

Python 程序设计 实验报告

Project 3: 鸢尾花数据集分类与可视化

姓名: 陈铄涵 学号: 2024140014 课程: Python 程序设计 日期: 2025/12/12

1 项目概述与工作内容

本项目基于经典的鸢尾花 (Iris) 数据集，**独立设计并实现了完整的分类与可视化系统**，涵盖四种分类任务：四特征三分类（数据探索）、两特征三分类（2D 决策边界）、三特征两分类（3D 二分类边界）、三特征三分类（3D 三分类边界）。



图 1: 项目工作流程图 (红框为本人重点实现: 决策边界可视化模块)

2 项目框架与系统架构

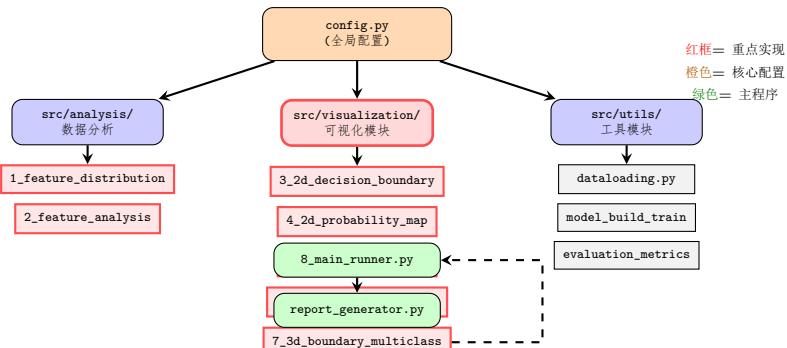


图 2: 项目架构图 (共 11 个核心文件, 本人重点实现 8 个)

3 核心代码与技术亮点

1. 深色样式

```

1 def set_dark_style():
2     plt.style.use('dark_background')
3     plt.rcParams.update({
4         'figure.facecolor': '#0d0d0d',
5         'axes.facecolor': '#0d0d0d',
6     })

```

2. 决策边界网格

```

1 def create_mesh(X, resolution=200):
2     xx, yy = np.meshgrid(
3         np.linspace(x_min, x_max, resolution),
4         np.linspace(y_min, y_max, resolution))
5     return xx, yy

```

3. 3D 等值面 (Marching Cubes)

```

1 from skimage import measure
2 verts, faces, _, _ = measure.marching_cubes(
3     prob_volume, level=0.5)
4 mesh = Poly3DCollection(verts[faces])
5 ax.add_collection3d(mesh)

```

4. 分类器配置

```

1 CLASSIFIERS = {
2     'LogisticRegression': LogisticRegression(),
3     'SVM(RBF)': SVC(kernel='rbf', probability=True),
4     'RandomForest': RandomForestClassifier(),
5     'GradientBoosting': HistGradientBoostingClassifier(),
6 }

```

4 可视化结果展示

本项目实现了四种分类任务的可视化，分别展示如下：

4.1 任务 1: 四特征三分类 (数据探索)

使用全部 4 个特征 (Sepal Length/Width, Petal Length/Width) 对 3 个鸢尾花品种进行探索性分析。图3左图展示四特征分布 (小提琴图 + 箱线图 + 散点图)，Petal 特征呈双峰分布，表明 Setosa 与其他类差异显著；右图为 Pearson 相关系数矩阵，Petal Length 与 Width 强正相关 ($r = 0.96$)。

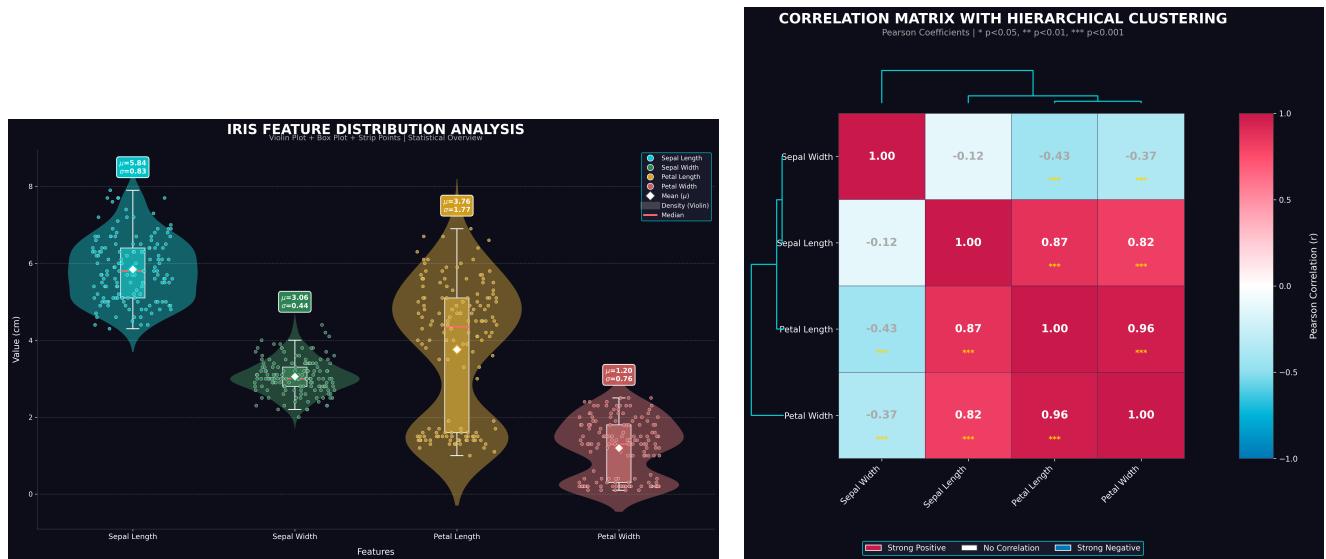


图 3: 四特征三分类数据探索: 左-特征分布分析; 右-相关性热力图

4.2 任务 2: 两特征三分类 (2D 决策边界)

选取最具区分力的 2 个特征 (Petal Length, Petal Width) 进行三分类。图4展示四种分类器的 2D 决策边界与概率分布对比: Logistic Regression 产生线性边界; SVM-RBF 和 Random Forest 捕捉非线性模式; 所有分类器对 Setosa 完美分类, Versicolor 与 Virginica 交界处存在不确定性区域。

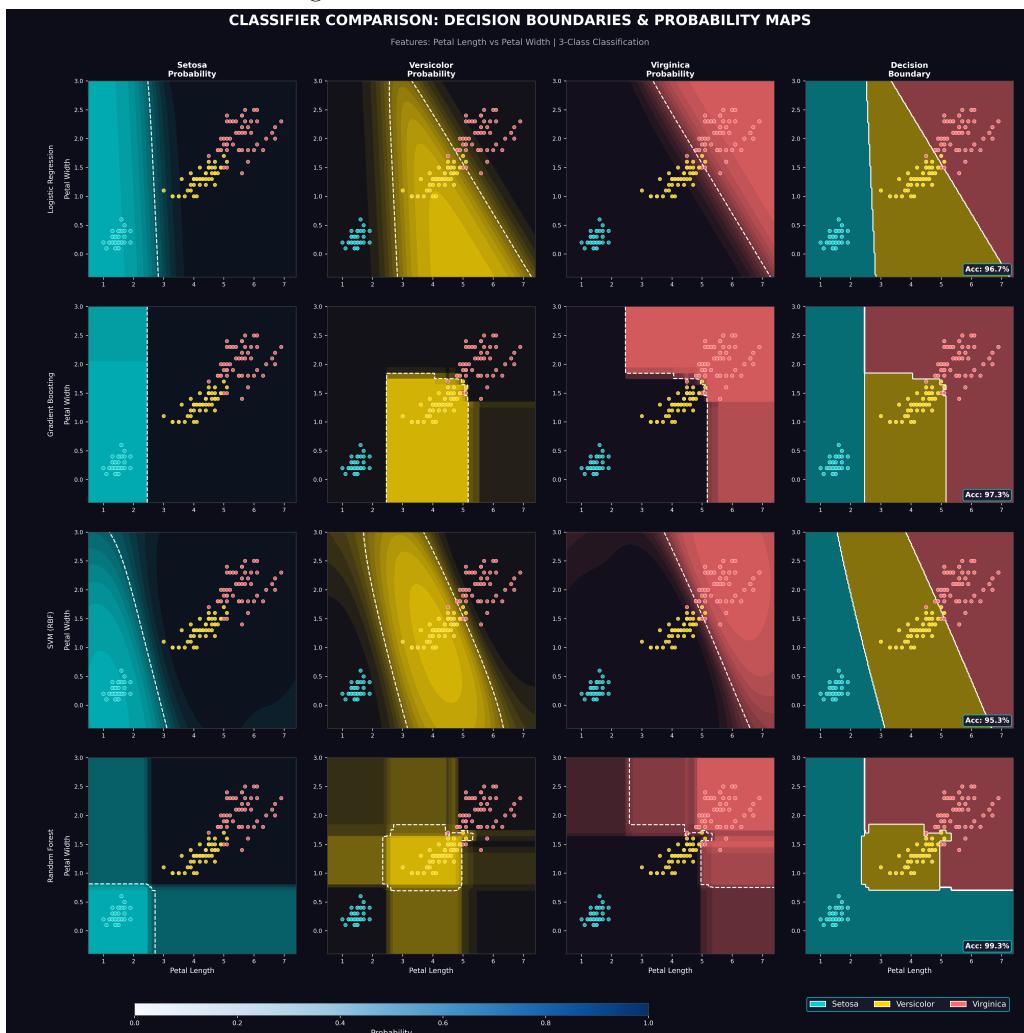
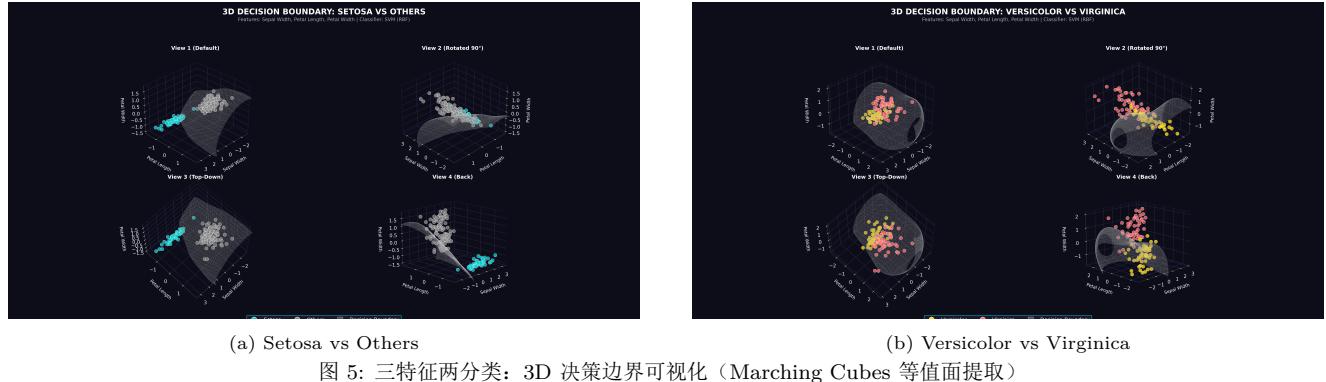


图 4: 两特征三分类: 四种分类器 2D 决策边界与概率图对比 (Petal Length vs Petal Width)

4.3 任务 3: 三特征两分类 (3D 二分类边界)

使用 3 个特征 (Sepal Width, Petal Length, Petal Width) 进行两分类任务。图5展示两个二分类场景: 左图为 Setosa vs Others (边界清晰, 线性可分); 右图为 Versicolor vs Virginica (边界复杂, 需非线性分类器)。采用 Marching Cubes 算法提取 3D 等值面。



4.4 任务 4: 三特征三分类 (3D 三分类边界)

使用 3 个特征同时对 3 个类别进行分类。图6展示 3D 三分类综合视图: 左上为 3D 散点图, 其他三图为不同平面切片, 清晰展示三类的空间分布与决策边界交汇区域。

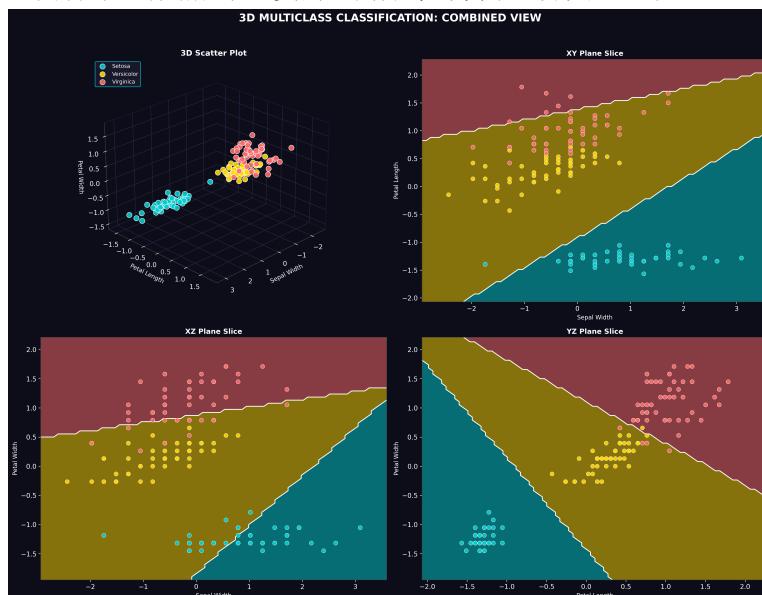


图 6: 三特征三分类: 3D 综合视图 (散点图 + 切片)

表 1: 分类器性能 (5 折交叉验证)

分类器	Acc	F1
Logistic Reg.	97.3%	97.3%
SVM (RBF)	96.7%	96.7%
Random Forest	96.7%	96.7%
Grad. Boost	94.7%	94.6%

所有分类器均达到 94%+ 准确率, 验证了可视化方法的有效性。

5 结论与总结

- 特征重要性: Petal Length/Width 是区分品种的关键特征, 仅用两特征即可达 95%+ 准确率。
- 类别可分性: Setosa 线性可分; Versicolor 与 Virginica 边界模糊, 是分类难点。
- 可视化价值: 2D/3D 决策边界可视化直观展示分类器决策逻辑, 识别不确定性区域。
- 技术创新: 采用 Marching Cubes 算法提取 3D 等值面, 实现科研级深色可视化风格。

技术特点: 模块化架构 | 四种分类任务 | Marching Cubes 3D 等值面 | 科研级可视化风格

6 运行说明

```
1 pip install -r requirements.txt      # 安装依赖
2 python src/8_main_runner.py        # 一键生成所有可视化
```

GitHub: <https://github.com/Csh0601/project3for-pythonhomework>

7 教师评语

评分: _____ 日期: _____