**开学第四次读书报告——郭昊南**

1. **自己提出的问题的理解：**
2. **提出的问题1：** 聚类和降维有什么区别和联系？

**讨论后的理解：**

联系：聚类是对类别变量做降维的方法之一，适用于无标注数据。

区别：降维是在数据预处理时为了防止过拟合时常用的方法，聚类是在整个数据集上按一个定义的拓扑距离进行分组的过程。

讨论结果：聚类是指将不同的实体分成一个一个类别，所以是对不同实体进行操作，而矩阵储存数据时，一般每一行一个特征，每一列一个样本，因此相当于对实体个数进行操作，所以书上称为对纵向结构进行操作。降维是指对于数据的特征数进行操作，降低数据的特征数，因此相当于对横向结构进行操作

1. **别人提出的问题的理解：**
2. **问题2：什么是无标注数据？他和有标注有数据的差异体现在哪里？**

**自己的理解：**

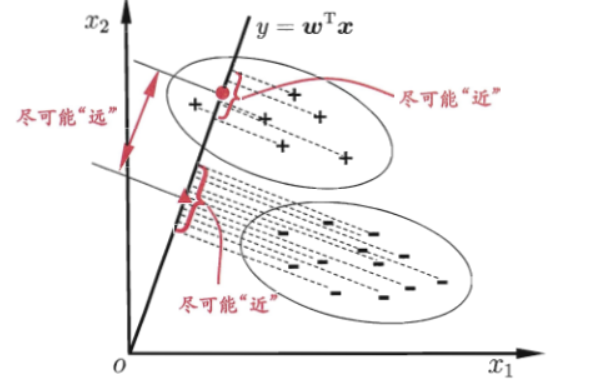
无标注数据是指未经人为处理过的数据，区别于有标注数据，有标注数据是经过认为处理的，保留学习需要的相关属性的数据。

讨论结果：无标注数据直观理解应该是没有人为进行标注过的数据，比如没有人为标定类别什么的，无标注数据可以理解为原始的数据。有标注数据的话可以用来训练分类器，因为它相当于给出了函数的输入和输出值，来求一个函数或者概率模型，适合有监督学习，但是无标注数据可以理解为仅仅有输入数据，所以实际上只能分析这些数据的模式，从而进行进一步的研究

1. **问题4：LDA是什么？**

**自己的理解：**

LDA线性判别分析是一种经典的降维方法，思想是给定训练集样例，设法将样例投影到一条直线上，使得同类样例的投影尽可能接近，异类样例的投影点尽可能原理；在对新的样本进行分类时，将其投影到同样的这条直线上，再根据投影点的位置来确定新样本的类别。

****

1. **读书计划**

1、本周完成的内容章节：《统计学习方法》第十二、十三、十四章