读书报告

1. 自己提出的问题的理解
2. 提出的问题1：怎么理解，从贝叶斯估计的角度来看，正则化项对应于模型的先验概率，可以假设复杂的模型有较小的先验概率，简单的模型有较大的先验概率。

讨论后的理解：正则化参数等价于对参数引入先验概率，使得模型复杂度变小（缩小解空间），对于噪声以及outiliers的鲁棒性增强（泛化能力），整个最优化问题从贝叶斯观点来看是一种贝叶斯最大后验估计，其中正则化项对应后验估计的先验信息，损失函数对应后验估计中的似然函数，两者的乘积对应贝叶斯最大后验估计的形式。经验风险其实就是样本本身带来的误差，结构风险就是学习器带来的误差，当假设空间、损失函数、训练集确定的情况下，经验风险可以确定。从贝叶斯角度看，有正则化就是最大后验概率的参数估计方法，无正则化就是最大似然概率的参数估计方法。

参考链接：https://www.jianshu.com/p/4964f9273070

1. 提出的问题2：怎么理解极大似然估计是经验风险最小化的一个例子（1.3.2）

讨论后的理解：极大似然估计是建立在极大似然原理的基础上的一个统计学习方法，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结果A，B，C，…，若在一次试验中，结果A出现了，那么可以认为实验条件对A的出现有利，也即出现的概率P(A)较大。极大似然原理的直观想法我们利用例子说明，设甲箱中有99个球，1个黑球；乙箱中有1个白球，99个黑球。现随机取出一箱，再从抽取的一箱中随机取出一球，结果是黑球，这一黑球从乙箱抽取的概率比从甲箱抽取的概率大得多，这时我们自然更多地相信这个黑球是取自乙箱的。一般来说，事件A发生的概率与某一未知参数θ有关，θ取值不同，则事件A发生的概率P(A|θ)也不同，当我们在一次试验中事件A发生了，则认为此时θ值应是t的一切可能取值中使P(A|θ)达到最大的那一个，极大似然估计法就是要选取这样的t值作为参数t 的估计值，使所选取的样本再被选的总体中出现的可能性最大，所以在这里极大似然估计就相当于是我们将上述的P取为负值就是经验风险的值，那么要求经验风险越小越好就是在做一个极大似然估计。

1. 别人提出的问题的理解
2. 问题3：1.7中，生成方法和判别方法分别适用于什么情况？

自己的理解：监督学习方法又分成生成方法和判别方法，所学到的模型又分别称为生成模型和判别模型。判别方法由数据直接学习决策函数Y=f(x)或者条件概率分布P(Y|X)作为预测的模型，基本思想是有限样本条件下建立判别函数，不考虑样本的产生模型，直接研究预测模型。而生成方法是由数据学习联合概率密度分布P(X,Y),然后求出条件概率分布P(Y|X)作为预测的模型，基本思想是首先简历样本的联合概率概率密度模型，然后得到后验概率P(Y|X)，再利用它进行分类。

1. 问题4：罚项J(f)有没有一个确定的公式呢？不太理解什么叫“定义在假设空间上的泛函”

自己的理解：罚函数是利用问题的目标函数和约束函数构造出带参数的增广目标函数，把约束非线性规划问题转化为一系列无约束非线性规划问题，进而用无约束最优化方法求解约束问题。泛函是一个约束优化的函数用于调整结构风险，是一个函数的函数

关于罚函数的具体知识详见

<https://wenku.baidu.com/view/3f4186373968011ca3009173.html>

1. 问题5：奥卡姆剃刀原理中，很好的解释已知数据评判标准是什么？

自己的理解：奥卡姆剃刀定理是机器学习选择算法时可参展的标准之一，其含义是：再其他条件一样的情况下，选择简单的那个，该定律的意义在于数据的拟合和低复杂度之间存在着折中。理论上假设的解决方案越复杂，就越能拟合数据，训练数据误差就会越低，泛化数据误差实际是训练数据误差与另一个名为过拟合误差的函数之和。再泛化误差最小的情况下，可获得最佳复杂性，用于计算过拟合误差的方法统称为贝叶斯方差方法，在现实生活中，通常只会获得训练数据误差，但实践表明，如果你不去选择能够使训练数据误差最小化的模型，而是选择复杂性低一点的模型，算法的表现往往会更好，过拟合是机器学习算法性能不佳的主因，这也是机器学习中应用奥卡姆剃刀定律的原因。

1. （必填）读书计划

1、本周完成的内容章节：对统计学习第一章节进行了整理，没有继续推进

2、下周计划：统计学习第二章