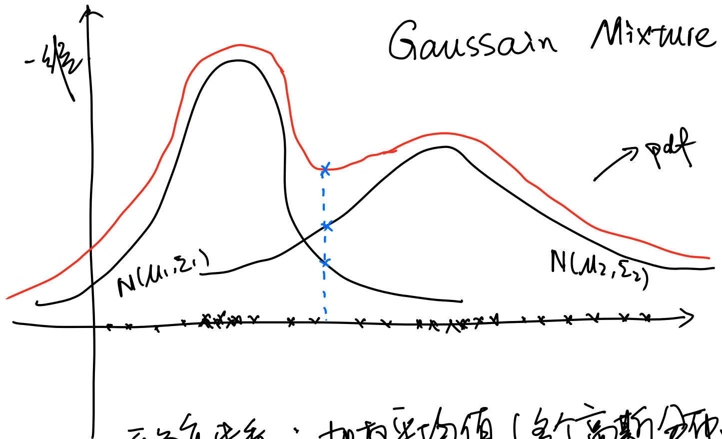
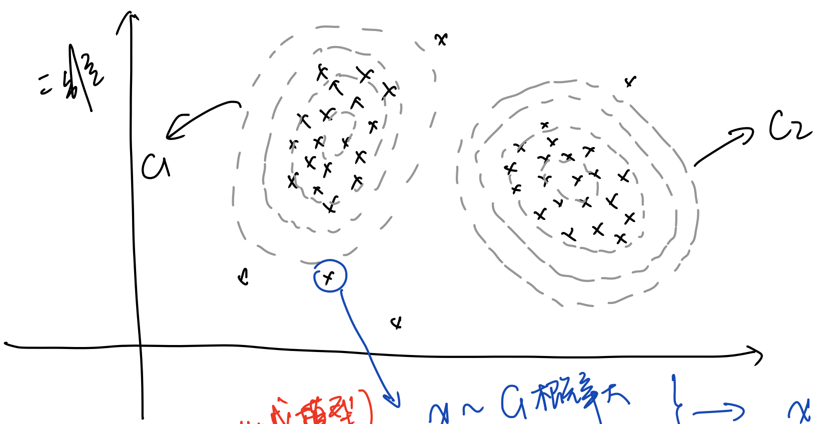
模型介绍

Gaussain Mixture Model 高斯混合模型



从几何角度来看：加权平均值（多个高斯分布叠加而成）

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



$x \sim C_1$ 概率大
$x \sim C_2$ 概率小 $\longrightarrow x \sim z$

从混合模型角度（生成模型）

x: observed variable

z: latent variable → 对应的样本 x 是属于哪一个高斯分布

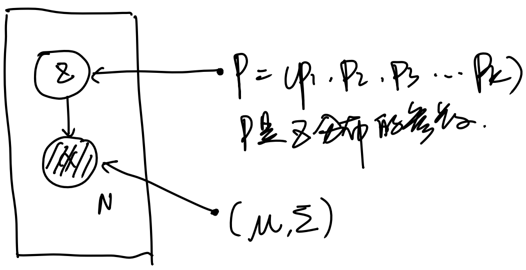
↳ 离散的随机变量

z	C_1	C_2	...	C_K
$p(z)$	p_1	p_2		p_k

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

生成样本 → 先通过 $x \sim z$ 确定属于哪一个高斯 → 再 $x \sim N(\mu_k,\Sigma_k)$ 在高斯中生成点

概率图



极大似然、

从几何角度看 $P(x)=\sum_{k=1}^{K}\alpha_k N(x|\mu_k,\Sigma_k)$，$\sum_{k=1}^{K}\alpha_k=1$

α_k → 权重

从混合模型角度看

x: observed variable

z: latent variable

z	C_1	C_2	...	C_K
P	P_1	P_2	...	P_K

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

$$P(x)=\sum_z P(x,z)=\sum_{k=1}^{K}P(x,z=C_k)$$
$$=\sum_{k=1}^{K}P(z=C_k)\cdot P(x|z=C_k)$$
$$=\sum_{k=1}^{K}P_k\cdot N(x|\mu_k,\Sigma_k)$$

P_k → 概率值

X: observed data → $X=(x_1,x_2,\dots,x_N)$ x_i,x_j ⇒ iid

(X,Z): complete data

θ: parameter → $\theta=\{P_1,P_2\cdots P_K,\mu_1,\mu_2\cdots\mu_K,\Sigma_1,\Sigma_2,\cdots\Sigma_K\}$

$\hat{\theta}_{MLE}=\arg\max_\theta \log P(X)$ → 对于GMM没有解析解，MLE算不出来.

$=\arg\max_\theta \log\prod_{i=1}^{N}P(x_i)=\arg\max\sum_{i=1}^{N}\log P(x_i)$

$=\underbrace{\arg\max_\theta\sum_{i=1}^{N}\log\sum_{k=1}^{K}P_k N(x_i|\mu_k,\Sigma_k)}$

$\log(\triangle+\triangle+\triangle)$ 形式，且里面有高斯分布

∴ 直接用MLE求解GMM，无法得出解析解

→ 可以用EM求解

EM : E-Step

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

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

离散的

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

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

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

先看第一项 $\boxed{\sum_{z} \underbrace{\log P(x_1,z_1|\theta)}_{z_1\text{相关}} \underbrace{\prod_{i=1}^{N} P(z_i|x_i,\theta^{(t)})}_{P(z_1|x_1,\theta^{(t)}) \cdot \prod_{i=2}^{N} P(z_i|x_i,\theta^{(t)})}}$

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

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

省略了θ

$$= \sum_{z_2 \cdots z_N} P(z_2|x_2)\, P(z_3|x_3) \cdots P(z_N|x_N)$$

$$= \underbrace{\sum_{z_2} P(z_2|x_2)}_{1} \cdot \underbrace{\sum_{z_3} P(z_3|x_3)}_{1} \cdots \underbrace{\sum_{z_N} P(z_N|x_N)}_{1}$$

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

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

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

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

z	c_1	c_2	c_3
p	p_1	p_2	p_3

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

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

代入 $Q(\theta,\theta^{(t)})=\sum_{i=1}^{N}\sum_{z_i}\log P_{z_i}N(x_i|\mu_{z_i},\Sigma_{z_i})\cdot\frac{P_{z_i}\cdot N(x_i|\mu_{z_i},\Sigma_{z_i})}{\sum_{k=1}^{K}P_k N(x_i|\mu_k,\Sigma_k)}$

EM：M-step

$$\begin{array}{c|cccc} z & c_1 & c_2 & \cdots & c_k \\ \hline p & p_1 & p_2 & & p_k \end{array}\qquad \sum_i p_i=1$$

E-step：

$$Q(\theta,\theta^{(t)})=\sum_{i=1}^{N}\sum_{z_i}\log P_{z_i}N(x_i|\underbrace{\mu_{z_i},\Sigma_{z_i}}_{\theta})\cdot\underbrace{\frac{P_{z_i}\cdot N(x_i|\mu_{z_i}^{(t)},\Sigma_{z_i}^{(t)})}{\sum_{k=1}^{K}P_k^{(t)}N(x_i|\mu_k^{(t)},\Sigma_k^{(t)})}}$$

$$=\sum_{i=1}^{N}\sum_{z_i}\log[P_{z_i}N(x_i|\mu_{z_i},\Sigma_{z_i})]\cdot\underbrace{P(z_i|x_i,\theta^{(t)})}_{常量}$$

$$=\sum_{z_i}\sum_{i=1}^{N}\log[P_{z_i}N(x_i|\mu_{z_i},\Sigma_{z_i})]P(z_i|x_i,\theta^{(t)})$$

$$=\sum_{k=1}^{K}\sum_{i=1}^{N}\log[\underbrace{P_k}_{z_i=C_k}\,N(x_i|\underbrace{\mu_k,\Sigma_k}_{新的\ \theta^{(t+1)}})]P(z_i=C_k|x_i,\theta^{(t)})$$

$$=\sum_{k=1}^{K}\sum_{i=1}^{N}[\log P_k+\log N(x_i|\mu_k,\Sigma_k)]P(z_i=C_k|x_i,\theta^{(t)})$$

M-Step:

$$\theta^{(t+1)}=\arg\max_{\theta}Q(\theta,\theta^{(t)})\ \to\ 求\ P_k,\mu_k,\Sigma_k$$

$$求\ P_k^{(t+1)}=\{p_1^{(t+1)},p_2^{(t+1)}\cdots p_k^{(t+1)}\}$$

约束优化问题："拉格朗日乘子法"

$$P_k^{(t+1)}=\arg\max_{P_k}\sum_{k=1}^{K}\sum_{i=1}^{N}\log P_k\cdot P(z_i=C_k|x_i,\theta^{(t)}),\quad s.t.\ \sum_{k=1}^{K}P_k=1$$

拉格朗日乘子法

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

$$\frac{\partial\mathcal{L}}{\partial P_k}=\sum_{i=1}^{N}\frac{1}{P_k}P(z_i=C_k|x_i,\theta^{(t)})+\lambda=0$$

$$\Rightarrow\sum_{i=1}^{N}P(z_i=C_k|x_i,\theta^{(t)})+P_k\lambda=0\quad(\lambda=-N)$$

$k = 1 \cdots K.$

$$\Rightarrow \sum_{i=1}^{N} \underbrace{\sum_{k=1}^{K} P(z_i = C_k \mid x_i, \theta^{(t)})}_{1} + \underbrace{\sum_{k=1}^{K} P_k}_{1} \cdot \overset{-N}{\lambda} = 0$$

$$\Rightarrow N + \lambda = 0 \Rightarrow \lambda \Rightarrow -N$$

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

$$P_k^{(t+1)} = (P_1^{(t+1)}, \cdots P_K^{(t+1)})$$