Advanced Probabilistic Methods

Marshal Sinaga - marshal.sinaga@aalto.fi

This note will be updated regularly until April 16, 2024. All materials are adopted from Pekka Martinen (CS-E4820).

1 Variational Inference

1.1 Variational Bayes for Simple Model

Suppose we have N independent observations $\mathbf{x} = (x_1, \cdots, x_N)$ from a two-component mixture of univariate Gaussian distributions.

$$p(x_n|\theta) = (1-\tau)N(x_n|0,1) + \tau N(x_n|\theta,1) \tag{1}$$

that is with probability $1-\tau$ the observation x_n is generated from the first component $N(x_n|0,1)$, and with probability τ from the second component $N(x_n|\theta,1)$. The model 1 has two unknown parameters (τ, θ), the mixture coefficient and the mean of the second component.

The goal is to carry out a full Bayesian analysis via mean-field variational Bayesian approximation. We place the following priors on the unknown parameters.

$$\begin{aligned} \tau &\sim Beta(\alpha_0, \alpha_0) \\ \theta &\sim N(0, \beta_0^{-1}) \end{aligned}$$

We formulate the model using latent variables $\mathbf{z} = (z_1, \cdots, z_N)$, which explicitly specify the component responsible for generating observation x_n. In detail,

$$z_n = (z_{n1}, z_{n2})^\top = \begin{cases} (1,0)^\top & x_n \text{ is from } N(x_n|0,1) \\ (0,1)^\top & x_n \text{ is from } N(x_n|\theta,1) \end{cases}$$

and place a prior on the latent variables

$$p(\mathbf{z}|\tau) = \prod_{n=1}^{N} \tau^{z_{n2}}(1-\tau)^{z_{n1}}$$

The likelihood in the latent variable model is given by

$$p(\mathbf{x}|\mathbf{z},\theta) = \prod_{n=1}^{N} N(x_n|0,1)^{z_{n1}} N(x_n|\theta,1)^{z_{n2}}$$

The joint distribution of all observed ($\mathbf{x}$) and unobserved variables ($\mathbf{z}, \tau, \theta$) factories as follows

$$p(\mathbf{x}, \mathbf{z}, \tau, \theta) = p(\tau)p(\theta)p(\mathbf{z}|\tau)p(\mathbf{x}|\mathbf{z}, \theta)$$

and the log joint distribution can correspondingly written as

$$\log p(\mathbf{x}, \mathbf{z}, \tau, \theta) = \log p(\tau) + \log p(\theta) + \log p(\mathbf{z}|\tau) + \log p(\mathbf{x}|\mathbf{z}, \theta)$$

We approximate the posterior distribution $p(\mathbf{z}, \tau, \theta|\mathbf{x})$ using the factorized variational distribution $q(\mathbf{z})q(\theta)q(\theta)$

Update factor $q(\mathbf{z})$ To compute the updated distribution $q^*(\mathbf{z})$, we first compute the expectation of the log of the joint distribution over all other unknowns in the model.

$$\begin{aligned}
&\log q^*(\mathbf{z}) = \mathbb{E}_{\tau,\theta}[\log p(\mathbf{x}, \mathbf{z}, \tau, \theta)] \\
&= \mathbb{E}_{\tau}[\log p(\mathbf{z}|\tau) + \mathbb{E}_{\theta}[\log p(\mathbf{x}|\mathbf{z}, \theta)]] + \text{ const} \\
&= \mathbb{E}_{\tau}[\sum_{n=1}^{N} z_{n2} \log \tau + z_{n1} \log(1-\tau)] + \mathbb{E}_{\theta}[\sum_{n=1}^{N} z_{n1} \log N(x_n|0,1) + z_{n2} \log N(x_n|\theta, 1)] + \text{const} \\
&= \sum_{n=1}^{N} z_{n2}\mathbb{E}_{\tau}[\log \tau] + z_{n1}\mathbb{E}_{\tau}[\log(1-\tau)] + \sum_{n=1}^{N} z_{n1} \log N(x_n|0,1) + z_{n2}\mathbb{E}_{\theta}[\log N(x_n|\theta, 1)] + \text{const} \\
&= \sum_{n=1}^{N} z_{n1} \left(\mathbb{E}_{\tau}[\log(1-\tau)] - \frac{1}{2}\log 2\pi - \frac{1}{2}x_n^2\right) + \sum_{n=1}^{N} z_{n2} \left(\mathbb{E}_{\tau}[\log \tau] - \frac{1}{2}\log 2\pi - \frac{1}{2}\mathbb{E}_{\theta}[(x_n-\theta)^2]\right) + \text{const} \\
&= \sum_{n=1}^{N} z_{n1} \log \rho_{n1} + z_{n2} \log \rho_{n2} + \text{const} \qquad (2)
\end{aligned}$$

Where we have defined ρ_{n1} and ρ_{n2} for all n as follows

$$\log \rho_{n1} = \mathbb{E}_{\tau}[\log(1-\tau)] - \frac{1}{2}\log 2\pi - \frac{1}{2}x_n^2 \qquad (3)$$

$$\log \rho_{n2} = \mathbb{E}_{\tau}[\log \tau] - \frac{1}{2}\log 2\pi - \frac{1}{2}\mathbb{E}_{\theta}[(x_n-\theta)^2] \qquad (4)$$

By exponentiating both sides of Equation 2, we obtain

$$q^*(\mathbf{z}) \propto \prod_{n=1}^{N} \prod_{k=1}^{2} \rho_{nk}^{z_{nk}}$$

which can be normalized to make a proper distribution

$$q^*(\mathbf{z}) = \prod_{n=1}^{N} \prod_{k=1}^{2} r_{nk}^{z_{nk}}$$

where

$$r_{nk} = \frac{\rho_{nk}}{\sum_{j=1}^{2} \rho_{nj}}$$

Note that computing r_{nk} requires $\mathbb{E}_{\tau}[\log \tau], \mathbb{E}_{\tau}[\log(1-\tau)]$, and $\mathbb{E}_{\theta}[(x_n-\theta)^2]$, where the expectations are computed over the distribution $q(\tau)$ and $q(\theta)$, which will be derived next.

Update factor $q(\tau)$

$$
\begin{aligned}
&\log q^*(\tau) = \mathbb{E}_{\mathbf{z},\theta}[\log p(\mathbf{x},\mathbf{z},\tau,\theta)] \\
&= \log p(\tau) + \mathbb{E}_{\mathbf{z}}[\log p(\mathbf{z}|\tau)] + \text{const} \\
&= \log p(\tau) + \sum_{n=1}^{N} r_{n2}\log\tau + r_{n1}\log(1-\tau) + \text{const} \\
&= \log\tau^{\alpha_0-1} + \log(1-\tau)^{\alpha_0-1} + \sum_{n=1}^{N}\log\tau^{r_{n2}} + \log(1-\tau)^{r_{n1}} + \text{const} \qquad (5)
\end{aligned}
$$

We exponentiate and recognize the exponentiated form as

$$q^*(\tau) = Beta(\tau|N_2+\alpha_0, N_1+\alpha_0)$$

i.e., τ has $Beta(a,b)$ with $a = N_2+\alpha_0$ and $b = N_1+\alpha_0$, where $N_k = \sum_{n=1}^{N} r_{nk}$ for $k = 1,2$. Using this distribution, we get the following formulas for the terms required when updating $q(\mathbf{z})$

$$\mathbb{E}_\tau[\log\tau] = \psi(N_2+\alpha_0) + \psi(N_1+N_2+2\alpha_0) \qquad (6)$$

$$\mathbb{E}_\tau[\log(1-\tau)] = \psi(N_1+\alpha_0) + \psi(N_1+N_2+2\alpha_0) \qquad (7)$$

where ψ is the digamma function. Formulas above follow from the basic property of Beta distribution and the fact that if $\tau \sim Beta(a,b)$ then $1-\tau \sim Beta(b,a)$

Update factor $q(\theta)$

$$
\begin{aligned}
&\log q^*(\theta) = \mathbb{E}_{\tau,\mathbf{z}}[\log p(\mathbf{x},\mathbf{z},\tau,\theta)] \\
&= \log p(\theta) + \mathbb{E}_{\mathbf{z}}[\log p(\mathbf{x}|\mathbf{z},\theta)] + \text{const} \\
&= -\frac{1}{2}\log\beta_0^{-1} - \frac{\beta_0}{2}\theta^2 + \mathbb{E}_{\mathbf{z}}\left[\sum_{n=1}^{N} z_{n1}\left(-\frac{1}{2}x_n^2\right) + z_{n2}\left(-\frac{1}{2}(x_n-\theta)^2\right)\right] + \text{const} \\
&= -\frac{1}{2}\log\beta_0^{-1} - \frac{\beta_0}{2}\theta^2 + \sum_{n=1}^{N} r_{n1}\left(-\frac{1}{2}x_n^2\right) + r_{n2}\left(-\frac{1}{2}(x_n-\theta)^2\right) + \text{const} \\
&= -\frac{\beta_0}{2}\theta^2 + \sum_{n=1}^{N} -\frac{r_{n2}}{2}\left(x_n^2 - 2x_n\theta + \theta^2\right) + \text{const} \\
&= -\frac{1}{2}\left(\left(\beta_0 + \sum_{n=1}^{N} r_{n2}\right)\theta^2 + \sum_{n=1}^{N} r_{n2}x_n^2 - 2\theta\sum_{n=1}^{N} r_{n2}x_n\right) \\
&= -\frac{\beta_0 + \sum_{n=1}^{N} r_{n2}}{2}\left(\theta - \frac{1}{\beta_0 + \sum_{n=1}^{N} r_{n2}}\sum_{n=1}^{N} r_{n2}x_n\right)^2 + \text{const} \qquad (8)
\end{aligned}
$$

Again, we exponentiate both sides of 8 and recognize this as

$$q^*(\theta) = N(\theta|m_2, \beta_2^{-1}) \qquad (9)$$

with

$$\beta_2 = \beta_0 + N_2 \quad \text{and} \quad m_2 = \beta_2^{-1}N_2\bar{x}_2$$

where we have defined

$$\bar{x}_2 = \frac{1}{N_2} \sum_{n=1}^{N} r_{n2} x_n$$

We can use the distribution 9 to compute $\mathbb{E}_\theta[(x_n - \theta)^2]$, needed when updating $q(\mathbf{z})$:

$$\begin{aligned}
\mathbb{E}_\theta[(x_n - \theta)^2] &= \mathbb{E}_\theta[(x_n - m_2 + m_2 - \theta)^2] \\
&= (x_n - m_2)^2 + 2(x_n - m_2)\mathbb{E}[m_2 - \theta] + \mathbb{E}[(m_2 - \theta)^2] \\
&= (x_n - m_2)^2 + \beta_2^{-1} \qquad (10)
\end{aligned}$$

The overall VB algorithm is obtained by cycling through updating:

- The responsibilities r_{nk} using formulas 3, 4, 5
- The terms 10 needed when computing the responsibilities
- The term 6 and 7 needed when computing the responsibilities