CMEE Masters: Miniproject Assessment

Assignment Objectives: To address on a model-fitting problem using computational methods, and produce a written report, all in a coherent, reproducible, modular workflow under version control.

Computing

Student's Name: Zongyi Hu

Overall Project Organization

All the directories were in place, you had files from a previous run in Code and Data.

You had a `readme` file listing most of the key files. You gave a detailed description of the workflow, indicating which files do what – good. You should also have listed packages and dependencies (and what they were for), as well as versions (for languages and dependencies). Your instructions were explicit, good!

You could have put the writeup LaTeX source files and pdf in a separate directory, such as a writeup directory – this is what you should aim to do for your final dissertation.

Overall your organisation and documentation is OK, you are missing some key parts of the documentation, and you had lots of files left from a previous run of the project.

The Code

Your choice of coding tools and number of packages used was more or less appropriate for this project. You used a sensible minimum of packages for this project – good, relying on too many packages is not good for reproducibility, nor for the development of your coding/quantitative abilities.

Your code is clean, and generally well commented, if a little sparingly. Your Python makes good use of modules and docstrings – excellent, this is well rationalised and Pythonic! Your R could use further rationalisation – think about where you could implement more functions to make your R code more compact and compartmentalised in general. Your use of Python vs R was fairly logical. Your project ran a somewhat slowly at approximately 21 minutes and five seconds. Consider profiling in the future for complex projects such as this; your Gompertz_model function took quite a long time to run, and the use of NLSM functions also ate up quite a lot of time. Try catch statements are a good way to do what you needed to here, but they are also costly computationally, which is a problem exacerbated by using a relatively slow language such as R. All of these bottlenecks might be solved by looking at using Python as opposed to R for these parts of a project like this, but there may also be ways in R to solve them, either way profiling is a good idea, as it tells you where your problems are!

Running the project script gave no errors – this is an achievement, well done! However, the LaTeX report compilation had issues, but these passed silently! Be wary of silent errors, these are particularly dangerous. The error was likely a bibtex error as these are very common.

Also, remember, you should write into your workflow commands that will delete all existing output files every time the workflow is run (they should be re-generated afresh). Also, put in checks so that the computational workflow aborts if any step in the analysis gives an error.

You wrote messages to the terminal about the progress of the workflow, and you clearly delineated these using special characters – well done! This is exactly how progress messages should be displayed. as asterisks.

You had some minor issues with your documentation, and the silent error is an issue. However, your code is clean and sensible. Overall, good!

Marks for the project and computational workflow: 76

Report

The Report

The report was well organized. Reasonably easy to read, but some issues with grammar, which can be ironed out in the main report with some peer- and supervisor inputs.

The Intro/background was not very deep. The objective were not at all clearly defined and grounded.

You achieved a reasonable technical depth with your model fitting and selection analysis. But there was scope for doing more (see comments in attached pdf).

Methods were not very well organized, and did not clearly explain the choices made. On the other hand, they went too much in depth about relatively unimportant elements such as some specific packages.

OK discussion as far as the scope of the results is concerned, but failing to provide a sufficiently deep discussion, a clear overview of the results and an account of the limitations of the study.

In your main project report, to the extent possible, consider using a results-focused title for your main dissertation. Basically, the main finding(s) of the paper can be indicated in the title. Obviously, if you have a lot of findings, you can't do that. But then, it means you have too many 'storylines' in your study. Also, there, make sure your Introduction is more explicit about the motivation for you objectives and each of your and hypotheses, and provide clear answers to them at the end, in the Discussion.

The figures were not uniformly well formatted and presented. Have a look at this for further guidelines on figures and tables for your main dissertation writeup: `http://abacus.bates.edu/~ganderso/biology/resources/writing/HTW_Guide_Table-Figures_9-30-08.pdf`.

(More specific feedback is in the attached pdf, and we can also discuss more aspects of your write-up in our 1:1 feedback meeting)

Marks for the Report: 64

Signed: Samraat Pawar

March 18, 2021

Notes on Assessment :

- This written feedback will be discussed in a 1:1 session scheduled after this assessment has been been given to you.
- The coursework marking criteria (included in this feedback at bottom) were used for both the computing and report components of the Miniproject Assessment. *In contrast*, Your final dissertation project marks are going to be based pretty much exclusively on the written report and viva (not code). Expect your final dissertation report to be marked more stringently, using the dissertation marking criteria (also included in this report).
- If there were technical errors made in the model fitting and selection, the points have been deducted from the report (not the computing) component's marks.
- In many cases, the marker would have contrasted what you have done with what you should do in your actual dissertation. *This does not mean that you were penalized* — one of the main goals of the miniproject is to provide feedback useful for your main dissertation. However, there may be cases where what you have done is just really bad practise (for example missing line numbers or abstract), irrespective of whether it is a mini- or main-project report – you will be penalized in that case.
- The markers for this assessment are playing the role of somebody trying to understand and use your project organization and workflow from scratch. So it will seem like the feedback is particularly pedantic in places — please take it in the right spirit!
- Ultimately, keep in mind that this mini-project was partly an exercise in reproducible workflow development — you may need to trade-off some computational elegance (but hopefully not reproducibility!), such as having everything run with one run_project command, in favor of a good written report — that's what matters most in the end. In this context, the main thing to keep in mind is that one or both of your markers will likely not be particularly quantitative, so you will need keep the explanations simple (but not patronizingly so!). In general, this advice holds while writing papers for more general (not narrow-subject focused) journals as well — keep it succinct and simple. Therefore, please also consider the report component mark separately from the computing component mark.