1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1：在ROC曲线中，为什么主对角线（代表随机猜测，即每个类分成正例为一个固定比例）上测试数据的假阳率FPR总是和真阳率TPR相等？

讨论后的理解：因为是随机猜测，所以对于每个分类，实际上为正例和实际上为负例的概率都是相同的，即：,设这个比值为k，则可以推出真阳率和假阳率相同，即：

1. 提出的问题2：在One Rule-1函数算法中，当BestCond为空集时，为什么涵盖了D中所有的样例？

讨论后的理解：因为在规则学习之后，如果还存在训练样例没有被学习到的规则所覆盖，则将这些案例归类到默认类。所以当bestcond为空集的时候，即规则集合为空，此时所有的样例都将被放在默认样例中，所以涵盖了D中所有的样例。

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3：P59如何理解后剪枝操作中“如果一个树节点的错误估计要比其子树的错误估计小，则其子树将被剪枝”？错误估计的思路是什么？

自己的理解：因为剪枝是为了解决过拟合的问题，子树比其树节点划分的更细，所以更容易产生过拟合，为了解决这个问题，可以将其子树剪枝。

1. 问题4：我没有理解60页的“剪枝之后的规则集合可能不再是互不相交且完全覆盖的”，因为我感觉示例里面P59 3.7(b)的剪枝后的决策树就是互不相交且完全覆盖的，而且我还觉得在这个图里对其他任意的树节点剪枝，得到的结果还是互不相交且完全覆盖的。

自己的理解：因为当决策树的某一分支样例点过少的时候，有可能会产生过拟合问题，如果这一分支是从决策树的根节点开始，则可以将整个分支从根节点剪枝掉，那么整个决策树就会缺少一些规则，则规则集合可能不再是互不相交且完全覆盖的了。

1. （必填）读书计划
2. 本周完成的内容章节：《web data mining》第三章监督学习3.1~3.5
3. 下周计划：《web data mining》第三章监督学习3.6~3.10