极大似然估计

七月算法 邹博

2015年10月17日

贝叶斯公式带来的思考 $P(A|D)=\frac{P(D|A)P(A)}{P(D)}$

- 给定某些样本D，在这些样本中计算某结论A_1、A_2……A_n出现的概率，即$P(A_i|D)$

$$\max P(A_i|D)=\max\frac{P(D|A_i)P(A_i)}{P(D)}=\max\left(P(D|A_i)P(A_i)\right)\to\max P(D|A_i)$$

$$\Rightarrow \max P(A_i|D)\to\max P(D|A_i)$$

  - 第一个等式：贝叶斯公式；
  - 第二个等式：样本给定，则对任何A_i,$P(D)$是常数；
  - 第三个箭头：若这些结论A_1、A_2……A_n的先验概率相等(或近似)，则得到最后一个等式：即第二行的公式。

极大似然估计

- 设总体分布为f(x, θ)，$X_1, X_2 \ldots X_n$为该总体采样得到的样本。因为$X_1,X_2\ldots X_n$独立同分布，于是，它们的联合密度函数为：

$$L(x_1,x_2,\cdots,x_n;\theta_1,\theta_2,\cdots\theta_k)=\prod_{i=1}^{n}f(x_i;\theta_1,\theta_2,\cdots\theta_k)$$

- 这里，θ被看做固定但未知的参数；反过来，因为样本已经存在，可以看成$x_1,x_2\ldots x_n$是固定的，L(x, θ)是关于θ的函数，即似然函数。
- 求参数θ的值，使得似然函数取极大值，这种方法就是极大似然估计。

极大似然估计的具体实践操作

- 在实践中，由于求导数的需要，往往将似然函数取对数，得到对数似然函数；若对数似然函数可导，可通过求导的方式，解下列方程组，得到驻点，然后分析该驻点是极大值点

$$\log L(\theta_1,\theta_2,\cdots\theta_k)=\sum_{i=1}^{n}\log f(x_i;\theta_1,\theta_2,\cdots\theta_k)$$

$$\frac{\partial L(\theta)}{\partial \theta_i}=0,\qquad i=1,2,\cdots k$$

极大似然估计

- 找出与样本的分布最接近的概率分布模型。
- 简单的例子
  - 10次抛硬币的结果是：正正反正正正反反正正
- 假设p是每次抛硬币结果为正的概率。则：
- 得到这样的实验结果的概率是：

$$
\begin{aligned}
P &= pp(1-p)ppp(1-p)(1-p)pp \\
&= p^7(1-p)^3
\end{aligned}
$$

极大似然估计MLE

- 目标函数：$\max P = \max\limits_{0 \le p \le 1} p^7 (1-p)^3$
- 最优解是：p=0.7
  - 思考：如何求解？

- 一般形式：$L_{\overline{p}} = \prod\limits_x p(x)^{\overline{p}(x)}$

$p(x)$模型是估计的概率分布

$\overline{p}(x)$是实验结果的分布

正态分布的极大似然估计

- 若给定一组样本$X_1,X_2\ldots X_n$，已知它们来自于高斯分布$N(\mu,\sigma)$，试估计参数μ,σ。

按照MLE的过程分析

- 高斯分布的概率密度函数：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- 将X_i的样本值x_i带入，得到：

$$L(x)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

化简对数似然函数

$$l(x) = \log \prod_i \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$= \sum_i \log \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$= \left(\sum_i \log \frac{1}{\sqrt{2\pi}\sigma} \right) + \left(\sum_i -\frac{(x_i-\mu)^2}{2\sigma^2} \right)$$

$$= -\frac{n}{2} \log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \sum_i (x_i-\mu)^2$$

参数估计的结论

- 目标函数 $$l(x) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_i (x_i - \mu)^2$$

- 将目标函数对参数μ，σ分别求偏导，很容易得到μ，σ的式子：

$$\mu = \frac{1}{n}\sum_i x_i$$

$$\sigma^2 = \frac{1}{n}\sum_i (x_i - \mu)^2$$

符合直观想象

$$\mu = \frac{1}{n}\sum_i x_i$$

$$\sigma^2 = \frac{1}{n}\sum_i (x_i - \mu)^2$$

- 上述结论和矩估计的结果是一致的，并且意义非常直观：样本的均值即高斯分布的均值，样本的伪方差即高斯分布的方差。
  - 注：经典意义下的方差，分母是n-1；在似然估计的方法中，求的方差是n
- 该结论将在EM(期望最大化算法)、GMM高斯混合模型中将继续使用。

贝叶斯公式 $P(A|B)=\frac{P(B|A)P(A)}{P(B)}$

- 给定某系统的若干样本x，计算该系统的参数，即

$$P(\theta|x)=\frac{P(x|\theta)P(\theta)}{P(x)}$$

  - P(θ)：没有数据支持下，θ 发生的概率：先验概率。
  - P(θ |x)：在数据x的支持下，θ 发生的概率：后验概率。
  - P(x| θ)：给定某参数 θ 的概率分布：似然函数。
- 例如：
  - 在没有任何信息的前提下，猜测某人姓氏：先猜李王张刘……猜对的概率相对较大：先验概率。
  - 若知道某人来自“牛家村”，则他姓牛的概率很大：后验概率——但不排除他姓郭、杨等情况。

思考题：概率计算

- 随机抽查发现“七月题库APP”的用户实际年龄，调查结果显示年龄均值25岁，标准差2，那么实际用户年龄在21-29岁的概率至少是多少？

思考：随机变量无法直接(完全)观察

- 随机挑选10000位志愿者，测量他们的身高：若样本中存在男性和女性，身高分别服从N(μ_1, σ_1)和N(μ_2, σ_2)的分布，试估计$\mu_1, \sigma_1, \mu_2, \sigma_2$。

- 无监督分类：聚类/EM

思考

- 给定两个随机变量X和Y，如何度量这两个随机变量的“距离”?
- 对称阵的不同特征值对应的特征向量，是否一定正交?
- 如何证明大数定理?
- 仿照指数分布的概率密度函数 $f(x)=\lambda e^{-\lambda x}$，猜测相对应的幂分布的概率密度函数，查阅关于幂律分布的相关文献。

$$f(x)=ax^{-r}，\ a,r\text{为正常数}$$

参考文献

- 王松桂，程维虎，高旅端编，概率论与数理统计，科学出版社，2000

我们在这里

七月算法 http://www.julyedu.com/

- 视频/课程/社区

- 七月题库APP：Android/iOS
  - http://www.julyapp.com/
- 微博
  - @研究者July
  - @七月题库
  - @邹博_机器学习
- 微信公众号
  - julyedu

感谢大家！

恳请大家批评指正！