

Second project of Data Mining

Mario Martin

November 10, 2016

This document describes rules and things to do for the second part of the project. The presentation of the project will consist in an oral presentation by all members of the group plus the delivery of the documentation. Presentation of the project will be on **22th of December** on time scheduled for the laboratory.

Documentation will be delivered in pdf format just the day before the presentation, that is the 21th of December until midnight. It has to be delivered by e-mail to mmartin@lsi.upc.edu.

Oral presentation will follow the same rules that were followed for the first project. That is:

- All members should participate in the oral presentation.
- Presentation will be done in front of your colleagues.
- Presentation will not last more that 20-25 minutes, including time for questions.

Member of the group could obtain different marks depending on the implication in the work and the presentation done.

1 What should be done in the project?

In this second project you will:

1. Select a non-trial dataset for a classification task
2. Do the pre-processing necessary for you data. Also, describe and justify it in the documentation.
3. Apply the methods explained in theory lectures to your dataset. If the algorithm requires it, find the best parameters for the algorithm. Tuning of parameters has to be done in a meaningful way. An explanation of the procedure followed and parameters tested should appear in the documentation.
4. Try to interpret the models returned by the algorithms, specially in the case of decision trees.
5. Evaluate and compare different methods applied.

6. Discuss why you think that one method works better than another in *your* dataset.

Of course, for the second project you can apply if necessary any technique learned in the first part of the course.

2 What should be specially considered?

Accuracy values are not the most relevant outcomes of your work. You should show that you are able to understand the data mining process, to reason, to take decisions, and to evaluate results.