实验 5

1（选做*）. 使用支持向量机完成手写数字的识别

使用 digits/trainingDigits 文件夹下的文件作为训练集（只包含了数字 1 和数字 9），digits/testDigits 文件夹下的文件作为测试集，使用完整版 Platt SMO 算法（使用径向基核函数）进行手写数字 1 和 9 的识别。测试不同径向基核函数到达率 σ=10，20，100 的测试错误率。（不使用 sklearn 库）

2. 使用 sklearn 库中的 make_circles 函数生成具有两个类别的圆形数据集，并使用核函数 SVM 对该数据集进行分类，并将结果可视化绘制决策边界。

3. 使用 AdaBoost 元算法进行病马死亡率的预测

使用 horseColicTraining2.txt 文件作为训练集，horseColicTest2.txt 文件作为测试集，使用基于单层决策树的 AdaBoost 算法（弱分类器数目为 40）进行病马死亡率的预测。（不使用 sklearn 库）

4. 对于 Iris 数据集（sklearn 库自带鸢尾花数据集），试采用 Bagging 方法如：随机森林以及 Boosting 方法如：Adaboost 和 SVM 分别进行分类（采用 sklearn 库或者自编 python 代码均可），对比几种算法的训练集误差、测试集误差和运行时间。

5.（选做）Kaggle 上的信用卡欺诈数据集（https://www.kaggle.com/mlg-ulb/creditcardfraud）是一个

非均衡数据集，请针对该数据集进行非均衡数据的处理（重采样、欠采样等），并使用数据集切分和交叉验证，分别训练和测试 AdaBoost 以及 XGBoost
模型的性能（需要包含 precision，recall，F1-score），绘制 ROC 曲线计算 AUC 值。