Bayesian Methods in Machine Learning
Seminar: 7

Evgenii Egorov

Skoltech

Mean Field Approximation

Recall, ELBO: $\mathcal{L}[q] = \langle \log p(x, z) \rangle_q - \langle \log q(z) \rangle_q$.
We need to optimize it over q. Let's assume $q(z) = \prod_j q(z_j)$. Hence, let's re-write all to adapt the optimization to this form.

$$p(x_{1:n}, z_{1:m}) = p(x_{1:n}) \prod_{j=1}^{m} p(z_j | z_{1:j-1}, x_{1:n}),$$

$$\mathcal{L}[q] = \sum_{j=1}^{m} \langle \log p(z_j | z_{-j}, x_{1:n}) \rangle_{\prod_{j=1}^{m} q(z_j)}, \sum_{j=1}^{m} \langle \mathbb{E}_{-j} \log p(z_j | z_{-j}, x_{1:n}) \rangle_{q(z_j)} - \sum_{j=1}^{m} \langle \log q(z_j) \rangle_{q(z_j)}.$$

Mean Field Approximation

$$\mathcal{F}[q] = \sum_{j=1}^{m} \langle \mathbb{E}_{-j} \log p(z_j | z_{-j}, x_{1:n}) \rangle_{q(z_j)} - \sum_{j=1}^{m} \langle \log q(z_j) \rangle_{q(z_j)} - \sum_{j=1}^{m} \lambda_j \Big(\int q(z_j) dz_j - 1 \Big),$$

$$\frac{\delta}{\delta q_j} F[q] = \mathbb{E}_{-j} \log p(z_j | z_{-j}, x_{1:n}) - \log q(z_j) - \lambda_j = 0.$$

Hence,

$$q_j \propto \exp\{\mathbb{E}_{-j} \log p(z_j | z_{-j}, x_{1:n})\} \propto \exp\{\mathbb{E}_{-j} \log p(z_j, z_{-j}, x_{1:n})\}.$$

Normal-Gamma Model

For $x_i \in \mathbb{R}$, $X = \{x_i\}_{i=1}^N$, $\theta = (\mu, \lambda)$

$$p(X, \mu, \lambda) = \left[\prod_{n=1}^{N} \mathcal{N}(x_n|\mu, \lambda^{-1})\right] \mathcal{N}(\mu|m_0, (\beta\lambda)^{-1}) G(\lambda|a_0, b_0).$$

It will be usefully to write its log:

$$\log p(X, \mu, \lambda) = \left[\sum_{n=1}^{N} \frac{1}{2}\log\lambda - \frac{\lambda}{2}(\mu - x_2)^2\right] + \frac{1}{2}\log(\beta_0\lambda) - \frac{\beta_0\lambda}{2}(\mu - m_0)^2 + (a_0 - 1)\log\lambda - b_0\lambda$$

Consider following approximation:

$$p(\mu, \lambda|X) = q(\lambda)q(\mu)$$

Recall general mean-field update equation:

$$\log q(\theta_j) = \langle \log p(X, \theta)\rangle_{q(\theta_{-j})}$$

Bayesian GMM

Consider the following model:

$$p(X, z, \pi, \mu, \Lambda) = \prod_{n,k} \left[\left(\mathcal{N}(x_n | \mu_k, \Lambda_k^{-1}) \pi_k \right)^{z_{nk}} \right] \mathrm{Dir}(\pi | \alpha_0) \prod_k \mathcal{N}(\mu_k | m_0, (\beta \Lambda_k)^{-1}) W(\Lambda_k | W_0, \mu_0).$$

We consider following approximation:

$$p(z, \pi, \mu, \Lambda | X) = q(z) q(\pi, \mu, \Lambda).$$