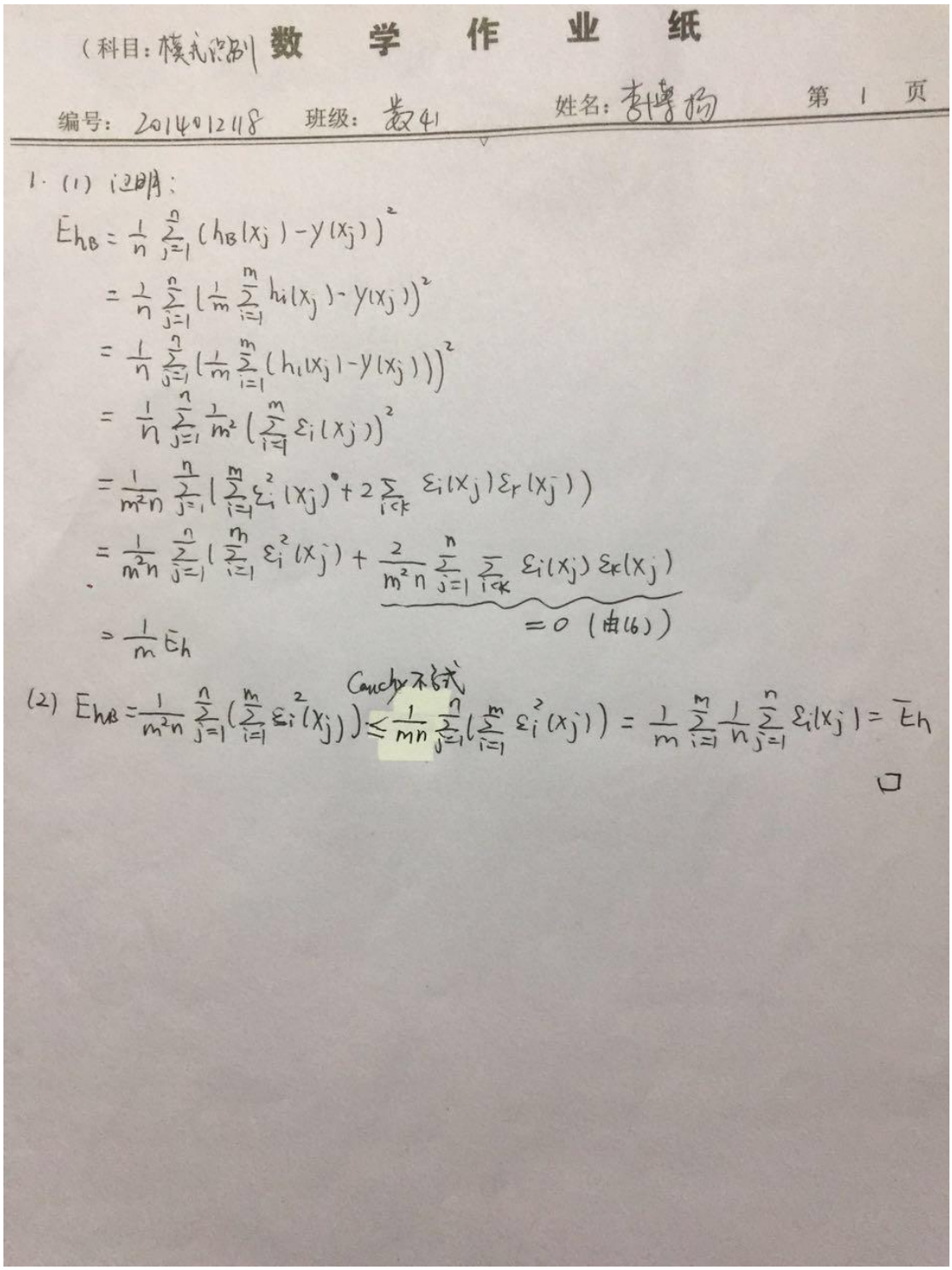
决策树与 Boosting 方法

数 41 李博扬 2014012118

1、



2．1 训练基于决策树桩的 adaboost（训练、测试错误率见 mat 文件）

2．2 训练决策树

最大分裂数	训练错误率	测试错误率
4	9.0%	10.15%

20	8.7%	9.49%
100	8.6%	9.49%

2．3 训练随机森林

分类器数	训练错误率	测试错误率
30	5.3%	8.12%
100	3.4%	4.58%
300	4.7%	4.98%

2．4

由实验结果可见，adaboost 算法随着迭代次数的增加，在足够大时，训练错误率可以达到 0，测试集错误率一直降低。而决策树与随机森林没有这么明显的效果。且在训练与测试错误率方面，随机森林要优于决策树，但不如 adaboost。建立决策树的关键在于当前状态下选择哪个属性作为分类依据。随机森林在 bagging 基础上做了修改，建立多棵 CART 决策树形成随机森林，通过投票决定分类。Adaboost 通过增加训练被前一分类器分错的样本的权重，降低错误率，但对噪声数据和异常数据很敏感。相对于其他分类方法，Adaboost 不容易出现过拟合。