HW4 集成模型

1 adaboost（50）

1.1 输入数据集 （10）

data1.mat为分类数据集，每一行为一个样本，前两列为特征，最后一列为目标值。按照7:3的比率划分训练集和验证集。

1.2 模型训练（20）

使用sklearn工具包，调用ensemble.AdaBoostClassifier接口对模型进行训练。

1.3 分析（20）

- 可视化决策边界，并输出验证集准确率
- 基于实验，分析不同的基分类器和基分类器数量对于模型性能的影响

2 随机森林（50）

1.1 输入数据集（10）

data1.mat为分类数据集，每一行为一个样本，前两列为特征，最后一列为目标值。按照7:3的比率划分训练集和验证集。

1.2 模型训练（10）

使用sklearn工具包，调用ensemble.RandomForestClassifier接口对模型进行训练。

1.3 分析（30）

- 换用不同的n_estimators、criterion、max_depth、min_samples_split ，分析其对于验证集准确率的影响。

3 Bonus（20）

3.1 使用Iris数据集分别对adaboost和随机森林进行训练。

Iris也称鸢尾花卉数据集，是一类多重变量分析的数据集。数据集包含150个数据样本，分为3类，每类50个数据，每个数据包含4个属性。可通过花萼长度，花萼宽度，花瓣长度，花瓣宽度4个属性预测鸢尾花卉属于（Setosa，Versicolour，Virginica）三个种类中的哪一类。

Iris数据集的调用

```
from sklearn.datasets import load_iris
X, y = load_iris(return_X_y=True)
```