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

前情回顾

- 从含有隐含变量的数据中计算极大似然估
  - 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
      - $Q(\theta,\theta^i)=E_Z[\log P(Y,Z|\theta)\,|Y,\theta^i]=\sum_Z \log P(Y,Z|\theta)\,P(Z|Y,\theta^i)$
    - 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}，
      - $\theta^{i+1}=argmax_\theta Q(\theta,\theta^i)$

如果参数已知，计算隐变量的“期望”分布

如果隐变量已知，对参数做极大似然估计

前情回顾

- $L(\theta) = \log P(Y|\theta) = \log \sum_Z P(Y, Z|\theta) = \log \sum_Z P(Y|Z,\theta)P(Z|\theta)$
- $L(\theta) - L(\theta^i) = \ log \sum_Z P(Y|Z,\theta)P(Z|\theta) - log P(Y|\theta^i)$

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

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

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

$$Q=\sum_Z P\left(Z\middle|Y,\theta^i\right)\log P(Y|Z,\theta)P(Z|\theta)$$

$$P\left(Z\middle|Y,\theta^i\right)$$

扔硬币例子：扔B（A硬币是正面π）或者C（A硬币是反面$1-\pi$）

$$P(Z=B|Y,\theta^{i-1})$$

$$=\frac{p\left(Y\middle|Z=B,\theta^{i-1}\right)p(Z=B|\theta^{i-1})}{p(Y|\theta^{i-1})}$$

$$=\frac{\pi^{k-1}\left(p^{k-1}\right)^{y_j}\left(1-p^{k-1}\right)^{1-y_j}}{\pi^{k-1}(p^{k-1})^{y_j}(1-p^{k-1})^{1-y_j}+(1-\pi^{k-1})(q^{k-1})^{y_j}(1-q^{k-1})^{1-y_j}}$$

混合高斯：观测变量y来自哪个高斯分模型

$$P\left(\gamma_{jk}=1\middle|y_j,\theta_k^{i-1}\right)$$

$$=\frac{P\left(y_j|\gamma_{jk}=1,\theta_k^{i-1}\right)P(\gamma_{jk}=1|\theta_k^{i-1})}{P(y_j|\theta_k^{i-1})}$$

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

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

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

扔硬币例子：扔B（A硬币是正面π）或者C（A硬币是反面 $1-\pi$ ）

$$= log\left[\pi p^{y_j}(1-p)^{1-y_j}\right]$$

or

$$= log\left[(1-\pi)q^{y_j}(1-q)^{1-y_j}\right]$$

混合高斯：观测变量y来自哪个高斯分模型

$$= \frac{1}{\sqrt{2\pi}\sigma_k} exp\left(-\frac{(y_j-\mu_k)^2}{2\sigma_k^2}\right) \alpha_k$$

$$Q=\sum_Z P\left(Z\middle|Y,\theta^i\right)\log P(Y|Z,\theta)P(Z|\theta)$$

扔硬币例子：扔B（A硬币是正面π）或者C（A硬币是反面 $1-\pi$ ）

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

$$\frac{\partial Q}{\partial \pi}=0 \qquad \frac{\partial Q}{\partial p}=0 \qquad \frac{\partial Q}{\partial q}=0$$

混合高斯：观测变量y来自哪个高斯分模型

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

$$\underbrace{\frac{\partial Q}{\partial \alpha_k}=0 \qquad \sum_{k=1}^{K}\alpha_k=1}_{\frac{\partial(Q+\lambda(\sum_{k=1}^{K}\alpha_k-1))}{\partial \alpha_k}=0} \qquad \frac{\partial Q}{\partial \mu_k}=0 \qquad \frac{\partial Q}{\partial \sigma_k^2}=0$$

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

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

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

宋祖儿/出席/活动/同框/任嘉伦/宣传/新剧

[1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0]

沙县小吃/进驻/中东/沙特/成为/传播/中华/美食/与/文化/的/窗口

[0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1]

出席，成为，传播，窗口，活动，进驻，美食，任嘉伦，宋祖儿，沙县小吃，沙特，同框，文化，宣传，新剧，中东

z^i $\quad \mathrm{p}(c_1|z^i) = \prod_{j=1}^{d} \boxed{p(x_j^i|c_1)} p(c_1)$

$p(x_j^i = 1|c_1)$

$p(x_j^i = 0|c_1)$

$1 - p(x_j^i = 1|c_1)$

定义参数θ：

$\phi_{z^i} = p(z^i = 1)$，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

- **EM算法**

$\phi_{z^i} = p(z^i = 1)$，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

$$Q = \sum_i \sum_{z^i} p\left(z^i \middle| x, \theta^k\right) \log p(x, z^i|\theta)$$

对于文档i，有

$$p\left(z^i = 1 \middle| x^i, \theta^k\right) = \frac{p\left(z^i = 1, x^i \middle| \theta^k\right)}{p\left(x^i, \theta^k\right)} = \frac{p\left(z^i = 1 \middle| \theta^k\right) \boxed{p\left(x^i \middle| z^i = 1, \theta^k\right)}}{p\left(x^i, \theta^k\right)}$$

$$p\left(x^i \middle| z^i = 1, \theta^k\right) = \prod_{j=1}^{d} p\left(x_j^i = 1 \middle| z^i = 1, \theta^k\right)^{x_j^i} p\left(x_j^i = 0 \middle| z^i = 1, \theta^k\right)^{1-x_j^i}$$

- **EM算法**

$\boldsymbol{\phi_{z^i}} = p(z^i = 1)$，文档$i$属于第1类

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

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

$$Q = \sum_i \sum_{z^i} p\left(z^i \middle| x, \theta^k\right) \log p(x, z^i|\theta)$$

对于文档i，有

$$p\left(z^i = 1 \middle| x^i, \theta^k\right) = \frac{\prod_{j=1}^d \varphi_{j|z^i=1}^{x_j^i} \left(1 - \varphi_{j|z^i=1}\right)^{1-x_j^i} \phi_{z^i}}{\prod_{j=1}^d \varphi_{j|z^i=1}^{x_j^i} \left(1 - \varphi_{j|z^i=1}\right)^{1-x_j^i} \phi_{z^i} + \prod_{j=1}^d \varphi_{j|z^i=0}^{x_j^i} \left(1 - \varphi_{j|z^i=0}\right)^{1-x_j^i} (1 - \phi_{z^i})}$$

$\boldsymbol{w^i}$

- **EM算法**

$\boldsymbol{\phi_{z^i}} = p(z^i = 1)$，文档$i$属于第1类

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

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

$$Q = \sum_i \sum_{z^i} p\left(z^i \middle| x, \theta^k\right) \log p(x, z^i|\theta)$$

对于文档i，有

$$\log p(x^i, z^i = 1|\theta) = \log p\left(x^i|z^i = 1, \theta\right) p(z^i = 1|\theta)$$

$$= \log \prod_{j=1}^{d} p\left(x_j^i = 1\middle|z^i = 1, \theta\right)^{x_j^i} p\left(x_j^i = 0\middle|z^i = 1, \theta\right)^{1-x_j^i} p(z^i = 1|\theta)$$

$$= \log \prod_{j=1}^{d} \varphi_{j|z^i=1}^{x_j^i} \left(1 - \varphi_{j|z^i=1}\right)^{1-x_j^i} \phi_{z^i} = \sum_{j=1}^{d} \left[x_j^i \log \varphi_{j|z^i=1} + (1 - x_j^i) \log(1 - \varphi_{j|z^i=1})\right] + \log \phi_{z^i}$$

- **EM算法**

$\boldsymbol{\phi}_{z^i} = p(z^i = 1)$，文档$i$属于第1类

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

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

$$Q = \sum_i \sum_{z^i} p\left(z^i \middle| x, \theta^k\right) \log p(x, z^i|\theta)$$

$$\begin{cases} \boldsymbol{w^i} \\ \sum_{j=1}^d \left[x_j^i \log \varphi_{j|z^i=1} + \left(1 - x_j^i\right) \log(1 - \varphi_{j|z^i=1})\right] + \log \phi_{z^i} \end{cases}$$

对于文档i，有

$$Q = \sum_{i=1}^m \left\{w^i \left[\sum_{j=1}^d \left[x_j^i \, log \, \varphi_{j|z^i=1} + (1 - x_j^i) \, log(1 - \varphi_{j|z^i=1})\right] + log \, \phi_{z^i}\right] + (1 - w^i)\left[\sum_{j=1}^d \left[x_j^i \, log \, \varphi_{j|z^i=0} + +(1 - x_j^i) \, log\left(1 - \varphi_{j|z^i=0}\right)\right] + log\left(1 - \phi_{z^i}\right)\right]\right\}$$