

《统计学习方法》读书报告

读书内容:

第 14 章第 2, 3 节

提出问题:

1、公式(14.21)中为什么会有 $(-1)^{(k-l)}$ 这一项? 难道分法的数目还能是负的吗?

2、为什么“夹角余弦”可以作为相似度的参考? 难道与原点共线的两个点, 不管距离多远, 都是十分“相似”的吗?

回答问题:

郑健雄问题 1:

两者差别挺大的。例如给学生按年级分类, 第一层次是: 小学, 初中, 高中; 第二层次是某个年级, 第三层次是某个班。而 k 均值聚类就是分成 k 个类, 类本身没有从属或包含关系。层次聚类更加人性化, 数据也更加结构化。

郑健雄问题 2:

感觉差别挺大第。一个是算范数 (距离), 一个算概率。两个共同之处可能就是都要找最大值吧。

王贵涛问题 1:

是没有意义, 所以会有终止条件