Bayesian Statistics and Probabilistic Programming – Spring 2021

Course project guidelines

Goals, extension, format

In addition to the merely academic purpose of grading, the course project should be an opportunity of practicing and orienting several skills to the common aim of gathering and sorting information on a specific subject and, then, composing a short, but self-contained and informative report on it.
An orientative figure for the expected length could be around fifteen pages, plus code. This is just, as stated, a *roughly orientative* estimate, meaning that a three page report, say, will certainly be deemed as much too short, whereas a forty five pages report would be an undue anticipation of your Master thesis.

Both individual or group projects are possible. In collective projects each author is responsible of and should be able to answer questions about the whole project. Meaning there should be no allocation of different parts to different team members.

Send email with your intended project topic (and, if applicable, the list of group members) so I can give you the go ahead.

Possible subjects

In principle, any Data Analysis topic is possible, provided it is connected with, or seen from a Bayesian and/or Probabilistic Programming perspective. Thus if you know of, or have seen some possible subject, you can propose it by sending me an email. As possible suggestions of lines along which you can search for a theme, I propose the following classification:

A) Methodology-oriented projects:

1. Sampling schemes/approximation methods. Variants and extensions of sampling or approximation methods seen -or to be seen this month- in class. Some of these: Integrated Nested Laplace Approximation (INLA), Variational Inference, Langevin Markov Chain Monte Carlo, Adaptive MCMC, Reversible Jump MCMC.
2. Other models: Bayesian versions of CART, or clustering, or neural networks, or other aspects of inference: Bayesian model selection (in particular predictor selection, e.g. *spike-slab* method). Model averaging.
3. Centered on a probabilistic programming language, e.g., Edward(2), Pyro, Tensorflow-probability, Turing.

B) Data-oriented projects:

Projects centered on analyzing a given dataset or collection of datasets. This variety of project should include treatment applying several methods, with a detailed comparison and discussion of their respective strengths and weaknesses. Code should be thoroughly documented.

If in your job or research work you are using Bayesian Statistical methods, a possible course project is an informative description of some relevant problem and its treatment.

Caution

Beware of too ambitious projects which, as stated above, should better go to your Master thesis.