DATA6810 - Probabilistic Models for Complex Data

Assignment 3

Due: Sunday 29 May 2022, 23:59AEST

1 General Description

This assignment for DATA6810 evaluates knowledge in the topic of Mixture Models. It consists of programming and technical exercises which are closely related to the topics covered in the respective tutorials. The expected completion time is 3 hours.

1.1 Deliverables

Submission will be through Microsoft Teams. Please complete the python notebooks / R markdown files for each section and upload them to the Teams assignment environment in the "your work" section.

Marking criteria involves a total of 100 points. A penalty of MINUS 5 points per each day after the due date will be applied.

2 Mixture Models

Complete the following questions, and make sure that your submitted code is well-commented.

Consider the following Gaussian mixture model for a dataset $\boldsymbol{y} = (y_1, \ldots, y_n)$, where each observation y_i has a one-dimensional covariate x_i:

$$L(y_i|x_i, \boldsymbol{\beta}, \boldsymbol{w}, \sigma^2, K) = \sum_{k=1}^{K} w_k \mathcal{N}(\beta_{0k} + \beta_{1k} x_i, \sigma^2),$$

where $\mathcal{N}(\mu, \sigma^2)$ is a Gaussian distribution with mean μ and variance σ^2. The mixture weights $(w_1, \ldots, w_K)$ are positive numbers summing to 1. Our goal is Bayesian inference of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K)$ where $\boldsymbol{\beta}_k = (\beta_{0k}, \beta_{1k})^\top$, $\boldsymbol{w} = (w_1, \ldots, w_{K-1})$ and σ^2. In this question we will fix $K = 3$ and choose the following prior distributions:

- $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3 \sim \text{MVN}(\boldsymbol{0}, 100\boldsymbol{I})$, where $\boldsymbol{I}$ is an identity matrix and $\text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

- $\sigma \sim$ half-t$(1, 5)$, or equivalently, $\sigma^2|a \sim \text{IG}(\frac{1}{2}, \frac{1}{a})$ and $a \sim \text{IG}(\frac{1}{2}, \frac{1}{5^2})$. We say that x has a half-t distribution with degree of freedom ν and scale A, written as $x \sim$ half-t(ν, A), if

$$g(x|\nu, A) \propto (1 + (x/A)^2/\nu)^{-(\nu+1)/2}, \quad \text{with } x > 0$$

and x has an inverse-Gamma distribution with shape a and rate b, written as $x \sim \text{IG}(a, b)$, if

$$g(x|a, b) \propto x^{-a-1} \exp(-b/x), \quad \text{with } x > 0.$$

(a) Choose a suitable prior for $\boldsymbol{w}$ and find the joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{w}, \sigma^2, a)$ given a sequence of independent observations $\boldsymbol{y}$, up to a normalising constant. [7 marks]

(b) What is the full conditional distribution of each parameter under this posterior distribution? Explain why Gibbs sampling from this distribution is infeasible. [10 marks]

(c) We will rectify the situation by introducing $\boldsymbol{z} = (z_1, \ldots, z_n)$ with z_i labelling the mixture component generating observation i. State a suitable prior for $\boldsymbol{z}$. [3 marks]

(d) Write down the likelihood $L(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{z}, \sigma^2, a, K)$ and hence find the densities/mass functions of the full conditionals [25 marks]

   (i) $z_i|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{z}_{-i}, \sigma^2, a, \quad i = 1, \ldots, n,$
   (ii) $\boldsymbol{\beta}_k|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{z}, \sigma^2, a, \quad k = 1, \ldots, K,$
   (iii) $\boldsymbol{w}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{z}, \sigma^2, a$
   (iv) $a|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{z}, \sigma^2$
   (v) $\sigma^2|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{z}, a$

(e) Implement a Gibbs sampler using your results in part (d). [30 marks]

(f) Consider the two datasets given in `.Rdata` format, which are named `dataset1` and `dataset2`, respectively. [20 marks]

   (i) Run your Gibbs sampler on the first dataset for 25 000 iterations. Produce a trace plot for $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, w_1$ and w_2, and also plot the evolution of their estimated posterior means.
   (ii) Is the empirical evidence you obtained in part (i) consistent with the idea that the Gibbs sampler is mixing well? If no, how might you fix this issue? If yes, do you believe this?
   (iii) Repeat parts (i)–(ii) for the second dataset. Comment and explain on any differences you observe.

(g) Explain how you would assess MCMC convergence for the model above. [5 marks]