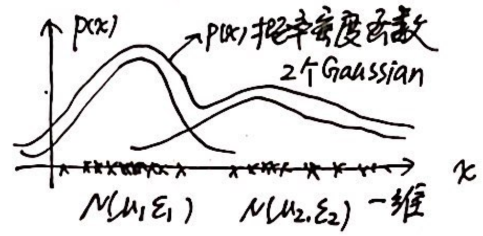
生成模型　白板(11)　Gaussian Mixture Model 高斯混合模型

一般数据，假设是高斯分布，是合理的(由中心极限定理)

几何角度看：概率是多个高斯叠加的加权平均

$$p(x)=\sum_{k=1}^{K}\alpha_k \underbrace{N(\mu_k,\Sigma_k)}_{N(x|\mu_k,\Sigma_k)},\quad \sum_{k=1}^{K}\alpha_k=1 \quad \alpha\text{是权重系数}$$

↳ 高斯混合模型



混合模型看：(生成模型)

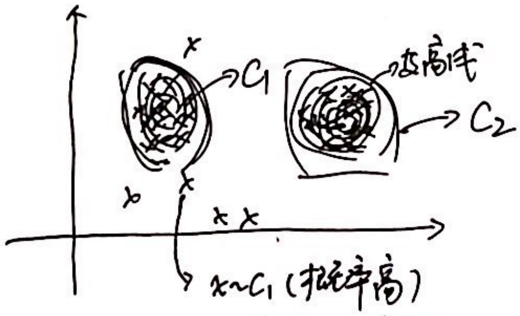
x: observed variable

z: latent variable

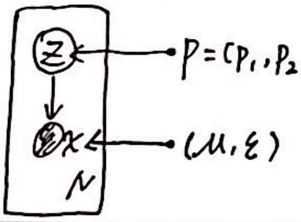
↳ 对应x样本是属于哪一个高斯分布 (离散z)

	C_1	C_2	……	C_k
z	1	2	……	k
$p(z)$	p_1	p_2	……	p_k

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



概率图.



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

极大似然，(做不出，引入EM)

GitHub: CBJerry993

X: observed data $X=\{x_1,x_2\cdots x_N\}$

(X,Z): complete data

θ: parameter $\theta=\{p_1,p_2\cdots p_k,\mu_1,\mu_2,\cdots\mu_k,\Sigma_1,\Sigma_2\cdots\Sigma_k\}$

$$\hat{\theta}_{MLE}=\arg\max_\theta \log p(X)\overset{iid}{=}\arg\max_\theta \log\prod_{i=1}^{N}p(x_i)=\arg\max_\theta\sum_{i=1}^{N}\log(p(x_i))=\arg\max_\theta\sum_{i=1}^{N}\log\sum_{k=1}^{K}p_k\cdot N(x_i|\mu_k,\Sigma_k)$$

，直接用MLE求解GMM，无解析解，只能近似解. ⟶ 用EM.

log(连加) 无解析解

白板(11) Gaussian Mixture Model

EM求解

GitHub: CBJerry993

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

E-step: 表示Q函数

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

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

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

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

看其中一项 $\sum_{z_1 z_2 \cdots z_n} \log p(x_1,z_1|\theta)\cdot \prod_{i=1}^{N} p(z_i|x_i,\theta^{(t)}) = \cancel{\sum_{z_1 z_2 \cdots z_n} \log p(x_1,z_1|\theta)\cdot \prod_{i=1}^{N} p(z_i|x_i,\theta^{(t)})}$

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

提一项

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

$$\sum_{z_2 \cdots z_n} p(z_2|x_2)\cdot p(z_3|x_3) \cdots p(z_n|x_N)$$

$$= \underbrace{\sum_{z_2} p(z_2|x_2)}_{1}\cdot \underbrace{\sum_{z_3} p(z_3|x_3)}_{1} \cdots \underbrace{\sum_{z_n} p(z_n|x_n)}_{1} = 1$$

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

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

已知 $p(x) = \sum_{k=1}^{K} p_k \cdot N(x|\mu_k,\Sigma_k)$, $x|z=c_1 \cdots c_k \sim N(x|\mu_k,\Sigma_k)$.

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

下标表示，当 $z=c_1$ 时 p_1，$z=c_2$ 时 p_2，$z=c_k$ 时 p_k

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

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

变量

记 $p(z_i|x_i,\theta^{(t)})$ 常量

白板 (11) Gaussian Mixture Model

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

$$\text{5}k = \sum_{k=1}^{K}\sum_{i=1}^{N}\left[\cancel{\ }\ \log\left[p_k \cdot N(x_i|\mu_k,\Sigma_k)\right]\cdot p(z_i = C_k|x_i,\theta^{(t)})\right) \qquad z_i = C_k$$

$$(\text{E-step}) \rightarrow = \sum_{k=1}^{K}\sum_{i=1}^{N}\left[\log p_k + \log N(x_i|\mu_k,\Sigma_k)\right]\cdot p(z_i = C_k|x_i,\theta^{(t)})$$

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

求 $p_k^{(t+1)} = \arg\max_{p_k} \sum_{k=1}^{K}\sum_{i=1}^{N} \log p_k \cdot p(z_i = C_k|x_i,\theta^{(t)})$, $\quad s.t. \overset{\text{约束}}{\sum_{k=1}^{K}} p_k = 1 \rightarrow$ M-step

(去掉 p_k 无关)

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

GitHub·CBJerry993

约束问题，拉格朗日乘子法。

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

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

$$\Rightarrow \sum_{i=1}^{N} p(z_i = C_k|x_i,\theta^{(t)}) + p_k\lambda = 0$$

$$\xRightarrow{k=1,\cdots K} \sum_{i=1}^{N}\underbrace{\sum_{k=1}^{K} p(z_i = C_k|x_i,\theta^{(t)})}_{=1} + \underbrace{\sum_{k=1}^{K}p_k}_{=1}\lambda = 0$$

$$\Longrightarrow N + \lambda = 0 \quad \therefore \lambda = -N$$

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