山东大学 计算机科学与技术 学院

机器学习与模式识别 课程实验报告

学号：202300130183	姓名： 宋浩宇	班级：23 级人工智能班
实验题目：Random Forest		
实验学时：2	实验日期：2025/4/15	

实验环境：

软件环境：

系统：Windows 11 家庭中文版 23H2 22631.4317

计算软件：MATLAB 版本: 9.8.0.1323502 (R2020a)

Java 版本: Java 1.8.0_202-b08 with Oracle Corporation Java HotSpot(TM) 64-Bit Server VM mixed mode

Python 解释器版本：Python 3.12.6

硬件环境：

CPU：13th Gen Intel(R) Core(TM) i9-13980HX 2.20 GHz

内存：32.0 GB (31.6 GB 可用)

磁盘驱动器：NVMe WD_BLACKSN850X2000GB

显示适配器：NVIDIA GeForce RTX 4080 Laptop GPU

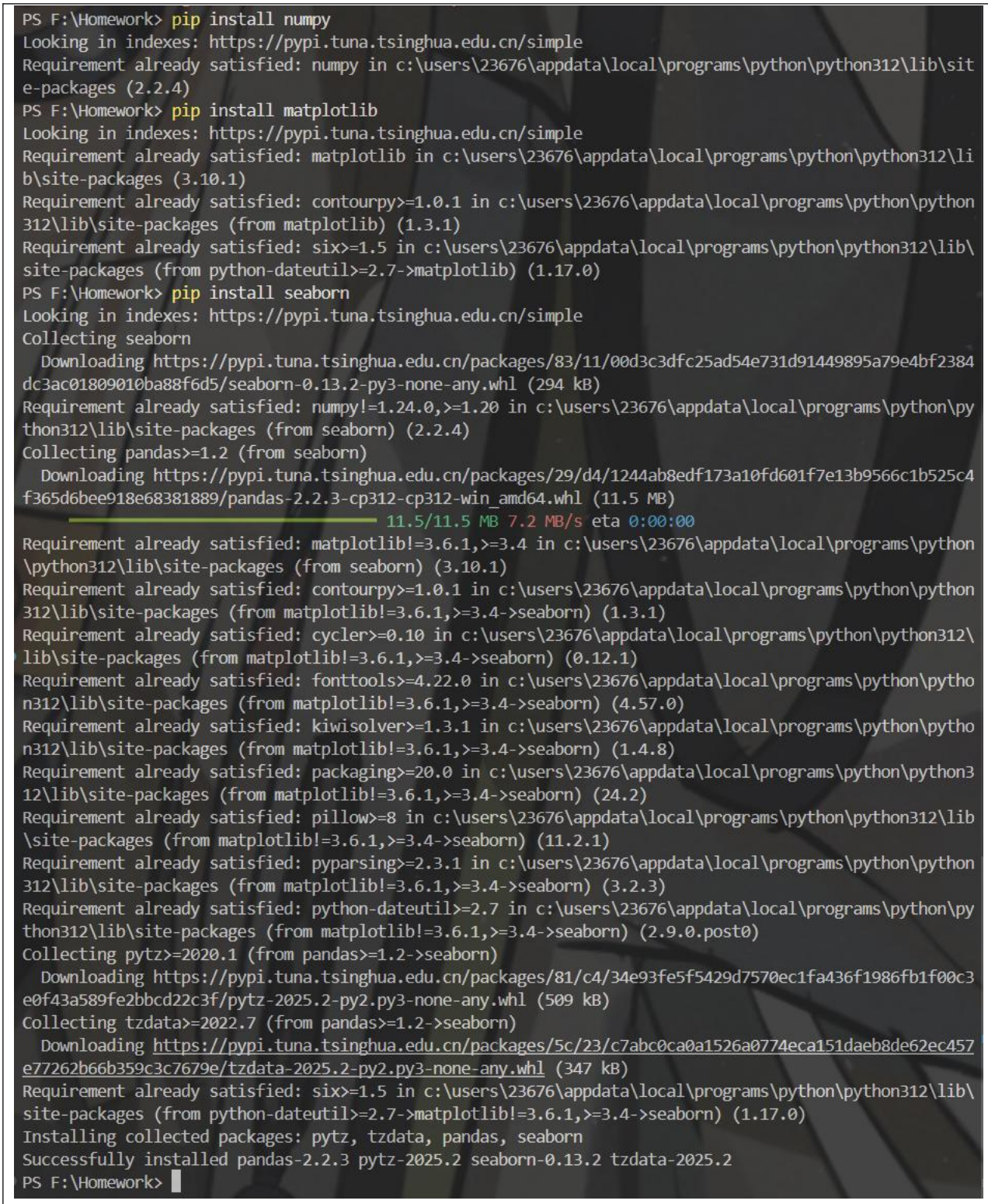
1. 实验内容

In this exercise, we'll take a look at motivating another powerful algorithm—a non-parametric algorithm called random forests.

2. 实验步骤

（1） 配置实验环境

安装基本数学库


PS F:\Homework> pip install numpy
Looking in indexes: https://pypi.tuna.tsinghua.edu.cn/simple
Requirement already satisfied: numpy in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (2.2.4)
PS F:\Homework> pip install matplotlib
Looking in indexes: https://pypi.tuna.tsinghua.edu.cn/simple
Requirement already satisfied: matplotlib in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (3.10.1)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (1.3.1)
Requirement already satisfied: six>=1.5 in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.17.0)
PS F:\Homework> pip install seaborn
Looking in indexes: https://pypi.tuna.tsinghua.edu.cn/simple
Collecting seaborn
Downloading https://pypi.tuna.tsinghua.edu.cn/packages/83/11/00d3c3dfc25ad54e731d91449895a79e4bf2384dc3ac01809010ba88f6d5/seaborn-0.13.2-py3-none-any.whl (294 kB)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (from seaborn) (2.2.4)
Collecting pandas>=1.2 (from seaborn)
Downloading https://pypi.tuna.tsinghua.edu.cn/packages/29/d4/1244ab8edf173a10fd601f7e13b9566c1b525c4f365d6bee918e68381889/pandas-2.2.3-cp312-cp312-win_amd64.whl (11.5 MB)
11.5/11.5 MB 7.2 MB/s eta 0:00:00
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (from seaborn) (3.10.1)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.3.1)
Requirement already satisfied: cycler>=0.10 in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.57.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.8)
Requirement already satisfied: packaging>=20.0 in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (24.2)
Requirement already satisfied: pillow>=8 in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (11.2.1)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0.post0)
Collecting pytz>=2020.1 (from pandas>=1.2->seaborn)
Downloading https://pypi.tuna.tsinghua.edu.cn/packages/81/c4/34e93fe5f5429d7570ec1fa436f1986fb1f00c3e0f43a589fe2bbcd22c3f/pytz-2025.2-py2.py3-none-any.whl (509 kB)
Collecting tzdata>=2022.7 (from pandas>=1.2->seaborn)
Downloading https://pypi.tuna.tsinghua.edu.cn/packages/5c/23/c7abc0ca0a1526a0774eca151daeb8de62ec457e77262b66b359c3c7679e/tzdata-2025.2-py2.py3-none-any.whl (347 kB)
Requirement already satisfied: six>=1.5 in c:\users\23676\appdata\local\programs\python\python312\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.17.0)
Installing collected packages: pytz, tzdata, pandas, seaborn
Successfully installed pandas-2.2.3 pytz-2025.2 seaborn-0.13.2 tzdata-2025.2
PS F:\Homework>

安装机器学习框架

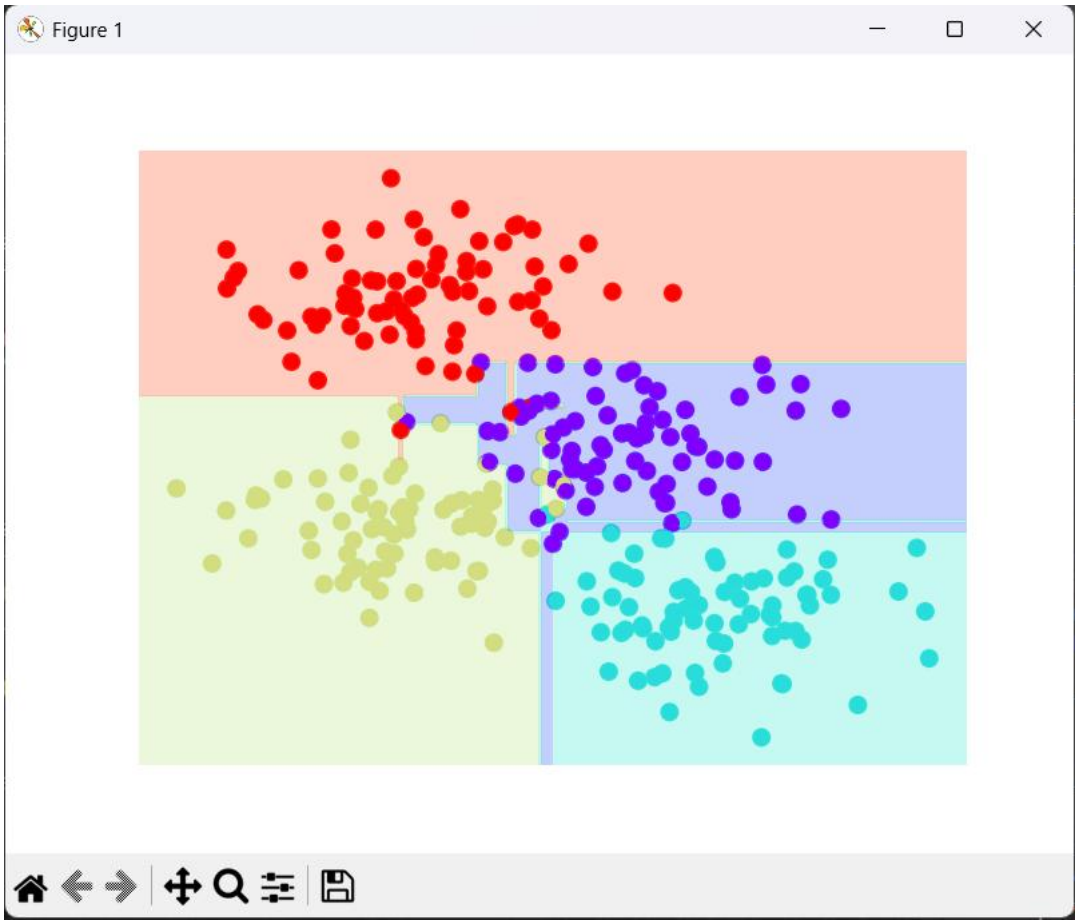
```
PS F:\Homework> pip install scikit-learn
Looking in indexes: https://pypi.tuna.tsinghua.edu.cn/simple
Collecting scikit-learn
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/62/27/585859e72e117fe861c2079bcba35591a84f801
e21bc1ab85bce6ce60305/scikit_learn-1.6.1-cp312-cp312-win_amd64.whl (11.1 MB)
     ---------------------------------------- 11.1/11.1 MB 7.5 MB/s eta 0:00:00
Requirement already satisfied: numpy>=1.19.5 in c:\users\23676\appdata\local\programs\python\python312
\lib\site-packages (from scikit-learn) (2.2.4)
Collecting scipy>=1.6.0 (from scikit-learn)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/f5/6f/e6e5aff77ea2a48dd96808bb51d7450875af154
ee7cbe72188afb0b37929/scipy-1.15.2-cp312-cp312-win_amd64.whl (40.9 MB)
     ---------------------------------------- 40.9/40.9 MB 7.3 MB/s eta 0:00:00
Collecting joblib>=1.2.0 (from scikit-learn)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/91/29/df4b9b42f2be0b623cbd5e2140cafcaa2bef075
9a00b7b70104dcfe2fb51/joblib-1.4.2-py3-none-any.whl (301 kB)
Collecting threadpoolctl>=3.1.0 (from scikit-learn)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/32/d5/f9a850d79b0851d1d4ef6456097579a9005b31f
ea68726a4ae5f2d82ddd9/threadpoolctl-3.6.0-py3-none-any.whl (18 kB)
Installing collected packages: threadpoolctl, scipy, joblib, scikit-learn
Successfully installed joblib-1.4.2 scikit-learn-1.6.1 scipy-1.15.2 threadpoolctl-3.6.0
```

（2） 获取测试数据并编写简单决策树

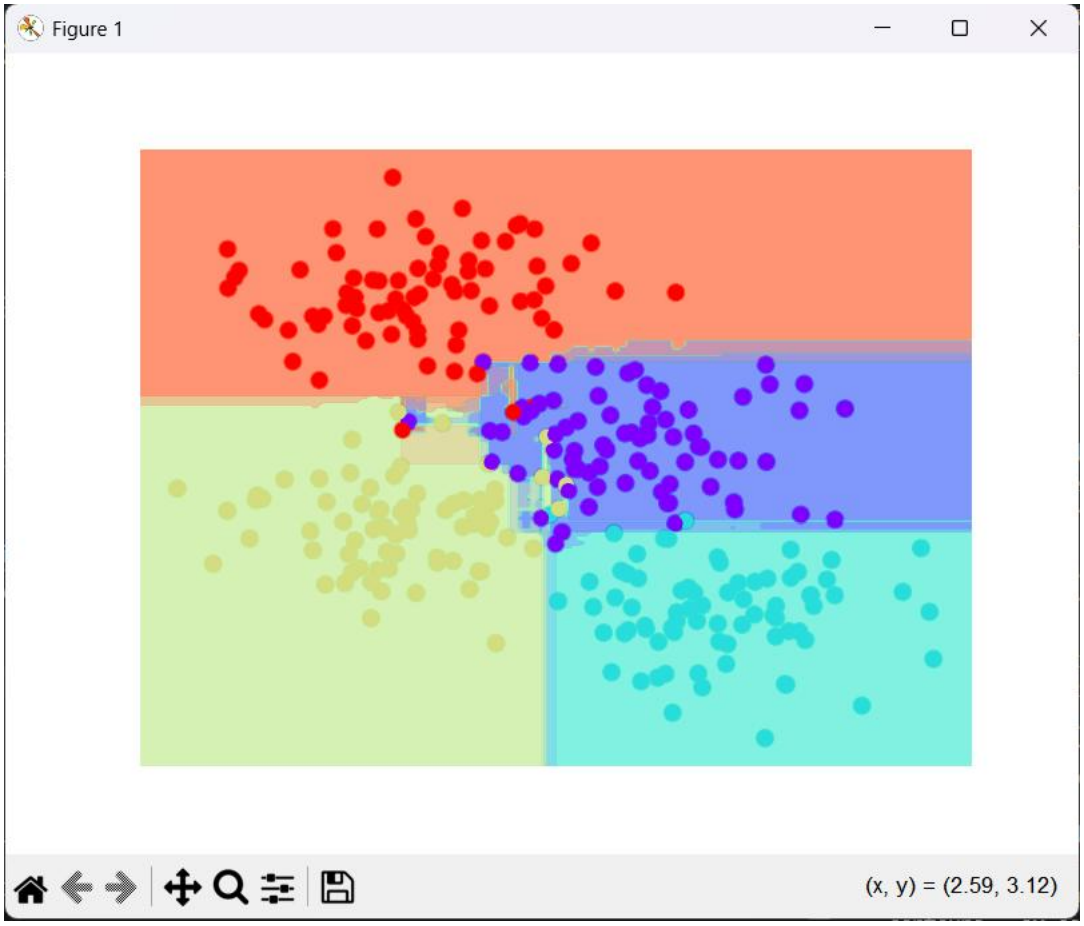
（3） 加入决策器并编写简单随机森林，训练模型

（4） 结果可视化

3. 测试结果



决策树分类结果如上



随机森林分类结果如上

4. 附录：实现源代码

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import make_blobs
import seaborn as sns
from sklearn.tree import DecisionTreeClassifier

from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import BaggingClassifier
from sklearn.ensemble import RandomForestClassifier
sns.set()
X, y = make_blobs(n_samples=300, centers=4, random_state=0,
cluster_std=1.0)
plt.scatter(X[:, 0], X[:, 1], c=y, s=50, cmap="rainbow")
tree = DecisionTreeClassifier().fit(X, y)
plt.show()
plt.cla()
def visualize_classifier(model, X, y, ax=None, cmap="rainbow"):
    ax = ax or plt.gca()
    # 绘制训练点
    ax.scatter(
        X[:, 0], X[:, 1], c=y, s=30, cmap=cmap, clim=(y.min(), y.max()),
```

```
zorder=3
    )
    ax.axis("tight")
    ax.axis("off")
    xlim = ax.get_xlim()
    ylim = ax.get_ylim()
    # 拟合估计器
    model.fit(X, y)
    xx, yy = np.meshgrid(np.linspace(*xlim, num=200), np.linspace(*ylim,
num=200))
    Z = model.predict(np.c_[xx.ravel(), yy.ravel()]).reshape(xx.shape)
    # 创建颜色图
    n_classes = len(np.unique(y))
    contours = ax.contourf(
        xx,
        yy,
        Z,
        alpha=0.3,
        levels=np.arange(n_classes + 1) - 0.5,
        cmap=cmap,
        # clim=(y.min(), y.max()),
        zorder=1,
    )
    ax.set(xlim=xlim, ylim=ylim)
visualize_classifier(DecisionTreeClassifier(), X, y)
tree = DecisionTreeClassifier()
bag = BaggingClassifier(tree, n_estimators=100, max_samples=0.8,
random_state=1)
bag.fit(X, y)
visualize_classifier(bag, X, y)
model = RandomForestClassifier(n_estimators=100, random_state=0)
visualize_classifier(model, X, y)
plt.show()
```