极大似然估计

前言

这个还需要我来介绍么？简单阐述一下吧，这个在《概率论与数理统计》中，宋浩老师已经讲过了，在前边的线性模型中也经常用到。那么极大似然估计(Maximum Likelihood Estimation, 简称MLE)是什么呢？它是在假设的分布中，***基于原有样本进行概率估计***。或者说的再简单点，你可以参考曾经在线性回归中介绍的例子。

极大似然估计

还记得我们前边说到的频率派么？你会发现极大似然其实是频率派的思想，你可以再看看上边的红字。说白了，就是根据样本直接估计概率！并且，概率模型的训练过程就是参数估计(parameter estimation)过程，也就是极大似然估计求参数(这在宋浩的概率课上也有讲)！

现有类别 c 的类条件概率为 $P(x|c)$ ，假设 $P(x|c)$ 具有确定的形式并且被参数向量 θ_c 唯一确定，则我们的任务就是利用训练集 D 估计参数 θ_c 。为明确起见，我们将 $P(x|c)$ 记为 $P(x|\theta_c)$ 。令 D_c 表示训练集 D 中第 c 类样本组成的集合，假设这这些样本独立同分布，那么参数 θ_c 对于数据集 D_c 的似然为：

$$P(D_c|\theta_c) = \prod_{x \in D_c} P(x|\theta_c), \quad (7.9)$$

也就是寻找能最大化似然 $P(D_c|\theta_c)$ 的参数值 $\hat{\theta}_c$ 。*直观上看，极大似然估计是试图在 θ_c 所有可能取值中，找到一个能使数据出现的最大可能性值。*

那么对于式(7.9)中的连乘操作，我们可以直接取对数，这是为了方便后续的计算，并且保证不下溢。于是我们有了对数似然(log-likelihood)

$$\begin{aligned} LL(\theta_c) &= logP(D_c|\theta_c) \\ &= \sum_{x \in D_c} logP(x|\theta_c), \quad (7.10) \end{aligned}$$

此时参数 θ_c 的极大似然估计 $\hat{\theta}_c$ 为：

$$\hat{\theta}_c = arg\ max_{\theta_c}\ LL(\theta_c), \quad (7.11)$$

需注意的是，这种参数化的方法虽能使类条件概率估计变得相对简单，但估计结果的准确性严重依赖于所假设的概率分布形式是否符合潜在的真实数据分布（你也看到了，我们需要假设分布）。在现实应用中，欲做出能较好地接近潜在真实分布的假设，往往需在一定程度上利用关于应用任务本身的经验知识，否则若仅凭“猜测”来假设概率分布形式，很可能产生误导性的结果。