Master M2 MVA 2019/2020 - Graphical models

Homework 3

These exercises are due on or before January 15th 2020 and should be submitted on Moodle. They can be done in groups of two students. The write-up can be in English or in French. Please submit your answers as a pdf file that you will name MVA DM1 your name.pdf if you worked alone or MVA DM1 name1 name2.pdf with both of your names if you work as a group of two. Indicate your name(s) as well in the documents. Please submit your code as a separate zipped folder and name it MVA DM1 .zip if you worked alone or MVA DM1 name1 name2.zip with both of your names if you worked as a group of two. Note that your files should weight no more than 16Mb.

Gibbs sampling and mean field VB for the probit model

Download the German credit dataset available in e.g. the UCI repository: https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/german/german.data-numeric

We would like to classify the last column (good vs bad credit) based on the previous columns (see the UCI repository for more information on each variable). To do so, we consider the probit model, where $y_i = \mathrm{sgn}(\beta^T x_i + \epsilon_i)$, $\epsilon_i \sim N(0,1)$ and a Gaussian prior $\beta \sim N(0, \tau I_p)$, with $p = \dim \beta$, and $\tau = 10^2$.

1. Start by adding a constant column, and normalising all the predictors (so that their average is 0 and their standard deviation is one). Why is it important to pre-process the predictors in this way? (think in particular about the choice of the prior distribution).

2. Explain why the variance of the ϵ_i is one, and not σ^2, with σ^2 an extra parameter that must be estimated.

3. From the calculations in the last lecture, implement a Gibbs sampler for this model (based on the introduction of latent variables $z_i = \beta^T x_i + \epsilon_i$, as explained during the course). Run your algorithm on the German dataset, and plot the (approximated) marginal posterior distribution of each parameter.

4. Now implement the mean field variational algorithm seen during the course, and compare the results with those of Gibbs; both in terms of speed, and accuracy. Comment. (Give details about your derivations of the distributions computed at every iteration.)
5. Bonus question: find formal arguments on why the posterior variance is under-estimated by the mean-field approach.
6. Complete separation occurs in a dataset if there exists β such that $y_i \beta^T x_i > 0$ for all the datapoints. Represent graphically complete separation. Is maximum likelihood estimation possible in such a case? (Explain.) Construct a simple dataset with full separation, run your Gibbs sampler on this dataset, and comment.