%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

第1题

比较boosting和bagging两种策略的区别

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

解：

Boosting：

1. 每一轮的训练集不变，只是训练集中每个样例在分类器中的权重发生变化。而权值是根据上一轮的分类结果进行调整。
2. 根据错误率不断调整样例的权值，错误率越大则权重越大。
3. 每个弱分类器都有相应的权重，对于分类误差小的分类器会有更大的权重。
4. 各个预测函数只能顺序生成，因为后一个模型参数需要前一轮模型的结果。

Bagging：

1. 训练集是在原始集中有放回选取的，从原始集中选出的各轮训练集之间是独立的。
2. 使用均匀取样，每个样例的权重相等。
3. 所有预测函数的权重相等。
4. 各个预测函数可以并行生成。

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

第2题 随机森林比普通决策树Bagging训练速度谁快？为什么？。

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

解：

随机森林的训练效率优于Bagging，因为在个体决策树的构建过程中，Bagging使用的是“确定型”的决策树，在选择特征划分时要对节点的所有特征进行考察，而随机森林使用的“随机型”的决策树则只需考察一个子集。

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

第3题

Adaboosting集成法用于西瓜数据集3.0alpha。

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

第4题

bagging集成法决策树，用于西瓜数据集3.0alpha.。

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

第5题

随机森林方法处理糖尿病数据集pima-indians-diabetes.data.csv。

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

