Matias Quiroz and Mattias Villani
Division of Statistics and Machine Learning
Dept. of Computer and Information Science
Linköping University
matias.quiroz@liu.se

Instructions for Bayesian Learning project

This is a document with some rough guidelines for the project in the Bayesian learning course. The instructions for the individual project report are intentionally vague. You are supposed to find a problem for analysis by yourself and then select an existing data set, or several data sets, that helps to shed light on your problem. It is also possible to took at a theoretical problem without any data, but most students prefer to analyze some data. Note that the project is **individual**. You can of course have the usual discussions with classmates about Bayesian learning in general, but don't ask for specific solutions to problems that you encounter in your project. It is YOUR project, YOU get the grade, YOU do the work.

- **Deadlines**
  Project proposal: No later than **May 19 at midnight**.
  Project: No later than **June 5 at midnight**.

- **Grade**
  The project is graded on the A-F scale for master students. I might need to complement the project report with oral examinations if my posterior about your grade is too wide. For Ph.D. students the grade is pass or fail, but note that pass corresponds to at least B (preferably closer to A) on the master student scale.

- **The product**
  You are expected to upload a written (well written!) report in PDF format (no obscure .doc files!) on Lisam before the deadline. Please attach computer code in an appendix of the project report. You should also upload an executable version of your code, together with the data sets you use. I should be able to execute your code straightforwardly if the need arises. Comment your code and please put an effort in making it look good. To deliver programming code in good shape is a good preparation exercise for your future life as a Bayesian statistician.
  It is impossible to say anything sensible about the extent (e.g. how many pages?) of the report. All I can say is that the project is 4 credit points and it is the only examination for the course. Note that using pre-packaged software rather than your own written program code increases the demands on the statistical analysis and modelling, and makes it even more important that you clearly interpret the results to show that you have understood the theory and concepts underlying the analysis performed by the software package.

- **The content of the project**
  You are free to work on any problem you like, but I do expect you to do a Bayesian analysis of the problem. It should be noted that it is not acceptable to basically redo what you did in the computer labs. As an example, it is not sufficient to implement a probit regression and redo computer lab 3 but on a different data set. However, if you add (for example) Bayesian variable selection and do a proper model evaluation,

perhaps comparing to other binary regression models, then the problem suddenly becomes much more interesting (and challenging!).

The typical project involves the following **four** steps:

- A well-defined **problem definition**. The problem definition can change when you start analyzing the data, but having an idea to start with helps. I would suggest that you pick a problem that YOU care about. Make it fun for you. There is no need to make any kind of novel scientific contribution in this course (but I don't mind if you do!).

- A **data set**. It is certainly ok to use several data sets, and even to divide the project report in more or less unrelated parts. But I think that it is easier to do a good project if you focus on a single data set. If you write your own code, I strongly recommend that you first try it out on simulated data. This gives you the possibility to check that the code works properly, as you know the values of the parameters you are trying to estimate. This is a powerful way to discover painful bugs in your code!

- **Models**. When you choose a problem definition and a dataset, make sure that it gives you the opportunity to analyze interesting models. You can try several models for your data set if you want to. Remember that almost any toy model becomes interesting (and not so toy anymore!) if it's considered in a hierarchical framework.

- **Analysis**. Once you have your data set and some estimated models, move on to posterior analysis. Maybe you can illustrate how to do predictions with the model (perhaps computing the predictive distribution by simulation)? How to evaluate the adequacy of the model? Analyze which covariates to include in the model, if you are in a regression-type of situation? Only your fantasy sets the limits.

- **Project proposal**
  Submit a project proposal to me (through Lisam, again pdf only!) no later than **May 19**. The project proposal is a very short document (not more than a page) where you explain what you intend to do in the project. Use the four bullet points above (problem definition, data set, models and analysis) to structure your project proposal. Keep it brief, just tell me enough so that I can judge if the project is worth pursuing.

GOOD LUCK AND MAY BAYES BE WITH YOU!