Machine learning Homework- Variational Inference

Abinav Ravi Venkatakrishnan - 03694216

February 5, 2019

1 KL Divergence

Problem 1:

Given Two gaussians $p(x)$ and $q(x)$ with means μ_1 and μ_2 and covariances $\sum_1$ and $\sum_2$. We know that for gaussians with diagonal covariances the gaussian function factorizes into product of gaussians which implies that

$$p(x) = \prod_j p_j(x_j) \tag{1}$$

also we know that the formula of KL divergence can be written as sum of individual divergence because of the above property.hence

$$KL(p||q) = \int p(x_j)log(\frac{p(x_j)}{q(x_j)}) \tag{2}$$
$$= \mathbb{E}_p(\log p(x)) - \mathbb{E}_p(\log q(x)) \tag{3}$$

Now $\mathbb{E}_p(\log p(x))$ is the entropy of the univariate gaussian which is given by $-\frac{1}{2}\ln(2\pi\sigma^2 \exp)$ also can be written as $-\frac{1}{2}\log(2\pi\sigma_1^2) - \frac{1}{2}$

$\mathbb{E}_p(\log q(x))$ is given by

$$\int p_j \log(q_j(x)) = \mathbb{E}_{pj}(-logq_j(x)) \tag{4}$$
$$= \mathbb{E}_{pj}[\frac{1}{2}log(2\pi\sigma_2 + \frac{(x-\mu_2)^2}{2\sigma_2^2})] \tag{5}$$
$$= \frac{1}{2}\log(2\pi\sigma_2^2) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} \tag{6}$$

adding both we get

$$KL(p||q) = -\frac{1}{2} - \frac{1}{2}\log(2\pi\sigma_1^2) + \frac{1}{2}\log(2\pi\sigma_2^2) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} \tag{7}$$
$$= -\frac{1}{2} + \log(\frac{\sigma_2}{\sigma_1}) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} \tag{8}$$

Problem 2:

The KL divergence is written as

$$KL(p||q) = -\int p(x)\log q(x)dx + \int p(x)logp(x)dx \tag{9}$$

we know that gaussian q(x) can be written as with Identity covariance matrix as

$$\log q(x) = -\frac{D}{2}\log 2\pi - \frac{1}{2}|I| - \frac{1}{2}(x-\mu)^T I(x-\mu) \tag{10}$$
$$-\int p(x)\log q(x)dx = -\int p(x) - \frac{D}{2}\log 2\pi - \frac{1}{2}|I| - \frac{1}{2}(x-\mu)^T I(x-\mu) \tag{11}$$
$$\tag{12}$$

whereas since the second term is not dependent on $q(x)$ it can be considered as a constant.Considering constants the equation becomes

$$KL(p||q) = \frac{1}{2}(\mathbb{E}_p[x] - \mu)^T(\mathbb{E}_p[x] - \mu) + const \tag{13}$$

computing gradient for above equation and after some calculations we get

$$\nabla_\mu KL(p||q) = \mu - \mathbb{E}_p[x] \tag{14}$$

Optimal parameter is obtained when the gradient is set to zero so we get

$$\mu = \mathbb{E}_p[x] \tag{15}$$

2 Mean field variational inference:

Problem 3:

We need to find the posterior$p(z|x)$ we know that posterior is proportional to the likelihood. so
$p(z|x) \propto p(x|z)$
we also know that from given gaussian distribution

$$\mathcal{N}(x|\theta^T Z, 1) \tag{16}$$

$$p(z|x) = p(z1)p(z2)p(x|z) \tag{17}$$

$$p(z|x) = exp(-\frac{1}{2}(z_1^2 + z_2^2 + (x - \theta_1 z1 - \theta_2 z2)^2)) \tag{18}$$

upon expanding we get a $2\theta_1\theta_2 z_1 z_2$ term due to which we cannot factorize.

Problem 4:

As seen in the previous problem we cannot factorize the true posterior. So $q(z)$ is not able to match the true posterior $p(z|x)$