

Project 2018

Emerging Technologies

Due: last commit on or before November 30th

This document contains the instructions for Project 2018 for Emerging Technologies. The project involves writing documentation, code, and comments in the programming language Python [2] and using the Jupyter notebook [5] software.

You must use the version control software git [1] to track your work and you will submit your project by providing a URL to your git repository. It is suggested you use GitHub [3] for this purpose and that you consider making your repository publicly available so that prospective employers may view it. However, should you wish to, you may restrict general public access to your repository so long as you give permission to the lecturer to view it. Furthermore, any git repository URL to which you provide access to the lecturer will suffice – you don't have to use GitHub.

Please be advised that all students are bound by the Quality Assurance Framework [4] at GMIT which includes the Code of Student Conduct and the Policy on Plagiarism. The onus is on the student to ensure they do not, even inadvertently, break the rules. A clean and comprehensive git history is the best way to demonstrate to the examiner that your submission is your own work. It is, however, expected that you draw on works that are not your own to build your submission and you should systematically reference those works to enhance your submission.

Content

Your submission will have five main components. Each component is either a Jupyter notebook or a Python script.

1. **numpy random notebook:** a jupyter notebook explaining the concepts behind and the use of the numpy random package, including plots of the various distributions.

2. **Iris dataset notebook:** a jupyter notebook explaining the famous iris data set including the difficulty in writing an algorithm to separate the three classes of iris based on the variables in the dataset.
3. **MNIST dataset notebook:** a jupyter notebook explaining how to read the MNIST dataset efficiently into memory in Python.
4. **Digit recognition script:** a Python script that takes an image file containing a handwritten digit and identifies the digit using a supervised learning algorithm and the MNIST dataset.
5. **Digit recognition notebook:** a jupyter notebook explaining how the above Python script works and discussing its performance.

Each file should be easily identified by anyone visiting your repository, and your README should clearly explain how to run them. I would suggest naming the five files `numpy-random.ipynb`, `iris-dataset.ipynb`, `mnist-dataset.ipynb`, `digitrec.py`, and `digit-recognition.ipynb` respectively.

Minimum Viable Project

The minimum standard for this project is a git repository containing the four notebooks and one Python script listed above, along with a README, LICENSE and gitignore. Your README should clearly document how to run the script and notebooks. Your repository should not contain any unnecessary files, such as backup or compiled files.

A better project will be well organised and contain detailed explanations. The quality of the submission will be immediately evident in the ease by which the concepts, code and setup are understood. The architecture of each notebook and script will be well conceived.

Please note that at this level there is no one right answer. Rather, it is up to you to take the assignment, make it your own by constructing your own narrative around it and then to justify that narrative through evidence.

Submission

The version control software git must be used to manage the development of the software and you must make your repository available to the lecturer by URL. Your git repository will form the main submission of the project. You must submit the URL of your git repository using the link on the course

Moodle page before the deadline. You can do this at any time, as the last commit before the deadline will be used as your submission for this project.

Any submission that does not have a full and incremental git history with informative commit messages over the course of the project timeline will be accorded a proportionate mark. It is expected that your repository will have at least tens of commits, with each commit relating to a reasonably small unit of work. In the last week of term, or at any other time, you may be asked by the lecturer to explain the contents of your git repository. While it is encouraged that students will engage in peer learning, any unreferenced documentation and software that is contained in your submission must have been written by you. You can show this by having a long incremental commit history and by being able to explain your code.

Marking scheme

This assignment is worth one-hundred percent of your marks for this module. It is broken into three distinct parts: forty percent for the first three Jupyter notebooks (on numpy, iris, and MNIST), forty percent for the Python script and associated Jupyter notebook on recognising digits, and finally twenty percent for your presentation of these ideas. However, the two forty percent sections will be marked using the same marking scheme and the mark for each will be the same unless there is justification for a difference. The following marking scheme will be used. Students should note that in certain circumstances the examiner's overall impression of the project may influence marks in each individual component and the mark adjusted accordingly.

Category	%	Description
Research	20%	Investigation of each notebook and script as demonstrated by references, background information, and approach.
Development	20%	Clear, well-written, and efficient code with appropriate comments.
Consistency	20%	Good planning and pragmatic attitude to work as evidenced by commit history.
Documentation	20%	Concise descriptions and explanations of theoretical and practical aspects of problems.
Presentation	20%	Natural and intuitive presentation of concepts and code in repository and through in-person discussion.

Advice for students

- Your git log history should be extensive. A reasonable unit of work for a single commit is a small function, or a handful of comments, or a small change that fixes a bug. If you are well organised you will find it easier to determine the size of a reasonable commit, and it will show in your git history.
- Using information, code and data from outside sources is sometimes acceptable – so long as it is licensed to permit this, you clearly reference the source, and the overall project is substantially your own work. Using a source that does not meet these three conditions could jeopardise your mark.
- You must be able to explain your submission while constructing it, and afterwards. Bear this in mind when you are writing your README. If you had trouble understanding something in the first place, you will likely have trouble explaining it a couple of weeks later. Write a short explanation of it in your submission, so that you can jog your memory later.
- Everyone is susceptible to procrastination and disorganisation. You are expected to be aware of this and take reasonable measures to avoid them. The best way to do this is to draw up an initial straight-forward project plan and keep it updated. You can show the examiner that you have done this in several ways. The easiest is to summarise the project plan in your README. Another way is to use a to-do list like GitHub Issues.
- Students have problems with projects from time to time. Some of these are unavoidable, such as external factors relating to family issues or illness. In such cases allowances can sometimes be made. Other problems are preventable, such as missing the submission deadline because you are having internet connectivity issues five minutes before it. Students should be able to show that up until an issue arose they had completed a reasonable and proportionate amount of work and took reasonable steps to avoid preventable issues.
- Go easy on yourself – this is one project in one module. It will not define you or your life. A higher overall course mark should not be determined by a single project, but rather your performance in all your work in all your modules. Here, you are just trying to demonstrate to yourself, to the examiners, and to prospective future employers, that

you can take a reasonably straight-forward problem and solve it within a few weeks.

References

- [1] Software Freedom Conservancy. Git.
<https://git-scm.com/>.
- [2] Python Software Foundation. Welcome to python.org.
<https://www.python.org/>.
- [3] Inc. GitHub. Github.
<https://github.com/>.
- [4] GMIT. Quality assurance framework.
<https://www.gmit.ie/general/quality-assurance-framework>.
- [5] Project Jupyter. Project jupyter home.
<http://jupyter.org/>.