Variational Inference

Machine Learning - Prof. Dr. Stephan Günnemann

Leonardo Freiherr von Lerchenfeld

January 25, 2018

Contents

1 KL divergence

1.1 Problem 1

$$\begin{aligned}
p(z) &= \mathcal{N}(z|\mu_1, \Sigma_1) && (1)\\
q(z) &= \mathcal{N}(z|\mu_2, \Sigma_2) && (2)\\
\mathbb{KL}(p||q) &= \int p(z) log \frac{p(z)}{q(z)} dz && (3)\\
&= \int p(z) log\ p(z) dz - \int p(z) log\ q(z) dz && (4)\\
&= H(p) - H(p,q) && (5)
\end{aligned}$$

We see that the first term is the entropy of the Gaussian and find in a smart book (e.g., Bishop p. 688) that the entropy of a Gaussian is

$$H(p) = \frac{1}{2} ln\ \sigma^2 + \frac{1}{2}(1 + ln(2\pi)) \tag{6}$$

Now let's calculate the cross-entropy of two Gaussians

$$\begin{aligned}
H(p,q) &= \int p(z) log \left((2\pi\sigma_2^2)^{-\frac{1}{2}} exp\left(-\frac{(z-\mu_2)^2}{2\sigma_2^2} \right) \right) dz && (7)\\
&= -\frac{1}{2} log(2\pi\sigma_2^2) + \int p(z) \left(-\frac{(z-\mu_2)^2}{2\sigma_2^2} \right) dz && (8)\\
&= -\frac{1}{2} log(2\pi\sigma_2^2) - \frac{\int p(z) z^2 dz - \int p(z) 2z\mu_2 dz + \int p(z)\mu_2^2 dz}{2\sigma_2^2} && (9)\\
&= -\frac{1}{2} log(2\pi\sigma_2^2) - \frac{\mathbb{E}(z^2) - 2\mathbb{E}(z)\mu_2 + \mu_2^2}{2\sigma_2^2} && (10)
\end{aligned}$$

We know from probability theory that

$$Var(z) = \mathbb{E}[z^2] - \mathbb{E}[z]^2 \tag{11}$$

Thus,

$$\begin{aligned}
\mathbb{E}[z^2] &= Var(z) + \mathbb{E}[z]^2 && (12)\\
\mathbb{E}[z^2] &= \sigma^2 + \mu^2 && (13)
\end{aligned}$$

$$\begin{aligned}
H(p,q) &= -\frac{1}{2} log(2\pi\sigma_2^2) - \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} && (14)\\
\mathbb{KL}(p||q) &= H(p) - H(p,q) && (15)\\
&= \frac{1}{2}(1 + log(2\pi\sigma_1^2)) + \frac{1}{2} log(2\pi\sigma_2^2) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} && (16)\\
&= \frac{1}{2} + \frac{1}{2} log(2\pi(\sigma_1^2 + \sigma_2^2)) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} && (17)
\end{aligned}$$

1.2 Problem 2

$$
\begin{aligned}
\mu^* &= \underset{\mu}{\operatorname{argmin}} \mathbb{KL}(p||q) && (18)\\
&= \underset{\mu}{\operatorname{argmin}} \left(H(p,q) - H(p)\right) && (19)\\
&= \underset{\mu}{\operatorname{argmin}} \left(H(p,q)\right) && (20)\\
&= \underset{\mu}{\operatorname{argmin}} \int p(x) log\ q(x) dx && (21)\\
H(p,q) &= \int p(x) log \left(\frac{1}{2\pi I} exp\left(-\frac{(x-\mu)^2}{2I}\right)\right) dx && (22)\\
&= \int p(x)\ log\ \left((2\pi)^{-\frac{1}{2}}\right) dx - \int p(x) \frac{1}{2}(x-\mu)^2 dx && (23)\\
&= -\frac{1}{2}\left(log(2\pi) + \int p(x)x^2 dx - \int p(x) 2\mu x dx + \int p(x)\mu^2 dx\right) && (24)\\
H(p,q) &= -\mu \mathbb{E}_p(x) + \frac{1}{2}\mu^2 + const. && (25)\\
\frac{\partial H(p,q)}{\partial \mu} &= -\mathbb{E}_p(x) + \mu && (26)\\
&= 0 && (27)\\
\mu^* &= \mathbb{E}_p(x) && (28)
\end{aligned}
$$

2 Mean-field variational inference

2.1 Problem 3

$$
\begin{aligned}
p(z|x) &\propto p(x|z)p(z) && (29)\\
&= \mathcal{N}(x|\theta^T z, 1)\mathcal{N}(z_1|0,1)\mathcal{N}(z_2|0,1) && (30)\\
&= \frac{1}{\sqrt{2\pi 1}}\left(exp\left(-\frac{(x-\theta^T z)^2}{2}\right) exp\left(-\frac{(z_1-0)^2}{2}\right) exp\left(-\frac{(z_1-0)^2}{2}\right)\right) && (31)\\
&= \frac{1}{\sqrt{2\pi}}\left(exp\left(-\frac{(x-\theta_1 z_1 + \theta_2 z_2)^2 + z_1^2 + z_2^2}{2}\right)\right) && (32)
\end{aligned}
$$

No, the posterior cannot be factorized over z_1 and z_2, because there is a coupling term.

2.2 Problem 4