浙江工业大学

人工智能原理及应用
实验报告

实验名称：SVM 鸢尾花分类

学　　号：202205240220

姓　　名：潘家航

专业班级：自动化 2304

学　　院：信息工程学院

指导教师：付明磊

目录

一、实验目的

通过使用支持向量机（SVM）算法对鸢尾花数据集进行分类实验，理解 SVM 的基本原理、核函数的作用及分类决策边界的形成过程。

二、实验设备

1.硬件设备： 游侠 G15。

2.软件环境： 主流浏览器（如 Chrome、Edge）、代码编辑器 pycharm、Python 编程环境。

3.开发平台： 百度 AI 开放平台（ai.baidu.com）的 EasyDL 产品。

三、实验内容

支持向量机（SVM，Support Vector Machine）是一种基于统计学习理论的监督学习模型，主要用于分类和回归任务。其核心思想是找到一个最优超平面，使不同类别的数据点到该平面的间隔最大。对于非线性可分问题，可以通过核函数（Kernel Function）将数据从低维空间映射到高维空间，实现线性可分。

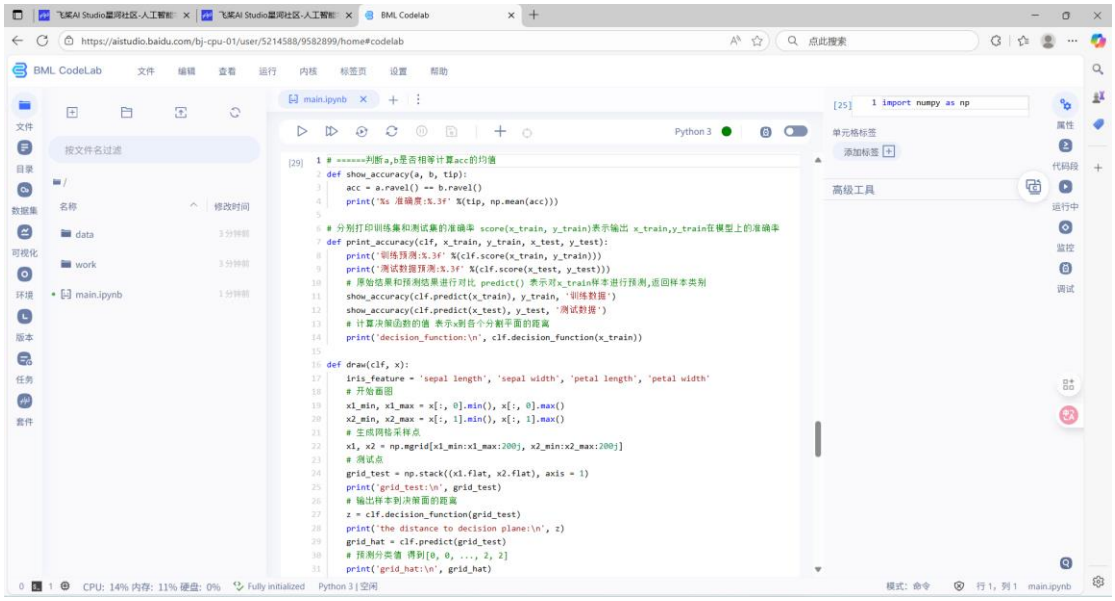
四、实验过程

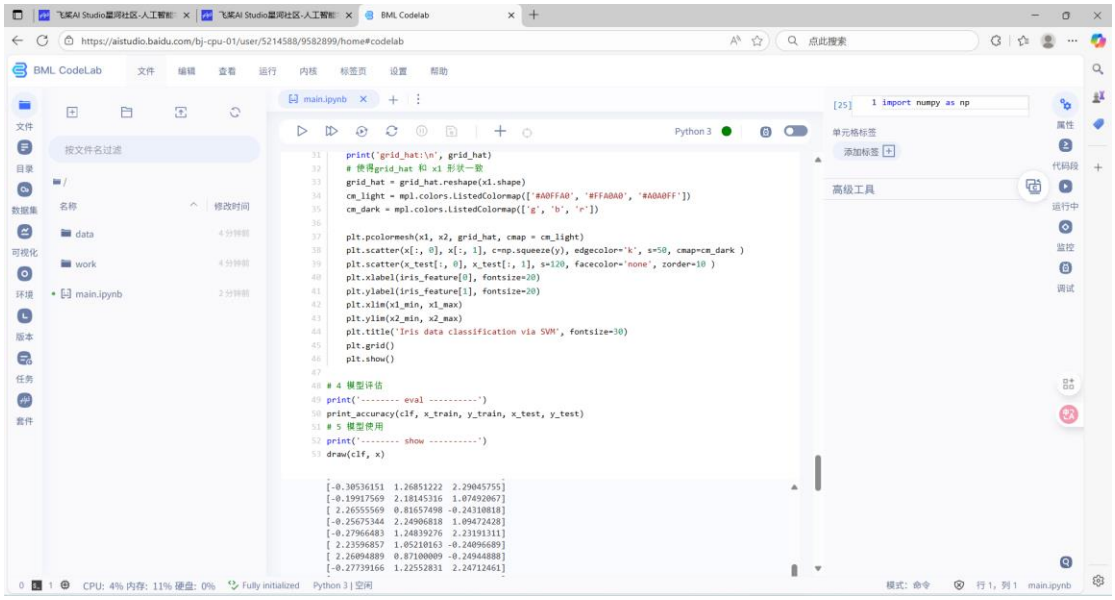
1. 导入所需库与数据集，并选取鸢尾花数据集的前两个特征用于可视化。
2. 使用 `train_test_split` 将数据划分为训练集与测试集。
3. 构建 SVM 模型，选择线性核函数（kernel="linear"）进行训练。
4. 对模型进行评估，并绘制二维特征下的决策边界。

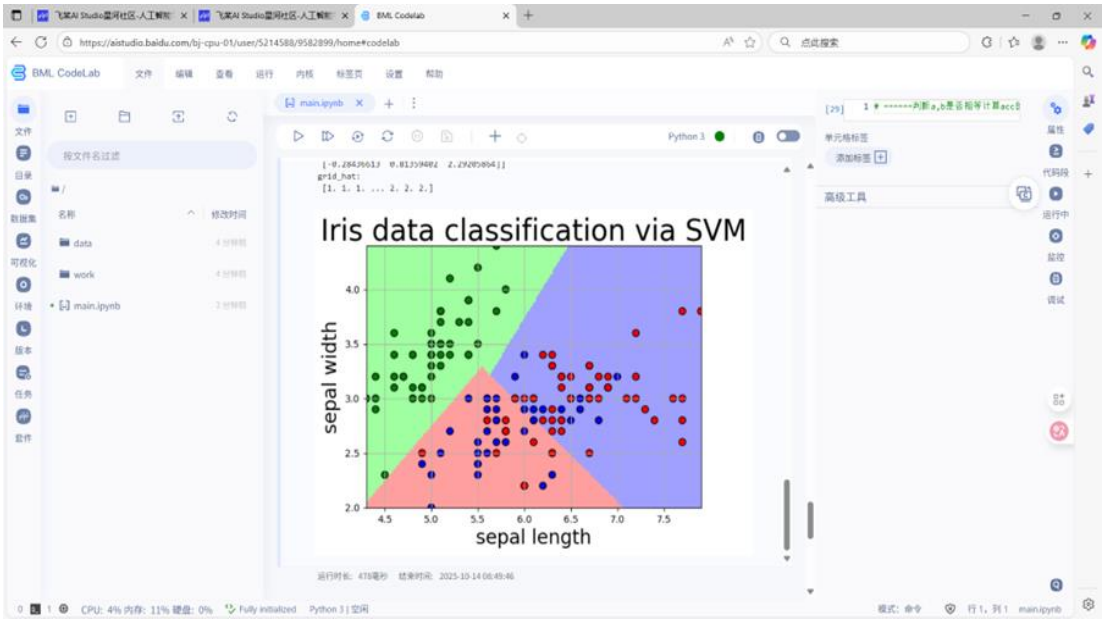
关键代码片段如下：

```
from sklearn import svm, model_selection
from sklearn.datasets import load_iris
iris = load_iris()
x = iris.data[:, :2]
y = iris.target
```

```
x_train, x_test, y_train, y_test = model_selection.train_test_split(x, y, test_size=0.2, random_state=1)
clf = svm.SVC(kernel="linear")
clf.fit(x_train, y_train)
```







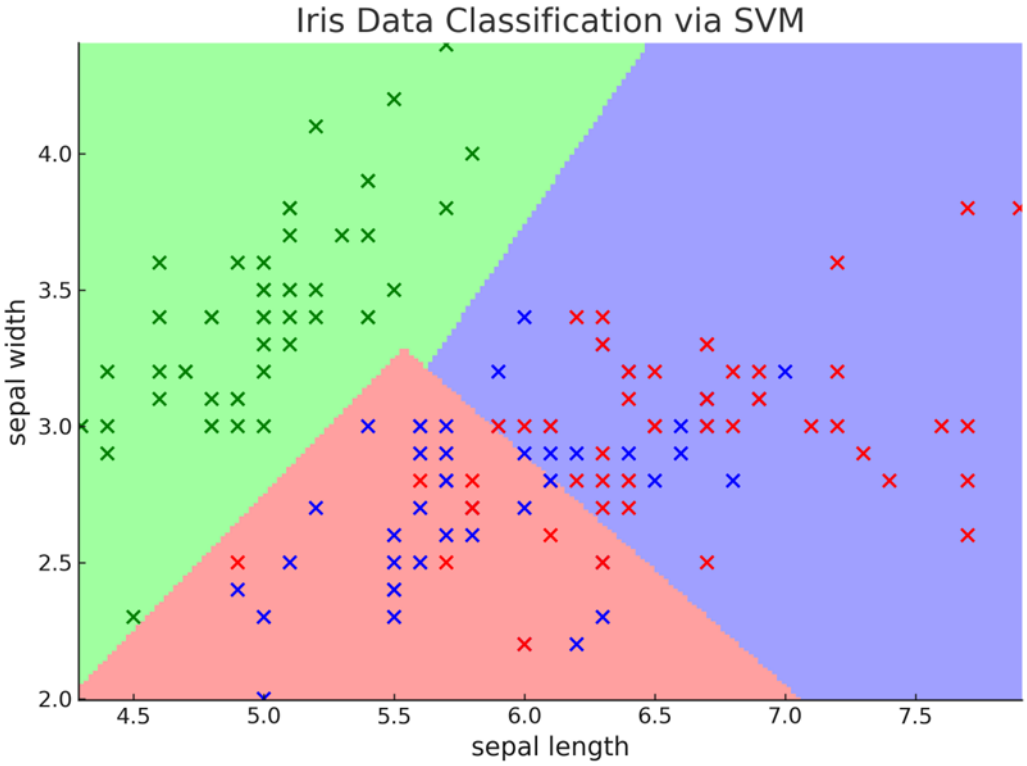
五、实验结果与分析

模型训练后，得到以下性能指标：

训练集准确率：0.817

测试集准确率：0.800

可视化结果如下图所示，展示了 SVM 在线性核函数下的分类边界：



鸢尾花数据集 SVM 分类决策边界图

六、实验小结

通过本次实验，掌握了 SVM 分类器的原理及其在实际数据集中的应用。实验结果表明，SVM 能够较好地区分鸢尾花的三种类别，并通过调整核函数或参数，可进一步提升分类性能。

七、其它

在本次基于 SVM 的鸢尾花分类实验中，我们虽成功构建了模型并实现了分类目标，但也遇到了一些挑战。实验初期，由于仅使用了数据集的前两个特征进行模型训练和可视化，导致模型无法充分利用全部特征信息，可能限制了其性能上限；同时，线性核函数在处理潜在的非线性关系时表现出的局限性，以及惩罚系数 C 等参数调优对模型效果较为敏感的问题，也都是实验中需要克服的难点。针对这些不足，未来的改进方向包括尝试引入如 RBF 核、多项式核等非线性核函数以捕捉更复杂的数据模式，利用所有特征或通过 PCA 等降维方法在保留更多信息的前提下进行训练，以及采用网格搜索（GridSearchCV）或交叉验证等系统化方法进行参数优化。通过本次实验，我们不仅加深了对 SVM 算法原理（如间隔最大化、核函数作用）

的理解，还熟练掌握了使用 scikit-learn 库构建、训练和评估 SVM 分类器的完整流程，包括数据预处理、模型训练、评估及决策边界可视化等关键步骤，并认识到了数据标准化、核函数与参数选择对模型性能的重要影响。展望未来，SVM 算法在图像识别（如手写数字识别）、文本情感分析、生物信息学中的基因分类以及金融风险控制等多个领域均有广泛的应用前景，本次实验为后续深入研究和应用复杂的机器学习模型奠定了坚实的基础。