Variational Inference 02 Algorithm

Chen Gong

30 November 2019

我们将 X：Observed data；Z：Latent Variable + Parameters。那么 (X,Z) 为 complete data。根据我们的贝叶斯分布公式：

$$p(X)=\frac{p(X,Z)}{p(Z|X)} \tag{1}$$

在两边同时取对数并且引入函数 $q(Z)$ 我们可以得到：

$$\begin{aligned}\log p(X) &= \log\frac{p(X,Z)}{p(Z|X)} \\ &= \log p(X,Z)-\log p(Z|X) \\ &= \log\frac{p(X,Z)}{q(Z)}-\log\frac{p(Z|X)}{q(Z)}\end{aligned} \tag{2}$$

1 公式化简

左边 $= p(X) = \int_Z log\ p(X)q(Z)dZ$。

右边 $=$

$$\int_Z q(Z)\log\ \frac{p(X,Z)}{q(Z)}dZ-\int_Z q(Z)\log\ \frac{p(Z|X)}{q(Z)}dZ \tag{3}$$

其中，$\int_Z q(Z)\log\ \frac{p(X,Z)}{q(Z)}dZ$ 被称为 Evidence Lower Bound (ELBO)，被我们记为 $\mathcal{L}(q)$，也就是变分。

$-\int_Z q(Z)\log\ \frac{p(Z|X)}{q(Z)}dZ$ 被称为 $KL(q||p)$。这里的 $KL(q||p)\geq 0$。

由于我们求不出 $p(Z|X)$,我们的目的是寻找一个 $q(Z)$,使得 $p(Z|X)$ 近似于 $q(Z)$,也就是 $KL(q||p)$ 越小越好。并且，$p(X)$ 是个定值，那么我们的目标变成了 $\arg\max_{q(z)}\mathcal{L}(q)$。那么，我们理一下思路，我们想要求得一个 $\widetilde{q}(Z)\approx p(Z|X)$。也就是

$$\widetilde{q}(Z)=\arg\max_{q(z)}\mathcal{L}(q)\Rightarrow\widetilde{q}(Z)\approx p(Z|X) \tag{4}$$

2 模型求解

那么我们如何来求解这个问题呢?我们使用到统计物理中的一种方法,就是平均场理论 (mean field theory)。也就是假设变分后验分式是一种完全可分解的分布：

$$q(z)=\prod_{i=1}^{M}q_i(z_i) \tag{5}$$

在这种分解的思想中，我们每次只考虑第 j 个分布，那么令 $q_i(1,2,\cdots,j-1,j+1,\cdots,M)$ 个分布 fixed。

那么很显然：

$$\mathcal{L}(q) = \int_Z q(Z)\log p(X,Z)dz - \int_Z q(Z)\log q(Z)dZ \tag{6}$$

我们先来分析第一项 $\int_Z q(Z)\log p(X,Z)dZ$。

$$\begin{aligned}
\int_Z q(Z)\log p(X,Z)dZ &= \int_Z \prod_{i=1}^M q_i(z_i)\log p(X,Z)dZ \\
&= \int_{z_j} q_j(z_j)\left[\int_{z_1}\int_{z_2}\cdots\int_{z_M}\prod_{i=1}^M q_i(z_i)\log p(X,Z)dz_1dz_2\cdots dz_M\right]dz_j \\
&= \int_{z_j} q_j(z_j)\mathbb{E}_{\prod_{i\neq j}^M q_i(x_i)}\left[\log p(X,Z)\right]dz_j
\end{aligned} \tag{7}$$

然后我们来分析第二项 $\int_Z q(Z)\log q(Z)dZ$，

$$\begin{aligned}
\int_Z q(Z)\log q(Z)dZ &= \int_Z \prod_{i=1}^M q_i(z_i)\sum_{i=1}^M \log q_i(z_i)dZ \\
&= \int_Z \prod_{i=1}^M q_i(z_i)\left[\log q_1(z_1) + q_2(z_2) + \cdots + q_M(z_M)\right]dZ
\end{aligned} \tag{8}$$

这个公式的计算如何进行呢？我们抽出一项来看，就会变得非常的清晰：

$$\begin{aligned}
\int_Z \prod_{i=1}^M q_i(z_i)\log q_1(z_1)dZ &= \int_{z_1z_2\cdots z_M} q_1q_2\cdots q_M\log q_1dz_1dz_2\cdots z_M \\
&= \int_{z_1} q_1\log q_1dz_1\cdot\int_{z_2} q_2dz_2\cdot\int_{z_3} q_3dz_3\cdots\int_{z_M} q_Mdz_M \\
&= \int_{z_1} q_1\log q_1dz_1
\end{aligned} \tag{9}$$

因为，$\int_{z_2} q_2dz_2$ 每一项的值都是 1。所以第二项可以写为：

$$\sum_{i=1}^M \int_{z_i} q_i(z_i)\log q_i(z_i)dz_i = \int_{z_j} q_j(z_j)\log q_i(z_i)dz_j + C \tag{10}$$

因为我们仅仅只关注第 j 项，其他的项都不关注。为了进一步表达计算，我们将：

$$\mathbb{E}_{\prod_{i\neq j}^M q_i(z_i)}\left[\log p(X,Z)\right] = \log\hat{p}(X,z_j) \tag{11}$$

那么 (8) 式可以写作：

$$\int_{z_j} q_j(z_j)\log\hat{p}(X,z_j)dz_j \tag{12}$$

这里的 $\hat{p}(X,z_j)$ 表示为一个相关的函数形式，假设具体参数未知。那么 (7) 式将等于 (13) 式减 (11) 式：

$$\int_{z_j} q_j(z_j)\log q_i(z_i)dz_j - \int_{z_j} q_j(z_j)\log\hat{p}(X,z_j)dz_j + C = -KL(q_j||\hat{p}(x,z_j)) + C \tag{13}$$

$\arg\max_{q_j(z_j)} -KL(q_j||\hat{p}(x,z_j))$ 等价于 $\arg\min_{q_j(z_j)} KL(q_j||\hat{p}(x,z_j))$。那么这个 $KL(q_j||\hat{p}(x,z_j))$ 要如何进行优化呢？我们下一节将回归 EM 算法，并给出求解的过程。