Note on variational infernece

June 26, 2023

In the question of Bayesian learning of a linear relation between the input $\mathbf{x}$ and output y, we assume the Gaussian likelihood,

$$p(y|\mathbf{x}, \mathbf{w}, \tau) = \mathcal{N}(y|\mathbf{w} \cdot \mathbf{x}, 1/\tau) \ , \tag{1}$$

where the variance being $1/\tau$. The prior over the parameters is

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|0, (\tau\alpha)^{-1}I) \ , \tag{2}$$

the prior over the noise parameter,

$$p(\tau) = G(\tau|a_0, b_0) \ , \tag{3}$$

and the prior over the sacle,

$$p(\alpha) = G(\alpha|c_0, d_0) \tag{4}$$

In the one-level variational Bayesian (VB) inference, one often needs to compute the following expectation,

$$\begin{aligned} -\int dx \mathcal{N}(x|\mu_1, \sigma_1^2) \log \mathcal{N}(x|\mu_2, \sigma_2^2) &= \mathbb{E}_{x\sim\mathcal{N}(\mu_1,\sigma_1^2)}\big[\frac{(x-\mu_1+\mu_1-\mu_2)^2}{2\sigma_2^2}\big] + \frac{1}{2}\log(2\pi\sigma_2^2) \\ &= \frac{\sigma_1^2 + (\mu_1-\mu_2)^2}{2\sigma_2^2} + \frac{1}{2}\log(2\pi\sigma_2^2) \ , \end{aligned} \tag{5}$$

which appears in the KL divergence. In the two-level VB, then one needs to compute the following integral,

$$\begin{aligned} & -\int dx d\tau G(\tau|a,b) \mathcal{N}(x|\mu_1, \sigma_1^2) \log \mathcal{N}(x|\mu_2, \tau^{-1}\sigma_2^2) \\ & = \mathbb{E}_{\tau\sim G(a,b)}\big[\tau\frac{\sigma_1^2 + (\mu_1-\mu_2)^2}{2\sigma_2^2} - \frac{1}{2}\log\tau\big] + \frac{1}{2}\log(2\pi\sigma_2^2) \\ & = \frac{1}{2}\left[\frac{a}{b}\frac{\sigma_1^2 + (\mu_1-\mu_2)^2}{\sigma_2^2} - \psi(a) + \log b + \log(2\pi\sigma_2^2)\right] . \end{aligned} \tag{6}$$

As for the Gamma distribution,

$$\int d\tau G(\tau|a,b)\log G(\tau|a,b) = -\log\Gamma(a) + (a-1)\psi(a) + \log b - a \tag{7}$$

In the variational inference, we have assumed the form of the variational distribution over the weight parameters $\mathbf{w}$ and scale parameter τ,

$$Q(\mathbf{w},\tau) = \mathcal{N}(\mathbf{w}|\mathbf{w}_{BV},\tau^{-1}\mathbf{V})G(\tau|a,b)\,, \tag{8}$$

with the variational parameters $\mathbf{w}_{VB}$, the diagonal covariance matrix $\mathbf{V}$ along with the parmaeters a and b in the Gamma distribution. The distribution for α is given by,

$$Q(\alpha) = G(\alpha|c,d)\,, \tag{9}$$

with variational parameters c and d. It can be shown that

$$\begin{aligned}\mathbb{E}_Q[\log Q(\mathbf{w},\tau)] = &-\frac{D}{2}\big[\frac{a}{b} - \psi(a) + \log b + \log|\mathbf{V}|\big]\\ &-\log\Gamma(a) + (a-1)\psi(a) + \log b - a\,,\end{aligned} \tag{10}$$

and

$$\mathbb{E}_Q[Q(\alpha|c,d)] = -\log\Gamma(c) + (c-1)\psi(c) + \log d - c \tag{11}$$

```python
import numpy as np

def incmatrix(genl1,genl2):
    m = len(genl1)
    n = len(genl2)
    M = None #to become the incidence matrix
    VT = np.zeros((n*m,1), int)  #dummy variable
```

Listing 1: Python example