DEIS - Departamento de Engenharia Informática e Sistemas
ISEC - Instituto Superior de Engenharia de Coimbra

KNOWLEDGE AND REASONING 2019/2020
Practical Assignment

To carry out the Practical Work, 4 topics are proposed. Below is the detailed description of each one and in Moodle you can find the data sets for Theme nº 1.
Each group has to select ONE theme to implement and study.

The choice of the topic to implement must be made using the referendum in Moodle. Only **one** of the students in each working group must select the desired topic.

The working groups must be formed by a maximum of 3 students.

The unique date of delivery of the work is **until 23.59 of the 28 of June of 2020**

The following elements must be submitted:

- Moodle: code and all necessary files for the execution and testing of the work; pdf of the report;
- Defenses will be combined later, remotely or in person, according to the evolution of the current situation.
- The defense of work is mandatory.

TOPIC 1 – NEURAL NETWORKS

In this theme, it is intended that students deepen their knowledge of neural networks. The objective is to implement a neuronal network capable of correctly classifying alphabetic characters - the five vowels (A, E, I, O, U).

In Moodle, black and white image files are provided, separated by three different folders that should be used in the tasks described below.

For this work, the following approach is suggested:

a) [20%]. Using Matlab's image manipulation functions convert the provided images into binary arrays. If you find it necessary, make a previous treatment to the images, such as resizing, or any other that you find relevant.
Start with a layered neuronal network of 10 neurons. Use the network to train character recognition in **Folder_1**. This folder contains one image of each vowel. Use all the examples in training. Test other topologies, activation and training functions, record and compare the results obtained.

b) [20%]. Using the base model implemented in paragraph a) make the necessary changes to implement and test various topologies and parameterizations of NN in order to obtain a good performance for the classification of the vowels provided in the folder **Folder_2**.

- Start by using a 70%, 15%, 15% dataset segmentation for training, validation and testing.
- Observe the matrix of confusion, training and test errors.
- Explore and compare various network configurations.
- Test different training / activation functions, different segmentations in the division of the examples. It is suggested to adapt the Excel file given in the practical classes, to record results and obtain the conclusions.
- Record the neuronal network (s) with the best performance (s).

c) [25%]. Now use the images from **Folder_3** that were not used in the previous training. Without training the network, check if the classification given by the NN is correct. Present the results obtained. Posteriorly:

- Retrain the network only with the examples in **Folder_3**. Test the network separately for the images from **Folder_1**, **Folder_2** and **Folder_3**. Compare and record the results obtained in each case.
- Re-train the network with all the images provided (**Folder_1** + **Folder_2** + **Folder_3**). Test the network for the **Folder_1**, **Folder_2** and **Folder_3** images separately. Compare and record the results obtained.

d) [15%]. Manually draw some vowels that are similar to the examples used in training the network. Transcribe the drawings to binary matrices. Develop a small program to read a file corresponding to one of these images and apply it to the best neural network obtained in c). What are the results?

e) [20%]. Develop a graphical application in Matlab that allows the user to do the tasks previously developed in an easy and intuitive way:
• Configure the topology of the neural network
• Choose training / activation functions
- Train the neural network
- Record a previously trained neuronal network
- Load a previously trained neuronal network and apply it to a dataset
- Draw a new letter, or load an image file where it is already drawn. Apply a neural network to classify the drawn letter
- View the classification results
- Generation / recording of result files

Make a report of the work done. Poor report quality can discount up to 50% of the total score obtained in the previous points.

TOPIC 2 – EXPERT SYSTEMS

The goal of this work is to develop an expert system for the selection of patients to be included in a medical trial related with oesophageal cancer.

- Inclusion Criteria:
  - Age ≥ 18 years;
  - World Health Organization (WHO) performance status score of 0 or 1;
  - Diagnosis of squamous cell carcinoma, undifferentiated carcinoma or adenocarcinoma of oesophageal or oesophagus-gastric junction (Siewert I e II);
  - Clinical Stage II or III disease, according to American Joint Committee on Cancer Staging classification, 7th edition, resectable disease;
  - Upper border of the tumour at least 3 cm below the upper oesophageal sphincter;
  - Signed informed consent.
- Exclusion Criteria:
  - Pregnant or lactating women;
  - Previous thoracic radiotherapy;
  - Impaired haematological, hepatic, renal or pulmonary function defined:
    - Neutrophils count < $1.5x10^9$/L,
    - Platelet count < $100x10^9$/L,
    - Serum concentration of total bilirubin > 1.5 x ULN (ULN = upper limit of normal),
    - Creatinine > 120 mcmol/L,
    - FEV1< 1,5 L;
  - Active infection or other medical condition that prevents the patient from receiving the planed treatment.

Selected patients must meet all inclusion criteria simultaneously and cannot be included in the study if they have one or more exclusion criteria. **The expert system must be able to identify patients eligible for the study and make that information available to the user.**

Selected patients will have to:
- perform a bronchoscopy, in the screening phase, if the tumour is related to the respiratory tree
- perform an endoscopy 6 weeks after starting treatment, if endoscopic lesions are detected

In any of the above extra exams is necessary, the user should be alerted.

If the stomach is included in the treatment volume, patients will have to fast for 6 hours before treatment. **This information should be made available to the user (healthcare professional) for him or her to inform the patient.**

The following steps should be considered:

a) [30%] Implement the rules that allow the system to identify patients eligible for treatment

b) [30%] Implement the rules that allow the system to alert for the need of extra exams and the need for fasting

c) [25%] Prepare a set of tests that will allow you to thoroughly test the functioning of the developed expert system. Make sure the tests cover all possible scenarios.

d) [15%] Complement the developed system with a graphical interface that allows the insertion of the necessary data and the presentation of the generated results.

Prepare a report of the performed work that, in addition to the technical details of the implementation and the analysis made to the tests carried out with the system, must include an inference tree or a diagram that allows a simple reading and understanding of the rules that the system implements. A poor-quality report can lead to discounts of up to 50% on the total score obtained in the previous points.

Note: Remember that you must use an expert system where the rules should not have "if" instructions.

TOPIC 3 – FUZZY INFERENCE SYSTEM (FIS)

1 - Introduction

In the touristic activity, due to the hotels changing occupation rates along the year, many new employees must be hired and many others must be fired or simply change their working place. This activity is carried out by the Human Resources Management (HRM). The goal of this practical work is to implement a fuzzy system for the evaluation of the HRM section based on the Mamdani Inference (FIS). With this system one tries to guarantee that the HRM work is more realistically evaluated, not based on simple Yes and No parameters used in the classical evaluation methodologies.

2 - Assignment

Using Matlab, create a model based on the Mamdani Inference for evaluating the HRM of a hotel company. The employees are evaluated according to the following 4 perspectives:

1. Hiring – about some of the characteristics of the hiring activity

2. Payment – about the costs of the hired people
3. Education – the costs associated with professional training
4. Retention – evaluation about how hired people stays or leaves the hotel, in a near future

The following table shows the input variables and domains for each one of these perspectives. As an orientation, the "Target" column, on the right, shows if an high evaluation corresponds to a high or low value of the input variable:

Activity	Input Variable	Domain	Target
Hiring	Hiring costs / Number of hired people	5 to 75 euros	Low
	Time average for a temporary job contract	3 to 120 days	Low
	Time average for a permanent job contract	3 to 90 days	Low
Payment	Personal costs / Total costs	0 to 45%	Low
	Average worker salary (without management costs)	400 to 1300 euros	Low
	Average worker salary (with management costs)	700 to 1800 euros	Low
	Bonus % over the base salary	0 to 25%	High
Education	Training costs (%) over the base salary	0 to 5%	Medium
	Number of training hours by worker, by year	0 to 45h	Medium
	Number of training hours by 1sr level worker, by year	0 to 100h	High
Retention	% of 1st level workers that leaved the hotel by their own wish	0 to 5%	Low
	% of workers that leaved the hotel by their own wish	0 to 15%	Low

a) [10%]. Define the linguistic terms for the 12 input variables. Choose the number of terms and the membership function type(s).

b) [20%]. Define the inference rules needed for this system:

   a. **Rule Block 1:** the output variables are Hiring, Payment, Education and Retention and all will have 3 linguistic terms only: low, medium, high, all defined in the domain [0,1].
   b. **Rule Block 2:** once the block 1 rules are defined, create a new block of rules that must be chained with the first ones. For this block the inputs are Hiring, Payment, Education and Retention and the output will be the classification of the HRM as Very Bad, Bad, Medium, High, Very High defined in the domain [0,1].

c) [50%]. Implement the system using Matlab

d) [20%]. Test the system and carefully look at the results.

3 – Final Report

Write a final report, clear and synthetic, about this assignment. A bad report quality may imply a **maximum of -50%** in the final assignment evaluation.

TOPIC 4 – CASE-BASED REASONING (CBR)

1 - Introduction

Please read the Introduction for TOPIC 3 – Fuzzy Logic.

2 - Assignment

In Matlab, create a model based on the CBR paradigm for the same task as TOPIC 3, i.e. evaluation of the HRM. The system must include the 4 phases *Retrieve*, *Reuse*, *Revise* and *Retain*:

a) [10%] Create a Case Library whose attributes will be the 12 input variables of the above table and the case solution will be the evaluation of each one, in [0, 1,] (from Bad to Very Good). Choose the number and type of cases. However, it must be enough for a good system startup.

b) In Matlab implement the CBR cycle:

   i) [15%]. *Retrieve*: use Linear Distance with weights > 1 for the attributes that you think are more important. The solution must show the more similar cases and not just the most similar.

   ii) [20%]. *Reuse*: this phase must implement the final evaluation adaptation if the new case is "distinct enough" from the previous cases. To implement this adaptation, use rules, interpolations or any other adequate approach.

   iii) [15%]. *Revise*: this phase must allow choosing a solution or confirm any purposed one, or even the manual introduction of a value, if needed.

   iv) [20%]. *Retain*: must retain the successfully adapted cases or new cases.

c) [20%]. Test the system and carefully look at the results.

3 – Final Report

Write a final report, clear and synthetic, about this assignment. A bad report quality may imply a **maximum of -50%** in the final assignment evaluation.