CEGE0004: Assignment (50%)

Module Coordinator: Dr. Aldo Lipani

1   Introduction

In this assignment you need to pick one or two datasets from one of those available at the UCI Repository[1] and perform supervised learning tasks (classification or regression) on it.

The report task is mandatory and counts a total of 20 marks. The remaining marks are equally split across 5 learning tasks. For each learning task you need to pick a learning algorithm from a family of algorithms: decision trees, instance-based learning, bayesian learning, neural networks, and model ensembles.

Note that, the learning algorithms that you can pick do not need to be among those that have been presented in the module. You can pick the ones that you like, and; whenever possible you should use as a target feature the one identified by the dataset. If this is not possible you can use another feature from those available in the dataset.

2   Instructions

To complete this assignment, you need to use Python. You are left free to use whatever technology you find useful. However, I would recommend you to make a Jupiter notebook for each task. This because notebooks will allow you to code, document your code, and present results all at the same time. No matter if you decide to follow this recommendation, please do not forget to state how to run your code.

3   Report (20)

You will be supplied with a report template containing further instructions and indications of what to write. Please follow these carefully. The report will be evaluated based on its general quality, task description, data analysis, and final comparison of the trained models on the test sets.

---

[1] https://archive.ics.uci.edu/ml/index.php

4   Learning Tasks $(5 \cdot 16)$

Each task will be evaluate based on the quality of its presentation, implementation, training, validation, and hyper-parameter tuning.

Note that you will not be evaluated based on the performance of the trained models; this assignment aims to evaluate your work and not the learning algorithms. The performance of these learned models is left to you to discuss in the report.

5   Submission

This assignment should be submitted as follows:

1. a zip file containing the project solution to the Assessment tab of the module Moodle page;

2. a pdf file of the project report to the Assessment tab of the module Moodle page;

3. as a GitHub repository by inviting me as a collaborator (aldolipani).

Failing to carefully follow these instructions may result in penalties.

6   Marking Scheme

The mark scheme is distributed in tasks as follows (total of 100):

n.	Task Description	Marks
1	Report	20
2	Decision Trees	16
3	Instance-based Learning	16
4	Neural Networks	16
5	Bayesian Learning	16
6	Model Ensembles	16

Any of the following is not allowed, their violation can result in penalties:

- Screenshots of your code pasted in your report. Please copy and paste your code as text;

- Copying code from other sources without referencing;

- Plagiarising is severely punished. Please read the assessment tab in Moodle for more details about it;

- Copying your classmate report or code is also considered plagiarism, and in this case all students involved are punished equally.