

Subject

Advanced Machine Learning

Self-directed Activity

Unit 1

Introduction on Machine Learning and

Deep Learning

Session 1

Professor: Alejandro Esteban Martínez

1 Objectives of the activity

Be able to analyze the classic Iris Dataset and the Breast Cancer Wisconsin Dataset with different plotting tools and compare the classification capabilities with conventional Machine Learning models and Deep Learning approaches.

2 Related Learning Results

You should be able to show me the combination of the knowledge you have gained during your studies and put them critically to test.

3 Activity Description

For this initial assignment I want you to be able to create a python script that plots the information inside the Iris Dataset and latter the Breast Cancer Wisconsin Dataset and to use different models to solve a classification problem of predicting the kind of iris flower from their attributes.

For the plotting initial part I want you to explore libraries that will be very important in your future as **pandas** to read the document containing the Dataset and **Matplotlib** and **Seaborn** to visually try to see how a human would classify the plants without any machine help. You are welcome to use any other library and tool for this process as long as you are able to explain what you are doing.

Afterwards I want you to explore different classification models from the scikit-learn library (one of the most used for conventional Machine Learning approaches). I want you to be able to use the models you have already learned in class like the **SVM**, **Decision Trees**, **Logistic Regression**... And then I want you to use the **Multilayer Perceptron (MLP)** from that same library. It is a very basic model and the scikit-learn implementation doesn't allow as much configuration as Keras or Pytorch. I want you to try with different numbers of Artificial Neurons in each layer and different number of layers.

Later I want you to compare the metrics of the different models used and their configurations and discuss why do you think some models performed better than others.

You should attempt to use the knowledge you've learned in the past subjects, so avoid unnecessary code repetition, comment accordingly throughout the entire script and write a small document explaining your findings and any problem you had during the development.



Final note, I would strongly recommend not using tools like ChatGPT or Github Copilot for this assignment. I'm not against the use of AI in my class but for this initial steps I want to be able to analyze your real level as a Computer Scientist and to know what I should maybe explain better during class or what information might be superfluous for you.

You can (and should!) search around on Internet how to do the assignment and read some information about the tools and libraries you will be using.

4 Grading

This assignment grade won't be part of the final subject grade, but I will give you some notes about your work that will help you improve.

Feel free to contact me over email, in class or during my tutoring hours for any questions, errors or reflections you may have.