EM 证明

2014 年 12 月 17 日

参数为θ，样本x_i的概率为$p(x_i;\theta)$，隐变量为z,则最大似然为：

$$
\begin{aligned}
\ell(\theta) &= \sum_{i=1}^{n} log(p(x_i;\theta)) \\
&= \sum_{i=1}^{n} log(\sum_{z} p(x_i, z;\theta))
\end{aligned}
$$

令$\Theta(z)$为隐变量z的分布，则：

$$
\begin{aligned}
\ell(\theta) &= \sum_{i=1}^{n} log(\sum_{z} p(x_i, z;\theta)) \\
&= \sum_{i=1}^{n} log(\sum_{z} (\Theta(z) * \frac{p(x_i, z;\theta)}{\Theta(z)})) \\
&\geq \sum_{i=1}^{n} \sum_{z} \Theta(z) * log(\frac{p(x_i, z;\theta)}{\Theta(z)})
\end{aligned}
$$

最后一步为Jesen不等式。有公式可以看到，最大似然$\ell(\theta)$的下界为$\frac{p(x_i,z;\theta)}{\Theta(z)}$的期望。

我们可以选择一个合适的隐状态的分布$\Theta(z)$使得不等式的等号成立，根据Jesen不等式成立的条件，对每个不同的i，log内部的$\frac{p(x_i,z;\theta)}{\Theta(z)}$为常数$C_i$,即$\frac{p(x_i,z;\theta)}{\Theta(z)}$不随$z$的变化而变化，得：$\frac{p(x_i,z;\theta)}{\Theta(z)} = C_i$ 而又有$\sum_z \Theta(z) = 1$，从而：

$$
\sum_{z} p(x_i, z;\theta) = \sum_{z} \Theta(z) * C_i = C_i = p(x_i;\theta)
$$

故而：

$$
\begin{aligned}
\Theta(z) &= \frac{p(x_i, z;\theta)}{p(x_i;\theta)} \\
&= p(z|x_i;\theta)
\end{aligned}
$$

即，当隐状态z的分布$\Theta(z)$取最大似然估计的结果$p(z|x_i;\theta)$时，不等式的等号成立：样本x的最大似然概率等于$\frac{p(x_i,z;\theta)}{\Theta(z)}$的期望。

目前为止，给定了样本x和参数θ，我们可以用最大似然来获得隐状态分布$\Theta(z)$，而给定了样本x和隐状态的分布$\Theta(z)$，我们又可以通过最大化样本的似然概率来获得参数θ。故而，EM训练的过程为:

E步：选择一个合适的隐状态的分布$\Theta(z)$，使得样本x的最大似然概率等于Jesen不等式得到的下界，即$\frac{p(x_i,z;\theta)}{\Theta(z)}$的期望。根据Jesen不等式成立的条件，给定样本$x$和参数$\theta$的情况下，隐状态的分布为：

$$
\Theta(z) = p(z|x_i;\theta)
$$

M步:最大化Jesen不等式得到的下界，来求得参数θ，即：

$$\theta = argmax\sum_{i=1}^{n}\sum_{z}\Theta(z)*log(\frac{p(x_i,z;\theta)}{\Theta(z)})$$

Jessen不等式理解：令$\frac{p(x_i,z;\theta)}{\Theta(z)} = x_z$,则有：

$$\begin{aligned}
log(\sum_{z}(\Theta(z)*\frac{p(x_i,z;\theta)}{\Theta(z)})) &= log(\sum_{z}(\Theta(z)*x_z)) \\
&= log(\Theta_0 x_0 + \Theta_1 x_1 + \Theta_2 x_2 + ... + \Theta_n x_n) \\
&\geq \Theta_0 log(x_0) + \Theta_1 log(x_1) + \Theta_2 log(x_2) + ... + \Theta_n log(x_n) \\
&= \sum_{z}\Theta(z)*log(x_z) \\
&= \sum_{z}(\Theta(z)*log(\frac{p(x_i,z;\theta)}{\Theta(z)}))
\end{aligned}$$

当$x_z = C$，且$sum_z\Theta(z) = 1$（概率分布的性质）时，

$$log(\sum_{z}(\Theta(z)*x_z)) = log(1*C) = log(C)$$

$$\sum_{z}\Theta(z)*log(x_z) = 1*log(C) = log(C)$$