Intelligent Data Analysis Computer Laboratory
Practice on classification and clustering software

26th of January – 3rd of February 2012

Arnaud Quirin <arnaud.quirin@softcomputing.es>
http://aquirin.ovh.org/mastersoft/

Exercise

The exercise can be realized during the laboratory, or at home. In the latter case, the deadline is the 31st of March. The estimated time of the exercise is ~5 hours of personal work. Here are the steps to follow.

Load one of the data present in the "exercise" folder of the ZIP file. These are all real-world data extracted from satellite (multispectral or hyperspectral) images of Strasbourg and/or Venice. The input are the values of the pixels in a N-dimensional space, and the output is the class of the pixels (nominal).

The goal of the exercise is to:

1. Train several models (classifiers) on the original and/or filtered data.
2. Save the trained models on the disk and apply it to the provided test set.
3. Keep only the model(s) having the best mean square error on the test set.
4. Explain the methodology used (the process followed) to select the winning model and write the obtained error in an email to me (email address in the first page).

Notes:

- Any software (Weka or KEEL) can be used.
- Exercises are *individuals*.
- The explanation should be sufficient to reproduce the methodology in a latter stage (include the algorithms used, the parameter values if the default one are not used, fix a random seed for the non-deterministic algorithms, etc).
- Notation is *not* based on the final obtained error, but on the methodology followed.
- Models, experiment configuration files and experiment result files can be send by email if inferior to 5MB when ZIPPED.