Pattern Recognition and Machine Learning II

Prof. Dr. Bernhard Sick
WS 2016/2017
Exercise 7

Task 1: Variational Inference – Univariate Gaussian

Assume you are given the dataset $\mathscr{D} = \{x_1, \ldots, x_N\}$. Use the variational inference framework to determine the posterior distribution for mean μ and precision τ. The likelihood function is given by a **univariate Gaussian**.

$$p(\mathscr{D}|\mu, \tau) = \prod_{n=1}^{N} \mathscr{N}(x_n|\mu, \tau^{-1})$$

Solution:

$$p(\mathscr{D}|\mu,\tau) = \prod_{n=1}^{N}\mathcal{N}(x_n|\mu,\tau^{-1}) \tag{1}$$
$$= \prod_{n=1}^{N}\left(\frac{\tau}{2\pi}\right)^{1/2}\exp\left\{-\frac{\tau}{2}(x_n-\mu)^2\right\} \tag{2}$$
$$= \left(\frac{\tau}{2\pi}\right)^{1/2}\exp\left\{-\frac{\tau}{2}(x_1-\mu)^2\right\}\cdot\ldots\cdot\left(\frac{\tau}{2\pi}\right)^{1/2}\exp\left\{-\frac{\tau}{2}(x_N-\mu)^2\right\} \tag{3}$$
$$= \left(\frac{\tau}{2\pi}\right)^{N/2}\exp\left\{-\frac{\tau}{2}\sum_{n=1}^{N}(x_n-\mu)^2\right\}. \tag{4}$$

We assume the following conjugate prior distributions for mean μ and precision τ:

$$p(\mu|\tau) = \mathcal{N}(\mu|\mu_0,(\lambda_0\tau)^{-1}), \tag{5}$$
$$p(\tau) = \mathrm{Gam}(\tau|a_0,b_0). \tag{6}$$

Together they constitute the Gaussian-Gamma conjugate prior distribution. To illustrate the variational approximation of the posterior distribution, the following factorization will be considered:

$$q(\mu,\tau) = q_\mu(\mu)q_\tau(\tau). \tag{7}$$

According to the lecture, we can obtain the optimum factor according to:

$$\ln q_j^\star(\mathbf{Z}_j) = \mathbb{E}_{i\neq j}[\ln p(\mathscr{D},\mu,\tau)] + \text{const} \tag{8}$$
$$= \mathbb{E}_{i\neq j}[\ln p(\mathscr{D}|\mu,\tau) + \ln p(\mu|\tau) + \ln p(\tau)] + \text{const} \tag{9}$$

Variational Distribution for the mean:

$$\ln q_\mu^\star(\mu) = \mathbb{E}_\tau[\ln p(\mathscr{D}|\mu,\tau) + \ln p(\mu|\tau)] + \text{const.} \tag{10}$$
$$= \mathbb{E}_\tau\left[\ln\left(\frac{\tau}{2\pi}\right)^{N/2}\exp\left\{-\frac{\tau}{2}\sum_{n=1}^{N}(x_n-\mu)^2\right\} + \ln\left(\frac{\lambda_0\tau}{2\pi}\right)\exp\left\{-\frac{\lambda_0\tau}{2}(\mu-\mu_0)^2\right\}\right] + \text{const} \tag{11}$$
$$= \mathbb{E}_\tau\left[-\frac{\tau}{2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{\lambda_0\tau}{2}(\mu-\mu_0)^2\right] + \text{const} \tag{12}$$
$$= \mathbb{E}_\tau\left[-\frac{\tau}{2}\left(\sum_{n=1}^{N}(x_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right)\right] + \text{const} \tag{13}$$
$$= -\frac{\mathbb{E}_\tau[\tau]}{2}\left(\sum_{n=1}^{N}(x_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right) + \text{const.} \tag{14}$$

From the fact that the distribution has one term that quadratically depends on the random variable and one term that linearly depends on it, it follows that it has to be a Gaussian distribution. By rearranging the Gaussian density, we only need to find the terms in front of the quadratic and linear terms of μ.

$$\mathcal{N}(\mu|\mu_N,\lambda_N^{-1}) = \left(\frac{\lambda_N}{2\pi}\right)^{1/2}\exp\left\{-\frac{\lambda_N}{2}(\mu-\mu_N)^2\right\} \tag{15}$$
$$\ln\mathcal{N}(\mu|\mu_N,\lambda_N^{-1}) = \frac{1}{2}\ln\frac{\lambda_N}{2\pi} - \frac{\lambda_N}{2}(\mu-\mu_N)^2 \tag{16}$$
$$= -\frac{\lambda_N}{2}(\mu^2 - 2\mu\mu_N + \mu_N^2) + \text{const} \tag{17}$$
$$= -\frac{\lambda_N}{2}\mu^2 + \lambda_N\mu_N\mu + \text{const} \tag{18}$$

Solution:

We can see that the blue terms can be used to specify the mean and precision of the variational distribution. Therefore, we can rearranging the inner part of the equation

$$\ln q_\mu^\star(\mu) = -\frac{\mathbb{E}_\tau[\tau]}{2}\left(\sum_{n=1}^{N}(x_n-\mu)^2+\lambda_0(\mu-\mu_0)^2\right)+\text{const} \tag{19}$$

according to

$$\begin{aligned}
& \sum_{n=1}^{N}(x_n-\mu)^2+\lambda_0(\mu-\mu_0)^2 && (20)\\
= & \sum_{n=1}^{N}(x_n^2-2\mu x_n+\mu^2)+\lambda_0(\mu^2-2\mu\mu_0+\mu_0^2) && (21)\\
= & \sum_{n=1}^{N}x_n^2-2\mu\sum_{n=1}^{N}x_n+N\mu^2+\lambda_0\mu^2-2\lambda_0\mu\mu_0+\lambda_0\mu_0^2 && (22)\\
= & N\mu^2+\lambda_0\mu^2-2\mu\sum_{n=1}^{N}x_n-2\lambda_0\mu_0\mu+\text{const} && (23)\\
= & (N+\lambda_0)\mu^2-\left(2\sum_{n=1}^{N}x_n+2\lambda_0\mu_0\right)\mu+\text{const.} && (24)
\end{aligned}$$

Inserting this term into Eq. (19), we obtain

$$\begin{aligned}
\ln q_\mu^\star(\mu) &= -\frac{\mathbb{E}_\tau[\tau]}{2}\left((N+\lambda_0)\mu^2-\left(2\sum_{n=1}^{N}x_n+2\lambda_0\mu_0\right)\mu\right)+\text{const} && (25)\\
&= -\frac{\mathbb{E}_\tau[\tau]}{2}(N+\lambda_0)\mu^2+\frac{\mathbb{E}_\tau[\tau]}{2}\left(2\sum_{n=1}^{N}x_n+2\lambda_0\mu_0\right)\mu+\text{const} && (26)\\
&= -\frac{\mathbb{E}_\tau[\tau]}{2}(N+\lambda_0)\mu^2+\mathbb{E}_\tau[\tau]\left(\sum_{n=1}^{N}x_n+\lambda_0\mu_0\right)\mu+\text{const.} && (27)
\end{aligned}$$

In Eq. (27), we identify the blue terms from Eq. (18) and it is easy to specify the mean μ_N and the precision λ_N^{-1}.

- For the mean μ_N:

$$\begin{aligned}
-\frac{\lambda_N}{2} &= -\frac{\mathbb{E}_\tau[\tau]}{2}(N+\lambda_0) && (28)\\
\lambda_N &= \mathbb{E}_\tau[\tau](N+\lambda_0) && (29)
\end{aligned}$$

- For the precision λ_N:

$$\begin{aligned}
\lambda_N\mu_N &= \mathbb{E}_\tau[\tau]\left(\sum_{n=1}^{N}x_n+\lambda_0\mu_0\right) && (30)\\
\mu_N &= \frac{\mathbb{E}_\tau[\tau]\left(\sum_{n=1}^{N}x_n+\lambda_0\mu_0\right)}{\mathbb{E}_\tau[\tau](N+\lambda_0)} && (31)\\
\mu_N &= \frac{\sum_{n=1}^{N}x_n+\lambda_0\mu_0}{N+\lambda_0} && (32)\\
\mu_N &= \frac{N\bar{x}+\lambda_0\mu_0}{N+\lambda_0} && (33)
\end{aligned}$$

The other variational distribution $q_\tau(\tau)$ can be determined analogously.[a]

[a] If there is a term that depends linearly on the random variable and one that depends on the random variable via a logarithm, it follows automatically that $q_\tau(\tau)$ has to be a gamma distribution.