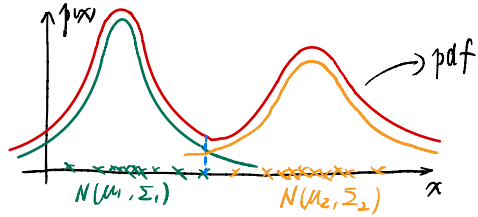
高斯混合模型（GMM）

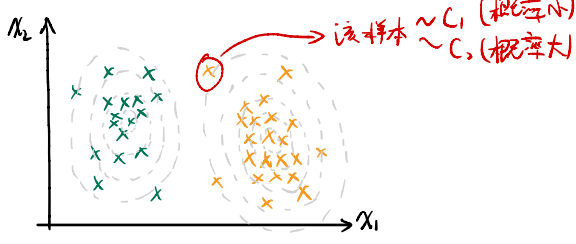
一. 模型介绍

一般来说，将数据假设为高斯分布是合理的。



从几何角度来看：加权平均值 ⟶ 多个高斯分布叠加而成

则 $p(x)=\sum_{k=1}^{K}\alpha_k N(x|\mu_k,\Sigma_k)$，$\sum_{k=1}^{K}\alpha_k=1$，$\alpha_k$为权重。



从混合模型角度来看：

x：观测变量

z：隐变量

⟶ 对应的样本x是属于哪一个高斯分布，因此它是离散的随机变量。

z	C_1	C_2	···	C_K
p(z)	p_1	p_2	···	p_K

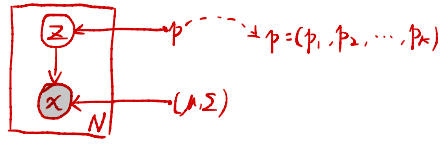
，$\sum_{k=1}^{K}p_k=1$

GMM为生成模型，其生成过程为：

① 假设有一个骰子，共每个面都有涂料，但涂料的重量不同，即投这个骰子每个面出现的概率不同，每个面出现的概率为p_k。

② 接下来根据出现概率为p_k的某个高斯分布，在这个高斯分布上采样。

其概率图模型可表示为：



混合模型角度 $P(X)$ 为：

$$\begin{aligned} P(X) &= \sum_{Z} P(X, Z) \\ &= \sum_{k=1}^{K} P(X, z=C_k) \\ &= \sum_{k=1}^{K} P(z=C_k) \cdot P(x \mid z=C_k) \\ &= \sum_{k=1}^{K} \pi_k \cdot N(x \mid \mu_k, \Sigma_k) \end{aligned}$$

△ 积分变求和.

求参数：

假设：X：观测数据，$X=(x_1, x_2, \cdots x_N)$

(X, Z)：完备数据.

θ：参数，$\theta=\{\pi_1, \pi_2, \cdots, \pi_k, \mu_1, \mu_2, \cdots \mu_k, \Sigma_1, \Sigma_2, \cdots \Sigma_k\}$

则：

$$\begin{aligned} \hat{\theta}_{MLE} &= \arg\max_{\theta} \log P(X) = \arg\max_{\theta} \log \prod_{i=1}^{N} P(x_i) = \arg\max_{\theta} \sum_{i=1}^{N} \log P(x_i) \\ &= \arg\max_{\theta} \sum_{i=1}^{N} \log \sum_{k=1}^{K} \pi_k \cdot N(x_i \mid \mu_k, \Sigma_k) \end{aligned}$$

由于 $\log(\triangle+\triangle+\triangle+\cdots+\triangle)$ 中是加和等号，因此，MLE不能求出其解析解.

∴用MLE求解GMM，无法得出解析解.

应用EM算法求解：（EM：$\theta^{(t+1)} = \arg\max_{\theta} \underbrace{E_{Z|X,\theta^{(t)}}[\log P(X, Z \mid \theta)]}_{Q(\theta, \theta^{(t)})}$）$\theta^{(t)}$ 是常数.

E-step：

$$\begin{aligned} Q(\theta, \theta^{(t)}) &= \int_Z \log P(X, Z \mid \theta) \cdot P(Z \mid X, \theta^{(t)}) \, dZ \\ &= \sum_{Z} \log \prod_{i=1}^{N} P(x_i, z_i \mid \theta) \cdot \prod_{i=1}^{N} P(z_i \mid x_i, \theta^{(t)}) \\ &= \sum_{z_1, z_2, \cdots z_N} \sum_{i=1}^{N} \log P(x_i, z_i \mid \theta) \cdot \prod_{i=1}^{N} P(z_i \mid x_i, \theta^{(t)}) \\ &= \sum_{z_1, z_2, \cdots z_N} \left[\log P(x_1, z_1 \mid \theta) + (\log P(x_2, z_2 \mid \theta) + \cdots + \log P(x_N, z_N \mid \theta) \right] \prod_{i=1}^{N} P(z_i \mid x_i, \theta^{(t)}) \end{aligned}$$

其中，项：$\sum_{z_1, \cdots z_N} \log P(x_1, z_1 \mid \theta) \prod_{i=1}^{N} P(z_i \mid x_i, \theta^{(t)})$

$$\begin{aligned} &= \sum_{z_1} \log P(x_1, z_1 \mid \theta) P(z_1 \mid x_1, \theta^{(t)}) \sum_{z_2, \cdots z_N} \prod_{i=2}^{N} P(z_i \mid x_i, \theta^{(t)}) \\ &= \sum_{z_1} \log P(x_1, z_1 \mid \theta) P(z_1 \mid x_1, \theta^{(t)}) \end{aligned}$$

$= \sum_{z_1} P(z_1 \mid x_1, \theta^{(t)}) \sum_{z_2} P(z_2 \mid x_2, \theta^{(t)}) \sum_{z_3} P(z_3 \mid x_3, \theta^{(t)}) \cdots \sum_{z_N} P(z_N \mid x_N, \theta^{(t)})$

$= 1$

因此：

$$\begin{aligned} &= \sum_{z_1} \log P(x_1, z_1 \mid \theta) \cdot P(z_1 \mid x_1, \theta^{(t)}) + \cdots + \sum_{z_N} \log P(x_N, z_N \mid \theta) \cdot P(z_N \mid x_N, \theta^{(t)}) \\ &= \sum_{i=1}^{N} \sum_{z_i} \log P(x_i, z_i \mid \theta) P(z_i \mid x_i, \theta^{(t)}) \end{aligned}$$

$$\because P(x)=\sum_{k=1}^{K}p_k\cdot N(x|\mu_k,\Sigma_k)$$

$$\therefore P(x,z)=P(z)\cdot P(x|z)$$

$$=p_z\cdot N(x|\mu_z,\Sigma_z)$$

$$\therefore P(z|x)=\frac{P(x,z)}{P(x)}=\frac{p_z\cdot N(x|\mu_z^{(t)},\Sigma_z^{(t)})}{\sum_{k=1}^{K}p_k^{(t)}N(x|\mu_k^{(t)},\Sigma_k^{(t)})}$$

则 $$Q(\theta,\theta^{(t)})=\sum_{i=1}^{N}\sum_{z_i}\log\left[p_{z_i}\cdot N(x_i|\mu_{z_i},\Sigma_{z_i})\right]\cdot\frac{p_{z_i}\cdot N(x_i|\mu_{z_i}^{(t)},\Sigma_{z_i}^{(t)})}{\sum_{k=1}^{K}p_k^{(t)}N(x_i|\mu_k^{(t)},\Sigma_k^{(t)})}$$

M-step：

由 E-step：

$$\begin{aligned}Q(\theta,\theta^{(t)})&=\sum_{i=1}^{N}\sum_{z_i}\log\left[p_{z_i}\cdot N(x_i|\mu_{z_i},\Sigma_{z_i})\right]\cdot P(z_i|x_i,\theta^{(t)})\\&=\sum_{z_i}\sum_{i=1}^{N}\log\left[p_{z_i}N(x_i|\mu_{z_i},\Sigma_{z_i})\right]\cdot P(z_i|x_i,\theta^{(t)})\\&=\sum_{k=1}^{K}\sum_{i=1}^{N}\log\left[p_k\cdot N(x_i|\mu_k,\Sigma_k)\right]\cdot P(z_i=C_k|x_i,\theta^{(t)})\\&=\sum_{k=1}^{K}\sum_{i=1}^{N}\left[\log p_k+\log N(x_i|\mu_k,\Sigma_k)\right]\cdot P(z_i=C_k|x_i,\theta^{(t)})\end{aligned}$$

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

求 $p_k^{(t+1)}$：

$$\begin{cases}p_k^{(t+1)}=\arg\max\sum_{k=1}^{K}\sum_{i=1}^{N}\log p_k\cdot P(z_i=C_k|x_i,\theta^{(t)})\\ \text{s.t. }\sum_{k=1}^{K}p_k=1\end{cases}$$

应用拉格朗日乘子法：

$$\mathcal{L}(p,\lambda)=\sum_{k=1}^{K}\sum_{i=1}^{N}\log p_k\cdot P(z_i=C_k|x_i,\theta^{(t)})+\lambda\left(\sum_{k=1}^{K}p_k-1\right)$$

此处只对 p_k → $\frac{\partial\mathcal{L}}{\partial p_k}=\sum_{i=1}^{N}\frac{1}{p_k}\cdot P(z_i=C_k|x_i,\theta^{(t)})+\lambda\triangleq 0$

求导，与 p_{k-1}，p_{k+1} 等式两边同乘 p_k.

…p_i 都无关. $$\Longrightarrow\sum_{i=1}^{N}P(z_i=C_k|x_i,\theta^{(t)})+p_k\lambda=0$$

$$\overset{k=1,\cdots,K}{\Longrightarrow}\underbrace{\sum_{i=1}^{N}\sum_{k=1}^{K}P(z_i=C_k|x_i,\theta^{(t)})}_{(\text{对}z_i\text{求和为}1)}+\underbrace{\sum_{k=1}^{K}p_k\lambda}_{1}=0$$

$$\Longrightarrow N+\lambda=0\Longrightarrow\lambda=-N.$$

故 $$p_k^{(t+1)}=\frac{1}{N}\sum_{i=1}^{N}P(z_i=C_k|x_i,\theta^{(t)})$$

$$p^{(t+1)}=(p_1^{(t+1)},p_2^{(t+1)},\cdots,p_K^{(t+1)})$$