# 《机器学习与深度学习》课程

# 实 验 报 告



**姓 名：**

**专 业：**  人工智能

**学 号：**

**江南大学人工智能与计算机学院**

# 集成学习

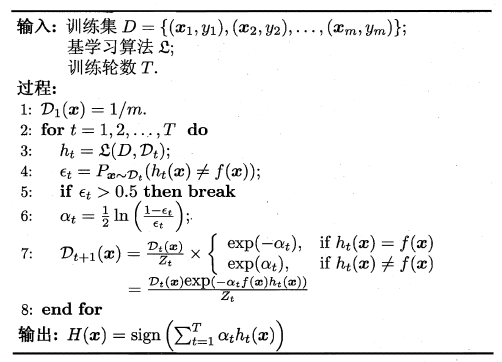
**1实验目的**

集成学习(ensemble learning) 通过构建并结合多个学习器来完成学习任务，常可获得比单一学习器显著优越的泛化性能。集成学习类型之一的Boosting是个体学习器间存在强依赖关系、串行生成的序列化方法。AdaBoost是Boosting 族算法最著名的代表。本实验的目的在于加深学生对AdaBoost 算法的理解，掌握算法的实现过程和使用方法。

**2实验原理**

AdaBoost的工作机制是:先从初始训练集训练出一个基学习器，再根据基学习器的表现对训练样本分布进行调整，使得先前基学习器做错的训练样本在后续受到更多关注，然后基于调整后的样本分布来训练下一个基学习器;如此重复进行，直至基学习器数目达到事先指定的值T， 最终将这T个基学习器进行加权结合。

AdaBoost算法的代码如下：



**3实验内容和要求**

疝病是描述马胃肠痛的术语。然而，这种病不一定源自马的胃肠问题，其他问题也可能引发马疝病。我们提供的马氙气数据集中包含了医院检测马疝病的一些指标和马的状态标签，我们要做的就是通过AdaBoost算法训练299个样本，形成一个以决策树为基学习器的集成学习器，并对67个样本进行测试。

horseColicTraining.txt为训练数据集，共299条训练数据，每条数据前21个为特征，最后一个为马的状况标签：1—仍存活，0—未能存活。

horseColicTest.txt为测试数据集，共67条测试数据，每条数据前21个为特征，最后一个为马的状况标签：1—仍存活，0—未能存活。

实验具体要求：

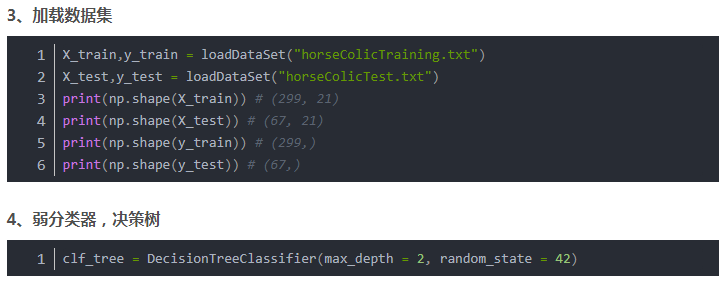
1）使用 Python 语言实现 AdaBoost 算法，在马氙气数据集（horseColicTest.txt，horseColicTraining.txt）上训练一个集成分类器，估计马疝气的死亡率。

2）画图输出混淆矩阵

3）计算并给出查准率、查全率和 F1 度量

4）绘制 P-R 曲线。

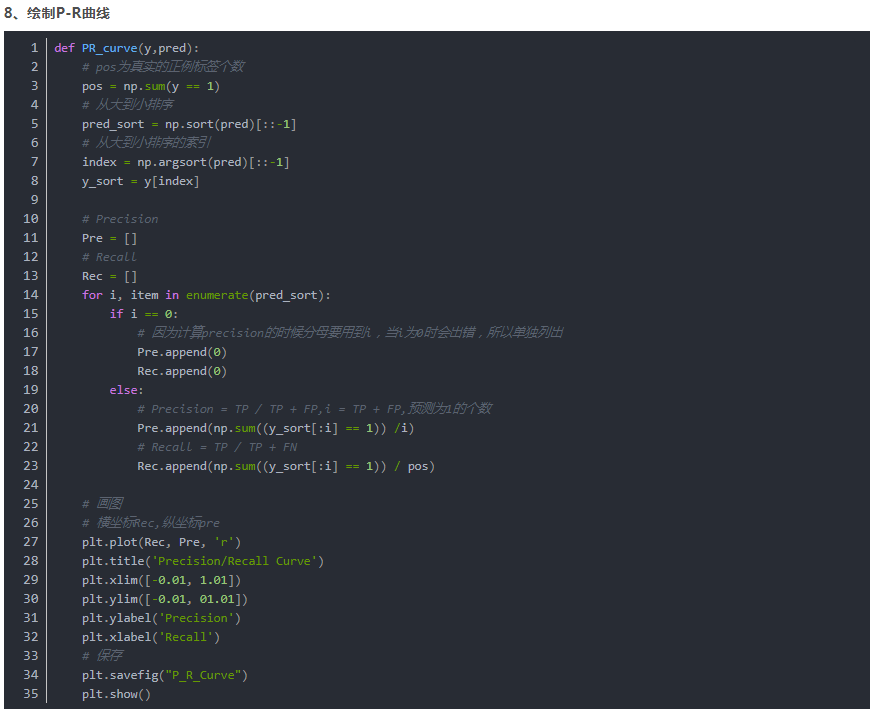
**4实验代码和结果**

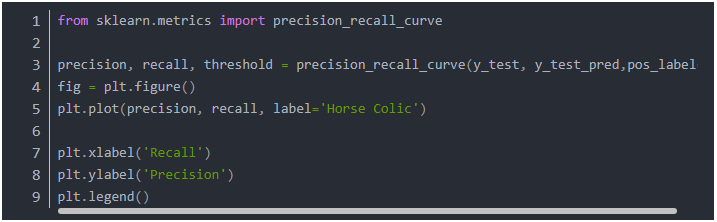


5、AdaBoost算法









## 四、实验结果截图及结果分析

