寒假第四周读书报告 3.23-3.29

61518426 周之遥

·读书进度

统计学习方法完成第14章：聚类方法余下部分内容

层次聚类有聚合聚类和分裂聚类，最终得到一个层次化、结构化的聚类结果。

k均值聚类将样本划分为k类，对各个样本就近选取，并不断调整类中心，直到收敛。聚类效果依赖于k值的选取，最优的k值出现在拐点。

·提出问题

1.为什么说k均值聚类中当k较大时平均直径将收敛？另外为什么拐点就是最优的k？

解答：各个类之间是区分开来，没有关联的，类间距的影响不大，无论两个类相聚远近，我们都认为它们是两个类，平均直径已经足够。当然也可以考虑将类间距最大化，如果有一个类距离其它类很远，我们会倾向于将其它类进行合并，这其实并不常见，而且会增加计算复杂度。

·下周计划

完成15章奇异值分解。