读书报告

09118119 黄一凡

# 自己提出的问题

#### 1.如何对无监督学习的结果进行评估？

实际上无监督学习的评价没有一个标准的方法，需要具体应用具体分析， 比如聚类的话就需要分析其最终的聚类是不是高内聚的，类别数目是否合理，降维就需要考虑保存的信息等等，在具体应用的时候可以具体分析。

# 别人提出的问题

#### 1.什么是无标注数据？它和有标注有数据的差异体现在哪里？

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。

#### 2.如何理解P248最后一段话？

这里给出的是一个概率分布，z代表类别，而x代表样本，对于z来说，z有多个值，代表多个类别，而此时x固定， 可以根据不同类别对应的概率大小将x分配到不同的聚类中，因为软聚类的定义里说一个样本可以属于多个类别

#### 3.降维的维数由什么决定？

这个可以人为进行选定，也可以划定一个标准来决定保存多少信息， 比如70%的信息，主成分分析中可以计算出至少保留多少个维度的情况下可以保存这么多的原始信息，这种情况下就能计算出维数。

#### 4.聚类和降维代表了从两种方向进行压缩的方法，那么这两种方法之间有什么相似性和差异？

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。

# 读书计划

#### 本周所读：

《统计学习方法》12-14.1

#### 下周计划：

《统计学习方法》14.2-14.3