Advanced Topics in Computational Intelligence
(ATCI – MAI)
March 2015
Practical Project I: FIR and GFS

This is the first practical project of the ATCI course that has to be developed in groups of **two** people and should be delivered by **April 10th**.

Introduction

In this project it is proposed to compare different hybrid fuzzy logic techniques to predict human wine test preferences based on physicochemical properties from wine analyses. The wine data used for this study comes from the Portugal north-west region, named Minho, and it is available from the UCI machine learning repository. It has been proposed for both, regression and classification, by Cortez et al. (2009).

In this work you should analyze the white variant from the mentioned demarcated region as a regression problem. To this end, you will use the FIR methodology and, at least, two different GFS algorithms of your choice. It is recommended to use the VisualFIR and KEEL software, since you become acquainted with them in the FIR+GFS class lab.

Data

The data were collected from May 2004 to February 2007. This dataset is much larger than others available as benchmarks in the same domain. The more common physicochemical tests are measured, and correspond to the following 11 input variables: Fixed acidity, Volatile acidity, Citric acid, Residual sugar, Chlorides, Free sulfur dioxide, Total sulfur dioxide, Density, pH, Sulphates and Alcohol.

Each one of the 4898 wine samples was evaluated by a minimum of three sensory assessors, by means of blind tastes, which graded the wine in a scale that ranges from 0 to 10, that matches to very bad to excellent quality, respectively. The final score is given by the median of these evaluations, which corresponds to the output variable. This target variable denotes a typical normal shape distribution, with minimum and maximum values of 3 and 9 for the white wine. You can find more details related to the data in (Cortez et al., 2009).

The data has been already prepared as both Visual-FIR and KEEL software require. You will use a 5 fold cross-validation as proposed in (Cortez et al., 2009). In this manner you would be able to compare your hybrid fuzzy techniques results with the ones shown in that paper from multiple regression, neural networks and support vector machine techniques.

The data sets for FIR are provided in the FIRdata folder. The data sets for KEEL software are already included in the platform, so it will appear in the list of possible data sets to be studied when you open the KEEL software and you start a new regression experiment. Its name is WineWhite. Once selected, you should edit the

dataset, remove the default 10-fold CV, and add a 5-fold CV. Then you can proceed as usual to create your new experiment.

Expectations

It is expected that you analyze the problem to be modelled and study different hybrid fuzzy techniques to perform this task. To this end, you should compare different configuration parameters of the selected methodologies in order to arrive to some conclusions. You must design the experiments that you consider appropriate to draw your own conclusions of this study.

You should for example:

- Compare different discretization parameters in FIR methodology
- Compare different masks an get conclusions out of it about the feature selection preformed by FIR
- Compare different GFS approaches
- Compare the computational time needed for each methodology
- etc.

It is expected that you compare the wine quality prediction results obtained by using at least the mean absolute deviation (MAD) error and the regression error characteristic (REC) curve. You can find a description of both in (Cortez et al., 2009).

It is expected that you report the experiments that you have performed, i.e. goals, decisions, results, comparisons and conclusions, in a document being it clear and reasoned.

Project delivery

The deadline for delivering your project report is April 10th. You should send it to angela@cs.upc.edu.

The final report should include:

1. A description of the work and its goals
3. The hybrid fuzzy methods considered, reasoning the choice
4. The results obtained with each method, along with the best set of parameters for each of them
5. Comparison of the results for each method and their significance
6. Scientific and personal conclusions
7. Possible extensions
8. Known strengths and weaknesses

References

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., Reis, J., 2009. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4), 547-553.