# 实验三 集成学习一

**一．简单介绍AdaBoost和Random Forest算法的原理**

**二．Breast Cancer数据实验**

1.对Breast Cancer数据进行探索性分析

2.数据预处理

3.分别以决策数、逻辑回归、SVM为基函数，利用网格搜索等方法寻找不同基函数下AdaBoost算法的最优参数。利用Precision、Recall、F1和Auc等指标评价模型，探究和对比不同基函数下的AdaBoost算法性能。

4.对比以决策数、逻辑回归、SVM为基函数的AdaBoost、Random Forest以及Lars算法在Breast Cancer分类数据上的重要特征，得出影响Breast Cancer分类的关键因素。

**三．Boston数据实验**

1.对Boston房价数据进行探索性分析

2.数据预处理

3.以R2、MSE、MAE等指标为评价标准，探究Random Forest算法的参数对模型性能的影响

4.对单棵决策树以及以决策树为基函数的集成算法(AdaBoost，Random Forest)进行性能对比，探索相较于单模型而言，集成学习的特点。

5.分别对决策树、以决策数为基函数的AdaBoost算法、Random Forest以及Lars算法得出的特征重要性进行对比分析，得出影响Boston房价的关键因素。