Assessment

Machine Learning and Statistics, Winter 2023/24
ian.mcloughlin@atu.ie

Last updated: 6 December 2023

These are the instructions for the assessment of Machine Learning and Statistics in Winter 2023/24. They cover the full 100% of the module marks. The deadline for submission of all components within the repository is **Sunday 7 January 2024**.

Purpose

The main purpose of this assessment is to ensure students can demonstrate they can do the following.

1. Describe the stochastic nature of real-world measurements.

2. Select an appropriate mathematical model of a real-world problem.

3. Select an appropriate cost function for a given machine learning task.

4. Apply an optimization technique to the parameters of a model.

The assessment is then used to determine a mark out of 100% based on the details in this document.

Instructions

- There are three components to this assessment.

  - Tasks throughout the semester (40%).
  - A project (40%).
  - A repository containing all above work (20%).

- All work must be in the `main` branch of one GitHub repository[1].

- Use the form on the module page to submit your repository[2].

- Commits in GitHub on or before the deadline will be considered[3].

- You start immediately, spreading the work throughout the semester. This will ensure you make regular commits to your repository.

- Your repository should contain the following.

  - A `README.md` giving an executive summary of the purpose and contents of the repository, and instructions for a competent person to clone and run any code or notebooks in it.

  - A `.gitignore` file to ignore any temporary files and folders that should not normally be committed to a repository.

[1] If you have any issues with GitHub branches please ask for help long before the deadline. Note that GitHub renamed their default `master` branch to `main` some time ago. This should not cause issues if you create your repository as described in lectures.

[2] You should set up the repository and submit the URL immediately.

[3] Once you have submitted your URL, you do not need to do anything other than commit to your repository and push the changes to GitHub.

- A notebook[4] called `tasks.ipynb` containing work as detailed in the Tasks section below.

- A notebook called `project.ipynb` containing work as detailed in the Project section below.

- Any supporting data, image, or other files that form part of your submission. These should be neatly arranged, using subfolders as appropriate.

- Once completed, your repository should be readily presentable in job interviews. A technically competent person should be able to understand what is in your repository and how to interact with it without any extra information. This criterion will be decisive in determining your repository mark (20%).

Tasks

- Each topic covered in lectures will have tasks associated with it.

- The tasks will be listed in an accompanying document.

- All tasks for all topics should be completed in a single notebook called `tasks.ipynb`.

- There will be five topics, about one every two weeks throughout the semester.

- The first topic will cover getting set up with the technology used in the module.

- The tasks should be largely completed and committed to your repository as they are covered or soon after.

- Any changes to the tasks during the semester will be flagged by the lecturer. Please double-check the tasks list near the deadline to ensure you have completed all of them.

Project

- The project is to create a notebook exploring classification algorithms applied on the iris flower data set associated with Ronald A Fisher.[5]

- In your notebook, you should first explain what supervised learning is and then explain what classification algorithms are.

- Describe at least one common classification algorithm and implement it using the `scikit-learn` Python library.

- Throughout your notebook, use appropriate plots, mathematical notation, and diagrams to explain the relevant concepts.

[4] All uses of the term notebook in this document refer to Jupyter notebooks, as used in lectures.

[5] *Iris – UCI Machine Learning Repository*. Aug. 17, 2023. URL: https://archive.ics.uci.edu/dataset/53/iris (visited on 08/17/2023).

Marking Scheme

Each of the three components of your submission will be marked using the four categories below. To receive a good mark in a category, your submission needs to provide evidence of meeting each of the criteria listed under it[6].

Research (25%)*:* evidence of research on topics; appropriate referencing; building on work of others; comparison to similar work.

Development (25%)*:* clear, concise, and correct code; appropriate tests; demonstrable knowledge of different approaches and algorithms; clean architecture.

Documentation (25%)*:* clear explanations of concepts in notebooks; concise comments in code and elsewhere; appropriate, standard README for a GitHub repository.

Consistency (25%)*:* tens of commits, each representing a reasonable amount of work; literature, documentation, and code evidencing work on the assessment; evidence of reviewing and refactoring.

Policies

- Please remember that you are bound by ATU policies and regulations.[7] You should familiarize yourself with these on the Student Hub.[8]

- Pay particular attention to the Policy on Plagiarism and the Student Code of Conduct.

- If you have any doubts about what is permissible, email me to ask[9].

Advice

- Students sometimes struggle with the freedom given in an open-style assessment.

- You must decide where and how to start, what is relevant content for your submission, how much is enough, and how to make the submission your own.

- This is by design — we assume you have a reasonable knowledge of programming and an ability to source your own information.

- Companies tell us they want graduates who can (within reason) take initiative, work independently, source information, and make design decisions without needing to ask for help.

- You need a plan, you cannot just start coding straight away.