Uppsala University

Bayesian Statistics and Data Analysis

Mini-Project Instructions

Contents

1 Project Instructions

The last two weeks will focus on a course project where 2-3 students choose data and will do a Bayesian analysis of a real-world dataset.

Requirements for the projects are:

- Your project should be a Bayesian Data Analysis using Stan (you can use `brms` or `rstanarm` if you like. Although, this is easier and will be considered when grading, i.e. it will be harder to get a VG in the mini-project.
- Real data should be used (see below for details).
- At least two different models should be estimated and compared.

Tip! Write the report so you can add your final report as an example of work you have done to future potential employers.

For PhD students: You can choose to make a small project related to your research interest instead. Although, it should still be a 4 page paper output.

1.1 Suggested Reading/Video material

The project will be a small practical exercise in Bayesian data analysis. To get some inspiration, see the Stan YouTube channel [here] or the Stan User Guide [here]

1.2 Project Group and Expected Workload

It is possible to have only one student in a group, although this is not recommended. One student group will, in practice, mean additional work due to the requirements of the project.

The project is expected to take 40h per student in the group. Hence a 3 group project should be the equivalent of a 120h project.

1.3 Data Sets and Methods Recommendations

We recommend that you find a dataset you are interested in using yourself, ideally in a field you find interesting. Feel free to discuss potential projects with the teacher.

If you have a hard time finding a dataset to use, there are a lot of available datasets (and problems) at:

- The UCI Machine Learning repository: [here]
- The machine learning competition site Kaggle: [here]

The following data sets should not be used in the project:

- Titanic (R data set)
- mtcars (R data set)

1.4 Project Proposal

Students need to turn in a one-page project proposal and data description by and get approval for the proposed project. The Project proposal should follow the ICML paper format that can be found [here]. The ICML format is also available in overleaf here: [here]. We recommend using Overleaf for writing the proposal.

The project proposal should include all the group members names!

1. Title
    - The title should describe the problem and be like a real article title, i.e. don't write "Mini-project:" or similar in the title.
2. Abstract (1 paragraph)
3. Introduction (roughly 0.5 page)
    - Description of the problem/area/idea.
4. Data (roughly 0.5 page)
    - Description of the data, e.g. the number of observations, the number of groups (if you intend to use a hierarchical model), what is the dependent variable etc.
5. Models (roughly 0.25 page)
    - Describe the most basic model you will use/start from in math (using LaTeX equations).

1.5 Project Report

The Project outcome is a report in the ICML paper format that can be found [here]. The ICML format is also available in overleaf here: [here]. We recommend using Overleaf for writing the report.

The paper should consist of *between three and a half (3.5) and four (4) pages*, excluding references and eventual appendices. Write the report as you would do in a real situation, i.e. do not refer to the paper as a "mini-project" or similar.

The project report should include all the group members names!

The paper should include the following parts/sections:

1. Title
    - The title should describe the problem and be like a real article title, i.e. don't write "Mini-project:" or similar in the title.
2. Abstract (1 paragraph)
3. Introduction (roughly 0.5 page)
    - Description of the problem.

4. Data (roughly 0.5 page)
    - Description of the data, e.g. the number of observations, the number of groups (if you intend to use a hierarchical model), what is the dependent variable etc.
5. Models (roughly 0.5 page)
    - Description of the models
    - Description of how the models were compared (LOO/WAIC)
6. Results (roughly 1.5-2 pages)
    - Results of the different models
    - Which model does seem to work the best, and why?
7. Conclusions (roughly 0.5-1 pages)
    - Conclusions from the results.
    - Discussion of problems and potential improvements and other models
8. Acknowledgements (optional)
    - If you are willing to let me use your project report as an example in the next course, please state that in the Acknowledgements. Just add the following sentence: "We hereby grant our consent for the utilization of this project report as a reference material within the context of future editions of the course."
    - Other people you might want to thank.

Additional requirements and hints for the report:

1. All Figures using color should have a color-blind friendly color palette. See here and here.
2. Before you turn in the project, do a language check with a tool such as Grammarly. A project will poor english (errors that would have been spotted with a tool such as Grammarly) will affect your grade downwards.
3. The final report should look like a research paper, i.e. try to avoid bullet list and get a good flow in the text. Also do not refer to the report as a mini-project. See it as a real report (or short paper).
4. You should use correct reference systems. A tip is to use
`citet`,
`citep`, and
`bibtex`. This will also simplify your future thesis work.

1.6 Project Presentations

Presentation details:

- Each project needs to be presented in addition to submitting the mini-project report
- The presentation should be high level, but sufficiently detailed information should be readily available to facilitate answering questions from the audience
- For 1-2 person groups, the presentation should be 10 minutes
- For three-person groups, the presentation should be 15 minutes
- Afterwards, questions will be asked first by other students and then by attending teachers.

Specific presentation recommendations:

- The first slide needs to include the project title and names of the group members.
- The chosen methods(s) should be explained and justified (you are *not* holding this presentation for a hypothetical customer who doesn't care about the details of your methods).
- Big enough font size for text and figure labels should be used to make it easy for the audience to read slides.
- A good rule of thumb is to expect one slide to take 2 minutes to present.
- The last/final slide needs to include your conclusion and names of the group members.

Missing the project presentation

Suppose you cannot attend the presentations due to sickness or similar. In that case, you will have to turn in a video presentation where you present the whole project (i.e. the entire presentation). The teacher then grades this presentation.

1.7 Project Grading

Below are the criteria used when grading the mini-projects. Some general comments on grading are:

1. The more students the higher the quality expected of the project, i.e. a better report is expected from a three-student report than a two-student report.

To pass the report (G), the following criterias should be fulfilled:

1. The report should be turned in and follow the general outline of Section 1.5.

2. show basic knowledge and understanding of the core concepts of the course by using concepts correctly
3. show an understanding in when certain methods should be used or not, and how
4. use at least two (2) different models and compare them in a correct way
5. state what has been done in the report with clarity, good english and rigour so it is easy for a reader to understand and follow the paper.
6. correctly use references in the report following the guideline of the template in Section 1.5

To pass the mini-project with distinction (VG), the following criterias also apply in addition to the criteria for passing the report above:

1. show deep knowledge knowledge and understanding of the core concepts and how to adapt them in a good way to a new situation
2. connect the analysis in the report with other areas in statistics or previous courses taken in the masters program, i.e. not just repeat what has been done in previous labs.
3. use models that has not been part of the BSDA course