Lab 11 (Project 1.2) – Classification using Fuzzy Systems and NN
(Week 5.2-6)

1 – Introduction

In the second part of the 1st project, you are going to use the same dataset (Lab6Dataset.csv), and you will also try to implement a classifier to detect if the room capacity as been exceeded. However, this time you will be using a Fuzzy System. After all the work you have done to preprocess the data, you should be an expert on the problem by now. I.e., you probably know which features are the most important to detect if there are more than 2 persons in the room, so the idea is to find a small set of features (or of new features generated from the original ones) from which you can build a Fuzzy Rule based system that performs satisfactorily in that task.

####compare it against a NN that uses the same features as the FS.

Let us remember the problem you must solve:

During the COVID-19 pandemic, a restriction on the maximum amount of people that could be simultaneously inside a room, was imposed by Técnico Lisboa. This capacity depended on several factors, including the room dimension, ventilation, etc. The need to automatically detect the number of persons inside a lab without affecting privacy, led to the implementation of an experimental lab based on low-cost, non-intrusive sensors.

The lab consists of a 13m^2 room where a Zigbee based wireless sensor network was installed. The lab has three workstations (a chair, and a desk with a dock station and a table lamp). There is a small window above workstation 3 and there is no heating/ventilation/AC system active in the room.

The wireless network is a Zigbee-based star network with six slave nodes feeding data to the master node. There is one CO2 sensor (MH-Z14A) in the center of the room, two digital infrared motion sensors (PIR) in opposed walls, and, in each workstation, a node containing a light sensor (BH1750) and a temperature sensor (LMT84LP) has been installed.

PIR sensor data indicates if movement was detected during the last 30s. For the remaining sensor nodes, the Arduino Uno microcontroller board sampled data from the sensors and transmitted it periodically via a Zigbee module every 30s.

Sensor measurements were taken over a period of several days. Each student manually annotated when entered and left the room during this period. Therefore, true occupancy was annotated during the measurement period.

The resulting dataset, Lab6Dataset.csv, is now available.

2 – Objectives

During the worse times of the pandemic, Técnico imposed a limit of 2 persons inside the above lab. However, the students that use the lab had frequent deadlines, and often ignored the 2-person limit. The objectives of this project are:

a) Develop a Fuzzy Rule based classifier that can detect when there are more than 2 persons inside the lab. To obtain a "manageable" Fuzzy System, you will need to somehow reduce the number of used features (see section 4 for more insights regarding this).
b) Develop a NN-based classifier that uses the same features of the Fuzzy Rule based system and see how it compares with the Fuzzy system.

2.1 – Submission details and Deadline

This Lab will be evaluated as the second part of project 1, and accounts for 20% of the Lab final grade. The final code and a comprehensive report must be submitted via Fenix until **Monday, June 20th, at 23:59**.

The students must submit:

a) All the developed code used to train and create the models.
b) A piece of code that will allow me to test your Fuzzy model using unseen data. This code, called TestMe, accepts as a parameter the name of a .csv file that has the same structure of Lab6Dataset.csv. The code must test the Fuzzy System you created on this new data. The output of the code must be the confusion matrix of the binary problem (is capacity exceeded or not), and the Precision, Recall and F1 of the "3 or more persons in the room" class.
c) A report where you indicate the options you made regarding the data preparation, the experimental setup, the construction and architecture of the models, the evaluation and validation process, and the results you obtained. Remember to take note of all decisions you make while checking and preparing the data and all new features that you used to build the system.

3 – Dataset

The "Lab6dataset.csv" file is composed of 10129 records, taken between and 11/01/2021 at 10:53, and 16/01/2021 at 9:04, approximately every 30s (some data points are missing). Each record contains the date/time, the data collected from the 9 sensors and the number of persons in the room.

Note that the dataset has not been preprocessed – it might contain noise, outliers, missing inputs, redundant features, etc.

The fields of the data set are:

- Date
- Time
- Si_Temp (°C, float)
- S_i_Light (Lux, int))
- CO_2 (PPM, int)
- PIR_i (Boolean)
- Number of persons in the room (0-3, int)

4 – Implementation, Evaluation and Validation

You can obviously start with the Preprocessed dataset you created in the 1st part of the project. However, if you removed more than a handful (or two) of datapoints from the original dataset during that phase, you should consider using a less aggressive outlier detection method.

To implement a manageable Fuzzy System that doesn't suffer from combinatorial rule explosion (and lack of interpretability), you will need to reduce the number of used features, to somehow generate new features that combine existing ones, or to use an hierarchical system that allows you to combine features in small groups (see the slides about fuzzy system complexity).

By now you probably have realized that some features are not as important for the problem, so you might be tempted to simply remove them. However, it might probably be more interesting to manipulate them into generating new features that might be relevant. For example, CO_2 level seems to be kind of useless since there is a delay until it accumulates into a level that indicates an excess number of people in the room (and it also takes some time to dissipate). So why not create a new feature that indicates how fast CO_2 is increasing or decreasing to estimate if the room capacity has been exceeded?

Another idea is to somehow include the time of day in your system. For example, in the evening, the total amount of light in the room can be a good indicator of how many people are in the room. The same can be said if the day is cloudy, and the amount of light is low even during the day.

To implement such ideas, you can manipulate the dataset (automatically) or include rules that define the "context" (Is it night, is it cloudy, is the CO2 increasing much, etc.) where you are trying to decide if there are too many people in the room. These rules can be fuzzy, but do not necessarily have to be. For example, probably you don't need Fuzzy to properly use the PIR information.

As an expert, it is up to you to analyze the problem, look at the data, see which features might or might not be helpful, check if new features might be useful, decide if the order and/or the date/time is relevant, etc. Be **creative** and show your Fuzziness. Remember that, since you have lots of annotated data, this problem is not the most adequate to be solved using a Fuzzy system. However, this will show how well you have assimilated the Fuzziness concepts you have learned during the classes.

Once again, try to avoid OVERFITTING. Even though there is no Training set when you are building a Fuzzy System, you should maintain a separate Test set that you will only apply once all the Fuzzy system has been completed. Use a Validation set to check your system's response while you are creating the Fuzzy System.