ST451 Bayesian Machine Learning
Week 7

Exercises

1. Consider an observation from a trinomial random variable $x = (x_1, x_2, x_3)$ and parameters $\theta = (\theta_1, \theta_2, \theta_3)$. The likelihood for this observation is proportional to

$$f(x|\theta) = f(x_1, x_2, x_3|\theta_1, \theta_2, \theta_3) = \frac{\Gamma(n+1)}{\Gamma(x_1+1)\Gamma(x_2+1)\Gamma(x_3+1)}\theta_1^{x_1}\theta_2^{x_2}\theta_3^{x_3},$$

for $0 < \theta_i < 1$, $x_i \in \{0, 1, \ldots, n\}$, $i \in \{1, 2, 3\}$, $x_1 + x_2 + x_3 = n$ and $\theta_1 + \theta_2 + \theta_3 = 1$. It is also known that $E(x_i) = n\theta_i$.

   (a) Assign the Dirichlet$(\alpha_1, \alpha_2, \alpha_3)$ distribution to θ with

$$\pi(\theta) = \pi(\theta_1, \theta_2, \theta_3) = \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)}\theta_1^{\alpha_1-1}\theta_2^{\alpha_2-1}\theta_3^{\alpha_3-1},$$

   for $0 < \theta_i < 1$, $\alpha_i > 0$, $i \in \{1, 2, 3\}$, $\theta_1 + \theta_2 + \theta_3 = 1$. Derive the posterior of θ.

   (b) Find the Jeffreys' prior for θ and use it to obtain the corresponding posterior distribution.

   (c) Let $y = (y_1, y_2, y_3)$ represent a future observation from the same model. Write down the posterior predictive distribution of y based on a prior of your choice. Describe a procedure to simulate from this posterior predictive distribution based on random samples from the posterior distribution of θ.

2. Assume that the data $x = (x_1, \ldots, x_n)$ are independent scalar random variables and the distribution of each x_i, is given by a mixture of $K = 2$ Normal distributions with parameters $(\mu, \sigma^2) = (\mu_k, \sigma_k^2)$, for $k = 0, 1$. In other words assume that z_i is a binary random variable being in category 1 with probability π and in category 0 with probability $1 - \pi$, and that

$$f(x_i|\mu, \sigma^2, \pi) = (1-\pi)f(x_i|\mu_0, \sigma_0^2) + \pi f(x_i|\mu_1, \sigma_1^2),$$

where $f(x_i|\lambda_k) = N(x_i|\mu_k, \sigma_k^2)$. Give the details of the EM algorithm that can be used to find maximum likelihood estimate of the $\theta = (\mu, \sigma^2, \pi)$. Define explicitly the E and the M steps.

3. Consider the Water Treatment Plant Data Set from the UCI repository. You can check computer class of week 4 on how to access data from the UCI repository. Choose 4 continuous variables from the dataset to analyse using Gaussian Mixture models. Select the optimal model from a set of models with up to 7 clusters and 4 covariance matrix types (spherical,tied, diagonal and full) and present its output. Also fit a Bayesian Gaussian mixture model and compare the results with a standard Gaussian mixture case.