ST451 Bayesian Machine Learning
Week 5

Exercises

1. Consider the model where we have independent observations $y = (y_1, \dots, y_n)$ from the $\mathrm{N}(\mu, \sigma^2)$ with the $\mathrm{N}(\mu_0, \omega^2\sigma^2)$ and $\mathrm{IGamma}(\alpha_0, \beta_0)$ being the priors for μ and σ^2 respectively. The aim is to perform variational inference for this model under the mean field approximation framework. Derive the explicit algorithm for this problem and present its steps.

   Answer: The likelihood for $\theta = (\pi, \mu_1, \mu_2, \sigma^2)$ based $(y_i, X_i)_{i=1}^n$ can be written as

$$f(x, y|\theta) = \prod_{i=1}^n \left[\pi N(\mu_1, \sigma^2)\right]^{y_i} \left[(1-\pi)N(\mu_0, \sigma^2)\right]^{1-y_i}$$

   To maximise with respect to π we write the log-likelihood keeping the terms that involve π

$$\log f(x, y|\pi) = c + \sum_{i=1}^n \{y_i \log \pi + (1-y_i)\log(1-\pi)\}$$

   After differentiating the above wrt π, setting equal to 0 and solving the equation we get

$$\hat{\pi} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{n_1}{n} = \frac{n_1}{n_0 + n_1},$$

   where n_0 denotes the number of individuals in category 0, n_1 denotes the number of individuals in category 1, and $n = n_1 + n_2$.
   " To maximise with respect to μ_1 we write the log-likelihood keeping the terms that involve μ_1:

$$\log f(x, y|\mu_1) = c + \sum_{i=1}^n y_i \log N(x_i|\mu_1, \sigma^2) = c - \frac{1}{2}\frac{\sum_{i=1}^n y_i(x_i - \mu_1)^2}{\sigma^2}$$

   After differentiating the above wrt μ_1, setting equal to 0 and solving the equation we get

$$\hat{\mu}_1 = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n y_i} = \frac{\sum_{i=1}^n y_i x_i}{n_1}$$

   Similarly we obtain

$$\hat{\mu}_0 = \frac{\sum_{i=1}^n (1-y_i)x_i}{\sum_{i=1}^n (1-y_i)} = \frac{\sum_{i=1}^n (1-y_i)x_i}{n_0}$$

   Finally for the common variance σ^2

$$\begin{aligned}
\log f(x, y|\sigma^2) &= c + \sum_{i=1}^n (1-y_i)\log N(x_i|\mu_0, \sigma^2) + \sum_{i=1}^n y_i \log N(x_i|\mu_1, \sigma^2) \\
&= c - (n_0/2)\log\sigma^2 - \frac{1}{2}\frac{\sum_{i=1}^n (1-y_i)(x_i-\mu_0)^2}{\sigma^2} - (n_1/2)\log\sigma^2 - \frac{1}{2}\frac{\sum_{i=1}^n y_i(x_i-\mu_1)^2}{\sigma^2} \\
&= c - (n/2)\log\sigma^2 - \frac{1}{2}\frac{\sum_{i=1}^n \left\{(1-y_i)(x_i-\mu_0)^2 + y_i(x_i-\mu_1)^2\right\}}{\sigma^2}
\end{aligned}$$

After differentiating the above wrt σ^2, setting equal to 0 and solving the equation we get

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{(1-y_i)(x_i-\mu_0)^2 + y_i(x_i-\mu_1)^2\right\}$$

2. Following the previous exercise, simulate data from this model, using parameters of your preference, and apply the algorithm for variational inference for the parameters you used. Provide evidence of agreement between the estimated and true values of the parameters used.

   Answer: See file ending in ...exercise2.ipynb

3. Consider the model for the VIX index mentioned in the lecture slides

   $$Y_t = Y_{t-1} + \kappa(\mu - Y_{t-1})\delta + \epsilon_t,$$

   where Y_t is VIX at time t, and ϵ_t are independent error terms following the student-t distribution with scale σ and 3 degrees of freedom. Download the VIX series (file vix_201518.csv) and fit this model using automatic Variational Bayes via RStan.

   Hint: In RStan the code $x \sim$ student_t(df,m,s) indicates that x follows the student-t distribution with df degrees of freedom, location mu and scale s.

   Answer: See file ending in ...exercise3.Rmd