窦一蒲-2020年2月23日-读书报告

读书报告内容：

1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1： 通过leave-one-out在已标注数据集中交叉验证精度的方法具体怎么实现的呢？

讨论后的理解：每次把一个labeled数据和unlabeled数据用这种方式训练，去不同的μ，然后再用一开始拿出来的验证精确度。

1. 提出的问题2：1.5.11中mixture model究竟指什么？意义何在？为什么要有这前提假设？

讨论后的理解：是为了假设数据同分布，从而生成模型，如果数据没有一个分布，感觉就生成不了模型。

1. 提出的问题3：请问一下5.1.1中种子的类别标识表示了聚类结果中各个簇的类别标识是什么意思呢？

讨论后的理解：先对labeled数据聚类，得到几个簇，然后再利用这些簇对无labeled的聚类，这些labeled数据就是种子。

1. （必填）别人提出的问题的理解（选择几个问题罗列，并给出理解）：
2. 问题4： 通过leave-one-out在已标注数据集中交叉验证精度的方法选取μ是什么意思？

自己的理解：为了得到μ的具体值，对不同的取值进行验证，看哪个取值的效果最好。

5、问题5：在5.1.2中，提到将U中标为类 的一定数量、分类可信度高的数据加入到L中，请问这里的分类可信度是怎么衡量的，怎么理解？

自己的理解：每一次都找概率最高的那几个。

6、问题6：there is a one-to-one correspondence between mixture components and classes.怎么理解？

自己的理解：可能是混合模型里面的其中一个模型。

1. 问题7：在Co-Training中提到，用来训练分类器的两个特征的子集is sufficient for learning the target classification function。但是在训练之前我们是怎么得到这个结论的？还是说这只是一个assumption?

自己的理解：这只要彼此之间无关（条件独立）即可，当然我们也可以按人的思维去思考哪些特征凑在一起更容易判定种类，这样子人工选取。