EM算法

Y表示能观测到的随机变量，Z表示隐变量。Y和Z连在一起的称为完全数据（complete-data），仅仅只有Y称为不完全数据（incomplete-data）。$P(Y|\Theta)$称为不完全数据的概率分布，$P(Y,Z|\Theta)$称为完全数据的概率分布。需要估计的模型参数是Θ，因此$P(Y|\Theta)$称为不完全数据的似然函数，$P(Y,Z|\Theta)$称为完全数据的似然函数。

EM算法的导出

对数似然函数$L(\Theta)=logP(Y|\Theta)$

$$= log\sum_z(P(Y,Z|\Theta)) \text{（应用边缘分布求和）}$$

$$= log\sum_z(P(Z|\Theta)P(Y|Z,\Theta)) \text{（应用条件概率：此时为全概率公式）}$$

因为原始的问题包含一个隐变量Z，所以这个对数似然函数公式中包含对Z求和的计算。因为log函数中存在加法运算，因此想要通过求导来直接求解极大似然会很困难。EM算法不是采用直接求解，而是迭代的方法一步一步极大化$L(\Theta)$，近似求解参数Θ的。记第i次迭代之后的参数为$\Theta^{(i)}$，我们计算$L(\Theta)$与$L(\Theta^{(i)})$的差值：

$$L(\Theta)-L(\Theta^{(i)}) = logP(Y|\Theta) - logP(Y|\Theta^{(i)})$$

$$= log\sum_z(P(Z|\Theta)P(Y|Z,\Theta)) - logP(Y|\Theta^{(i)})$$

$$= log\sum_z\left\{P(Z|Y,\Theta^{(i)})\frac{P(Z|\Theta)P(Y|Z,\Theta)}{P(Z|Y,\Theta^{(i)})}\right\} - logP(Y|\Theta^{(i)})$$

上式中红色的部分是我们新补充的项，构造了一个新的表达式。因为$log(x)$是凹函数，且$\sum_z P(Z|Y,\Theta^{(i)})=1$。这样我们构造出的表达式就可以利用Jenson不等式：

$$L(\Theta)-L(\Theta^{(i)}) = log\sum_z\left\{P(Z|Y,\Theta^{(i)})\frac{P(Z|\Theta)P(Y|Z,\Theta)}{P(Z|Y,\Theta^{(i)})}\right\} - logP(Y|\Theta^{(i)})$$

$$\geq \sum_z\left\{P(Z|Y,\Theta^{(i)})log\frac{P(Z|\Theta)P(Y|Z,\Theta)}{P(Z|Y,\Theta^{(i)})}\right\} - logP(Y|\Theta^{(i)})$$

$$= \sum_z\left\{P(Z|Y,\Theta^{(i)})log\frac{P(Z|\Theta)P(Y|Z,\Theta)}{P(Z|Y,\Theta^{(i)})logP(Y|\Theta^{(i)})}\right\} \dots eq.(1)$$

令$B(\Theta,\Theta^{(i)}) = L(\Theta^{(i)}) + eq.(1)$

$$= L(\Theta^{(i)}) + \sum_z\left\{P(Z|Y,\Theta^{(i)})log\frac{P(Z|\Theta)P(Y|Z,\Theta)}{P(Z|Y,\Theta^{(i)})logP(Y|\Theta^{(i)})}\right\}$$

可以根据上面的不等式得到：$L(\Theta)\geq B(\Theta,\Theta^{(i)})$，且易证 $L(\Theta^{(i)})=B(\Theta^{(i)},\Theta^{(i)})$，因此我们找到了$L(\Theta)$的下确界。只要通过调整参数$\Theta$增大$B(\Theta,\Theta^{(i)})$我们就能同时使$L(\Theta)$增大。选择$\Theta^{(i+1)}=\arg\max_\Theta B(\Theta,\Theta^{(i)})$：

$$\Theta^{(i+1)} = \arg\max_\Theta B(\Theta,\Theta^{(i)})$$

$$= \arg\max_\Theta\left\{L(\Theta^{(i)} + \sum_z\left[P(Z|Y,\Theta^{(i)})log\frac{P(Z|\Theta)P(Y|Z,\Theta)}{P(Z|Y,\Theta^{(i)})logP(Y|\Theta^{(i)})}\right]\right\}$$

$$= \arg\max_\Theta \sum_z P(Z|Y,\Theta^{(i)})\left\{log\left[P(Z|\Theta)P(Y|Z,\Theta)\right] - log\left[P(Z|Y,\Theta^{(i)})logP(Y|\Theta^{(i)})\right]\right\}$$

$$= \arg\max_\Theta\left\{\sum_z P(Z|Y,\Theta^{(i)})log\left[P(Z|\Theta)P(Y|Z,\Theta)\right]\right\}$$

$$= \arg\max_\Theta\left\{\sum_z P(Z|Y,\Theta^{(i)})logP(Y,Z|\Theta)\right\}$$

$$= \arg\max_\Theta Q(\Theta,\Theta^{(i)})$$

因此我们就得到了Q函数。

EM算法的单调性

定理9.1希望证明在EM算法的求解过程中，$P(Y|\Theta^{(i)})$是单调递增的。

备注：我个人觉得其实在EM算法的导出中，已经保证了$P(Y|\Theta^{(i)})$是单调递增的？

证明：

因为 $P(Y|\Theta)=\dfrac{P(Y,Z|\Theta)}{P(Z|Y,\Theta)}$，两边取对数可以得到$logP(Y|\Theta)=logP(Y,Z|\Theta)-logP(Z|Y,\Theta)$。

Q函数的定义为：$Q(\Theta,\Theta^{(i)})=\sum_z P(Z|Y,\Theta^{(i)})logP(Y,Z|\Theta)$

我们构造一个H函数：$H(\Theta,\Theta^{(i)})=\sum_z P(Z|Y,\Theta^{(i)})logP(Z|Y,\Theta)$

此时可以发现

$$
\begin{aligned}
Q(\Theta,\Theta^{(i)})-H(\Theta,\Theta^{(i)}) &= \sum_z P(Z|Y,\Theta^{(i)})\left[logP(Y,Z|\Theta)-logP(Z|Y,\Theta)\right] \\
&= logP(Y|\Theta)\sum_z P(Z|Y,\Theta^{(i)}) \\
&= logP(Y|\Theta)
\end{aligned}
$$

因为log函数严格单调递增，为了证明$P(Y|\Theta^{(i)})$是单调递增的，我们只需要证明$logP(Y|\Theta^{(i)})$是单调递增的：

$$
\begin{aligned}
&logP(Y|\Theta^{(i+1)})-logP(Y|\Theta^{(i)}) \\
&=[Q(\Theta^{(i+1)},\Theta^{(i)})-H(\Theta^{(i+1)},\Theta^{(i)})]-[Q(\Theta^{(i)},\Theta^{(i)})-H(\Theta^{(i)},\Theta^{(i)})] \\
&=[Q(\Theta^{(i+1)},\Theta^{(i)})-Q(\Theta^{(i)},\Theta^{(i)})]-[H(\Theta^{(i+1)},\Theta^{(i)})-H(\Theta^{(i)},\Theta^{(i)})]
\end{aligned}
$$

根据之前推导出Q函数的过程，我们可以知道$Q(\Theta^{(i+1)},\Theta^{(i)})-Q(\Theta^{(i)},\Theta^{(i)})\geq 0$

$$
\begin{aligned}
&H(\Theta^{(i+1)},\Theta^{(i)})-H(\Theta^{(i)},\Theta^{(i)}) \\
&=\sum_z P(Z|Y,\Theta^{(i)})[logP(Z|Y,\Theta^{(i+1)})-logP(Z|Y,\Theta^{(i)})] \\
&=\sum_z P(Z|Y,\Theta^{(i)})log\frac{P(Z|Y,\Theta^{(i+1)})}{P(Z|Y,\Theta^{(i)})} \\
&\leq log(\sum_z P(Z|Y,\Theta^{(i)})\frac{P(Z|Y,\Theta^{(i+1)})}{P(Z|Y,\Theta^{(i)})}) \quad (Jenson\ Inequality) \\
&=log(\sum_z P(Z|Y,\Theta^{(i+1)})) \\
&=0
\end{aligned}
$$

综上，可以得到$logP(Y|\Theta^{(i+1)})-logP(Y|\Theta^{(i)})\geq 0$