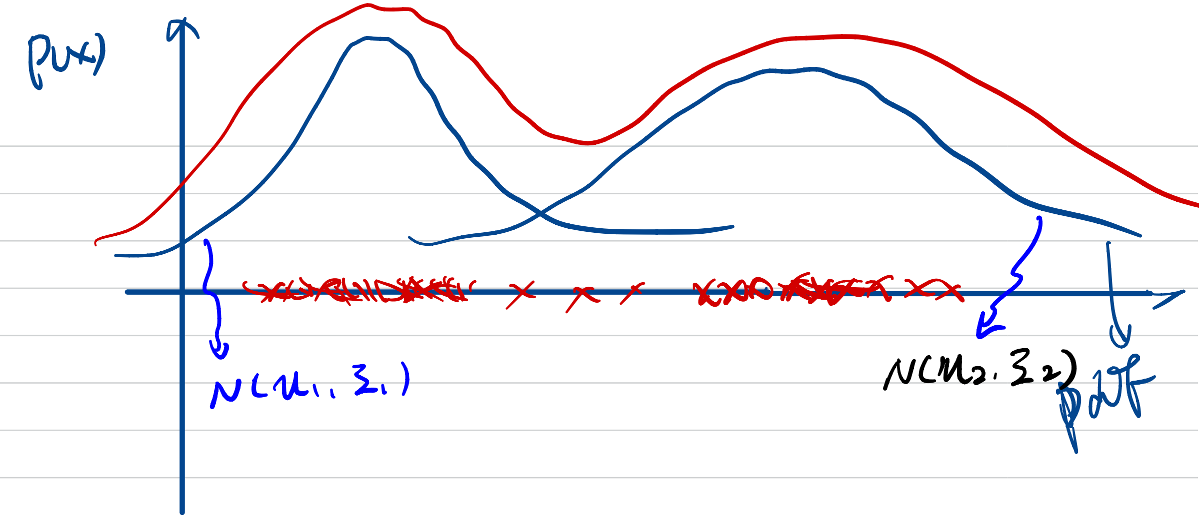
Gaussian mixture model

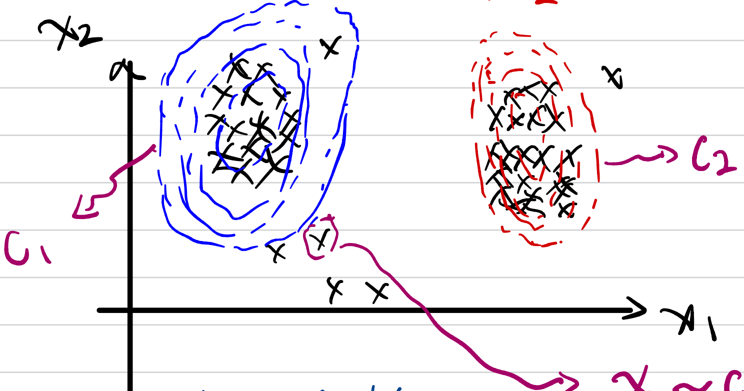
高斯 混合模型



从几何角度来看：加权平均 → (多个高斯分布叠加而成)

$$P(x)=\sum_{k=1}^{K} \alpha_k N(x|\mu_k,\Sigma_k), \quad \sum_{k=1}^{K} \alpha_k = 1$$

↳ Red curve. ↳ 权重



$x \sim C_1$ (概率大)

$\sim C_2$ (概率小)

→ $x \sim z$

从混合模型角度来看 (生成模型)

生成角度

x: observed variable

z: latent variable

→ 对应的样本 x 是属于哪个高斯分布.

→ 离散随机变量

z	C_1	C_2	......	C_k
$P(z)$	P_1	P_2		P_k

$$\sum_{k=1}^{K} p_k = 1$$

生成模型中样本的生成过程：

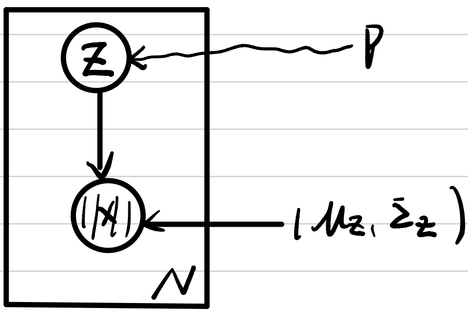
① 假设有一个骰子（K个面）

② 得到第K个面 ~ K Gaussian

③ 从第K个高斯分布中采样。

④ Repeat ①~③

概率图：



混合模型的PDF：

$$
\begin{aligned}
P(x) &= \sum_{z} P(x, z) \\
&= \sum_{k=1}^{K} P(x, z = C_k) \\
&= \sum_{k=1}^{K} P(z = C_k) \cdot P(x \mid z = C_k) \\
&= \sum_{k=1}^{K} \underbrace{p_k}_{\text{概率值}} \cdot N(x \mid \mu_k, \Sigma_k)
\end{aligned}
$$

Equals to Geometry's view.

GLM - MLE (Learning)

X: observed data — $X = (x_1, x_2, \dots, x_N)$

(X, Z): complete data — $(x_1, z_1)(x_2, z_2) \cdots (x_N, z_N)$

θ: parameters — $\theta = \{p_1, \dots, p_K, \mu_1 \dots \mu_K, \Sigma_1 \dots \Sigma_K\}$

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

直接用MLE求解GMM，无法得出解析解。

用EM算法求解.

EM求解 GMM - MLE

$$EM: \theta^{(t+1)} = \arg\max \underbrace{E_{z|x,\theta^{(t)}}[\log P(x,z|\theta)]}_{Q(\theta,\theta^{(t)})}$$

$$Q(\theta,\theta^{(t)}) = \int_z \underline{P(z|x,\theta^{(t)}) \cdot \log P(x,z|\theta)\, dz}$$

GMM中 Z 为离散的随机变量

$$= \sum_Z \log \prod_{i=1}^{N} P(x_i,z_i|\theta) \cdot \prod_{i=1}^{N} P(z_i|x_i,\theta^{(t)})$$

$$= \sum_{z_1 z_2 \cdots z_N} \sum_{i=1}^{N} \log P(x_i,z_i|\theta) \cdot \prod_{i=1}^{N} P(z_i|x_i,\theta^{(t)})$$

$$= \sum_{z_1 \cdots z_N} [\log P(x_1,z_1|\theta) + \log(x_2,z_2|\theta) + \cdots \log(x_N,z_N|\theta)] \prod_{i=1}^{N} P(z_i|x_i,\theta^{(t)})$$

提一项出来

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

$$= \sum_{z_1 \cdots z_N} \log P(x_1,z_1|\theta)\, P(z_1|x_1,\theta^{(t)}) \cdot \prod_{i=2}^{N} P(z_i|x_i,\theta^{(t)})$$

$$= \sum_{z_1} \log P(x_1,z_1|\theta)\, P(z_1|x_1,\theta^{(t)}) \underbrace{\sum_{z_2 \cdots z_N} \prod_{i=2}^{N} P(z_i|x_i,\theta^{(t)})}_{1}$$

$$= \sum_{z_1} \log P(x_1,z_1|\theta)\, P(z_1|x_1,\theta^{(t)})$$

$$= \sum_{z_1} \log P(x_1,z_1|\theta) P(z_1|x_1,\theta^{(t)}) + \cdots + \sum_{z_N} \log P(x_N,z_N|\theta) P(z_N|x_N,\theta^{(t)})$$

$$= \underline{\sum_{i=1}^{N} \sum_{z_i} \log P(x_i,z_i|\theta)\, P(z_i|x_i,\theta^{(t)})}$$

将参数置换为GMM中的值:

z: latent variable

z	C_1	C_2	$\cdots$	C_k
P_z	P_1	P_2	$\cdots$	P_k

, $\sum_{k=1}^{K} P_k = 1$

$$x \mid z = C_k \sim N(x \mid \mu_k, \Sigma_k)$$

$$\therefore P(x, z) = P(z) P(x \mid z)$$

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

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

代入到EM式中:

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

U

EM中的E step

E-step:

$$Q(\theta,\theta^{(t)}) = \sum_{i=1}^{N} \sum_{z_i} \log\left[p_{z_i}\cdot N(x_i\mid\mu_{z_i},\Sigma_{z_i})\right] P(z_i\mid x_i,\theta^{(t)})$$

$$\sum_{z_i} \log\left[p_{z_i}\cdot N(x_i\mid\mu_{z_i},\Sigma_{z_i})\right]\cdot P(z_i\mid x_i,\theta^{(t)})$$

$$= \sum_{k=1}^{K} \log\left[p_k\cdot N(x_i\mid\mu_k,\Sigma_k)\right]\cdot P(z_i=C_k\mid x_i,\theta^{(t)})$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K}\left[\log p_k + \log N(x_i\mid\mu_k,\Sigma_k)\right]\cdot P(z_i=C_k\mid x_i,\theta^{(t)})$$

$$\theta^{(t+1)} = \arg\max_{\theta} Q(\theta,\theta^{(t)}) \rightsquigarrow \text{M-step}$$

Here we take $p^{(t+1)}$ for example:

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

$$\begin{cases} \max_{p} \sum_{k=1}^{K}\sum_{i=1}^{N} \log p_k\, P(z_i=C_k\mid x_i,\theta^{(t)}) \\ \text{s.t. } \sum_{k=1}^{K} p_k = 1 \end{cases}$$

拉格朗日乘子法求解

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

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

$$\xrightarrow{\times p_k} \sum_{i=1}^{N} P(z_i=C_k\mid x_i,\theta^{(t)}) + p_k\lambda = 0 \quad ①$$

$$\xrightarrow{k=1\cdots K} \underbrace{\sum_{i=1}^{N}\sum_{k=1}^{K} P(z_i=C_k\mid x_i,\theta^{(t)})}_{1} + \lambda\underbrace{\sum_{k=1}^{K} p_k}_{1} = 0 \quad ②$$

So $N + \lambda = 0$

$$\boxed{\lambda = -N}$$

take it into ①:

$$\sum_{i=1}^{N} P(z_i = C_k \mid x_i, \theta^{(t)}) - N p_k = 0$$

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

$$p^{t+1} = (p_1^{t+1} \cdots\cdots p_k^{t+1})$$

Also for u_k^{t+1}:

$$u^{t+1} = (u_1^{t+1}, u_2^{t+1} \cdots u_k^{t+1})$$

$$Q(\theta, \theta^{(t)}) = \sum_{k=1}^{K} \sum_{i=1}^{N} \log[p_k \cdot N(x_i \mid u_k, \Sigma_k)] \cdot P(z_i = C_k \mid x_i, \theta^{(t)})$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{N} [\log p_k + \underbrace{\log N(x_i \mid u_k, \Sigma_k)] \cdot P(z_i = C_k \mid x_i, \theta^{(t)})}_{\text{related to } u_k}$$

So the goal is: $\underset{u}{\operatorname{argmax}} \underbrace{\sum_{k=1}^{K} \sum_{i=1}^{N} \log N(x_i \mid u_k, \Sigma_k) \cdot P(z_i = C_k \mid x_i, \theta^{(t)})}_{g(\theta)}$

$$\log N(x_i \mid u_k, \Sigma_k) = \log[(2\pi\Sigma_k)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(x_i - u_k)^T \Sigma_k^{-1} (x_i - u_k)}]$$

$$= -\frac{1}{2}\log(2\pi\Sigma_k) - \frac{1}{2}(x_i - u_k)^T \Sigma_k^{-1} (x_i - u_k)$$

$$= -\frac{1}{2}[\log(2\pi\Sigma_k) + (x_i - u_k)^T \Sigma_k^{-1} (x_i - u_k)]$$

$$\therefore \frac{\partial g(\theta)}{\partial u_k} = \sum_{i=1}^{N} -\frac{1}{2} \cdot (-\Sigma_k^{-1}(x_i - u_k) - \Sigma_k^{-T}(x_i - u_k)) \cdot P(z_i = C_k \mid x_i, \theta^{(t)})$$

$$= \sum_{i=1}^{N} \frac{1}{2}\left(\Sigma_k^{-1}(x_i - u_k) + \Sigma_k^{-T}(x_i - u_k)\right) \cdot P(z_i = c_k \mid x_i, \theta^{(t)})$$

$$= \frac{1}{2}\sum_{i=1}^{N}\left(\Sigma_k^{-1}(x_i - u_k) + \Sigma_k^{-1}(x_i - u_k)\right) \cdot P(z_i = c_k \mid x_i, \theta^{(t)})$$

left $\times\ \Sigma_k$

$$\Rightarrow \sum_{i=1}^{N} (x_i - u_k) P(z_i = c_k \mid x_i, \theta^{(t)}) = 0$$

$$\sum_{i=1}^{N} x_i P(z_i = c_k \mid x_i, \theta^{(t)}) = u_k \underbrace{\sum_{i=1}^{N} P(z_i = c_k \mid x_i, \theta^{(t)})}_{1}$$

Therefore: $u_k^{t+1} = \sum_{i=1}^{N} x_i P(z_i = c_k \mid x_i, \theta^{(t)})$

So: $u^{t+1} = (u_1^{t+1} \cdots u_k^{t+1})$

Σ^{t+1} is derivated in a simmilar way.

Summary:

1. GMM → 多个高斯加权均 → 引入隐变量 z 表示 Data 所属的 Gaussian Distribution.
2. GMM Learning (MLE) 不能直接得到解析解，原因是存在 $\log\Sigma$。
3. 使用迭代算法 EM 求解 GMM.