窦一蒲-2020年2月17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：怎么理解3.2.3节中，任何的threshold value在vi和vi+1之间有同样的效果？

讨论后的理解：虽然属性是连续的，但数据集中的具体值是离散的，所以两个离散值之中的任意值都能将它分成两个一样的数据集。

1. 提出的问题2：不利用validation set怎么判断estimated error？

讨论后的理解：用train set，当前节点错误比子树少就把子树剪掉。

1. 提出的问题3：比起单纯看模型中的accuracy来说，计算recall和precision可以提供哪些额外的信息？

讨论后的理解：可以提供精确度和完整度，如果分类不完整，模型有可能有问题。

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： 如何理解3.2.4节中处理决策树中to rank the new cases according to how likely they may be intrusions.

自己的理解：因为不同类型数据占比差可能很大，排序后看最有可能的。

5、问题5：cross-validation中，k次实验后，去k个accuracy的平均值有何意义？

自己的理解：代表了这种模型的平均精度。

6、问题6：breakeven point的计算对于分类器评估的意义何在？

自己的理解：当recall和precision相等时，F-score最大，最准确。

1. （必填）读书计划

1、本周完成的内容章节：如3.1-3.3

2、下周计划：第四章4.1-4.3