

# 四川大学计算机学院、软件学院

## 实验报告

学号：2022141460176 姓名：杨一舟 专业：计算机科学与技术

课程名称	数据挖掘导引	实验课时	
实验项目	使用 SVM 算法进行数据挖掘实验	实验时间	2024. 5. 20
实验目的	基于 IRIS 数据集通过编程实现 SVM 算法的实践		
实验环境	Visual studio Code		
实 验 内 容 (算法、程序、步骤和方法)	<h3>一、实验步骤</h3> <p>SVM (Support Vector Machine) 算法实验步骤:</p> <p><b>数据准备:</b></p> <p>收集带有标签的训练样本数据集, 其中每个样本都有对应的特征和标签, 本实验中选用 IRIS 数据集。</p> <p><b>选择核函数:</b></p> <p>根据问题的特点选择合适的核函数, 如线性核、多项式核、高斯核等。</p> <p><b>构建模型:</b></p> <p>将数据集映射到高维空间 (如果使用非线性核函数), 并搜索满足最优化目标函数的最优分离超平面。</p>		

**训练模型：**

使用训练数据训练 SVM 模型，找到最佳的超平面或超平面集合。

**预测：**

使用训练好的模型对新的数据进行预测，判断其类别。

**评估模型：**

使用测试集评估模型的性能，计算准确率、召回率、F1 分数等指标。

**调参：**

根据评估结果调整模型参数，如惩罚项系数  $C$ 、核函数参数等，以优化模型性能。

**优化和迭代：**

如果需要，可以进行特征选择、降维等操作，进一步提高模型的性能和效率。

## 二、实验源代码

```
import matplotlib.pyplot as plt
import numpy as np
from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC
from sklearn.metrics import classification_report, confusion_matrix

# 加载 iris 数据集
iris = datasets.load_iris()
X = iris.data[:, :2] # 只选择前两个特征: sepal length 和 sepal width
y = iris.target
feature_names = iris.feature_names[:2] # 更新特征名
target_names = iris.target_names

# 划分数据集为训练集和测试集
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# 数据标准化
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# 初始化 SVM 分类器
svm = SVC(kernel='linear', C=1, random_state=42)

# 训练模型
svm.fit(X_train, y_train)

# 预测测试集
y_pred = svm.predict(X_test)

# 打印分类报告和混淆矩阵
print("Classification report:")
print(classification_report(y_test, y_pred, target_names=target_names))
print("\nConfusion matrix:")
print(confusion_matrix(y_test, y_pred))
```

```
# 可视化决策边界
plt.figure(figsize=(10, 8))

# 创建网格用于评估模型
x_min, x_max = X_train[:, 0].min() - 1, X_train[:, 0].max() + 1
y_min, y_max = X_train[:, 1].min() - 1, X_train[:, 1].max() + 1
xx, yy = np.meshgrid(np.arange(x_min, x_max, 0.1),
                     np.arange(y_min, y_max, 0.1))
Z = svm.predict(np.c_[xx.ravel(), yy.ravel()])
Z = Z.reshape(xx.shape)

# 绘制决策边界和样本点
plt.contourf(xx, yy, Z, alpha=0.4)
plt.scatter(X_train[y_train == 0, 0], X_train[y_train == 0, 1], c='red',
            label=target_names[0], alpha=0.8)
plt.scatter(X_train[y_train == 1, 0], X_train[y_train == 1, 1], c='green',
            label=target_names[1], alpha=0.8)
plt.scatter(X_train[y_train == 2, 0], X_train[y_train == 2, 1], c='blue',
            label=target_names[2], alpha=0.8)

# 添加图例和标题
plt.xlabel(feature_names[0])
plt.ylabel(feature_names[1])
plt.legend()
plt.title('SVM Decision Boundary with Iris Dataset')
plt.show()
```

### 三、实验结果

```
(mountaintorch) C:\Users\MountainMist\Desktop\2024春\数据挖掘\数据挖掘导引期末报告>python SVM算法.py
Classification report:
      precision    recall  f1-score   support

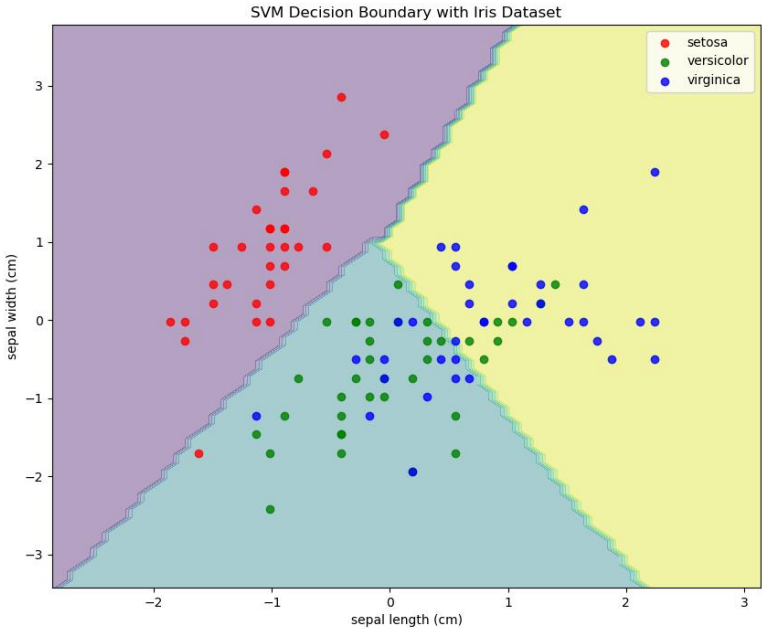
   setosa         1.00        1.00        1.00         19
  versicolor    0.54        0.54        0.54         13
   virginica    0.54        0.54        0.54         13

 accuracy         0.73         45
  macro avg       0.69        0.69        0.69         45
 weighted avg     0.73        0.73        0.73         45

Confusion matrix:
[[19  0  0]
 [ 0  7  6]
 [ 0  6  7]]
```

以上为分类报告和混淆矩阵

（接上）  
实验内容  
（算法、程  
序、步骤和  
方法）



以上为结果可视化展示

数据记录 和计算	<p>实验结果如上述所示</p> <p>结果分析：</p> <p>本代码的可视化结果展示了一个SVM分类器在 IRIS 数据集上的分类效果，包括决策边界和各类别的样本点。通过图形，可以直观地看到不同类别之间的分隔。</p>
结 论 (结 果)	<p>成功完成了 SVM 算法的设计与实践</p>
小 结	<p>在 SVM 算法设计与实践实验中，我深入理解了 SVM 的核心原理，包括超平面的选择与优化。实验结果表明，SVM 在 IRIS 数据集上能够准确划分不同类别的样本。这次实践加深了我对 SVM 算法的理解，并提高了我在分类问题上的建模能力。</p>
指导老师 评 议	<p>成绩评定：</p> <p>指导教师签名：</p>

# 实验报告说明

专业实验中心

**实验名称** 要用最简练的语言反映实验的内容。如验证某程序、定律、算法，可写成“验证×××”；分析×××。

**实验目的** 目的要明确，要抓住重点，可以从理论和实践两个方面考虑。在理论上，验证定理、公式、算法，并使实验者获得深刻和系统的理解，在实践上，掌握使用实验设备的技能技巧和程序的调试方法。一般需说明是验证型实验还是设计型实验，是创新型实验还是综合型实验。

**实验环境** 实验用的软硬件环境（配置）。

**实验内容（算法、程序、步骤和方法）** 这是实验报告极其重要的内容。这部分要写明依据何种原理、定律算法、或操作方法进行实验，要写明经过哪几个步骤。还应该画出流程图（实验装置的结构示意图），再配以相应的文字说明，这样既可以节省许多文字说明，又能使实验报告简明扼要，清楚明白。

**数据记录和计算** 指从实验中测出的数据以及计算结果。

**结论（结果）** 即根据实验过程中所见到的现象和测得的数据，作出结论。

**小结** 对本次实验的体会、思考和建议。

**备注或说明** 可写上实验成功或失败的原因，实验后的心得体会、建议等。

**注意：**

- 实验报告将记入实验成绩；
- 每次实验开始时，交上一次的实验报告，否则将扣除此次实验成绩。