering your answers to the questions (see the next section for the criteria). The report should follow the Springer Lecture Notes in Computer Science template (you can easily find it if you Google it, there are both LateX and Word templates) and should be at most 8 pages. You can submit via Canvas.

3 Grading

The assignments will be graded using certain criteria. The theoretical component counts for 40% while the practical component counts for the remaining 60%. For the theoretical part the following criteria are used:

- correctness (40%)

- explanation (40%)
- references (where applicable) (20%)

For the practical part these criteria are applied:

- description of setup (20%)
- rationale for choices (15%)
- description of results (40%)
- interpretation of results (25%)