假设训练集：$\{x^{(1)}, x^{(2)}, \cdots, x^{(m)}\}$，

通过联合分布建模数据：

$p(x^{(i)}, z^{(i)}) = p(x^{(i)} | z^{(i)}) \cdot p(z^{(i)})$，

其中，$z^{(i)} \sim$ 服从多项分布，以 ϕ 为参数，

且 $\sum_{j=1}^{k} \phi_j = 1$，$\phi_j > 0$，表示第 j 类出现的概率；

$p(z^{(i)} = j) = \phi_j$.

混合高斯模型 GMM 可以写为：

$$l(\phi, \mu, \Sigma) = \sum_{i=1}^{m} \log p(x^{(i)}; \phi, \mu, \Sigma)$$

$$= \sum_{i=1}^{m} \log \cdot \sum_{j=1}^{k} p(x^{(i)} | z^{(i)}; \mu, \Sigma) \cdot p(z^{(i)}; \phi).$$

这里，$x^{(i)} | z^{(i)} = j \sim N(\mu_j, \Sigma_j)$，表示 $x^{(i)}$ 服从第 j 个高斯分布的概率，$z^{(i)}$ 被称作隐变量

目标：求解 ϕ、μ、Σ 参数，使其满足给定数据 $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$ 下的最大似然。

直接求解（令导数为0）不可行。

由于 $z^{(i)}$ 表示每个 $x^{(i)}$ 来自于 k 个高斯中的哪个，假设已知 $z^{(i)}$，则问题其实很简单：

$$l(\phi, \mu, \Sigma) = \sum_{i=1}^{m} \log p(x^{(i)} | z^{(i)}; \mu, \Sigma) + \log p(z^{(i)}; \phi)$$

最大似然计算为：

$$\phi_j = \frac{1}{m} \sum_{i=1}^{m} \{z^{(i)} = j\}$$

$$\mu_j = \frac{\sum_{i=1}^{m} \{z^{(i)} = j\} \cdot x^{(i)}}{\sum_{i=1}^{m} \{z^{(i)} = j\}}$$

$$\Sigma_j = \frac{\sum_{i=1}^{m} \{z^{(i)} = j\} (x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^{m} \{z^{(i)} = j\}}$$

则EM算法为：（迭代以下两步直到收敛）

E步：猜测 $z^{(i)}$ 的值

对每个 i, j，令：

$$w_j^{(i)} = p(z^{(i)}=j \mid x^{(i)}; \phi, \mu, \Sigma)$$

其中：

$$p(z^{(i)}=j \mid x^{(i)}; \phi, \mu, \Sigma) = \frac{p(x^{(i)} \mid z^{(i)}=j; \mu, \Sigma)\, p(z^{(i)}=j; \phi)}{\sum_{l=1}^{k} p(x^{(i)} \mid z^{(i)}=l; \mu, \Sigma) \cdot p(z^{(i)}=l; \phi)}$$

$$= \frac{N(x^{(i)}; \mu_j, \Sigma_j) \cdot \phi_j}{\sum_{l=1}^{k} N(x^{(i)}; \mu_l, \Sigma_l) \cdot \phi_l}$$

M步：基于猜测更新模型参数：

$$\phi_j = \frac{1}{m} \sum_{i=1}^{m} w_j^{(i)}$$

$$\mu_j = \frac{\sum_{i=1}^{m} w_j^{(i)} x^{(i)}}{\sum_{i=1}^{m} w_j^{(i)}}$$

$$\Sigma_j = \frac{\sum_{i=1}^{m} w_j^{(i)} (x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^{m} w_j^{(i)}}$$

EM算法和k-mean很像，区别为：

① k-mean 为"硬聚类" $c^{(i)}$

② GMM 为"软聚类" $w_j^{(i)}$

关于EM算法一个更general的解释

EM算法可用于任何包含隐变量的参数估计。

引子：Jensen不等式：

设f为凸函数（即$f''(x)>0$，$\forall x\in\mathbb{R}$）。

当x为向量时，条件变为海森矩阵H为半正定：$H\geq 0$.

Jensen不等式表述为：

f为凸函数，X为随机变量，则：

$$E[f(x)]\geq f(E[x])$$

当且仅当$x=E[x]$以概率1成立，即x为常数.

EM算法：

给定训练集$\{x^{(1)},x^{(2)},\dots,x^{(m)}\}$，据此数据拟合

带参模型$p(x,z)$，定义似然：

$$l(\theta)=\sum_{i=1}^{m}\log p(x;\theta)$$

$$= \sum_{i=1}^{m} \log \sum_{z} p(x, z; \theta)$$

其中，$z^{(i)}$ 为隐变量，当 $z^{(i)}$ 已知时，通常上面的方程很好解。

直接优化似然 $l(\theta)$ 很困难，于是通过迭代：

① 构造 l 上的下界

② 优化（最大化）下界

对每个 i，定义 Q_i 为 z 上的分布，即：

$$\sum_{z} Q_i(z) = 1, \quad Q_i(z) \geq 0$$

考虑下式：

$$\sum_{i=1}^{m} \log p(x^{(i)}; \theta) = \sum_{i=1}^{m} \log \sum_{z^{(i)}} p(x^{(i)}, z^{(i)}; \theta)$$

$$= \sum_{i=1}^{m} \log \sum_{z^{(i)}} Q_i(z^{(i)}) \frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})}$$

$$\geq \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})}$$

最后一步用了 Jenson 不等式，且 $f(x) = \log(x)$，

随机变量 X 为 $\log\frac{p(x^{(i)},z^{(i)};\theta)}{Q_i(z^{(i)})}$（关于 $z^{(i)}$）的，

满足 $Q_i(z^{(i)})$ 的分布，则有：

$$f\left(E_{z^{(i)}\sim Q_i}\left[\log\frac{p(x^{(i)},z^{(i)};\theta)}{Q_i(z^{(i)})}\right]\right) \geq E_{z^{(i)}\sim Q_i}\left[f\left(\log\frac{p(x^{(i)},z^{(i)};\theta)}{Q_i(z^{(i)})}\right)\right]$$

注：由于 $f(x)=\log(x)$ 为 concave，所以符号反转了。

则对于每个分布，上式都给出了 $l(\theta)$ 的一个下界。

~~并且~~这里的 Q_i 还是未知的，不过如果我们已经估计了当前的参数 θ，则可以在这个 θ 下优化下界，即让上面的不等式取等号，即：

$$\frac{p(x^{(i)},z^{(i)};\theta)}{Q_i(z^{(i)})}=c$$

而又因为 $\sum_z Q_i(z^{(i)})=1$，则：

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

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

可以看到，这还是一个关于 $z^{(i)}$ 的分布。

总结一下就是：

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

M步： $\theta = \arg\max \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})}$.