

计算机与信息技术学院

本科生《机器学习》课程大作业

实验报告

**实验名称 有监督模型综合对比分析**

**小组成员 18281050 庞金果**

**18281046 龙敏**

**18281048 马润生**

**日 期 2020.12.27**

目录

[一、成员分工 3](#_Toc60237835)

[二、实现AdaBoost算法 3](#_Toc60237836)

[三、实现实现Bagging集成方法 4](#_Toc60237837)

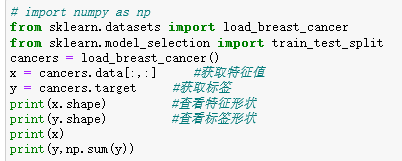
# 成员分工

庞金果完成实现AdaBoost算法部分

龙敏完成Bagging算法部分

# 二、实现AdaBoost算法

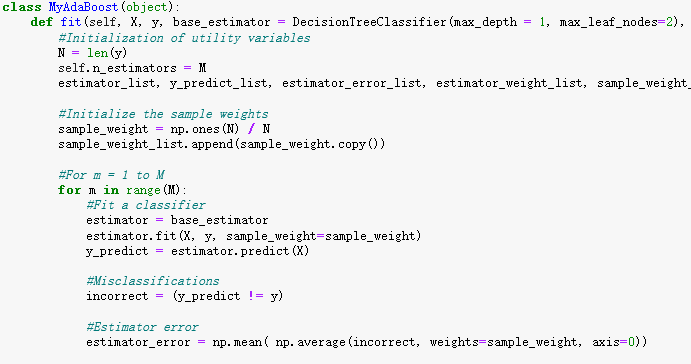
1. 数据集：乳腺癌数据集

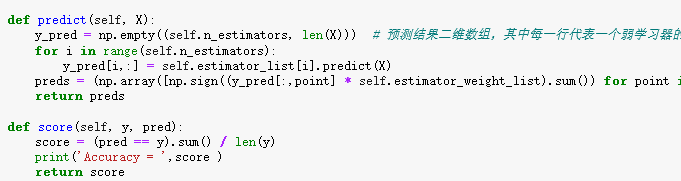


2. 实现AdaBoost算法、k折交叉验证

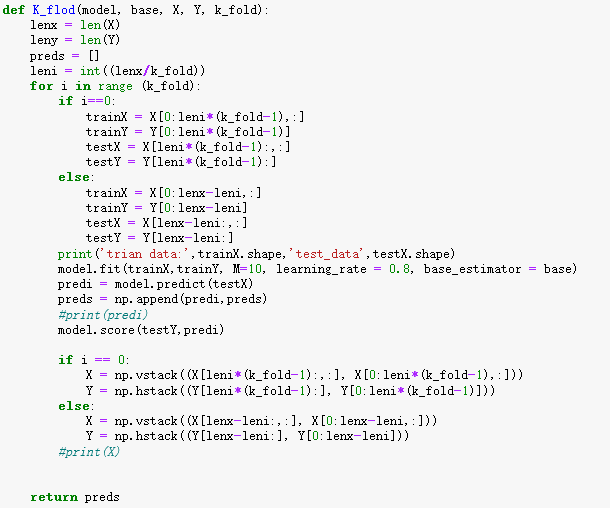
AdaBoost类：包括fit(), predict(), score()方法

fit(): 通过base\_estimator参数选择基分类器





k折交叉验证：

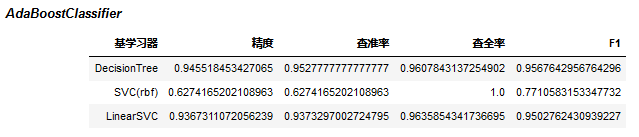


3. 选用不同的基学习器进行十折交叉验证，计算评价指标



从分类指标来看，LinearSVC模型的拟合效果最好，决策树拟合效果较差，SVC(rbf)模型拟合效果最差，精度仅达到0.62。LinearSVC模型的效果明显优于SVC(rbf)，这是因为本实验采用的数据集线性可分，特征数量多。

4. 与AdaBoostClassifier对比

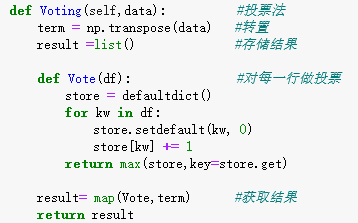


Myadaboost与AdaBoostClassfier在使用SVC(rbf)和LinearSVC作为基分类器时效果基本一致。但使用决策树作为基分类器时MyAdaboost分类效果明显不如AdaBoostClassfier，原因可能在于MyAdaboost的代码实现可能并不完全准确，具体原因待进一步探究。

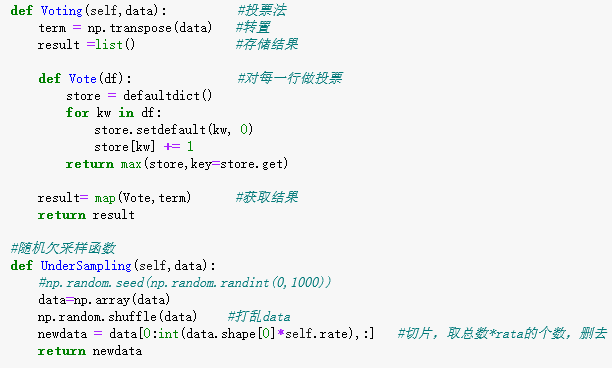
# 实现实现Bagging集成方法

1. Bagging算法实现

包括投票模块：



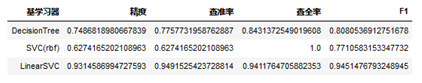
随机采样模块



综合训练模块。



1. 基于myBagging的Accuracy, Precision, Recall, F1



1. Scikit-learn 的BaggingClassifier得到的结果



1. 详细分析不同的基学习器的Bagging的表现差异以及可能的原因

从分类指标来看，DecisionTree模型的拟合效果最好，LinearSVC拟合效果较差，SVC(rbf)模型拟合效果最差。这是因为个体学习器不存在强依赖关系，差异尽可能地具有较大

1. AdaBoost的结果进行对比，分析两种集成方法的优劣

这两种方法都是把若干个分类器整合为一个分类器的方法，只是整合的方式不一样，最终得到不一样的效果，将不同的分类算法套入到此类算法框架中一定程度上会提高了原单一分类器的分类效果，但是也增大了计算量