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

2016-05-03

Bayesian Learning, 6 hp

Computer lab 3

You can use any programming language for the labs, but the hints, help and solutions will be in R.
You are supposed to work and submit your labs in pairs, but do make sure that both of you are contributing. Submit your solutions by Lisam no later than **May** 13 at **midnight**. You should submit the report (pdf only) and an executable file containing your code.

1. *Normal model, mixture of normal model with semi-conjugate prior.*
   The data `rainfall.dat` consist of daily records, from the beginning of 1948 to the end of 1983, of precipitation (rain or snow in units of $\frac{1}{100}$ inch, and records of zero precipitation are excluded) at Snoqualmie Falls, Washington. Analyze the data using the following two models.

   (a) *Normal model.*
   Assume the daily precipitation $\{y_1, ..., y_n\}$ are independent normally distributed, $y_1, ..., y_n|\mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2)$ where both μ and σ^2 are unknown. Let $\mu \sim \mathcal{N}(\mu_0, \tau_0^2)$ independently of $\sigma^2 \sim Inv\text{-}\chi^2(\nu_0, \sigma_0^2)$.

      i. Implement (code!) a Gibbs sampler that simulates from the joint posterior $p(\mu, \sigma^2|y_1, ..., y_n)$. [Hint: the full conditional posteriors are given on the slides from Lecture 8].

      ii. Analyze the daily precipitation using your Gibbs sampler in (a)-i. Investigate the convergence of the Gibbs sampler by suitable graphical methods. Is the Gibbs sampler efficient?

   (b) *Mixture normal model.*
   Alternatively, let's us now assume that the daily precipitation $\{y_1, ..., y_n\}$ follow an i.i.d two-component mixture of normals model:

$$p(y_i|\mu, \sigma^2, \pi) = \pi\mathcal{N}(y_i|\mu_1, \sigma_1^2) + (1 - \pi)\mathcal{N}(y_i|\mu_2, \sigma_2^2),$$

   where

$$\mu = (\mu_1, \mu_2) \quad \text{and} \quad \sigma^2 = (\sigma_1^2, \sigma_2^2).$$

      i. Use the Gibbs sampling data augmentation algorithm in Mattias' code (or code your own!) in `NormalMixtureGibbs.R` (available under Lecture 7 on the course page) to analyze the daily precipitation data. Set the prior hyperparameters suitably.

ii. Investigate the convergence of the Gibbs sampler by suitable graphical methods. Is this Gibbs sampler efficient?

(c) *Graphical comparison.*
Plot the following densities in one figure: 1) a histogram or kernel density estimate of the data. 2) Normal density $\mathcal{N}(\mu, \sigma^2)$ in (a); 3) Mixture of normals density $p(y_i|\mu, \sigma^2, \pi)$ in (b). Use the posterior mean value for all the parameters. Which model is better in terms of fitting?

2. *Binary regression models*

   (a) Use the code `OptimizeSpamR.zip` from the course web page (under Lecture 6) to analyze the spam data set using logistic regression. Approximate the posterior of the regression coefficients β by $\mathcal{N}(\tilde{\beta}, J^{-1})$, where J is the observed information matrix evaluated at the posterior mode $\tilde{\beta}$. All of this is given by the code. Use the prior $\beta \sim \mathcal{N}(0, \tau^2 I)$, with $\tau = 10$.

   (b) Implement (code!) a data augmentation Gibbs sampler for the probit regression model. [Hint: `rtnorm` function in `the msm`-package + Lecture 8.]

   (c) Analyze the spam data using your code from b), again using the prior $\beta \sim \mathcal{N}(0, \tau^2 I)$, with $\tau = 10$.

   (d) Compare the results in a) and c).

MAY BAYES BE WITH YOU!