模式识别 第三次作业　　自博16 陈斯杰　2016310721

EM and Gradient Descent

共有K个cluster 和n个样本点

$k=1\ldots K$,　$i=1\ldots n$

μ_k 和 σ_k^2 的似然函数为:

$$l(\{\mu_k,\sigma_k^2\}_{k=1}^K)=\sum_{i=1}^n \ln\left(\sum_{k=1}^K \pi_k \frac{1}{(2\pi)^{\frac{p}{2}}\cdot(\sigma_k^2)^{\frac{p}{2}}}\exp\left[-\frac{\|x_i-\mu_k\|^2}{2\sigma_k^2}\right]\right)$$

(p为x_i的维度)

$$=\sum_{i=1}^n \ln\left(\sum_{k=1}^K \pi_k \frac{\exp(-\frac{1}{2\sigma_k^2})(x_i-\mu_k)^T I(x_i-\mu_k)}{(2\pi\sigma_k^2)^{p/2}}\right)$$

将l对μ_k求偏导:

$$\nabla_{\mu_k} l=\sum_{i=1}^n \frac{\pi_k\cdot N(x_i\mid\mu_k,\sigma_k^2 I)}{\sum_{j=1}^K \pi_j N(x_i\mid\mu_j,\sigma_j^2 I)}\cdot I\cdot\frac{(x-\mu_k)}{\sigma_k^2}$$

$$=\sum_{i=1}^n \boxed{\frac{\pi_k\cdot N(x_i\mid\mu_k,\sigma_k^2 I)}{\sum_{j=1}^K \pi_j N(x_i\mid\mu_j,\sigma_j^2 I)}}\ \frac{(x-\mu_k)}{\sigma_k^2}$$

这一部分正是下面EM算法中E-step的$\tilde{z}_{ik}$

$$\tilde{z}_{ik}=Pr\left(x_i\in cluster_i \,\middle|\, \{\mu_j,\sigma_j^2\}_{j=1}^K, x_i\right)$$

$$=\frac{\pi_k\cdot N(x_i\mid\mu_k,\sigma_k^2 I)}{\sum_{j=1}^K \pi_j\cdot N(x_i\mid\mu_j,\sigma_j^2 I)}$$

$$=\sum_{i=1}^n \frac{\tilde{z}_{ik}}{\sigma_k^2}(x_i-\mu_k)$$

$$\Rightarrow \nabla_{\mu_k} l^{(t)}=\frac{1}{(\sigma_k^2)^{(t)}}\left[\sum_{i=1}^n \tilde{z}_{ik}^{(t+0.5)}x_i-\mu_k\sum_{i=1}^n \tilde{z}_{ik}^{(t+0.5)}\right]$$

令 $\eta_k^{(t)}=\dfrac{(\sigma_k^2)^{(t)}}{\sum_{i=1}^n \tilde{z}_{ik}^{(t+0.5)}}$

$$\mu_k^{(t+1)}=\mu_k^{(t)}+\eta_k^{(t)}\cdot\nabla_{\mu_k} l^{(t)}=\frac{\sum_{i=1}^n \tilde{z}_{ik}^{(t+0.5)}x_i}{\sum_{i=1}^n \tilde{z}_{ik}^{(t+0.5)}}$$ 与M-step中μ_k相同

同理，可对 σ_k^2 求偏导

$$\nabla_{\sigma_k^2} l = \sum_{i=1}^{n} \left(\underbrace{\frac{\pi_k N(x_i|\mu_k, \sigma_k^2 I)}{\sum_{j=1}^{k} \pi_j N(x_i|\mu_j, \sigma_j^2 I)}}_{\uparrow \text{ E-step中的 } \tilde{z}_{ik}} \cdot \left[\frac{(x_i-\mu_k)^T I (x_i-\mu_k)}{2(\sigma_k^2)^2} - \frac{p}{2\sigma_k^2} \right] \right)$$

$$= \sum_{i=1}^{n} \tilde{z}_{ik} \left[\frac{(x_i-\mu_k)^T I (x_i-\mu_k)}{2(\sigma_k^2)^2} - \frac{p}{2\sigma_k^2} \right]$$

令 $S_k^{(t)} = \dfrac{2\left[(\sigma_k^2)^{(t)}\right]^2}{p \cdot \sum_{i=1}^{n} \tilde{z}_{ik}^{(t+1)}}$

$$(\sigma_k^2)^{(t+1)} = (\sigma_k^2)^{(t)} + S_k^{(t)} \cdot \nabla_{\sigma_k^2} l^{(t)}$$

$$= \frac{\sum_{i=1}^{n} \tilde{z}_{ik}^{(t+1)} (x_i-\mu_k^{(t+1)})^T I (x_i-\mu_k^{(t+1)})}{p \cdot \sum_{i=1}^{n} \tilde{z}_{ik}^{(t+1)}}$$

与M-step中 σ_k^2 相同

Mixture of multinomial Variables

1. Write down the likelihood function

$X \sim Multinomial(\mu_1, \mu_2, \cdots, \mu_N)$

$X = (x_1, x_2, \cdots, x_N)$

$f(X \mid \mu_1, \cdots, \mu_n) = \prod_{i=1}^{N} \mu_i^{x_i}$

记 $\Theta = (\mu_1, \cdots, \mu_N)$, $\mu_i = \Theta[i]$

在混合模型中，有K种类别，记为 $\{1, 2, \cdots, K\}$,
每种类别都各自有自己的参数 Θ_{class} , $class \in \{1, \cdots K\}$
引入隐变量Y表示X对应的类别， $Y \in \{1, 2, \cdots, K\}$

考察样本容量为M的一个样本 $\mathbb{X} = (X_1, X_2, X_3, \cdots, X_M)$,
对应的隐变量向量记为 $\mathcal{Y} = (Y_1, Y_2, Y_3, \cdots, Y_M)$,
对应的样本参数集合记为 $\Theta_s = (\Theta_1, \Theta_2, \cdots, \Theta_M)$

$$f(\mathbb{X} \mid \Theta_s) = f(X_1 \mid \Theta_1) \cdot f(X_2 \mid \Theta_2) \cdots f(X_M \mid \Theta_M)$$

$$= \prod_{id=1}^{M} f(X_{id} \mid \Theta_{id})$$

$$= \prod_{id=1}^{M} \sum_{class=1}^{K} \pi_{class} \prod_{dim=1}^{N} (\Theta_{class}[dim])^{X_{id}[dim]}$$

对数似然为：

$$l(\pi, \mu) = \ln f(\mathbb{X} \mid \pi, \mu) = \sum_{id=1}^{M} \ln \sum_{class=1}^{K} \pi_{class} \prod_{dim=1}^{N} (\Theta_{class}[dim])^{X_{id}[dim]}$$

2. Write down the Q-function

$$\ln L(\Theta_1, \Theta_2, \cdots, \Theta_M \mid \mathbb{X}, \mathcal{Y}) = \ln P(\mathbb{X}, \mathcal{Y} \mid \Theta_1, \Theta_2, \cdots, \Theta_M)$$

其中：$P(\mathbb{X}, \mathcal{Y} \mid \Theta_s) = P_{\Theta_s}(\mathbb{X}, \mathcal{Y}) = P_{\Theta_s}(\mathbb{X} \mid \mathcal{Y}) P_{\Theta_s}(\mathcal{Y})$

$$= \prod_{id=1}^{M} P_{\Theta_s}(X_{id} \mid Y_{id}) \cdot P_{\Theta_s}(Y_{id})$$

$$= \prod_{id=1}^{M} f(X_{id} \mid \Theta_{Y_{id}}) \cdot \pi_{Y_{id}}$$

$$= \prod_{id=1}^{M} \left\{ \left[\prod_{dim=1}^{N} (\Theta_{Y_{id}}[dim])^{X_{id}[dim]} \right] \cdot \pi_{Y_{id}} \right\}$$

记 $\pi_s = (\pi_1, \pi_2, \cdots, \pi_K)$

$$P(Y_{id} \mid X_{id}, \Theta_s, \Pi_s) = \frac{\pi_{y_{id}} f(X_{id} \mid \Theta_{y_{id}})}{\sum_{k=1}^{K} \pi_k f(X_{id} \mid \Theta_k)}$$

$$P(Y \mid \mathbb{X}, \Theta_s, \Pi_s) = \prod_{id=1}^{M} \frac{\pi_{y_{id}} f(X_{id} \mid \Theta_{y_{id}})}{\sum_{k=1}^{K} \pi_k f(X_{id} \mid \Theta_k)}$$

$$Q[(\Theta_s, \Pi_s), (\Theta_s, \Pi_s)^g]$$
$$= \sum_{class=1}^{K} \sum_{id=1}^{M} \ln(\pi_{class}) P(class \mid X_{id}, \Theta_s^g, \Pi_s^g)$$
$$+ \sum_{class=1}^{K} \sum_{id=1}^{M} \ln f(X_{id} \mid \Theta_{class}) P(class \mid X_{id}, \Theta_s^g, \Pi_s^g)$$

3、本小题没证，pass

4、 Derive the updating rule in M-step

记 Q Function 的第一项为A

$$A=\sum_{class=1}^{K}\sum_{id=1}^{M}\ln(\pi_{class})P(class|X_{id},\Theta_s^g,\pi_s^g)$$

使用拉格朗日极值法求其在 $\sum_{class=1}^{K}\pi_{class}=1$ 时的条件极值

$$\frac{\partial}{\partial\pi_{class}}\left\{A+\lambda\left[\left(\sum_{class=1}^{K}\pi_{class}\right)-1\right]\right\}=0$$

$$\Rightarrow \sum_{id=1}^{M}\frac{1}{\pi_{class}}P(class|X_i,\Theta_s^g)+\lambda=0$$

$$\Rightarrow \pi_{class}=\frac{1}{M}\sum_{id=1}^{M}P(class|X_{id},\Theta_s^g,\pi_s^g)$$

$$=\frac{1}{M}\sum_{id=1}^{M}\frac{\pi_{class}^g\cdot f(X_{id}|\Theta_{class}^g)}{\sum_{k=1}^{K}\pi_k^g f(X_{id}|\Theta_k^g)}$$

$$=\frac{1}{M}\sum_{id=1}^{M}\frac{\pi_{class}^g\cdot\prod_{dim=1}^{N}\left(\Theta_{class}^g[dim]\right)^{X_{id}[dim]}}{\sum_{k=1}^{K}\pi_k^g\cdot\prod_{dim=1}^{N}\left(\Theta_k^g[dim]\right)^{X_{id}[dim]}}$$

记 Q Function 的第二项为B

$$B=\sum_{class=1}^{K}\sum_{id=1}^{M}\ln f(X_{id}|\Theta_{class})\cdot P(class|X_{id},\Theta_s^g,\pi_s^g)$$

$$=\sum_{class=1}^{K}\sum_{id=1}^{M}\left(\ln\prod_{dim=1}^{N}(\Theta_{Y_{id}}[dim])^{X_{id}[dim]}\right)\cdot\frac{\pi_{class}^g\cdot f(X_{id}|\Theta_{Y_{id}})}{\sum_{k=1}^{K}\pi_k^g\left[f(X_{id}|\Theta_k^g)\right]}$$

$$=\sum_{class=1}^{K}\sum_{id=1}^{M}\left[\sum_{dim=1}^{N}X_{id}[dim]\ln(\Theta_{class}[dim])\right]\cdot\underbrace{\frac{\pi_{class}^g f(X_{id}|\Theta_{Y_{id}}^g)}{\sum_{k=1}^{K}\pi_k^g f(X_{id}|\Theta_k^g)}}_{\text{常量系数}}$$

由多项分布性质 $\sum_{dim=1}^{N}X_{id}[dim]=1$，$X_{id}[dim]\in\{0,1\}$

$\therefore \exists!\ dim\in\{1,\dots,N\}$, s.t. $X_{id}[dim]=1$

因此，让 $X_{id}[dim]=1$ 的项对应的 $\Theta_{class}[dim]=1$ 即可

$$\Theta_{class}[dim]=\mathbb{I}(X_{id}[dim]=1)$$

EM Application

3.1

① 考察 $y^{(pr)} \sim N(\mu_p, \sigma_p^2)$, $z^{(pr)} \sim N(v_r, \tau_r^2)$, $\xi^{(pr)} \sim N(0, \sigma^2)$ 三者独立

易知 $x^{(pr)} \sim y^{(pr)} + z^{(pr)} + \xi^{(pr)}$

随机向量 $(y^{(pr)}, z^{(pr)}, x^{(pr)})$ 的均值为:

$$(\mu_p, v_r, \mu_p + v_r)$$

协方差矩阵为:

$$\begin{bmatrix} \sigma_p^2 & 0 & \sigma_p^2 \\ 0 & \tau_r^2 & \tau_r^2 \\ \sigma_p^2 & \tau_r^2 & \sigma_p^2+\tau_r^2+\sigma^2 \end{bmatrix}$$

②

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

$$x_1 | x_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \quad \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Q函数有如下形式:

$$Q_{pr}(y^{(pr)}, z^{(pr)}) = p(y^{(pr)}, z^{(pr)} | x^{(pr)})$$

$$= N\left([\mu_p, v_r] + \frac{x^{(pr)} - (\mu_p + v_r)}{\sigma_p^2 + \tau_r^2 + \sigma^2}[\sigma_p^2, \tau_r^2], \quad \begin{bmatrix} \sigma_p^2 & \\ & \tau_r^2 \end{bmatrix} + \frac{1}{\sigma_p^2+\tau_r^2+\sigma^2}\begin{bmatrix} \sigma_p^4 & \sigma_p^2\tau_r^2 \\ \sigma_p^2\tau_r^2 & \tau_r^4 \end{bmatrix}\right)$$

3.2 极大化Q函数可推得:

$$(y_{(t+1)}^{(pr)}, z_{(t+1)}^{(pr)}) = (\mu_p^{(t)}, v_r^{(t)}) + \frac{[(\sigma_p^2)^{(t)}, (\tau_r^2)^{(t)}](x^{(pr)} - \mu_p^{(t)} - v_r^{(t)})}{(\sigma_p^2)^{(t)} + (\tau_r^2)^{(t)} + \sigma^2}$$

$$\mu_p^{(t+1)} = \frac{1}{R}\sum_{r=1}^{R} y_{(t+1)}^{(pr)}$$

$$v_r^{(t+1)} = \frac{1}{P}\sum_{p=1}^{P} z_{(t+1)}^{(pr)}$$

$$(\sigma_p^2)^{(t+1)} = \frac{1}{R}\sum_{r=1}^{R}(y_{(t+1)}^{(pr)} - \mu_{(p)}^{(t+1)})^2$$

$$(\tau_r^2)^{(t+1)} = \frac{1}{P}\sum_{p=1}^{P}(z_{t+1}^{(pr)} - \tau_r^{(t+1)})^2$$