

R语言

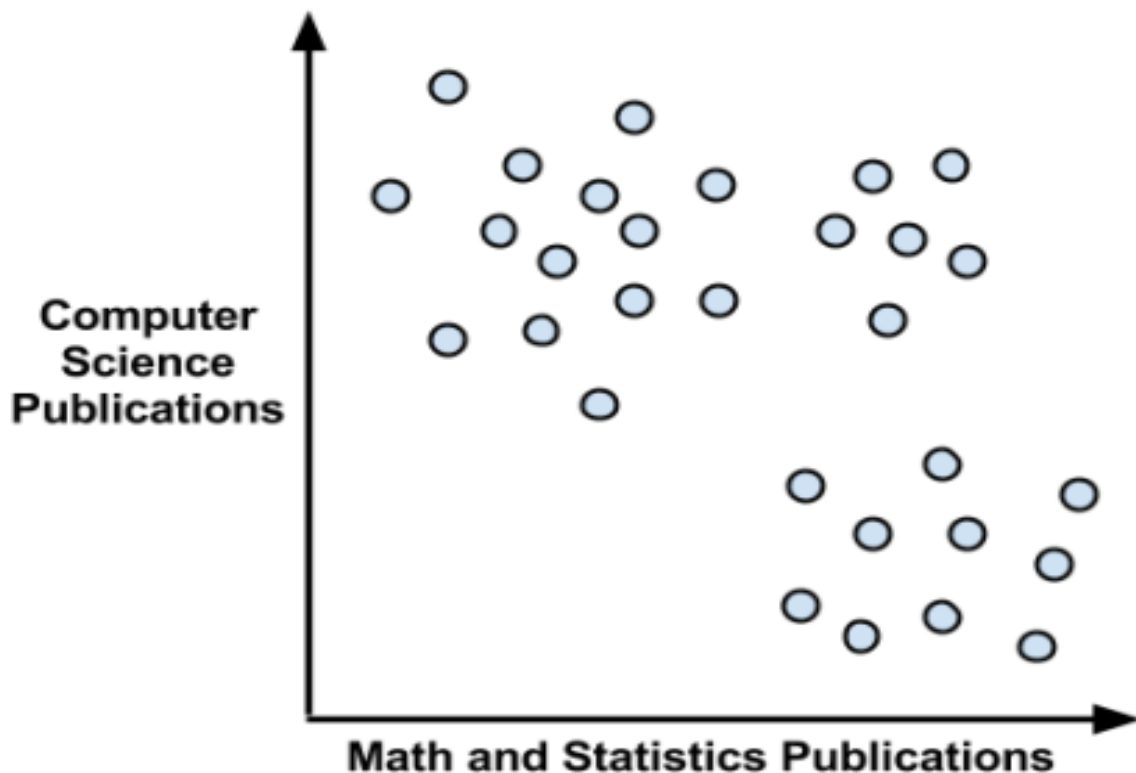
讲师：陈博

K均值聚类

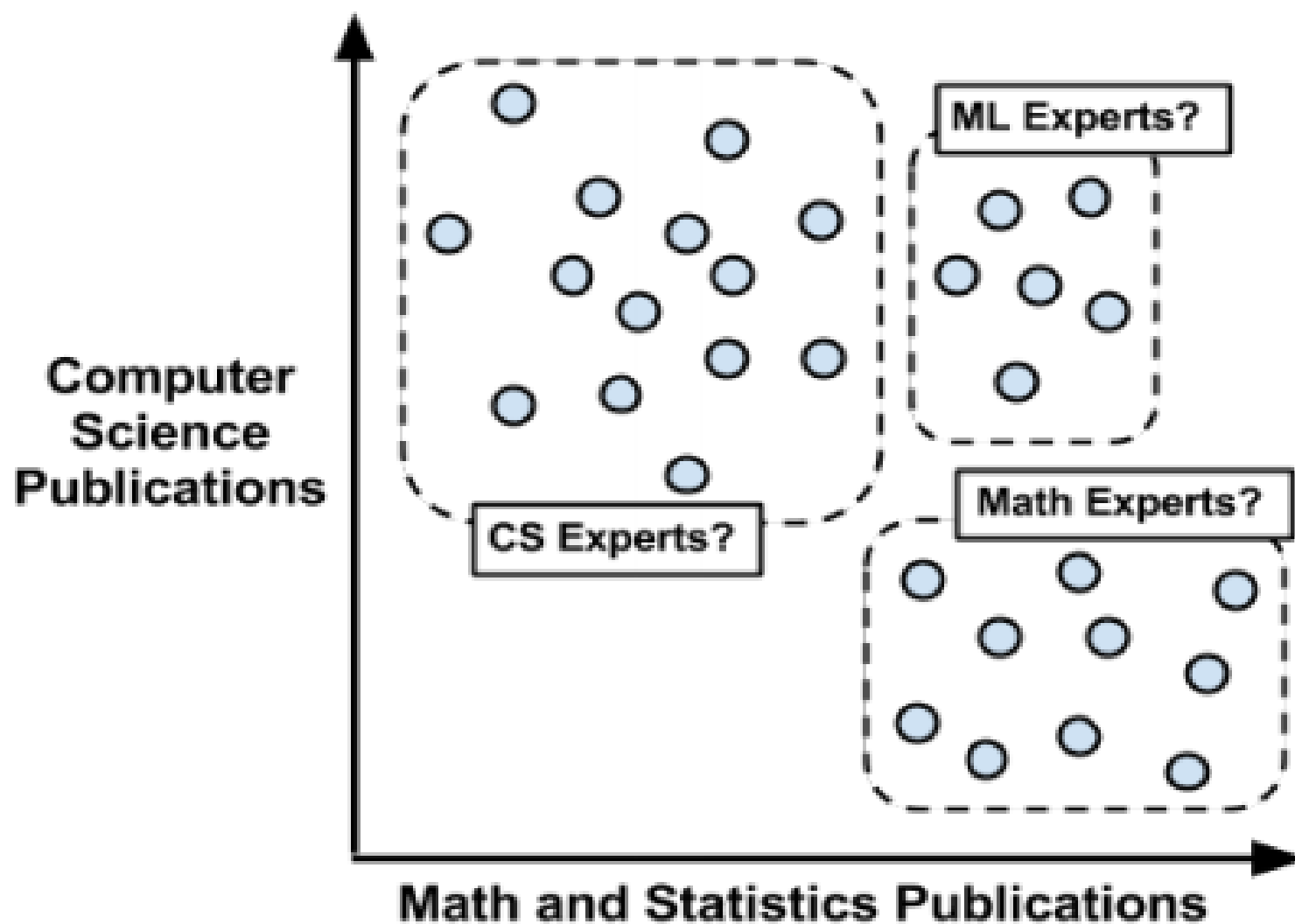
- 聚类：给事物打标签，寻找同一组内的个体之间的一些潜在的相似模式。力图找到数据的自然分组
- 学完本章后，你将理解：
 - K均值的基本方法，它是一种经典的并且容易理解的聚类算法
 - 如何将聚类应用到现实世界的市场细分任务，例如青少年社交媒体用户的市场细分

理解聚类

- 聚类是一种无监督的机器学习任务，它可以自动将数据划分成类cluster。因此聚类分组不需要提前被告知所划分的组应该是什么样的。因为我们甚至可能都不知道我们再寻找什么，所以聚类是用于知识发现而不是预测

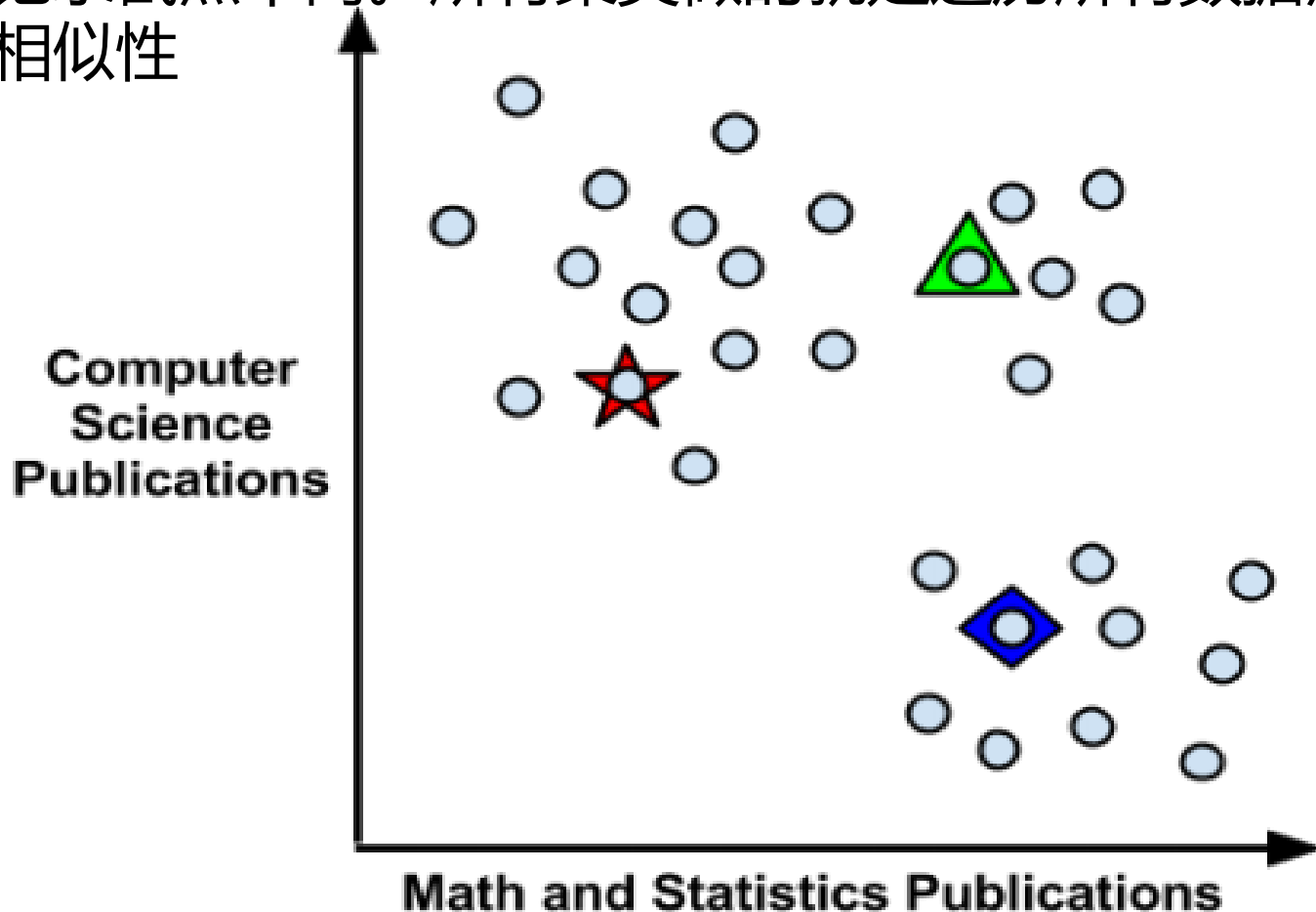


理解聚类



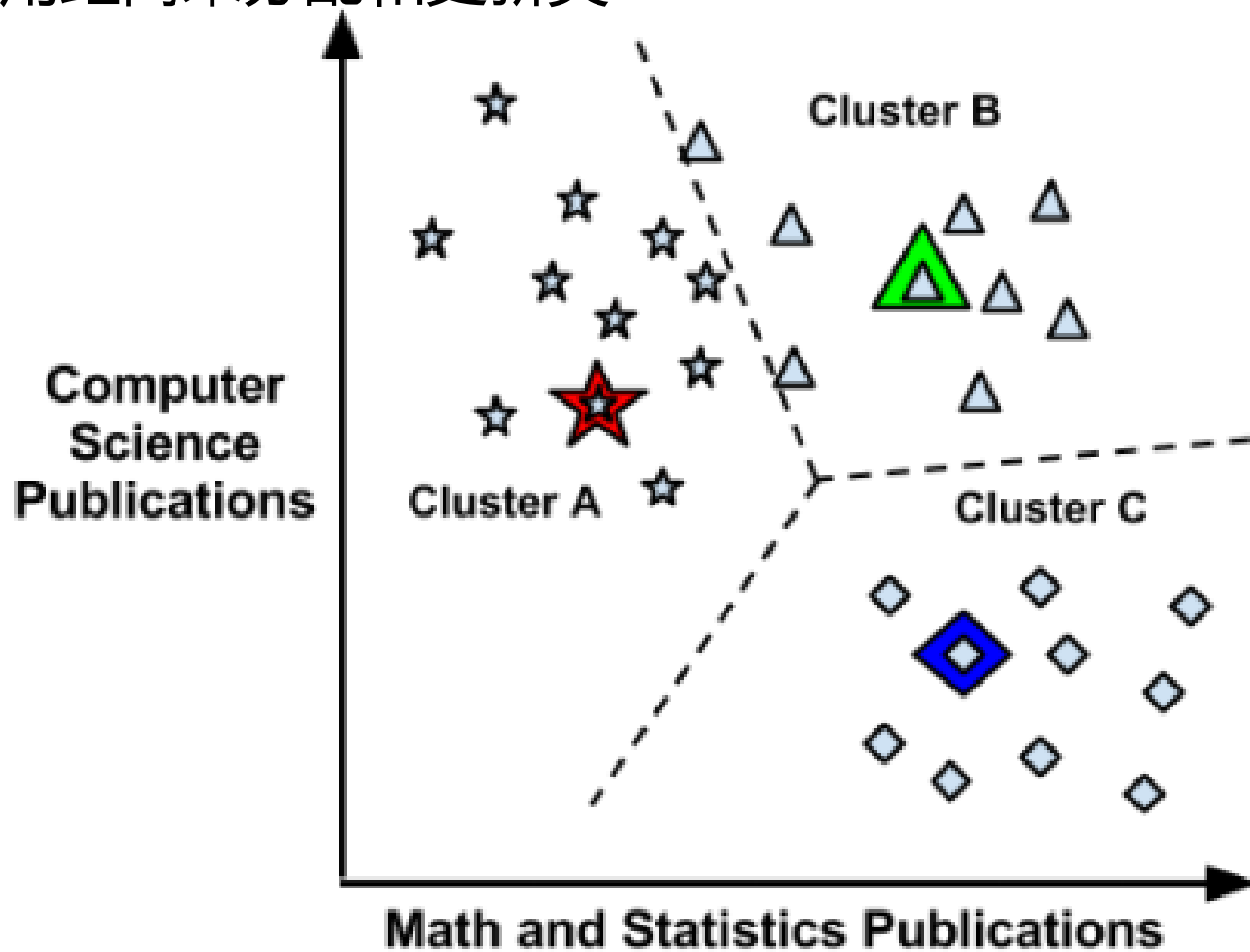
理解聚类

- 聚类原则是一个组内的记录彼此必须非常相似，而与该组之外的记录截然不同。所有聚类做的就是遍历所有数据然后找到这些相似性



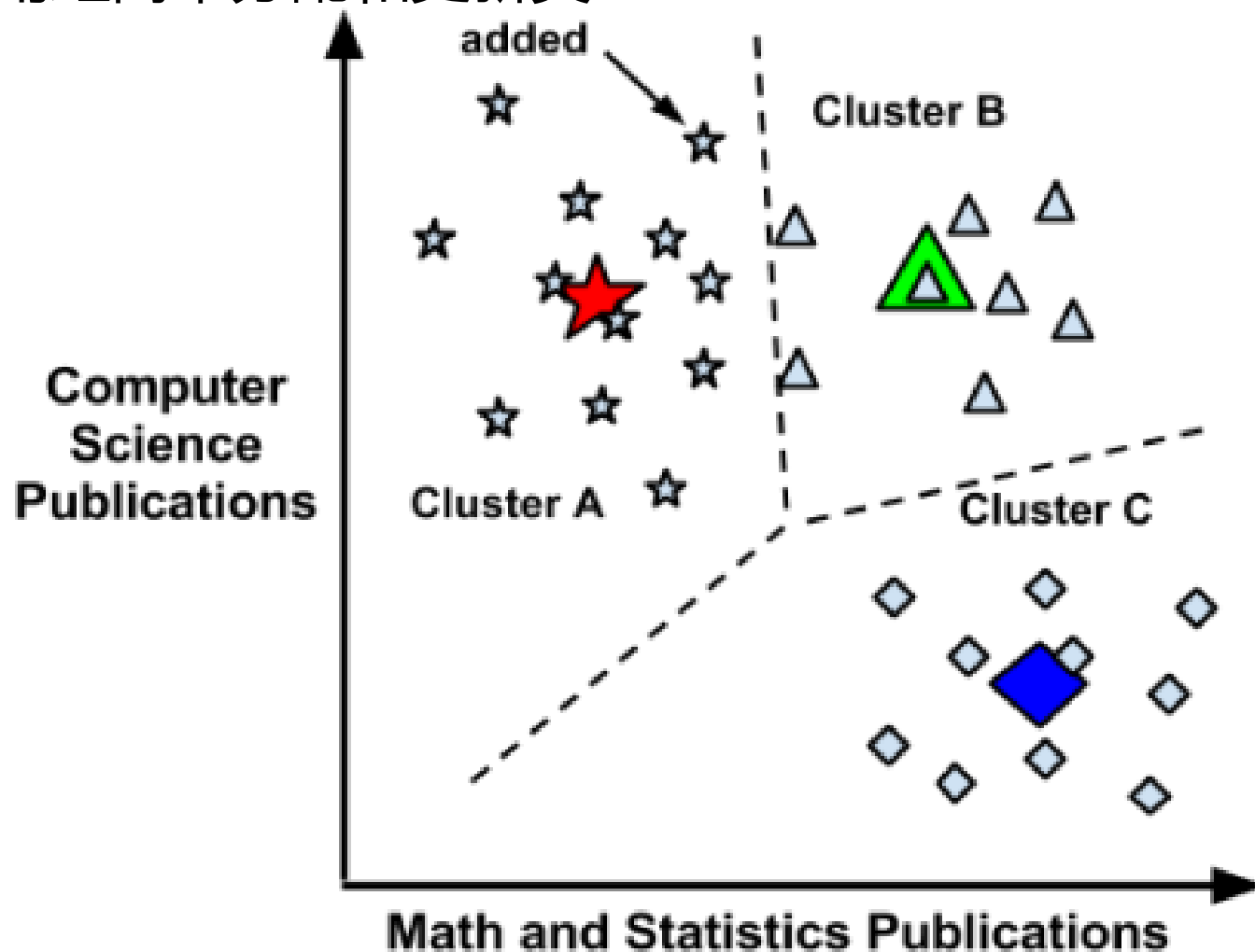
理解聚类

- 使用距离来分配和更新类

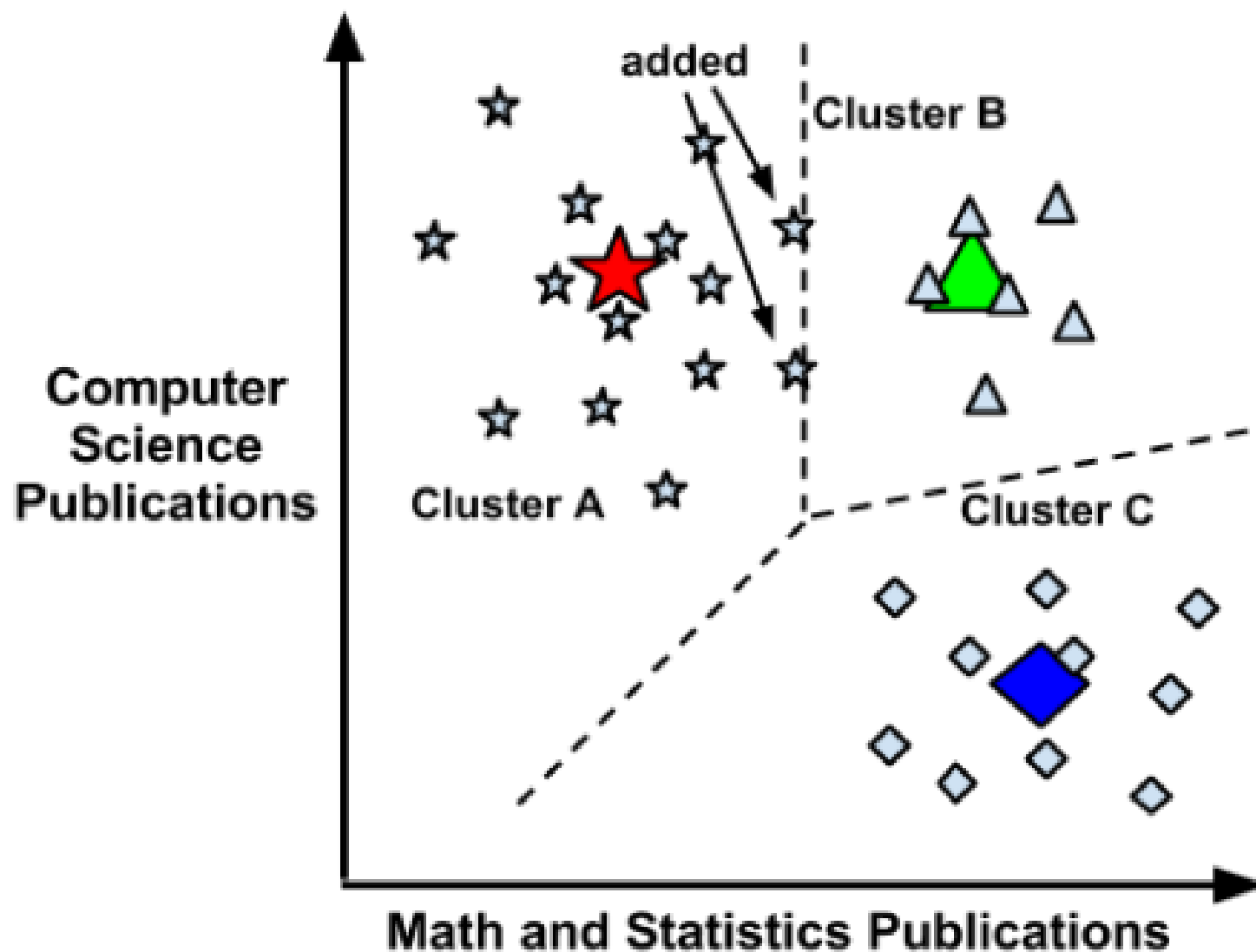


理解聚类

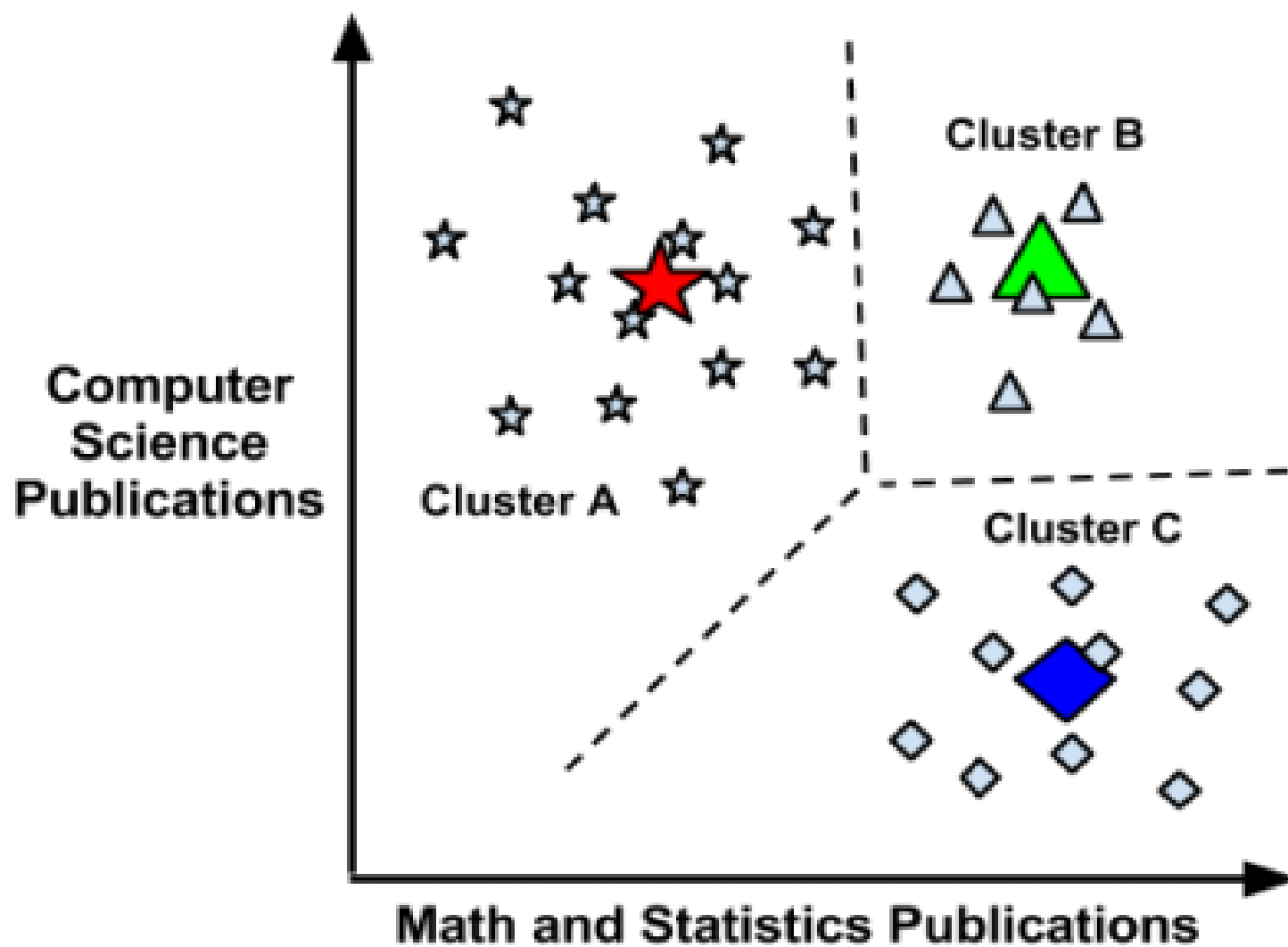
- 使用距离来分配和更新类



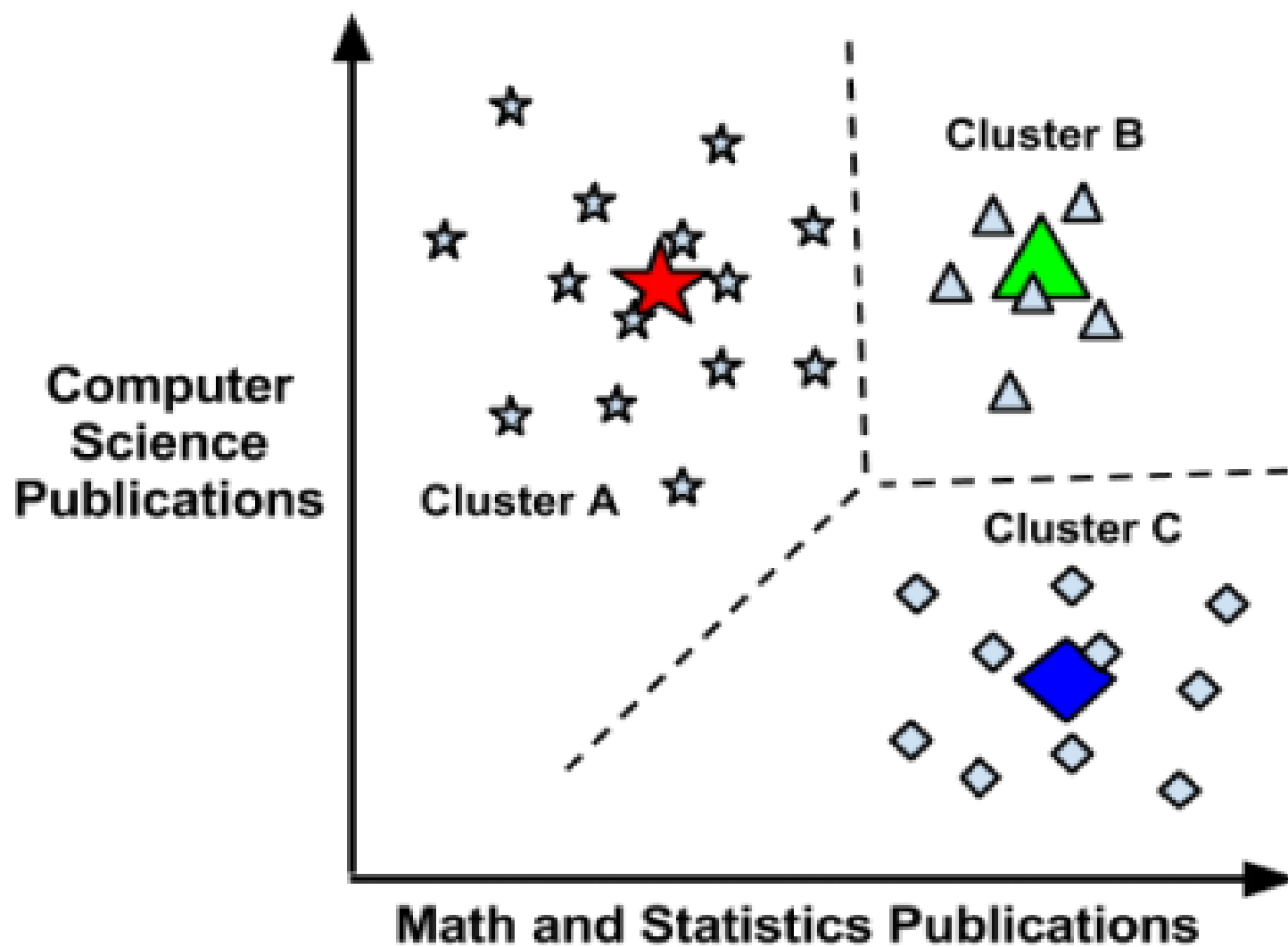
理解聚类



理解聚类



理解聚类



选择适当的聚类数

- 肘部法

