## 2020.3.22 第八次读书报告

09018330 孙毅远

### 一、自己提出的问题的理解

#### 1. k值的选择与数据有没有什么联系，除了交叉验证法多次计算选取最优的k值之外有无别的方法

理解：k值的选择可以先分析数据做出一个大概估计，然后在一定范围内进行交叉验证法得到最终k值

### 二、别人提出的问题的理解

#### 1. 如何理解超矩形区域

理解：超平面指的是n维欧式空间中的n-1维度的线性子空间

#### 2.在最近邻的搜索算法中，当目标点和当前最近点的距离为半径构成的圆与另一子结点的区域相交时，会移动到另一子节点当中递归地进行最近邻搜索，如何进行？

理解：我的理解时把搜索目标移动到另一子节点，然后遍历回溯

#### 3. 为什么用较小的k值学习的近似误差会减小，反之是估计误差会减小？

理解：近似误差是训练误差，估计误差是测试集的测试误差，k比较小的时候邻域较小，得到的训练集误差小，但测试集误差就会大，k较大的时候相反。

### 三、读书计划

#### 本周 统计学习方法学至5.1

#### 下周 学完第五章

### 四、读书笔记

#### 4.1 朴素贝叶斯法的学习与分类

* 朴素贝叶斯法通过训练数据集学习联合概率分布，具体学习先验概率分布及条件概率分布
* 先验概率分布：
* 条件概率分布：
* 由于条件概率分布具有指数级数量的参数，其估计实际是不可行的，朴素贝叶斯法对条件概率分布作了条件独立性的假设（较强的假设）
* 朴素贝叶斯法实际学习到生成数据的机制，所以属于生成模型。条件独立假设使其变得简单，但有时会牺牲一定的分类准确率
* 朴素贝叶斯法分类时，对给定的输入x，通过学习的模型计算后验概率分布，将后验概率最大的类作为x的类输出，具体计算根据贝叶斯定理进行
* 代入得到分类器

#### 4.2 朴素贝叶斯法的参数估计

* 极大似然估计
  + 先验概率 的极大似然估计是
  + 条件概率的极大似然估计是
* 学习与分类算法
  + 朴素贝叶斯算法
* 贝叶斯估计
  + 条件概率的贝叶斯估计
  + 先验概率的贝叶斯估计

#### 5.1 决策树模型与学习

* 分类决策树模型是一种描述对实例进行分类的树形结构，可以认为是定义在特征空间与类空间上的条件概率分布
* 也可看成一个if-then规则的集合