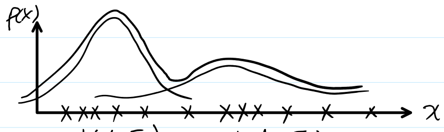
Gaussian Mixture Model

2021年6月24日 1:40

P66 GMM Introduction

一、Geometric Perspective：weighted average of multiple Gaussian distributions

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



$N(\mu_1,\Sigma_1)$ $N(\mu_2,\Sigma_2)$

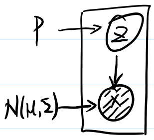
二、Mixture Model (Generative Model) Perspective：

x：observed variable $\quad z$：latent variable

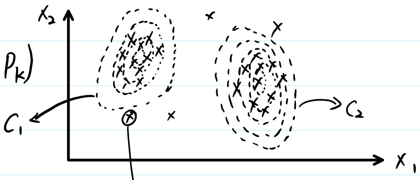
z is discrete random variable (The corresponding x belongs to a certain Gaussian distribution)

z	C_1	C_2	$\cdots$	C_K
P	P_1	P_2	$\cdots$	P_K

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



$P=(P_1,P_2,\dots,P_K)$



$x\in C_1(P_1)\quad x\in C_2(P_2)\quad P_1>P_2$

$x\in z$

P67 GMM MLE

$$P(x)=\sum_z P(x,z)=\sum_{k=1}^{K}P(x,z=C_k)=\sum_{k=1}^{K}P(z=C_k)P(x\mid z=C_k)=\sum_{k=1}^{K}P_kN(x\mid\mu_k,\Sigma_k)$$

X：observed data $\Rightarrow X=(x_1,x_2,\cdots,x_N)$

(X,Z)：complete data

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

$$\hat\theta_{MLE}=\arg\max_\theta\sum_{i=1}^{N}\log P(x_i)=\arg\max_\theta\sum_{i=1}^{N}\log\sum_{k=1}^{K}P_k\,P(x_i\mid\mu_k,\Sigma_k)$$

$\log\Sigma$ is hard to find an analystical solution

∴ Solving GMM directly by MLE doesn't lead to an analystical solution

P68 GMM EM-E-Step

$P(x,z)=P(z)P(x\mid z)=P_zN(x\mid\mu_z,\Sigma_z)$ $\quad z$ is not determined

$$P(x)=\sum_{k=1}^{K}P_kN(x\mid\mu_k,\Sigma_k)$$

$$P(z\mid x)=\frac{P(x,z)}{P(x)}=\frac{P_zN(x\mid\mu_z,\Sigma_z)}{\sum_{k=1}^{K}P_kN(x\mid\mu_k,\Sigma_k)}$$

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

$$Q(\theta,\theta^{(t)})=\sum_z P(z\mid x,\theta^{(t)})\log P(X,z\mid\theta)\quad\Leftarrow E\text{-step}$$

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

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

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

$$\because\sum_z P(z\mid x,\theta^{(t)})=1$$

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

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

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

$$=\sum_{i=1}^{N}\sum_{z_i}\frac{P_{z_i}^{(t)}N(x_i\mid\mu_i^{(t)},\Sigma_i^{(t)})}{\sum_{k=1}^{K}P_k^{(t)}N(x_i\mid\mu_k^{(t)},\Sigma_k^{(t)})}\log P_{z_i}N(x_i\mid\mu_{z_i},\Sigma_{z_i})$$

P69 GMM EM-M-Step

$$\frac{P_{z_i}^{(t)}N(x_i\mid\mu_i^{(t)},\Sigma_i^{(t)})}{\sum_{k}P_k^{(t)}N(x_i\mid\mu_k^{(t)},\Sigma_k^{(t)})}\Rightarrow P(z_i\mid x_i,\theta^{(t)})$$

For GMM EM-M step

$$\frac{P_{z_i}^{(t)} N(x_i|\mu_i^{(t)},\Sigma_i^{(t)})}{\sum_{k=1}^{K} P_k^{(t)} N(x_i|\mu_k^{(t)},\Sigma_k^{(t)})} \Rightarrow P(z_i|x_i,\theta^{(t)})$$

E-step $\Rightarrow$ $Q(\theta,\theta^{(t)}) = \sum_{i=1}^{N}\sum_{z_i} P(z_i|x_i,\theta^{(t)}) \log P_{z_i}\cdot N(x_i|\mu_{z_i},\Sigma_{z_i})$

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

$\theta^{(t+1)} = \arg\max_{\theta} Q(\theta,\theta^{(t)})$ $\qquad \theta=(P_1\cdots P_k,\mu_1\cdots\mu_k,\Sigma_1\cdots\Sigma_k)$

E.g. find $P^{(t+1)}=(P_1^{(t+1)},\cdots,P_k^{(t+1)})$ $\quad \begin{cases} \max_{P_k} \sum_{k=1}^{K}\sum_{i=1}^{N} P(z_i=c_k|x_i,\theta^{(t)}) \log P_k \\ s.t. \sum_{k=1}^{K} P_k=1 \end{cases}$

$$\mathcal{L}(P_1,\cdots,P_k,\lambda) = \sum_{k=1}^{K}\sum_{i=1}^{N} P(z_i=c_k|x_i,\theta^{(t)}) \log P_k + \lambda\left(\sum_{k=1}^{K} P_k - 1\right)$$

$$\frac{\partial \mathcal{L}}{\partial P_k} = \frac{1}{P_k}\sum_{i=1}^{N} P(z_i=c_k|x_i,\theta^{(t)}) + \lambda = 0 \Rightarrow \sum_{i=1}^{N}\underbrace{\sum_{k=1}^{K} P(z_i=c_k|x_i,\theta^{(t)})}_{=1} + \underbrace{\sum_{k=1}^{K}\lambda P_k}_{=\lambda} = 0$$

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

$\mu,\Sigma = \cdots\cdots$