期望-极大算法与高斯混合模型
Expection-Maximization Algorithm and Gaussian Mixture Model

西安交通大学管理学院
信息管理与电子商务系
智能决策与机器学习研究中心
刘佳鹏

EM算法

- EM 算法是一种迭代算法，用于含有隐变量(hidden variable)的概率模型参数的极大似然估计，或极大后验概率估计
- EM 算法的每次迭代由两步组成：E步，求期望(expectation)；M 步，求极大(maximization)
- 所以该算法称为期望极大算法(expectation maximization algorithm)，简称EM算法

EM算法

引例：（三硬币模型）假设有三枚硬币，分别记为A，B，C。这些硬币正面向上的概率分别是π，p和q。进行如下掷硬币试验：先掷硬币A，根据其结果选择硬币B或硬币C，正面选硬币B，反面选硬币C；然后掷选出的硬币，出现正面记为1，出现反面记为0；独立地重复n次试验（这里$n = 10$），观测结果如下：

$$1, 1, 0, 1, 0, 0, 1, 0, 1, 1$$

假设只能观测到掷硬币的结果，不能观测到掷硬币的过程。问如何估计三硬币正面出现的概率，即三硬币模型的参数

$$\theta = (\pi, p, q)$$

EM算法

- 引入随机变量$x \in \{0,1\}$表示一次试验观测的结果是1或者0，x是观测变量，可以观测
- 引入随机变量$z \in \{0,1\}$表示未观测到的掷硬币A的结果，z是隐变量，不可观测
- $\theta = (\pi, p, q)$是模型参数

三硬币模型可以写作

$$\begin{aligned} P(x \mid \theta) &= \sum_z P(x, z \mid \theta) = \sum_z P(z \mid \theta) P(x \mid z, \theta) \\ &= \pi p^x (1-p)^{1-x} + (1-\pi) q^x (1-q)^{1-x} \end{aligned}$$

该模型是以上数据的生成模型

EM算法

将观测数据表示为$X=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，未观测数据表示为$Z=(z_1,z_2,\cdots,z_n)^{\mathrm{T}}$，观测数据的似然函数为

$$P(X\mid\theta)=\sum_{Z}P(Z\mid\theta)P(X\mid Z,\theta)$$

即

$$P(X\mid\theta)=\prod_{j=1}^{n}\left[\pi p^{x_j}(1-p)^{1-x_j}+(1-\pi)q^{x_j}(1-q)^{1-x_j}\right]$$

考虑求模型参数$\theta=(\pi,p,q)$的极大似然估计，即

$$\hat{\theta}=\arg\max_{\theta}\log P(X\mid\theta)$$

这个问题没有解析解，只有通过迭代的方法解决。EM算法就是可以用于求解这个问题的一种迭代算法。

EM算法的推导

- 问题描述：
- 观测变量x，隐变量z，参数θ
- 观测数据$X=(x_1,x_2,\ldots,x_n)$
- 未观测数据$Z=(z_1,z_2,\ldots,z_n)$

观测数据的似然函数

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} P\left(x_i \mid \theta\right) \\
&= \prod_{i=1}^{n}\left(\sum_{z_i} P\left(x_i, z_i \mid \theta\right)\right) \\
&= \prod_{i=1}^{n}\left(\sum_{z_i} P\left(x_i \mid z_i, \theta\right) P\left(z_i \mid \theta\right)\right)
\end{aligned}
$$

EM算法的推导

对数似然函数

$$\begin{aligned} LL(\theta) &= \sum_{i=1}^{n} \log P(x_i \mid \theta) \\ &= \sum_{i=1}^{n} \log \left(\sum_{z_i} P(x_i, z_i \mid \theta) \right) \end{aligned}$$

由于$\log(\cdot)$中出现了求和符号，使得极大化对数似然函数变得十分困难

EM算法的推导

引入隐变量z的某种分布，满足（1）$Q(z)>0$（2）$\sum_{z} Q(z)=1$

对数似然函数可以写为

$$
\begin{aligned}
LL(\theta) &= \sum_{i=1}^{n} \log \left(\sum_{z_i} P\left(x_i, z_i \mid \theta\right) \right) \\
&= \sum_{i=1}^{n} \log \left(\sum_{z_i} Q\left(z_i\right) \cdot \frac{P\left(x_i, z_i \mid \theta\right)}{Q\left(z_i\right)} \right)
\end{aligned}
$$

EM算法的推导

Jensen不等式：如果$f(\cdot)$是凸函数，X是随机变量，那么$E[f(X)] \geq f(E(X))$。特别地，如果$f(\cdot)$是严格凸函数，当且仅当X是常量时，上式取等号。

EM算法的推导

$$LL(\theta)=\sum_{i=1}^{n}\log\left(\sum_{z_i}Q\left(z_i\right)\cdot\frac{P\left(x_i,z_i\mid\theta\right)}{Q\left(z_i\right)}\right)$$

将$\frac{P(x_i,z_i|\theta)}{Q(z_i)}$看成是随机变量，那么$\sum_{z_i}Q\left(z_i\right)\cdot\frac{P(x_i,z_i|\theta)}{Q(z_i)}$就是随机变量$\frac{P(x_i,z_i|\theta)}{Q(z_i)}$的期望。

又因为$f(x)=\log x$是严格凹函数，所以有

$$\log\left(\sum_{z_i}Q\left(z_i\right)\cdot\frac{P\left(x_i,z_i\mid\theta\right)}{Q\left(z_i\right)}\right)\geq\sum_{z_i}Q\left(z_i\right)\log\left(\frac{P\left(x_i,z_i\mid\theta\right)}{Q\left(z_i\right)}\right)$$

EM算法的推导

进一步有

$$
\begin{aligned}
LL(\theta) &= \sum_{i=1}^{n} \log\left(\sum_{z_i} Q(z_i) \cdot \frac{P(x_i, z_i \mid \theta)}{Q(z_i)}\right) \\
&\geq \sum_{i=1}^{n} \sum_{z_i} Q(z_i) \log\left(\frac{P(x_i, z_i \mid \theta)}{Q(z_i)}\right)
\end{aligned}
$$

将$J(\theta, Q(z)) = \sum_{i=1}^{n} \sum_{z_i} Q(z_i) \log\left(\frac{P(x_i, z_i|\theta)}{Q(z_i)}\right)$看成是对数似然函数$LL(\theta)$的下界，可以通过优化下界$J(\theta, Q(z))$来极大化对数似然函数$LL(\theta)$。注意$J(\theta, Q(z))$是关于参数$\theta$和隐变量$z$的分布$Q(z)$的函数。

EM算法的推导

$$
\begin{aligned}
LL(\theta) &= \sum_{i=1}^{n} \log \left(\sum_{z_i} Q\left(z_i\right) \cdot \frac{P\left(x_i, z_i \mid \theta\right)}{Q\left(z_i\right)} \right) \\
&\geq \sum_{i=1}^{n} \sum_{z_i} Q\left(z_i\right) \log \left(\frac{P\left(x_i, z_i \mid \theta\right)}{Q\left(z_i\right)} \right) \\
&= J(\theta, Q(z))
\end{aligned}
$$

通过Jensen不等式，将对数似然函数$LL(\theta)$中的“对和求对数”变成了“对对数求和”，这样求导就变得容易了。

EM算法的推导

- 如何通过优化下界$J(\theta, Q(z))$来实现极大化$LL(\theta)$呢？
- EM算法：
- 初始化参数值$\theta^{(1)}$，然后交替迭代以下两步骤直至收敛
- for $t = 1, 2, \ldots$
- E步：
$$Q^{(t)}(z) = \underset{Q(z)}{\arg\max}\, J\left(\theta^{(t)}, Q(z)\right)$$
- M步：
$$\theta^{(t+1)} = \underset{\theta}{\arg\max}\, J\left(\theta, Q^{(t)}(z)\right)$$

EM算法的推导

E步：

$$Q^{(t)}(z) = \arg\max_{Q(z)} J\left(\theta^{(t)}, Q(z)\right)$$

- 如何更新$Q(z)$使得$J(\theta^{(t)}, Q(z))$在$\theta^{(t)}$处实现极大化？
- 回忆：$J(\theta^{(t)}, Q(z))$是对数似然函数$LL(\theta^{(t)})$的下界，而且当且仅当随机变量$\frac{P\left(x_i,z_i|\theta^{(t)}\right)}{Q(z_i)}$是常量时，$LL(\theta^{(t)})$与$J(\theta^{(t)}, Q(z))$相等
- 因此，当随机变量$\frac{P\left(x_i,z_i|\theta^{(t)}\right)}{Q(z_i)}$是常量时，$J(\theta^{(t)}, Q(z))$达到当前$\theta^{(t)}$处的最大值，且最大值为$LL(\theta^{(t)})$
- 当随机变量$\frac{P\left(x_i,z_i|\theta^{(t)}\right)}{Q(z_i)}$是常量时，有

$$\frac{P\left(x_i, z_i \mid \theta^{(t)}\right)}{Q(z_i)} = c$$

c为常数

EM算法的推导

$$Q(z_i) = \frac{P\left(x_i, z_i \mid \theta^{(t)}\right)}{c}$$

因为$\sum_{z_i} Q(z_i) = 1$，所以有

$$\sum_{z_i} \frac{P\left(x_i, z_i \mid \theta^{(t)}\right)}{c} = 1$$

即

$$\sum_{z_i} P\left(x_i, z_i \mid \theta^{(t)}\right) = c$$

进一步有

$$\begin{aligned} Q(z_i) = \frac{P\left(x_i, z_i \mid \theta^{(t)}\right)}{c} &= \frac{P\left(x_i, z_i \mid \theta^{(t)}\right)}{\sum_{z_i} P\left(x_i, z_i \mid \theta^{(t)}\right)} \\ &= \frac{P\left(x_i, z_i \mid \theta^{(t)}\right)}{P\left(x_i \mid \theta^{(t)}\right)} = P\left(z_i \mid x_i, \theta^{(t)}\right) \end{aligned}$$

EM算法的推导

$$Q\left(z_i\right)=P\left(z_i \mid x_i, \theta^{(t)}\right)$$

上式告诉我们，只要将$Q^{(t)}\left(z_i\right)$设定为当前参数$\theta^{(t)}$下隐变量z_i的后验分布$P\left(z_i \mid x_i, \theta^{(t)}\right)$，即可使得$J(\theta^{(t)}, Q(z))$在当前$\theta^{(t)}$处实现极大化

EM算法的推导

M步：

$$\theta^{(t+1)} = \arg\max_{\theta} J\left(\theta, Q^{(t)}(z)\right)$$

由于M步中$Q^{(t)}(z)$是常量，所以有

$$\begin{aligned}
\theta^{(t+1)} &= \arg\max_{\theta} J\left(\theta, Q^{(t)}(z)\right) \\
&= \arg\max_{\theta} \sum_{i=1}^{n} \sum_{z_i} Q^{(t)}(z_i) \log\left(\frac{P(x_i, z_i|\theta)}{Q^{(t)}(z)}\right) \\
&= \arg\max_{\theta} \sum_{i=1}^{n} \sum_{z_i} Q^{(t)}(z_i) \log P(x_i, z_i|\theta) \\
&= \arg\max_{\theta} \sum_{i=1}^{n} \sum_{z_i} Q^{(t)}(z_i) \log P(x_i, z_i|\theta) \\
&= \arg\max_{\theta} \sum_{i=1}^{n} \sum_{z_i} Q^{(t)}(z_i) \log\left(P(x_i|z_i, \theta) P(z_i|\theta)\right)
\end{aligned}$$

EM算法的推导

- EM算法：
- 初始化参数值 $\theta^{(1)}$，然后交替迭代以下两步骤直至收敛
- for $t = 1, 2, \ldots$
- E步：计算当前参数 $\theta^{(t)}$ 下隐变量 z_i 的后验分布

$$P\left(z_i \mid x_i, \theta^{(t)}\right), \qquad i = 1, \cdots, n$$

并计算完全数据的对数似然函数关于隐变量的后验分布的期望

$$\sum_{i=1}^{n} \sum_{z_i} P\left(z_i \mid x_i, \theta^{(t)}\right) \log P\left(x_i, z_i | \theta\right)$$

- M步：上述期望关于参数 θ 求极大，更新参数 θ

$$\theta^{(t+1)} = \arg\max_{\theta} \sum_{i=1}^{n} \sum_{z_i} P\left(z_i \mid x_i, \theta^{(t)}\right) \log P\left(x_i, z_i | \theta\right)$$

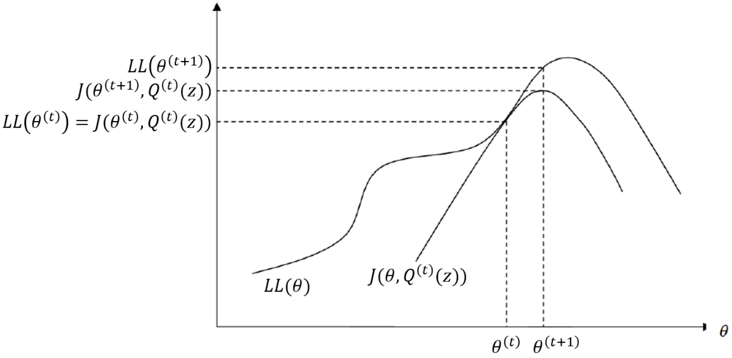
EM算法的推导

- EM算法的理论保证

$$
\begin{aligned}
LL\left(\theta^{(t+1)}\right) &\geq \sum_{i=1}^{n} \sum_{z_i} Q^{(t)}(z_i) \log \frac{P\left(x_i, z_i \mid \theta^{(t+1)}\right)}{Q^{(t)}(z_i)} \\
&\geq \sum_{i=1}^{n} \sum_{z_i} Q^{(t)}(z_i) \log \frac{P\left(x_i, z_i \mid \theta^{(t)}\right)}{Q^{(t)}(z_i)} \\
&= LL\left(\theta^{(t)}\right)
\end{aligned}
$$

说明：（1）Jensen不等式保证了第一个不等式成立；（2）M步保证了第二个不等式成立；（3）E步保证了最后的等式成立

EM算法的推导


$LL(\theta^{(t+1)})$
$J(\theta^{(t+1)}, Q^{(t)}(z))$
$LL(\theta^{(t)}) = J(\theta^{(t)}, Q^{(t)}(z))$
$LL(\theta)$
$J(\theta, Q^{(t)}(z))$
$\theta^{(t)}$
$\theta^{(t+1)}$
θ

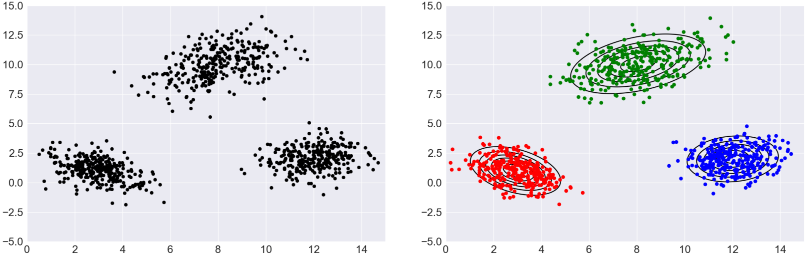
EM算法的推导

EM算法的注意事项

- EM算法可以看成是坐标上升法
- EM算法不能保证收敛到全局最优解。实际上，含有隐变量的优化问题一般是非凸问题，优化难度大
- EM算法是对初始值敏感的。为了保证求解效果，往往需要从不同初始值出发，然后对比不同解的质量

高斯混合模型

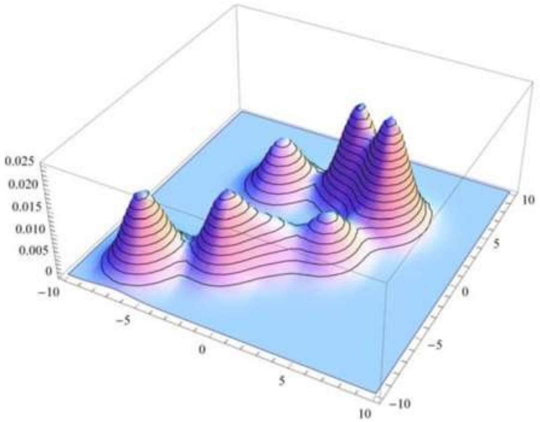
- 高斯混合模型(Gaussian mixture model, GMM)应用极其广泛，它可以看成是一种聚类算法，也可以用于概率密度估计
- EM算法是学习高斯混合模型的有效方法



高斯混合模型用于聚类[1]

[1] https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html#sphx-glr-auto-examples-cluster-plot-cluster-comparison-py

高斯混合模型



高斯混合模型用于概率密度估计[2]

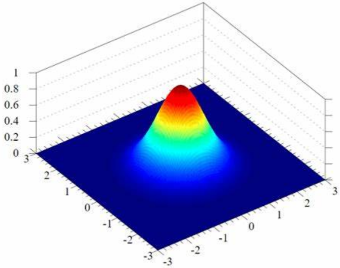
[2] https://scikit-learn.org/stable/auto_examples/mixture/plot_gmm_pdf.html#sphx-glr-auto-examples-mixture-plot-gmm-pdf-py

高斯混合模型

多元高斯（正态）分布

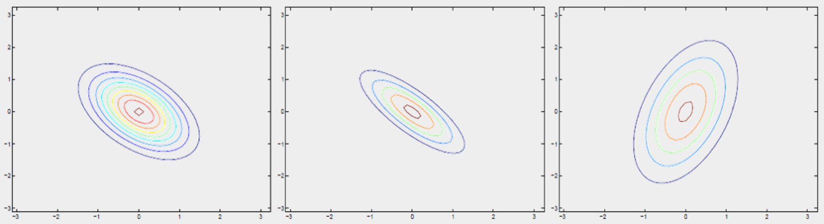
$$P(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

其中$\boldsymbol{x}$是m维随机变量，$\boldsymbol{\mu}$是m维均值向量，$\boldsymbol{\Sigma}$是$m \times m$的协方差矩阵



高斯混合模型

多元高斯（正态）分布



$$\Sigma = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 3 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

高斯混合模型

高斯混合模型是由一系列高斯分布组成的混合模型，其概率分布为

$$P(\mathbf{x}) = \sum_{k=1}^{K} \alpha_k P(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

其中α_k是系数，满足$\alpha_k \geq 0$，$\sum_{k=1}^{K} \alpha_k = 1$，$P(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$是第$k$个高斯混合成分的概率分布

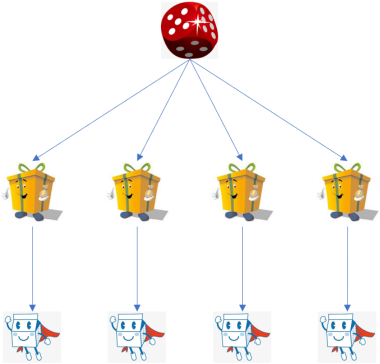
$$P(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)\right\}$$

其中$\boldsymbol{\mu}_k$和$\boldsymbol{\Sigma}_k$是第k个高斯混合成分的均值向量和协方差矩阵

高斯混合模型可以无限逼近任何连续的概率密度函数!

高斯混合模型

高斯混合模型是生成式模型!



（1）假设每个高斯混合成分k是一个生成器，它以参数μ_k和Σ_k产生样本
（2）通过投骰子决定由哪个高斯混合成分产生样本

高斯混合模型

- 在上述过程中，我们仅能观测到最终生成的样本$\boldsymbol{x}_i$，而无法观测到样本$\boldsymbol{x}_i$是由哪个生成器（高斯混合成分）产生的
- 用$z_i \in \{1, ..., K\}$表示生成样本$\boldsymbol{x}_i$的生成器（高斯混合成分）编号
- 因此，高斯混合模型是含有隐变量的生成式模型，可以用EM算法学习高斯混合模型
- 观测变量$\boldsymbol{x}_i$
- 隐变量z_i
- 参数$\theta = \{\alpha_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^{K}$
- 观测数据$X = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$
- 未观测数据$Z = \{z_1, ..., z_n\}$

高斯混合模型

单个样本的概率为

$$\begin{aligned} P(\boldsymbol{x}) &= \sum_{z \in \{1,\dots,K\}} P(\boldsymbol{x}, z) \\ &= \sum_{z \in \{1,\dots,K\}} P(\boldsymbol{x}|z) P(z) \\ &= \sum_{k=1}^{K} \alpha_k P(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \end{aligned}$$

高斯混合模型

观测数据$X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$的概率（似然函数）为

$$
\begin{aligned}
L(\theta) = P(X) &= \prod_{i=1}^{n} P(\mathbf{x}_i) \\
&= \prod_{i=1}^{n} \left(\sum_{z_i \in \{1,\dots,K\}} P(\mathbf{x}_i, z_i) \right) \\
&= \prod_{i=1}^{n} \left(\sum_{z_i \in \{1,\dots,K\}} P(\mathbf{x}_i | z_i) P(z_i) \right) \\
&= \prod_{i=1}^{n} \left(\sum_{k=1}^{K} \alpha_k P(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right)
\end{aligned}
$$

对数似然函数为

$$
LL(\theta) = \sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} \alpha_k P(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right)
$$

高斯混合模型

使用EM算法学习高斯混合模型

E步：计算隐变量z_i的分布$Q(z_i)$（即z_i的后验分布$P(z_i = k|\boldsymbol{x}_i)$）

$$
\begin{aligned}
Q(z_i) &= P(z_i = k|\boldsymbol{x}_i) \\
&= \frac{P(\boldsymbol{x}_i|z_i = k)\, P(z_i = k)}{P(\boldsymbol{x}_i)} \\
&= \frac{P(\boldsymbol{x}_i|z_i = k)\, P(z_i = k)}{\sum\limits_{k=1}^{K} P(\boldsymbol{x}_i|z_i = k)\, P(z_i = k)} \\
&= \frac{\alpha_k P(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum\limits_{k=1}^{K} \alpha_k P(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \\
&= \gamma_{ik}
\end{aligned}
$$

高斯混合模型

M步：关于参数θ求以下函数的极大化问题

$$J(\theta, Q(z)) = \sum_{i=1}^{n} \sum_{z_i} Q(z_i) \log \left(\frac{P(\boldsymbol{x}_i, z_i|\theta)}{Q(z_i)} \right)$$

即

$$\theta^* = \arg\max_{\theta} \sum_{i=1}^{n} \sum_{z_i} P(z_i|\boldsymbol{x}_i) \log P(\boldsymbol{x}_i, z_i|\theta)$$

高斯混合模型

$$
\begin{aligned}
\theta^* &= \arg\max_{\theta} \sum_{i=1}^{n}\sum_{k=1}^{K} \gamma_{ik} \log\left(P\left(\mathbf{x}_i | z_i = k, \theta\right) P\left(z_i = k|\theta\right)\right) \\
&= \arg\max_{\theta} \sum_{i=1}^{n}\sum_{k=1}^{K} \gamma_{ik} \left\{\log P\left(\mathbf{x}_i | z_i = k, \theta\right) + \log P\left(z_i = k|\theta\right)\right\} \\
&= \arg\max_{\theta} \sum_{i=1}^{n}\sum_{k=1}^{K} \gamma_{ik} \left\{\log P\left(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) + \log \alpha_k\right\} \\
&= \arg\max_{\theta} \sum_{i=1}^{n}\sum_{k=1}^{K} \gamma_{ik} \left\{\log\left\{\frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_k)^{\mathsf{T}}\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\right\}\right\} + \log \alpha_k\right\} \\
&= \arg\max_{\theta} \sum_{i=1}^{n}\sum_{k=1}^{K} \gamma_{ik} \left\{-\frac{1}{2}\log|\boldsymbol{\Sigma}_k| - \frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_k)^{\mathsf{T}}\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k) + \log \alpha_k\right\} \\
&= \arg\min_{\theta} \sum_{i=1}^{n}\sum_{k=1}^{K} \gamma_{ik} \left\{\frac{1}{2}\log|\boldsymbol{\Sigma}_k| + \frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_k)^{\mathsf{T}}\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k) - \log \alpha_k\right\}
\end{aligned}
$$

高斯混合模型

$$\boldsymbol{\mu}_k = \frac{\sum\limits_{i=1}^{n} \gamma_{ik} \boldsymbol{x}_i}{\sum\limits_{i=1}^{n} \gamma_{ik}}$$

$$\boldsymbol{\Sigma}_k = \frac{\sum\limits_{i=1}^{n} \gamma_{ik} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^{\mathrm{T}}}{\sum\limits_{i=1}^{n} \gamma_{ik}}$$

$$\alpha_k = \frac{1}{n} \sum_{i=1}^{n} \gamma_{ik}$$

高斯混合模型

高斯混合模型的EM算法

- 输入：样本集合$X = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$，混合成分个数$K$
- 过程：
- （1）初始化高斯混合模型的参数$\{(\alpha_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) | 1 \leqslant k \leqslant K\}$
- （2）交替迭代以下两步直至收敛：
- （2.1）E步：根据模型当前参数$\{(\alpha_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) | 1 \leqslant k \leqslant K\}$计算各样本$\boldsymbol{x}_i$属于各成分$k$的后验概率

$$\gamma_{ik} = P(z_i = k | \boldsymbol{x}_i) = \frac{\alpha_k P(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{k=1}^{K} \alpha_k P(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}$$

- （2.2）M步：根据以下公式更新模型参数

$$\boldsymbol{\mu}_k = \frac{\sum_{i=1}^{n} \gamma_{ik} \boldsymbol{x}_i}{\sum_{i=1}^{n} \gamma_{ik}} \quad \boldsymbol{\Sigma}_k = \frac{\sum_{i=1}^{n} \gamma_{ik} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^{\mathrm{T}}}{\sum_{i=1}^{n} \gamma_{ik}} \quad \alpha_k = \frac{1}{n} \sum_{i=1}^{n} \gamma_{ik}$$

高斯混合模型

高斯混合模型的EM算法

- （3）根据以下公式确定各样本 $\boldsymbol{x}_i,\ i=1,\cdots,n$ 的簇标记ρ_i

$$\rho_i = \underset{k\in\{1,\ldots,K\}}{\arg\max}\ \gamma_{ik}$$

- （4）分别确定各个簇 $C_k,\ k=1,\cdots,K,$ 中包含的样本

$$C_k = \{\boldsymbol{x}_i|\rho_i = k\}$$

- 输出：簇划分$\mathcal{C}=\{C_1,\ldots,C_K\}$

高斯混合模型

如何确定混合成分的个数？

- （1）根据任务要求预先指定混合成分的个数
- （2）AIC、BIC[3]

$$\text{AIC} = -LL(\theta) + \Phi$$

$$\text{BIC} = -LL(\theta) + \frac{1}{2}\Phi \log n$$

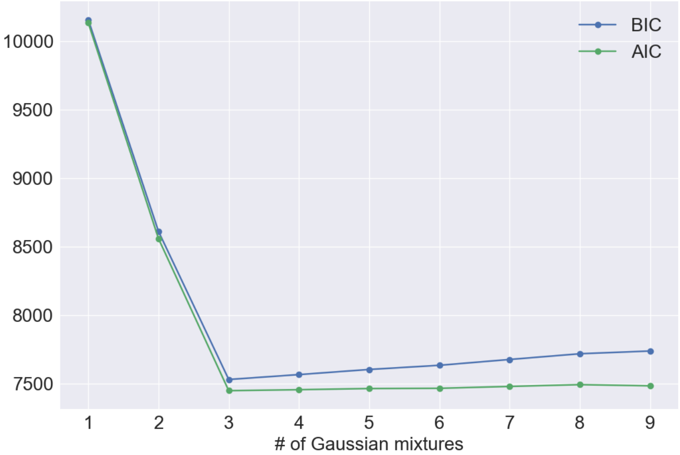
Φ是高斯混合模型中自由参数的个数

$$\Phi = Km + K\frac{(1+m)\,m}{2} + (K-1)$$

[3] https://scikit-learn.org/stable/auto_examples/mixture/plot_gmm_selection.html;https://zhuanlan.zhihu.com/p/81255623

高斯混合模型

- 使用AIC、BIC判断最优的混合成分个数



高斯混合模型

- （3）非参数模型：预先不指定混合成分的个数，使用Dirichlet Process作为先验[4]
  - Dirichlet Process 在计量模型中的应用: Li, Y., & Ansari, A. (2014). A Bayesian semiparametric approach for endogeneity and heterogeneity in choice models. Management Science, 60(5), 1161-1179.

[4] https://scikit-learn.org/stable/modules/mixture.html