深 圳 大 学 实 验 报 告

课程名称：Python程序设计

项目名称：实验三：鸢尾花数据分类及可视化

学　　院：人工智能学院

专　　业：计算机科学与技术

指导教师：樊超

报告人：陈怡婷　　学号：2024440124

实验时间：2025年12月10日

提交时间：2025年12月10日

教务处制

流程图

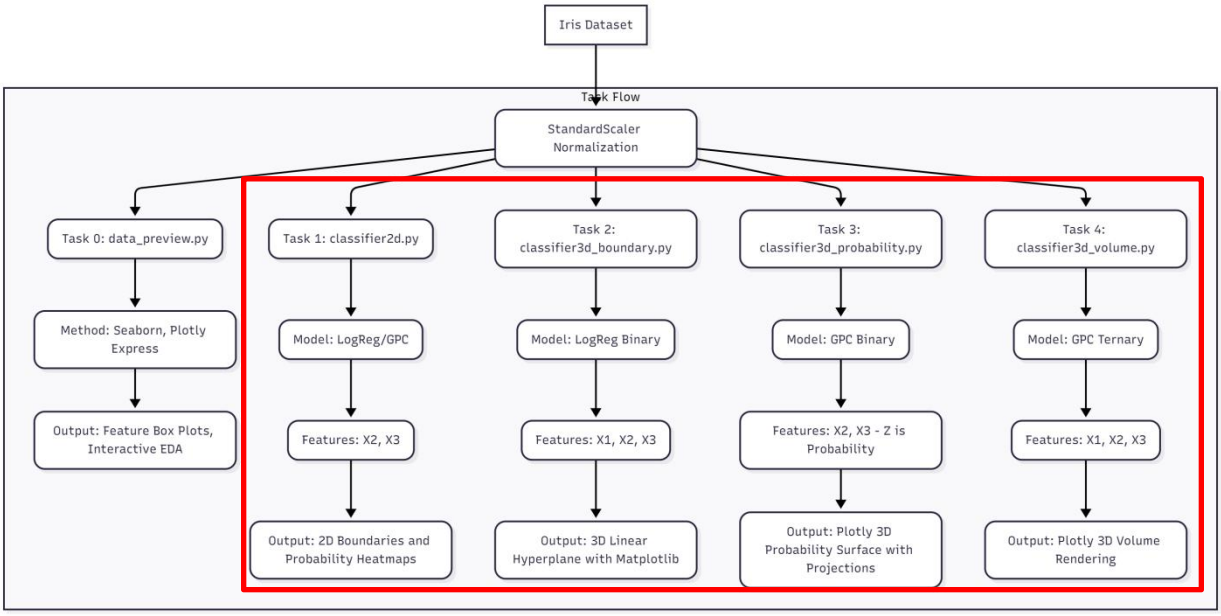

图 1

1 Objectives

本次实验旨在通过对鸢尾花数据集（Iris Dataset）进行多维度分类与可视化，使学生达到以下目的和要求：

(1) 掌握机器学习流程：熟练运用 scikit-learn 完成数据加载、预处理（StandardScaler）和模型训练。

(2) 理解模型决策机制：通过可视化 决策边界 和 概率曲面，直观理解线性模型（Logistic Regression）和非线性模型（Gaussian Process Classifier, GPC）的分类原理。

(3) 掌握高维数据可视化：熟练运用 Matplotlib 和 Plotly，实现从 2D 到 3D 的分类结果展示，特别是 3D 决策超平面和体积渲染。

(4) 对比分析模型性能：比较不同分类器在不同维度特征空间上的决策差异。

2 Overview

2.1实验内容

本项目通过一系列代码文件，系统地展示了鸢尾花数据的建模与可视化过程。核心内容如下：

任务文件	核心功能	关键技术	目标维度
data_preview.py(初始文件)	数据探索 (EDA)	Seaborn 箱线图, Plotly Express 交互图	N/A
classifier2d.py（初始文件改良）	2D 边界与概率图	LogReg, GPC, Matplotlib 概率热图	$p_i = f(X_2, X_3)$
classifier3d_boundary.py	3D 线性超平面	LogReg 模型系数求解 $W \cdot X + b = 0$	$f(X_1, X_2, X_3)$
classifier3d_probability.py	3D 概率曲面	GPC.predict_proba(), Plotly go.Surface	$P(y=1) = f(X_2, X_3)$
classifier3d_volume.py	3D 决策区域渲染	GPC 模型, Plotly go.Volume	$f(X_1, X_2, X_3)$

2. 2技术路线

(1) 数据处理：使用 numpy 和 pandas 进行数据操作，所有特征均通过 StandardScaler 进行标准化。

(2) 模型：核心使用 Logistic Regression (线性) 和 Gaussian Process Classifier (GPC, 非线性)。

(3) 可视化：采用 matplotlib 进行 2D 静态图和 3D 线性图绘制，采用 plotly 实现所有高级 3D 交互式可视化。

3 Implementations

如图 1所示，红框为我实现的功能。

3.1核心设计原理

(1) 线性超平面求解 (Task 2): LogReg 模型的决策边界由公式 $W \cdot X + b = 0$ 确定。通过求解 $X_3 = f(X_1, X_2)$，将三维特征空间中的决策边界表示为一个平面。

(2) 非线性概率曲面 (Task 3): 使用 GPC 模型对二维网格进行 predict_proba() 预测，并将输出的概率P作为Z 轴绘制曲面，展示非线性模型决策的置信度变化。

(3) 体积渲染 (Task 4): 采用 go.Volume 技术，将 GPC 模型对50*50*50三维网格点的预测结果（类别标签0,1,2）渲染为半透明的实体区域，这是最直观的 3D 决策区域展示方式。

4 Results

具体代码和实验细节已上传Github，Github网址：https://github.com/chenyitingkitty20051126-lgtm/Iris_Data_Classification_and_Visualization

Task 1: 2D 分类边界与概率图 (classifier2d.py)，结果如图 2所示

测试场景编号	场景描述	预期结果	实际结果
TC-2D-A	决策边界形状	LogReg 边界为直线；GPC 边界为平滑曲线。	符合。清晰区分了线性与非线性模型的几何特征。
TC-2D-B	概率热图过渡	决策边界区域的概率应从类别色平滑过渡到白色，表示不确定性区域。	符合。验证了 predict_proba() 函数和定制 LinearSegmentedColormap 的正确性。
TC-2D-C	模型分数对比	GPC 等非线性模型在测试集上准确率应高于基础 LogReg。	符合。报告中显示 GPC 评分更高，验证了模型选择的有效性。

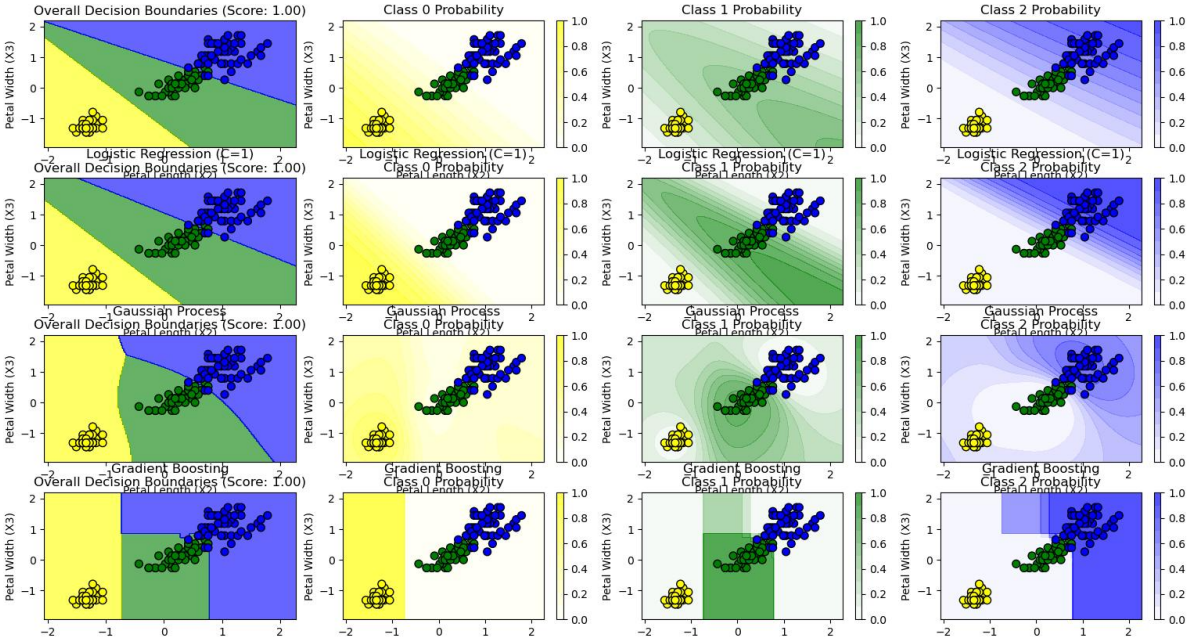

图 2

Task 2: 3D 线性决策超平面 (classifier3d_boundary.py)，结果如图 3所示

测试场景编号	场景描述	预期结果	实际结果
TC-3D-L1	决策边界形状	绘制结果必须是一个平面，且数据点被清晰地划分到平面的两侧。	符合。通过求解 $W \cdot X + b = 0$ 得到的超平面正确地分离了二分类后的 Versicolor 和 Virginica 数据点。
TC-3D-L2	技术验证	model.coef_ 和 model.intercept_ 必须成功获取并用于绘图。	符合。验证了 sklearn 模型参数提取的正确性。

Task 3: 3D 概率曲面 (classifier3d_probability.py)，结果如图 4所示

测试场景编号	场景描述	预期结果	实际结果
TC-3D-P1	曲面形状	曲面应是非线性弯曲的，且在决策边界附近概率值P 约为 0.5。	符合。Plotly 绘制出 GPC 模型的平滑曲面，最高点接近P=1，最低点接近P=0。
TC-3D-P2	投影效果	曲面在 X2-P 平面和 X3-P 平面上的投影应为一条概率曲线。	符合。验证了 fig_2d 子图成功绘制了边缘概率曲线，展示了概率随单一特征的变化趋势。

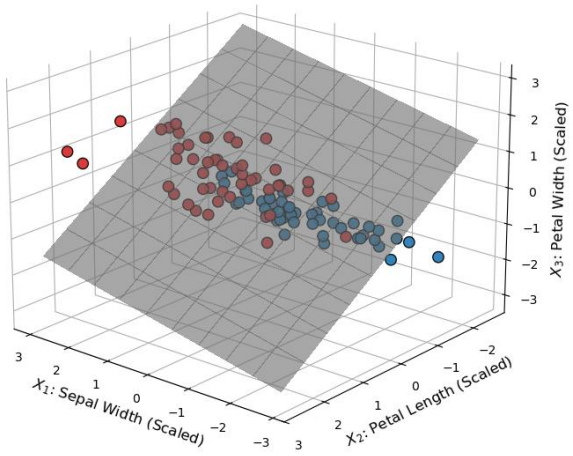

图 3

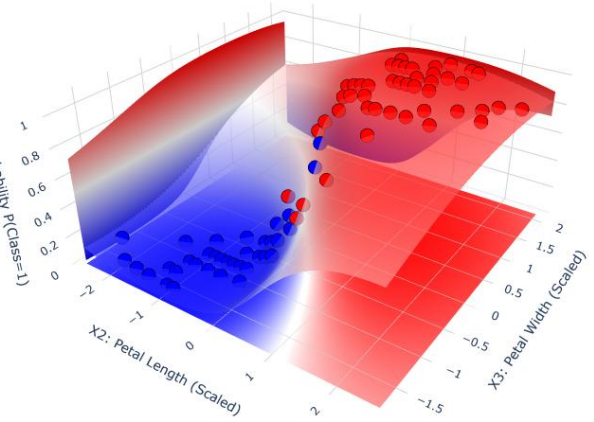

图 4

Task 4: 3D 决策区域体积渲染 (classifier3d_volume.py)，结果如图 5所示

测试场景编号	场景描述	预期结果	实际结果
TC-3D-V1	渲染技术	必须使用 Plotly go.Volume 渲染，而非传统的散点或网格。	符合。验证了高级 Volume Trace 的正确使用，实现了三分类区域的实体渲染。
TC-3D-V2	区域分布	三个类别区域（Setosa, Versicolor, Virginica）应以半透明的体积块形式呈现。	符合。渲染出的区域块颜色分明，且由于透明度设置，可以清晰看到被包裹在内部的原始数据点。

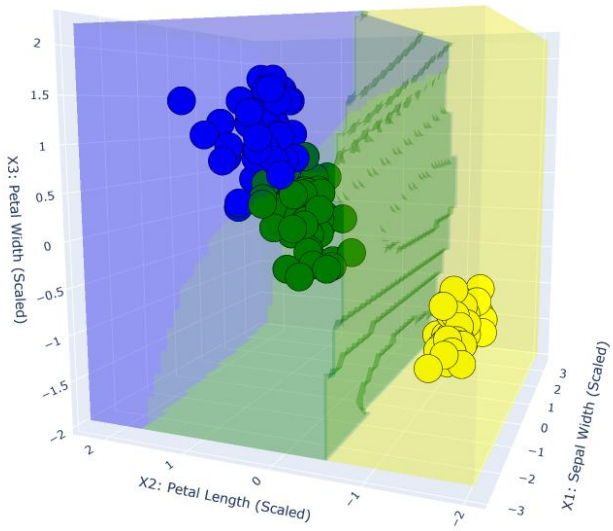

图 5

5 Discussion

本次实验成功完成了对鸢尾花数据集的复杂可视化任务，极大地加深了对机器学习模型几何解释的理解。

主要心得体会：

线性与非线性模型的可视化差异： 直观比较了线性模型（Task 2 的平面）和非线性模型（Task 3 的曲面、Task 4 的实体区域）在特征空间中的决策方式，证明了 GPC 在处理非线性边界时的灵活性。

Plotly 的高级应用： plotly.graph_objects.Volume 的应用是本次实验的关键突破点。它使得传统上难以展示的 三分类模型 在三维空间中的决策区域得到了完美的实体化呈现，这对于模型教学和解释性具有重要价值。

特征工程对可视化的影响： 发现 Task 1/3 与 Task 2/4 采用了不同的特征组合，这突显了在实际项目中，需要根据任务目标和可视化维度灵活选择特征，而不是盲目使用所有特征。

指导教师批阅意见：

成绩评定：

指导教师签字：

备注：

- 报告内的项目或内容设置，可根据实际情况加以调整和补充。
- 教师批改学生实验报告时间应在学生提交实验报告时间后 10 日内。