

统计学习第一次读书报告

报告人：71117415-喻泽弘

读书时间：3.1-3.8

读书进度：统计学习第一章

问题与解答

- 个人提出的问题

- 如何理解泛化公式上界的推导过程？

书中这一步推导 $P(R(f) - \hat{R}(f) \geq \epsilon) \leq \exp(-2N\epsilon^2)$,单纯那这个推导式和Hoeffding推导式进行对比可以发现不同, Hoeffding推导式是这样的

$P(S_n - ES_n \geq t) \leq \exp(\frac{-2t^2}{\sum_{i=1}^n (b_i - a_i)^2})$,对于书中的推导式, 由于是二分类问题, 因此

$[a_i, b_i] = [0, 1]$,因此 $\sum_{i=1}^n (b_i - a_i)^2 = N$,因此, 我们可以发现如果是Hoeffding推导式这个形式的话, 那么N应该出现在分母的下面而不是分子的位置, 相差了 N^2 , 观察两式发现, Hoeffding公式中的 S_n 是求和, 而 $R(f)$ 表示的是均值, 因此, 把均值代入Hoeffding表达式中, 就不能使用 b_i 以及 a_i 了, 而是要使用 $\frac{b_i}{n}$ 以及 $\frac{a_i}{n}$,从而就完成了推导

- 别人提出的问题

- 怎么理解极大似然估计是经验最小化的一个例子？

它是建立在极大似然原理的基础上的一个统计方法, 极大似然原理的直观想法是, 一个随机试验如有若干个可能的解A,B,C...,若在一次实验中, 如果A出现了, 那么可以认为实验条件对A的出现有利, 即出现的概率 $P(A)$ 较大。一般来说, 事件A发生的概率与某一参数 θ 有关, θ 的取值不同, 则事件A发生的概率 $P(A|\theta)$ 也不同, 我们在一次实验中事件A发生了, 则认为参数 θ 的值应是 t 的一切可能取值使 $P(A|\theta)$ 达到最大的那个, 极大似然估计选取这样的 t 值作为参考值, 使所选取的样本在被选的总体中出现的可能性最大。

从极大似然的定义中, 我们可以发现, 极大似然法, 就是从假设空间中选取最优解, 是的发生的概率最大, 从经验最小化上来理解, 便是误差最小化, 因此, 极大似然估计是经验最小化的一个例子。

- 结构风险最小化中的复杂项 $J(f)$ 是怎么确定的？

对于结构最小化中的泛化 $J(f)$,我们认为它是一种与算法复杂度有关的函数, 因为算法复杂度越高, 他的先验概率也就越小, 泛化能力不强, 但是往往复杂度较高的算法, 对训练数据的效果拟合的不错, 因此, 这里引入泛函 $J(f)$ 评估结构风险

读书计划

下一周完成统计学习第二章的阅读

读书收获

- 泛化误差上界：学习方法的泛化性能往往是通过研究泛化误差的概率上界进行的, 简称为泛化误差上界。具体来说, 就是通过比较两种学习方法的泛化误差上界的大小来比较他们的优劣。泛化误差上界通常具有以下性质：它是样本容量的函数, 当样本容量增加时, 泛化上界趋于0; 它是假设空间容量的函数, 假设空间越大, 模型越难学, 泛化误差上界就越大
- 正则化：模型选择的典型方法就是正则化。正则化是结构风险最小化策略的实现, 是在经验风险的基础上添加一个正则化项或罚项。正则化项一般是模型复杂度的单调递增函数
- 统计学习三要素：

1. 模型：所要学习的条件概率分布或决策函数。模型的假设空间包括所有可能的条件概率分布或决策函数。
2. 策略：引入损失函数以及风险函数从假设空间中选出最优模型
3. 算法：算法值学习模型的具体计算方法