机器学习

苏州大学计算机科学与技术学院
自然语言处理实验室
主讲：周夏冰
邮箱：zhouxiabing@suda.edu.cn

EM

EM算法简介

- 期望最大化(Expectation Maximization)算法。
  - 迭代算法，在概率模型中寻找参数极大似然估计的算法，
  - 概率模型依赖于无法观测的隐含变量。
  - 主要用于从含有隐含变量的数据中计算极大似然估计，是解决存在隐含变量优化问题的有效方法。

回顾

- **极大似然估计**

似然函数$L(\theta)=\prod_{i=1}^{n} p(x_i;\theta)$是参数$\theta$的函数。显然随着$\theta$在参数空间$\Theta$的变化，似然函数的值也要发生变化。而极大似然估计的目的就是在样本点$\{x_1,\ldots x_n\}$固定的情况下，寻找最优的θ来极大化似然函数，即

$$\theta^* = \arg\max_{\theta\in\Theta} L(\theta)$$

回顾

- **极大似然估计**

想要求θ^*，只需要使θ的似然函数$L(\theta)$极大化，极大值对应的θ^*就是我们的估计值。

因$L(\theta)$与$\ln L(\theta)$在同一θ处取到极值，所以对数化似然函数

$$\ln(L(\theta)) = \sum_{i=1}^{n} \ln p(x_i;\theta)$$

因此θ的极大似然估计θ^*也可从下述方程解得：

$$\frac{d}{d\theta} \ln L(\theta) = 0.$$

一个例子

- **例1-1：** 考虑投硬币的实验，现在我们有两枚硬币A和B，这两枚硬币和普通的硬币不一样，他们投掷出正面的概率和投掷出反面的概率不一定相同。我们将A和B投掷正面的概率分别设为θ_A和θ_B。独立地做5次试验，随机的从这两枚硬币抽1枚，投掷10次，统计出现正面的概率，得到结果

试验代号	投掷的硬币	出现正面的次数
1	B	5
2	A	9
3	A	8
4	B	4
5	A	7

有完整数据的参数估计

一个例子

- **例1-2：** 考虑投硬币的实验，现在我们有两枚硬币A和B，这两枚硬币和普通的硬币不一样，他们投掷出正面的概率和投掷出反面的概率不一定相同。我们将A和B投掷正面的概率分别设为θ_A和θ_B。独立地做5次试验，随机的从这两枚硬币抽1枚，投掷10次，统计出现正面的概率，得到结果

不完整数据的参数估计

试验代号	投掷的硬币	出现正面的次数
1	B	5
2	A	9
3	A	8
4	B	4
5	A	7

EM算法

$$P(Y|\theta)=\prod_{i=1}^{n}[\pi p^{y_i}(1-p)^{1-y_i}+(1-\pi)q^{y_i}(1-q)^{1-y_i}]$$

- **例2**：3枚硬币，A，B，C，这些硬币出现正面的概率π, p, q。进行如下掷硬币试验：先掷A，如果正面，$\boxed{\text{选硬币B}}$，$\boxed{\text{反面选C}}$；然后投选出来的硬币，出现正面记为1，出现反面记为0；独立……验，求三枚硬币出现正面的概率

先选择硬币

- **隐变量**：观测变量 是 隐变量

在硬币下表现出的正反面

- $P(y|\theta)=\sum_z P(y,z|\theta)=\sum_z P(z|\theta)P(y|z,\theta)=$

$$\boxed{\pi p^{y}(1-p)^{1-y}}+\boxed{(1-\pi)q^{y}(1-q)^{1-y}}$$

最大似然估计—存在隐变量

- $P(Y|\theta)=\prod_{i=1}^{n}[\pi p^{y_i}(1-p)^{1-y_i}+(1-\pi)q^{y_i}(1-q)^{1-y_i}]$
- $lnP(Y|\theta)=\sum_{i=1}^{n}ln[\pi p^{y_i}(1-p)^{1-y_i}+(1-\pi)q^{y_i}(1-q)^{1-y_i}]$
- $\frac{\partial lnP(Y|\theta)}{\partial\pi}=\sum_{i=1}^{n}\frac{p^{y_i}(1-p)^{1-y_i}-q^{y_i}(1-q)^{1-y_i}}{\pi p^{y_i}(1-p)^{1-y_i}+(1-\pi)q^{y_i}(1-q)^{1-y_i}}$

最大似然估计—不存在隐变量

明确知道仍的哪个硬币

- $P(Y|\theta) = [q^{y_1}(1-q)^{1-y_1}][p^{y_2}(1-p)^{1-y_2}]\cdots$

- $lnP(Y|\theta) = \sum_{i=1}^{n1} lnp^{y_i}(1-p)^{1-y_i} + \sum_{j=1}^{n2} lnq^{y_j}(1-q)^{1-y_j}$

- $\frac{\partial lnP(Y|\theta)}{\partial p} = \sum_{i=1}^{n1} \frac{(1-\cancel{p})^{1-y_j} \cdot y_j \cdot \cancel{p^{y_j}} \cdot p^{-1} + \cancel{p^{y_j}} \cdot (y_j-1) \cdot (1-\cancel{p})^{1-y_j} \cdot (1-p)^{-1}}{\cancel{p^{y_j}}(1-\cancel{p})^{1-y_j}}$

EM算法

- EM算法包含两步：
  - E：求期望（隐变量的条件概率分布的期望）$Q(\theta,\theta^i)$
  - M：求极大化 $\frac{Q}{\partial\theta}=0$
- （E-step）如果参数已知，根据训练数据推断出最优隐变量的值
- （M-step）如果隐变量已知，对参数做极大似然估计

EM算法过程

- **EM算法包含两步：**
  - E-step: 记θ^i为第i次迭代参数θ的估计值，在第$i+1$次迭代的E步，计算

    $$Q(\theta,\theta^i)=E_Z\left[\log P(Y,Z|\theta)\,\middle|Y,\theta^i\right]=\sum_Z P(Z|Y,\boxed{\theta^i})\log P(Y,Z|\theta)$$

  - M-step: 求使$Q(\theta,\theta^i)$极大化的θ，确定第i+1次迭代的参数的估计值θ^{i+1}，

    $$\theta^{i+1}=argmax_\theta Q(\theta,\theta^i)$$

EM算法的导出

- 目的：求$L(\theta) = \log P(Y|\theta)$
  - $L(\theta) = \log P(Y|\theta) = \log \sum_Z P(Y, Z|\theta)$

$$= \log \sum_Z P(Y|Z, \theta) P(Z|\theta)$$

EM算法的导出

$$P(Y|\theta) = \prod_{i=1}^{n}[\pi p^{y_i}(1-p)^{1-y_i} + (1-\pi)q^{y_i}(1-q)^{1-y_i}]$$

$$\theta = (\pi, p, q)$$

- 中心目标：极大化
  - $L(\theta) = log\sum_Z P(Y|Z,\theta)P(Z|\theta)$
- 新的θ，希望能使$L(\theta)$增加
  - 在第i次的参数为θ^i
  - $L(\theta) - L(\theta^i) = log\sum_Z P(Y|Z,\theta)P(Z|\theta) - logP(Y|\theta^i)$
  - $= log\sum_Z P(Z|Y,\theta^i)\frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^i)} - logP(Y|\theta^i)$

EM算法的导出

Jesen不等式：$\log\sum_j \lambda_j y_j \geq \sum_j \lambda_j \log y_j, \sum_j \lambda_j = 1$

- $log\boxed{\sum_Z P(Z|Y,\theta^i)}\frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^i)}$
- $\geq \sum_Z P(Z|Y,\theta^i)log\frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^i)}$

EM算法的导出

- 中心目标：极大化
  - $L(\theta) = log \sum_Z P(Y|Z,\theta)P(Z|\theta)$
- 新的θ，希望能使$L(\theta)$增加
  - 在第i次的参数为θ^i
  - $L(\theta) - L(\theta^i) = log \sum_Z P(Y|Z,\theta)P(Z|\theta) - logP(Y|\theta^i)$
  - $= log \sum_Z P(Z|Y,\theta^i) \frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^i)} - logP(Y|\theta^i)$
  - $\geq \sum_Z P(Z|Y,\theta^i) log \frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^i)P(Y|\theta^i)}$

令$B(\theta,\theta^i) = L(\theta^i) + \sum_Z P(Z|Y,\theta^i) log \frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^i)P(Y|\theta^i)}$

$L(\theta) \geq B(\theta,\theta^i)$

EM算法的导出

- $L(\theta) \geq B(\theta, \theta^i)$
- $B(\theta, \theta^i)$是一个下界
- 如果使$B(\theta, \theta^i)$增大的θ，也可以使$L(\theta)$增大
- 为了使$L(\theta)$尽可能增大，有$\theta^{i+1} = \underset{\theta}{\operatorname{argmax}} B(\theta, \theta^i)$
  - $\underset{\theta}{\operatorname{argmax}} L(\theta^i) + \sum_Z P(Z|Y, \theta^i) log \frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^i)P(Y|\theta^i)}$

EM算法的导出

- $\arg\max_{\theta} L(\theta^i) + \sum_Z P(Z|Y,\theta^i) log \frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^i)P(Y|\theta^i)}$
- $= \arg\max_{\theta} \sum_Z P(Z|Y,\theta^i) \log P(Y|Z,\theta)P(Z|\theta)$
- $= \arg\max_{\theta} \sum_Z P(Z|Y,\theta^i) \log P(Y,Z|\theta)$

$$Q(\theta,\theta^i) = E_Z[log\, P(Y,Z|\theta)\,|Y,\theta^i]$$
$$= \sum_Z log\, P(Y,Z|\theta)\, P(Z|Y,\theta^i)$$

EM算法

- EM算法包含两步：
  - E-step: 记$\boldsymbol{\theta}^{i}$为第$\boldsymbol{i}$次迭代参数$\boldsymbol{\theta}$的估计值，在第$\boldsymbol{i}+\mathbf{1}$次迭代的E步，计算

$$\boldsymbol{Q}\left(\boldsymbol{\theta},\boldsymbol{\theta}^{i}\right)=\boldsymbol{E}_{\boldsymbol{Z}}\left[\log \boldsymbol{P}(\boldsymbol{Y},\boldsymbol{Z}|\boldsymbol{\theta})\,|\boldsymbol{Y},\boldsymbol{\theta}^{i}\right]=\sum_{\boldsymbol{Z}}\boldsymbol{P}(\boldsymbol{Z}|\boldsymbol{Y},\boldsymbol{\theta}^{i})\log \boldsymbol{P}(\boldsymbol{Y},\boldsymbol{Z}|\boldsymbol{\theta})$$

给定当前观测变量和参数$\boldsymbol{\theta}^{i}$, Z的条件概率分布

  - M-step: 求使$\boldsymbol{Q}\left(\boldsymbol{\theta},\boldsymbol{\theta}^{i}\right)$极大化的$\boldsymbol{\theta}$

$$\boldsymbol{\theta}=\underset{\boldsymbol{\theta}}{\mathbf{argmax}}\,\boldsymbol{Q}\left(\boldsymbol{\theta},\boldsymbol{\theta}^{i}\right)$$

EM算法

- **说明：**
  - 对初值敏感
  - 局部极值
  - 停止迭代条件
    - $\|\theta^{i+1} - \theta^{i}\| < \varepsilon$
    - $\|Q(\theta^{i+1}, \theta^{i}) - Q(\theta^{i}, \theta^{i})\| < \varepsilon$

例子—扔硬币

$$P(事件A|观测y_j)=\frac{P(观测y_j|事件A)P(事件A)}{P(观测y_j)}$$

- 例2：3枚硬币，A，B，C，这些硬币出现正面的概率 π,p,q 。进行如下掷硬币试验：先掷A，如果正面，选硬币B，反面选C；然后投选出来的硬币，出现正面记为1，出现反面记为0；独立反复做n次试验，求三枚硬币出现正面的概率

- 隐变量：投掷的是B还是C

K表示第几轮参数值

- 硬币来自B: $\mu_j^k=P(Z=B|Y,\theta^{k-1})$

$$=\frac{\underbrace{\pi^{k-1}}_{硬币来自B的概率}\underbrace{(p^{k-1})^{y_j}(1-p^{k-1})^{1-y_j}}_{硬币来自B，观测y_j的概率}}{\underbrace{\pi^{k-1}(p^{k-1})^{y_j}(1-p^{k-1})^{1-y_j}+(1-\pi^{k-1})(q^{k-1})^{y_j}(1-q^{k-1})^{1-y_j}}_{全概率}}$$

- 硬币来自C: $1-\mu_j^k$

常数

例子—扔硬币

- $Q(\theta,\theta^i) = \sum_Z P(Z|Y,\theta^k)\, log\, P(Y,Z|\theta)$
- $= \sum_Z P(Z|Y,\theta^k)\, log\, P(Z|\theta)P(Y|Z,\theta)$
- $\sum_Z$包含了两种情况（来自B硬币or来自C硬币）
- $\sum_{j=1}^{N}\{\mu_j^k\, log[\pi p^{y_j}(1-p)^{1-y_j}] + (1-\mu_j^k)\, log[(1-\pi)q^{y_j}(1-q)^{1-y_j}]\}$
- 最大化：对参数求导

例子—扔硬币

- $\sum_{j=1}^{N}\{\mu_j^k\, log\left[\pi p^{y_j}(1-p)^{1-y_j}\right] + \left(1-\mu_j^k\right) log\left[(1-\pi)q^{y_j}(1-q)^{1-y_j}\right]\}$
- $\sum_{j=1}^{N}\{\mu_j^k\, log\left[\pi p^{y_j}(1-p)^{1-y_j}\right] + \left(1-\mu_j^k\right) log\left[(1-\pi)q^{y_j}(1-q)^{1-y_j}\right]\}$
- $\frac{\sum_{j=1}^{N}\mu_j^k}{\pi} - \frac{\sum_{j=1}^{N}\left(1-\mu_j^k\right)}{1-\pi} = 0$
- $\sum_{j=1}^{N}\mu_j^k - \sum_{j=1}^{N}\mu_j^k\pi - n\pi + \sum_{j=1}^{N}\mu_j^k\pi = 0$
- $\pi = \frac{1}{n}\sum_{j=1}^{N}\mu_j^k$

例子—扔硬币

- $\sum_{j=1}^{N}\{\mu_j^k \, log[\pi p^{y_j}(1-p)^{1-y_j}] + (1-\mu_j^k)\, log[(1-\pi)q^{y_j}(1-q)^{1-y_j}]\}$

- $\sum_{j=1}^{N}\mu_j^k \cdot \frac{\cancel{(1-p)^{1-y_j}}\cdot y_j\cdot \cancel{p^{y_j}}\cdot p^{-1}+\cancel{p^{y_j}}\cdot(y_j-1)\cdot\cancel{(1-p)^{1-y_j}}\cdot(1-p)^{-1}}{\cancel{p^{y_j}}\cancel{(1-p)^{1-y_j}}} = 0$

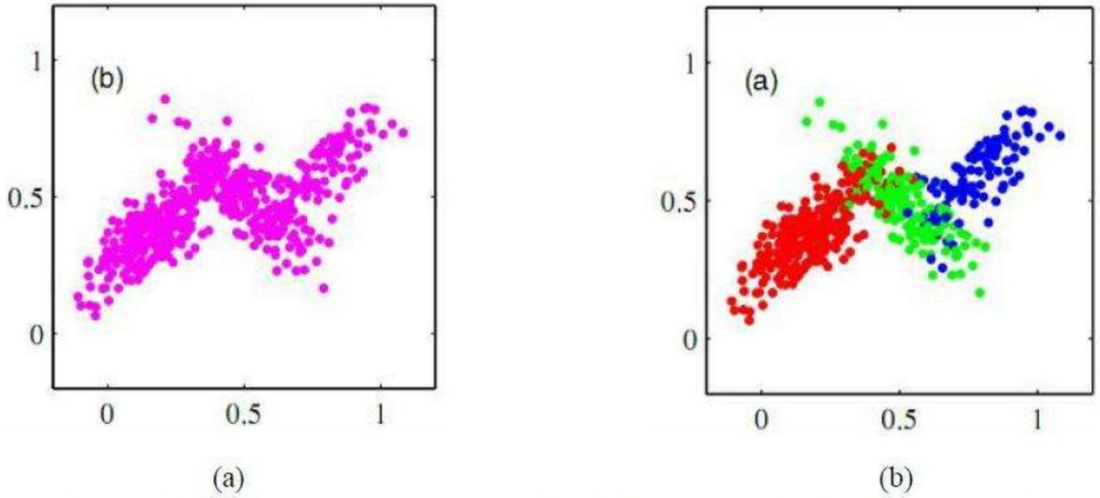
- $\sum_{j=1}^{N}\mu_j^k \cdot \frac{y_j\,(1-p)+(y_j-1)p}{p(1-p)} = 0$

- $\sum_{j=1}^{N}\mu_j^k y_j = \sum_{j=1}^{N}\mu_j^k p$

- $p = \frac{\sum_{j=1}^{N}\mu_j^k y_j}{\sum_{j=1}^{N}\mu_j^k}$

高斯混合模型

- 高斯混合模型（Gussian misture model）
  - 一个事物分解为若干的基于高斯概率密度函数（正态分布曲线）形成的模型
  - 公式表达：$P(y|\theta) = \sum_{k=1}^{K} \alpha_k \emptyset(y|\theta_k)$
  - $\theta = (\alpha_1, \alpha_2, \cdots, \alpha_k; \theta_1, \theta_2, \cdots, \theta_k)$
  - $\emptyset(y|\theta_k) = \frac{1}{\sqrt{2\pi}\sigma_k} exp(-\frac{(y-\mu_k)^2}{2\sigma_k^2})$



(a) (b)

图 1 高斯混合模型图示，(a)表示所有样本数据；(b)表示已经明确了样本的分类

高斯混合模型

$$P(\text{事件}A|\text{观测}y_j)=\frac{P(\text{观测}y_j|\text{事件}A)P(\text{事件}A)}{P(\text{观测}y_j)}$$

- 定义隐变量

$$\theta^{i-1}=\{\theta_1^{i-1},\theta_2^{i-1},\cdots,\theta_K^{i-1}\}$$

$$\theta_k^{i-1}=\{\alpha_k^{i-1},\sigma_k^{i-1},\mu_k^{i-1}\}$$

已知量

- $\gamma_{jk}=\begin{cases}1,\text{第}j\text{个观测来自第}k\text{个分模型}\\0,\text{否则}\end{cases}$

- $\hat{\gamma}_{jk}=P(\gamma_{jk}=1|y_j,\theta_k^{i-1})$

i-1表示第i-1轮参数

- $P(\gamma_{jk}=1|y_j,\theta_k^{i-1})=\dfrac{P(\gamma_{jk}=1,y_j|\theta_k^{i-1})}{P(y_j|\theta_k^{i-1})}=\dfrac{\overbrace{\boxed{P(y_j|\gamma_{jk}=1,\theta_k^{i-1})}}^{\emptyset(y|\theta_k)}\overbrace{\boxed{P(\gamma_{jk}=1|\theta_k^{i-1})}}^{\alpha_k}}{\boxed{P(y_j|\theta_k^{i-1})}}$

$$P(y_j|\theta)=\sum_{k=1}^{K}\alpha_k\emptyset(y_j|\theta_k)$$

$$=\frac{\emptyset(y_j|\theta_k^{i-1})\cdot\alpha_k}{\sum_{k=1}^{K}\alpha_k\emptyset(y_j|\theta_k^{i-1})}$$

高斯混合模型

$$\sum_{k=1}^{K}\left\{\sum_{j=1}^{N}\hat{\gamma}_{jk}\log\left(\alpha_k\cdot\frac{1}{\sqrt{2\pi}\sigma_k}exp(-\frac{(y_j-\mu_k)^2}{2\sigma_k^2})\right)\right\}$$

- $Q(\theta,\theta^i)=\sum_Z P(Z|Y,\theta^{i-1})\,log\,P(Y,Z|\theta)$
- $P(Z|Y,\theta^{i-1})=P(\gamma_{jk}|y_j,\theta_k^{i-1})=\hat{\gamma}_{jk}$
- $log\,P(Y,Z|\theta)=log\,P(Z|\theta)P(Y|Z,\theta)$
- $P(Z|\theta)=P(\gamma_{jk}|\theta_k)=\alpha_k$
- $P(Y|Z,\theta)=P(y_j|\gamma_{jk},\theta_k)=\frac{1}{\sqrt{2\pi}\sigma_k}exp(-\frac{(y_j-\mu_k)^2}{2\sigma_k^2})$

高斯混合模型

- $\sum_{k=1}^{K}\left\{\sum_{j=1}^{N}\hat{\gamma}_{jk}\log\left(\alpha_k\cdot\frac{1}{\sqrt{2\pi}\sigma_k}exp\left(-\frac{\left(y_j-\mu_k\right)^2}{2\sigma_k^2}\right)\right)\right\}$

- $\sum_{k=1}^{K}\left\{\sum_{j=1}^{N}\hat{\gamma}_{jk}\,log(\alpha_k)+\sum_{j=1}^{N}\hat{\gamma}_{jk}\left(\log\frac{1}{\sqrt{2\pi}\sigma_k}-\frac{\left(y_j-\mu_k\right)^2}{2\sigma_k^2}\right)\right\}$

高斯混合模型

$$\sum_{j=1}^{N} \hat{\gamma}_{jk}\, log(\alpha_k) + \lambda(\alpha_k - 1)$$

- $\sum_{k=1}^{K}\left\{\sum_{j=1}^{N} \hat{\gamma}_{jk}\, log(\alpha_k) + \sum_{j=1}^{N} \hat{\gamma}_{jk}{}_{jk}\left(\log\frac{1}{\sqrt{2\pi}\sigma_k} - \frac{(y_j-\mu_k)^2}{2\sigma_k^2}\right)\right\}$
- 注意约束： $\sum_{k=1}^{K} \alpha_k = 1$
- $J = \sum_{k=1}^{K}\sum_{j=1}^{N} \hat{\gamma}_{jk}\, log(\alpha_k) + \lambda\left(\sum_{k=1}^{K} \alpha_k - 1\right)$
- $\frac{\partial J}{\partial \alpha_k} = \sum_{j=1}^{N} \hat{\gamma}_{jk} + \lambda\alpha_k = 0$ $\Rightarrow$ $\sum_{k=1}^{K}\left(\sum_{j=1}^{N} \hat{\gamma}_{jk} + \lambda\alpha_k\right) = 0$ $\Rightarrow$ $\boxed{\lambda = -N}$
- $\alpha_k = \frac{\sum_{j=1}^{N} \hat{\gamma}_{jk}}{N}$

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

$$\sum_{k=1}^{K} \hat{\gamma}_{jk} = 1$$

高斯混合模型

- $\sum_{k=1}^{K}\left\{\sum_{j=1}^{N}\hat{\gamma}_{jk}\, log(\alpha_k)+\sum_{j=1}^{N}\hat{\gamma}_{jk}(\log\frac{1}{\sqrt{2\pi}\sigma_k}-\frac{(y_j-\mu_k)^2}{2\sigma_k^2})\right\}$
- $J=\sum_{k=1}^{K}\left\{\sum_{j=1}^{N}\hat{\gamma}_{jk}(-\frac{(y_j-\mu_k)^2}{2\sigma_k^2})\right\}$
- $\frac{\partial J}{\partial \mu_k}=2\sum_{j=1}^{N}\hat{\gamma}_{jk}(y_j-\mu_k)=0$
- $\mu_k=\frac{\sum_{j=1}^{N}\hat{\gamma}_{jk}y_j}{\sum_{j=1}^{N}\hat{\gamma}_{jk}}$

$$\sum_{j=1}^{N}\hat{\gamma}_{jk}(-(y_j-\mu_k)^2)$$

高斯混合模型

- $\sum_{k=1}^{K}\left\{\sum_{j=1}^{N}\hat{\gamma}_{jk}\,log(\alpha_k)+\sum_{j=1}^{N}\hat{\gamma}_{jk}(\log\frac{1}{\sqrt{2\pi}}-\frac{1}{2}\log\sigma_k^2-\frac{(y_j-\mu_k)^2}{2\sigma_k^2})\right\}$
- $J=\sum_{k=1}^{K}\left\{\sum_{j=1}^{N}\hat{\gamma}_{jk}(-\frac{1}{2}\log\sigma_k^2-\frac{(y_j-\mu_k)^2}{2\sigma_k^2})\right\}$
- $\frac{\partial J}{\partial\sigma_k^2}=\sum_{j=1}^{N}\hat{\gamma}_{jk}\left(-\frac{1}{2}\cdot\frac{1}{\sigma_k^2}+\frac{1}{2}\cdot\frac{(y_j-\mu_k)^2}{(\sigma_k^2)^2}\right)=0$
- $\sigma_k^2=\frac{\sum_{j=1}^{N}\hat{\gamma}_{jk}(y_j-\mu_k)^2}{\sum_{j=1}^{N}\hat{\gamma}_{jk}}$

$$\sum_{j=1}^{N}\left(\hat{\gamma}_{jk}-\frac{(y_j-\mu_k)^2}{\sigma_k^2}\right)=0$$

EM

- **EM 算法可以用到朴素贝叶斯法的无监督学习。**
  - **考虑无监督文本二分类问题，即训练数据包含m个文档$\{z^1, z^2, \cdots, z^m\}$，无监督场景下每个文档的类别未知，已知词表总数为d，每个文档数据均由d维向量表示**

$$(\mathbf{z^i} = \{x_1^i, x_2^i, \cdots, x_d^i\}, \quad x_j^i = \begin{cases} 0 & z^i\text{内不包含第}j\text{个词} \\ 1 & z^i\text{内包含第}j\text{个词} \end{cases})$$

考虑朴素贝叶斯判断公式： $p(c|x) \approx \prod_{j=1}^{d} p(x_j|c)p(c)$

定义如下模型参数：

$\phi_{z^i} = p(z^i = 1)$，文档i属于第1类

$\varphi_{j|z^i=1} = p(x_j^i = 1|z^i = 1)$，当$i$文档属于第1类时，第$j$个特征出现

$\varphi_{j|z^i=0} = p(x_j^i = 1|z^i = 0)$，当$i$文档属于第2类时，第$j$个特征出现