11753 Computational Intelligence
Master in Intelligent Systems
Universitat de les Illes Balears

Handout # 2: Self-Organizing Maps (SOM)

NOTE 1: Problem P2 requires *training* and *test datasets*. They are, respectively, stored in `dsxx1tr.txt` and `dsxx1te.txt` files:

```
import numpy as np
group = '01' # assuming group 1
ds = 2       # assuming problem 1
data = np.loadtxt('ds'+group+str(ds)+'tr.txt')
X_train = data[:, 0:2]
y_train = data[:, 2]
data = np.loadtxt('ds'+group+str(ds)+'te.txt')
X_test = data[:, 0:2]
y_test = data[:, 2]
```

Class labels are 1 for ω_1, 2 for ω_2, 3 for ω_3, etc.

NOTE 2: Problem P2 also requires the use of the libraries `minisom` (`https://github.com/JustGlowing/minisom`), `scikit-learn` (`https://scikit-learn.org`) and `matplotlib` (`https://matplotlib.org/`).

P2. **Given datasets `dsxx2tr.txt` and `dsxx2te.txt`**, you have to find suitable $N \times M$ SOMs for a number of cases. For each, you have to define and train the network, use it to classify the data and check the resulting SOM performance.

Consider the two following cases:

a) using the training data as they are, without normalization, and

b) normalizing the training data to ensure zero mean and unit variance. (Consider the pre-processing functions of `https://scikit-learn.org/stable/modules/preprocessing.html`, in particular the **StandardScaler**.)

For each case:

1) Define the size of the SOM according to the recommendations.

2) Train the SOM using the on-line version for 5 attempts keeping the SOM with the lowest quantization and topographic errors (if the lowest ones do not match for the same training, give priority to the quantization error).

3) Train the SOM using the batch algorithm and provide the quantization and topographic errors.

4) Provide the following performance data for each SOM:

   i) The U-matrix and the class representation map.

   ii) The confusion matrix and the classification accuracy for the test set using the SOM as a classifier.

- A report of the work done has to be released by March 29, 2022 in electronic form as a notebook file (.ipynb).
- Provide the requested data and plots/figures at each point above. For figures, use appropriate titles, axis labels and legends to clarify the results reported.
- Suitable comments are expected in the source code.
- This work has to be done individually (see the number of group in *Aula Digital*).