INTELLIGENT SYSTEMS

1st Semester 2024/2025

Assignment 1

Instructions

- This assignment consists of two exercises regarding fuzzy modelling of a regression problem and of a classification problem.
- Use Python, MATLAB or other language to solve the assignment, but do your project in a collaborative way using GitHub.
- Deadline for submission of the report is September 27th, 2024.
- Deliver all data files in GitHub, where the best models that you have developed should be included. Deliver also a script file in the language that you have chosen, which performs the required steps to obtain the developed model. Further, deliver a file `G**_A1_report.pdf` in which you explain how the model parameters were selected and present the main results. This report can be a notebook or a live script as long as it includes the above explanations and is converted to pdf. Make sure that the names and student numbers of your team members are indicated clearly.

The report must be submitted in the Fenix system (use you group number and your ISTIDs to name the report submitted on Fenix) and it is mandatory to include the link to the work on GitHub.

Hair Dryer Dataset

In this exercise you will model a single input single output (SISO) fuzzy rule-based system using the hair dryer dataset (Feedback's Process Trainer PT326; See also page 525 in Ljung, 1999).

The process works as follows: Air is fanned through a tube and heated at the inlet. The air temperature is measured by a thermocouple at the outlet. The input is the voltage over the heating device, which is just a mesh of resistor wires. The output is the outlet air temperature represented by the measured thermocouple voltage.

The dataset is available at "./data/hairdryer.csv". Pre-process the data if necessary and divide the dataset into a training set (to derive the model) and a test set (to validate the model) accordingly.

Wisconsin Breast Cancer Original Dataset

In this exercise you will develop a fuzzy rule-based model for classifying the classes in the Wisconsin Breast Cancer Original Dataset. The dataset is binary, meaning that there are two classes (0 and 1) for the existence (or not) of breast cancer in each sample (patient measurements).

The model must predict the class label using multiple integer input variables describing different measures which are relevant for cancer diagnosis. Specifically, the input variables are:

- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape
- Marginal Adhesion
- Single Epithelial Cell Size
- Bland Chromatin
- Normal Nucleoli
- Mitoses

The dataset is available at "./data/wbco.csv". Pre-process the data if necessary and divide the dataset into a training set (to derive the model) and a test set (to validate the model) accordingly. The two data sets should preserve the class distribution of both class samples.

Hints

You are advised to do all the steps by writing a script or a Jupyter notebook, since this facilitates batch processing for your experiments and save much "clicking" time. Note that the model performance can be evaluated by using an error measure, such as the mean squared error (MSE) or variance accounted for (VAF) for the regression model, or the accuracy for the classifier; see http://en.wikipedia.org/wiki/Accuracy_and_precision. You are free to use any other evaluation measure that you think is suitable for the exercise. A discussion on the appropriateness of the evaluation measures is also welcome.