

2017 机器学习第一次作业要求

1 作业

阅读周志华《机器学习》第八章，然后完成下面两个题目：

- 试析随机森林为何比决策树 Bagging 集成的训练速度更快。
- MultiBoosting 算法 [2] 将 AdaBoost 作为 Bagging 的基学习器，Iterative Bagging 算法 [1] 则是将 Bagging 作为 AdaBoost 的基学习器。试比较二者的优缺点。

2 作业要求

- 阅读上题中的两篇论文，理解并实现算法，从理论和实践两方面进行分析
- 分析后将分析结果和实现过程写成 pdf 文档与程序打包提交，在文中写清楚每个人的分工以及每个人的姓名学号等信息，命名格式为组号_作业次数.zip，如本次 1 号组的命名为 1_1.zip
- 提交的截止日期为 12 月 1 日晚 23 点 59 分，迟交者本次作业记 0 分，请大家在截止日期之前将作业发送到 ML_PKU_2017@163.com
- 数据集为 <https://archive.ics.uci.edu/ml/datasets/Chess+%28King-Rook+vs.+King%29>，之后数据集文件和论文文件会放在网盘中供大家下载

参考文献

- [1] Leo Breiman. Using iterated bagging to debias regressions. *Machine Learning*, 45(3):261–277, 2001.

- [2] Geoffrey I Webb. Multiboosting: A technique for combining boosting and wagging. *Machine learning*, 40(2):159–196, 2000.