深度之眼
deepshare.net

法律声明

本课件包括演示文稿、示例、代码、题库、视频和声音等内容，深度之眼和讲师拥有完全知识产权；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或者机构不得盗版、复制、仿造其中的创意和内容，我们保留一切通过法律手段追究违反者的权利。

课程详情请咨询

- 微信公众号：深度之眼
- 客服微信号：deepshare0920

公众号　　微 信

deepshare.net
深度之眼

西瓜书公式推导

导师：Sm1les（Datawhale南瓜书项目负责人）

deepshare.net
深度之眼

EM算法公式推导

Derivation of Expectation-maximization algorithm

本节大纲

Outline

先修内容：西瓜书7.6、统计学习方法9.1

1.EM算法的引入

- 为什么需要EM算法

2.EM算法的例子

- 三硬币模型

3.EM算法的导出

- Jensen（琴生）不等式
- EM算法的推导

4.EM算法求解例子

- 用EM算法求解三硬币模型

EM算法的引入

Introduction of EM algorithm

为什么需要EM算法：

概率模型有时既含有观测变量，又含有隐变量或者潜在变量。如果概率模型的变量都是观测变量，那么给定数据，可以直接用极大似然估计法，或者贝叶斯估计法估计模型参数。但是，当模型含有隐变量时，就不能简单地使用这些估计方法。EM算法就是含有隐变量的概率模型参数的极大似然估计法。

EM算法的例子

Example of EM algorithm

deepshare.net 深度之眼

《统计学习方法》例9.1（三硬币模型）：

假设有3枚硬币，分别记作A，B，C。这些硬币正面出现的概率分别是 π ， p 和 q 。进行如下掷硬币试验： 先掷硬币A，根据其结果选出硬币B或硬币C，正面选硬币B，反面选硬币C；然后掷选出的硬币，掷硬币的结果，出现正面记作1，出现反面记作0；独立地重复n次实验（这里，n=10），观测结果如下：

1，1，0，1，0，0，1，0，1，1

假设只能观测到掷硬币的结果，不能观测掷硬币的过程。问如何估计三硬币正面出现的概率，即三硬币模型的参数。

《统计学习方法》例9.1（三硬币模型）：

假设有3枚硬币，分别记作A，B，C。这些硬币正面出现的概率分别是 π， p 和q 。进行如下掷硬币试验： 先掷硬币A，更具其结果选出硬币B或硬币C，正面选硬币B，反面选硬币C；然后掷选出的硬币，掷硬币的结果，出现正面记作1，出现反面记作0；独立地重复n次实验（这里，n=10），观测结果如下：

1，1，0，1，0，0，1，0，1，1

假设只能观测到掷硬币的结果，不能观测掷硬币的过程。问如何估计三硬币正面出现的概率，即三硬币模型的参数。

解：对每一次试验可以进行如下建模

$$
\begin{aligned}
P(y|\theta) &= \sum_z P(y,z|\theta) = \sum_z P(z|\theta)P(y|z,\theta) \\
&= P(z=1|\theta)P(y|z=1,\theta) + P(z=0|\theta)P(y|z=0,\theta) \\
&= \begin{cases} \pi p + (1-\pi)q, & \text{if } y=1; \\ \pi(1-p) + (1-\pi)(1-q), & \text{if } y=0; \end{cases} \\
&= \pi p^y (1-p)^{1-y} + (1-\pi)q^y(1-q)^{1-y}
\end{aligned}
$$

其中，随机变量y是观测变量，表示一次试验观测的结果是1或0；随机变量z是隐变量，表示未观测到的掷硬币A的结果；θ=(π,p,q)是模型参数。将观测数据表示为 $Y=(Y_1,Y_2,...,Y_n)^T$，未观测数据表示为 $Z=(Z_1,Z_2,...,Z_n)^T$

则观测数据的似然函数为：

$$
P(Y|\theta) = \sum_Z P(Z|\theta)P(Y|Z,\theta) = \prod_{j=1}^{n} P(y_j|\theta) = \prod_{j=1}^{n} \left[\pi p^{y_j}(1-p)^{1-y_j} + (1-\pi)q^{y_j}(1-q)^{1-y_j}\right]
$$

$$P(Y|\theta)=\prod_{j=1}^{n}\left[\pi p^{y_j}(1-p)^{1-y_j}+(1-\pi)q^{y_j}(1-q)^{1-y_j}\right]$$

考虑求模型参数θ=(π,p,q)的极大似然估计，即使用对数似然函数来进行参数估计可得：

$$\begin{aligned}\hat{\theta}&=\arg\max_{\theta}\ln P(Y|\theta)\\&=\arg\max_{\theta}\ln\prod_{j=1}^{n}\left[\pi p^{y_j}(1-p)^{1-y_j}+(1-\pi)q^{y_j}(1-q)^{1-y_j}\right]\\&=\arg\max_{\theta}\sum_{j=1}^{n}\ln\left[\pi p^{y_j}(1-p)^{1-y_j}+(1-\pi)q^{y_j}(1-q)^{1-y_j}\right]\end{aligned}$$

上式没有解析解，也就是没办法直接解出π,p,q恰好等于某个常数，只能用迭代的方法来进行求解。

结 语

在这次课程中，我们学习了

EM算法的引入和例子

那么在下次课程中，我们将会学习

EM算法的导出