七、EM算法与最大熵

一、最大熵原理

最大熵的原理核心是去求条件熵的最大值

$$
\begin{aligned}
\max_{P\in \mathbf{C}} \quad & H(P) = -\sum_{x,y}\tilde{P}(x)P(y\,|\,x)\log P(y\,|\,x) \\
\text{s.t.} \quad & E_P(f_i) = E_{\tilde{P}}(f_i), \quad i=1,2,\cdots,n \\
& \sum_{y} P(y\,|\,x) = 1
\end{aligned}
$$

P~(X)是X的经验分布

二、EM算法（无监督学习算法）

EM算法的朴素引进

- 随机挑选10000位志愿者，测量他们的身高。
- 样本中存在男性和女性，身高分别服从N(μ1,σ1)和N(μ2,σ2)的分布。
- μ1,σ1,μ2,σ2未知
- 根据身高，判断其属于男性还是女性。

- 随机变量X是有K个高斯分布混合而成，取各个高斯分布的概率为$\pi_1 \pi_2 \ldots \pi_K$，第i个高斯分布的均值为μ_i，方差为Σ_i。若观测到随机变量X的一系列样本$x_1,x_2,\ldots,x_n$，试估计参数π，μ，Σ。

- 对数似然函数

$$l_{\pi,\mu,\Sigma}(x)=\sum_{i=1}^{N}\log\left(\sum_{k=1}^{K}\pi_k N(x_i \mid \mu_k,\Sigma_k)\right)$$

第一步：估算数据来自哪个组份

- 估计数据由每个组份生成的概率：对于每个样本x_i，它由第k个组份生成的概率为

$$\gamma(i,k)=\frac{\pi_k N(x_i \mid \mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j N(x_i \mid \mu_j,\Sigma_j)}$$

- 上式中的μ和Σ也是待估计的值，因此采样迭代法：在计算γ(i,k)时假定μ和Σ已知；
  - 需要先验给定μ和Σ。
  - γ(i,k)亦可看成组份k在生成数据x_i时所做的贡献。

第二步：估计每个组份的参数

- 对于所有的样本点，对于组份k而言，可看做生成了$\{\gamma(i,k)x_i \mid i=1,2,\cdots N\}$这些点。组份k是一个标准的高斯分布，利用上面的结论：

$$\begin{cases}\mu=\frac{1}{n}\sum_i x_i\\ \sigma^2=\frac{1}{n}\sum_i (x_i-\mu)^2\end{cases}$$

$$\begin{cases}N_k=\sum_{i=1}^{N}\gamma(i,k)\\ \mu_k=\frac{1}{N_k}\sum_{i=1}^{N}\gamma(i,k)x_i\\ \Sigma_k=\frac{1}{N_k}\sum_{i=1}^{N}\gamma(i,k)(x_i-\mu_k)(x_i-\mu_k)^T\\ \pi_k=\frac{N_k}{N}=\frac{1}{N}\sum_{i=1}^{N}\gamma(i,k)\end{cases}$$

三、经典EM算法

取对数似然函数

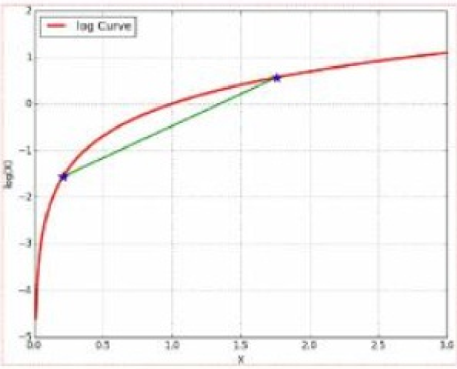
$$
\begin{aligned}
\ell(\theta) &= \sum_{i=1}^{m} \log p(x;\theta) \\
&= \sum_{i=1}^{m} \log \sum_{z} p(x,z;\theta)
\end{aligned}
$$

其中，Z是随机隐变量

Jensen不等式

令Qi是Z的某一个分布，Qi⩾0，有：

$$
\begin{aligned}
\sum_i \log p(x^{(i)};\theta) &= \sum_i \log \sum_{z^{(i)}} p(x^{(i)},z^{(i)};\theta) \\
&= \sum_i \log \sum_{z^{(i)}} Q_i(z^{(i)}) \frac{p(x^{(i)},z^{(i)};\theta)}{Q_i(z^{(i)})} \\
&\geq \sum_i \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)},z^{(i)};\theta)}{Q_i(z^{(i)})}
\end{aligned}
$$



$Q_i(z^{(i)}) \propto p(x^{(i)}, z^{(i)}; \theta)$ $\quad\quad$ $\sum_z Q_i(z^{(i)}) = 1$

$$\begin{aligned} Q_i(z^{(i)}) &= \frac{p(x^{(i)}, z^{(i)}; \theta)}{\sum_z p(x^{(i)}, z; \theta)} \\ &= \frac{p(x^{(i)}, z^{(i)}; \theta)}{p(x^{(i)}; \theta)} \\ &= p(z^{(i)} | x^{(i)}; \theta) \end{aligned}$$

EM算法整体框架

Repeat until convergence {

(E-step) For each i, set

$$Q_i(z^{(i)}) := p(z^{(i)} | x^{(i)}; \theta).$$

(M-step) Set

$$\theta := \arg\max_\theta \sum_i \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})}$$

}