Lab of Applied Computational Intelligence (CInte)

IST

2022/2023

Experimental Setup and Classification using Neural Networks

Guide 5

October, 4th, 2022

(Week 3.1)

1 – Objectives

In this work you will use Scikit-learn (https://scikit-learn.org/) to implement a classification task using NN and gain insights on the configuration of MLP hyperparameters and the process of experimental setup.

In addition to the previous lab (and before this lab class), I suggest that you follow the DataQuest Scikitlearn tutorial, by Satyabrata Pal:

https://www.dataquest.io/blog/sci-kit-learn-tutorial/

2 – Data Libraries

For this work we are going to use the datasets from the previous lab, "iris" and "Haberman". The first has information about 3 types of iris flowers and the second about survivability of cancer. You will see that the file does not have the name of the columns, so you will have to add manually those names. Read the datafiles with pandas.

3 – Data Preprocessing and Experimental Setup

Use the previously acquired knowledge to preprocess the data as you deem necessary, and create your **Train**, **Validation** and **Test** sets (check the lecture slides!). Take advantage of what Scikit-learn offers you, but don't use it blindly. Don't forget that you should only use your Test set after you are satisfied with all NN parametrization and hyperparametrization. Consider if using cross-validation is necessary.

4 – Classification task

Use Scikit-learn to implement NN MLP classifiers that predict the datasets outputs (**from sklearn.neural_network import** MLPClassifier)

You will need to configure and optimize your NNs in order to obtain good results. Do your best... but try to avoid overfitting.

5 – Evaluation and Validation

Use the previously acquired knowledge to properly evaluate and validate the performance of your classifiers.