02936 Bayesian Data Analysis 2021: Project work

Revised Oct 28, 2021

Project work
The second half of the Bayesian Data Analysis course is project work. The purpose of the project is to apply the techniques from the course on some data of your own choosing. For Ph.D. students, it would be obvious to analyze a dataset relevant to your own research. If you don't have a dataset at hand, you can find one online, e.g. on the UCI Machine Learning Repository or ask me for suggestions.

Groups
The projects should preferably be done in groups of 2-3 students, but you can also work on your own if your prefer. If you do not have a group, ask in the Teams channel. If you have a project idea or a dataset, but no group, announce it on Teams.

As soon as you have found a group and decided on a project topic, let me know by email (send one email per group). You should also state you when you would prefer to present the project (see details below). The deadline for forming groups and choosing topic is November 14th.

Project delivarables
The project delivarables is a short report and a short project presentation in class. It should be emphasized that the project is meant to be a small, so we do not expect you to do very advanced modelling work. The important thing is that you demonstrate that you understand the Bayesian workflow. However, we do, of course, expect more from a group of 3 compared to a group of 1.

The report
The report (one per group) should contain a brief description of the problem, the data, the models and the resulting data analysis. The introduction must be written in a way that allows a person outside your field to understand the gist of the problem. We encourage you to use Stan for the analysis, but this is not a strict requirement.

You can use the following as a checklist for the report:

- Short description and motivation for: the problem, the overall modelling goal and the dataset
- Model description: you need to describe at least two different models (e.g. hierarchical vs non-hierarchical, linear vs. non-linear, feature selection etc)
- Description and motivation of the priors you used
- Convergens diagnostics and posterior predictive checking
- Model comparisons using LOO-CV and predictive accuracy if relevant
- Prior sensitivity analysis
- Conclusion
- Maximum 6 pages, but additional figures, table and code can be put in the appendix if needed
- The code must be provided in the appendix

The project presentation
The (prelimary) results of the project must be presented in class on either the 9th of December from 8-12 or the 16th of December from 8-12. The exact timeslots depends on how many groups we will have. If you are following the course remotely, you can make the presentation on Zoom. The presentation must contain a description of the problem and the preliminary results. Groups of 1 and 2 students will have 10 minutes for presentation, while groups of 3 will have 15 minutes for the presentation. Each presentation will be followed by approximately 5 minutes of questions. Once the groups are formed, I will send out a schedule for the presentations.

Help and supervision
We will have project sessions every Thursday from 10-12 (the usual time) on DTU Campus, where you can get help and guidance with the project work if needed.

Important deadlines

- Group formation: 14th of November 2021

- Presentations in class: 9th and 16th of December 201

- Report: 9th of January, 2022

The report must be handed-in via DTU Learn.

Questions
If you have any questions: come to the Thursday sessions, ask on Teams, or send me an email on `miri@dtu.dk`.