方骏-2020年5月24日-读书报告

1. 自己提出的问题的理解：
2. 别人提出的问题的理解：

1、 问题1：如何对无监督学习的结果进行有效评估？

自己的理解：分情况讨论，对于聚类和降维，方法应该不同于评估聚类的方法，有很多相关的指标进行统计，总体上说无法直接根据模型预测的准确率进行评估，因为无监督学习算法不做出预测，也不存在y值。第一是评估有无聚类趋势，可以用的指标有霍普金斯统计量；第二是判定数据簇数，可以用的方法有肘状图和gap 统计量等；第三就是题主问的聚类质量的评估，可以用的指标有簇内的凝聚度，簇间的分离度以及同时考虑了二者的轮廓系数。

1. 读书计划

1、本周完成的内容章节：《统计机器学习》第十二章

2、下周计划：《统计机器学习》第十三章、十四章

四、读书摘要及理解

1、监督学习可以认为是学习一个模型，使它能对给定的输入预测响应的输出。监督学习包括分类、标注、回归。对于分类问题，感知机、k近邻法、朴素贝叶斯法、决策树、逻辑斯谛回归与最大熵模型、支持向量机、提升方法是分类方法。就是从实例的特征向量到类标记的预测问题，标注问题是从观测序列到标记序列的预测问题。

2、分类问题与标准问题的预测模型都可以认为是表示从输入空间到输出空间的映射，它们可以写成条件概率分布或决策函数的形式。有时，模型更直接地表示为概率模型，或者非概率模型。朴素贝叶斯法、隐马尔可夫模型是概率模型，感知机、k近邻法、支持向量机、提升方法是非法概率模型。而决策树、逻辑斯谛回归与最大熵模型、条件随机场既可以看作是概率模型，又可以看作是非概率模型。

3、学习策略就是优化以下结构风险函数：



第一项是经验风险，第二项为正则化项。支持向量机用L2范数表示模型的复杂度，原始的逻辑斯谛回归与最大熵模型没有正则化项。概率模型的学习可以形式化为极大似然估计或贝叶斯估计的极大后验概率估计。这时，学习的策略是极小化对数似然损失或极小化正则化的对数似然损失，对数似然损失可以写成：



4、这些学习算法的学习方法都是有区别的，最优化的最优解的求解方法多种多样，比如朴素贝叶斯法就是极大似然估计值求解最优解，感知机、逻辑斯谛回归等都是用的梯度下降和拟牛顿法。支持向量机可以用序列最小最优算法，决策树学习是启发式算法求解等。