

# 1BM120 - Computational Intelligence

## Assignment 2

**Ir. Caro Fuchs - Dr. Marco S. Nobile**

Assignment 2 is composed of two tasks: knowledge-driven modeling and data-driven modeling. That is, in the former case you must create two fuzzy inference systems (FIS) using domain knowledge, while in the latter case the model will be generated automatically using a machine learning approach.

### Fuzzy models for clinical decision support (10 points)

When a patient enters the Emergency Department of a hospital, his vitals and blood levels are immediately checked by a nurse. The following measurements are deemed important and are recorded by the nurse:

- Heart beats (per minute)
- Systolic blood pressure (in mmHg)
- Glucose level (mmol/L in the blood)
- Leukocyte count (per nL in the blood)

These combined measurements give insights in the health status of the patient and indicate how severe his condition is. Therefore, the head nurse decides based on these measurements in which order the patients should be examined by the doctors. However, there is only one head nurse available in the hospital, which leads to problems when multiple patients enter the Emergency Department at the same time. This also means that this nurse can not leave her working spot for breaks.

The management of the Emergency department has asked you to come up with an automated system to order the patients based on the severity of their condition and to build a prototype of this system to show the viability. Unfortunately, there is no data available, so you will have to go the 'knowledge-driven' way.

You have decided to develop a fuzzy model that gives each patient a health status score based on their bloodwork. What this score looks like (for example the range) is up to you. You can use any source to determine the parameters of the fuzzy sets. For example, for the blood level variables you could use: [https://www.catharinaziekenhuis.nl/files/Verwijzer/Specialismen\\_en\\_afdelingen/Laboratorium\\_algemeen\\_klinisch/Referentiewaarden/20210406\\_appreferentiewaardenlijst\\_V37.pdf](https://www.catharinaziekenhuis.nl/files/Verwijzer/Specialismen_en_afdelingen/Laboratorium_algemeen_klinisch/Referentiewaarden/20210406_appreferentiewaardenlijst_V37.pdf). Since 'healthy' values can be different for different demographic groups, you can assume your model has to work for adult males.

Document your modeling decisions and assumptions well and refer to the sources you used to determine the parameter settings. Show with some examples your model works well and orders the patient logically. In order to do that, create the data of three patients with varying health statuses, and show that your model orders them properly.

To summarize, for the first part of the Assignment you must:

- Implement a fuzzy inference system (FIS) to support the clinical decisions using the Simpyful python library (2 point);
- Define and implement the fuzzy sets and show them in graphs (2 points);
- Justify the choice of your parameters (2 points);
- Define the fuzzy rules with proper operator usage and explain their rationale (2 points);
- Show that the model works and does what it needs to do with some examples (2 points).

## Data-driven modeling (10 points)

An investment company would like to sell part of their real estate property. They would like to know what characteristics determine the house prices, and what would be a good price for the houses they wish to sell. Therefore, the management asks you to develop a model that predicts price based on historical data.

The historical data (which you can find in the file *previous\_house\_sales.csv*) contains data on 496 previously done transactions, of which the following variables are recorded::

1. **crim**: per capita crime rate by town.
2. **zn**: proportion of residential land zoned for lots over 25,000 sq.ft.
3. **indus**: proportion of non-retail business acres per town.
4. **nox**: nitrogen oxides concentration (parts per 10 million).
5. **rm**: average number of rooms per dwelling.
6. **age**: proportion of owner-occupied units built prior to 1940.
7. **dis**: weighted mean of distances to five Boston employment centres.
8. **rad**: index of accessibility to radial highways.
9. **tax**: full-value property-tax rate per \$10,000.
10. **prratio**: pupil-teacher ratio by town.
11. **lstat**: lower status of the population (percent).
12. **price**: median value of the house in \$1000s.

There are 10 houses that are meant to be sold by the investment company. Their characteristics are contained in the dataset *houses\_to\_be\_sold.csv*. What should their prices be, based on your fuzzy model?

For the second part of the Assignment you must:

- Use the data set *previous\_house\_sales.csv* to build a predictive Takagi-Sugeno FIS able to predict the price of houses given the other variables (3 points). Provide argumentation for your model's setting (for example, first - or zero order model etc.);
- Investigate the impact to performances (rely on pyFUME's cross-validation functionality) using 2, 3 and 4 clusters (3 points);
- Provide an interpretation of the rules (2 points);
- Predict the prices of the houses in the *houses\_to\_be\_sold.csv* dataset<sup>1</sup> (1 point).

---

<sup>1</sup> The group that comes closest to the actual values wins some bars of Tony's chocolate, to be picked up at TU/e's Atlas building.

# Final information

You have to produce and submit on Canvas **four** files:

- 1) the **python source code for the first part** of the assignment (knowledge-driven modeling with Simpful);
- 2) the **python source code for the second part** of the assignment (data-driven modeling with pyFUME);
- 3) the **executable file containing the Simpful model/code** (as produced by pyFUME) for the second part of the assignment;
- 4) a **report (in .pdf-format)** containing all the information and comments as required by the assignments above.

The page limit (mandatory) is **6 pages**. Try to aggregate multiple results in the same figure, where possible.

The source code can be **either a python script or a jupyter notebook**. A SINGLE python file for the first part, a TWO python files for the second part (one containing the pyFUME code, one containing the Simpful model).