vEM

基于Jensen不等式的EM

对于似然函数$l(x;\theta)$，有

$$
\begin{aligned}
l(x;\theta) &= \log p(x;\theta) \\
&= \log \int_Z p(x,z;\theta)dz \\
&= \log \int_Z q(z)\frac{p(x,z;\theta)}{q(z)}dz \\
&= \log E[\frac{p(x,z;\theta)}{q(z)}]
\end{aligned}
$$

由Jensen不等式，对于任何凸函数，有 $E[f(x)] \geq f(E[x])$ 当且仅当$x = E(x)$时取等，因$\log(x)$ 为凹函数，有

$$
\begin{aligned}
l(x;\theta) = \log p(x;\theta) = \log E[\frac{p(x,z;\theta)}{q(z)}] \\
\geq E[\log\left(\frac{p(x,z;\theta)}{q(z)}\right)]
\end{aligned}
$$

取等条件为

$$
\begin{aligned}
\frac{p(x,z;\theta)}{q(z)} = c \\
q(z;\theta) = c \cdot p(x,z;\theta)
\end{aligned}
$$

因$\sum q(z;\theta) = 1$，有

$$
\begin{aligned}
q(z;\theta) = \frac{p(x,z;\theta)}{\sum_z p(x,z;\theta)} = \frac{p(x,z;\theta)}{p(x;\theta)} \\
= p(z|x;\theta)
\end{aligned}
$$

代入原方程，有

$$
\begin{aligned}
l(x;\theta) == \log E[\frac{p(x,z;\theta)}{q(z)}] \\
\geq E[\log\left(\frac{p(x,z;\theta)}{q(z)}\right)]
\end{aligned}
$$

$$
= \int_z p(z|x;\theta)\log\left(\frac{p(x,z;\theta)}{p(z|x;\theta)}\right) = Q(\theta;\theta^{old})
$$

因此，从某种意义上讲EM算法先确定似然函数下界，然后通过最大似然法不断提高似然函数下界。

ELBO

刚才的推导可以看出EM算法通过构建辅助函数$q(z;\theta) = p(z|x;\theta)$来确定函数下界，该下界也被称为Evidence Lower Bond，我们定义其为 $ELBO = \int_z q(z)\log\left(\frac{p(x,z;\theta)}{q(z)}\right)$ 并且定义

$$
\begin{aligned}
\log p(x;\theta) - \int_z q(z)\log\left(\frac{p(x,z;\theta)}{q(z)}\right) &= \int_z q(z)\log p(x;\theta)dz - \int_z q(z)\log\left(\frac{p(x,z;\theta}{q(z)}\right. \\
&= \int_z q(z)\log\left(\frac{q(z;\theta)p(x;\theta)}{p(x,z;\theta)}\right) \\
&= -\int_z q(z)\log\left(\frac{p(z|x;\theta)}{q(z;\theta)}\right) \\
&= KL(q||p)
\end{aligned}
$$

为分布q与p之间的Kullback-Leibler散度，用来衡量两个分布之间的偏差

变分期望最大化

$p(z|x;\theta)$无法求得，我们可以通过构建一个其他分布用来逐步逼近的方式间接求解；并用KL散度来衡量二者的误差，同时通过ELBO最大化完成参数的估计。

因此我们构建辅助函数

$$q_\psi(\mathbf{Z}) \approx p_\theta(\mathbf{Z}|\mathbf{Y})$$

计算其与目标分布的KL散度来逼近目标分布

$$KL\left(q(\mathbf{Z}), p(\mathbf{Z}|\mathbf{Y})\right) = \mathbb{E}_q\left[\log \frac{q(z)}{p(z)}\right] = \int_{\mathcal{Z}} q(z) \log \frac{q(z)}{p(z)} \mathrm{d}z.$$

并通过最大化ElBO完成参数估计

$$J(\theta, \psi) = \log p_\theta(\mathbf{Y}) - KL[q_\psi(\mathbf{Z})||p_\theta(\mathbf{Z}|\mathbf{Y})] = \mathbb{E}_q[\log p_\theta(\mathbf{Y}, \mathbf{Z})] + \mathcal{H}[q_\psi(\mathbf{Z})]$$

vEM迭代

M步

更新参数$\theta = \{\alpha, \pi\}$

$$\theta^h = \arg\max J(\theta, \psi^h) = \arg\max_\theta \mathbb{E}_{q_{\psi^h}}[\log p_\theta(\mathbf{Y}, \mathbf{Z})]$$

VE步

更新辅助函数$q_\psi(\mathbf{Z})$

$$\psi^h = \arg\max J(\theta^h, \psi) = \arg\min_\psi KL[q_\psi(\mathbf{Z}) \,||\, p_{\theta^h}(\mathbf{Z} \mid \mathbf{Y})]$$