

《模式识别与深度学习》

实验 1——SVM 分类算法实验指导书

实验时间：2022 年 9 月 20 日-2022 年 9 月 28 日

1. 实验目的

- 1) 熟练掌握 SVM 分类算法原理及代码实现，能够对已实现算法进行灵活调用获取分类结果。
- 2) 熟练掌握线性和非线性 SVM 算法，并且根据实验数据的特性调节线性函数或核函数。

2. 实验环境

- 1) Python3
- 2) Numpy （科学计算，数据处理）
- 3) Scikit-learn （开源机器学习库）

3. 实验要求

本次实验包含两部分：

- 1) 在理解并掌握 SVM 鸢尾花分类算法原理的基础上，调节不同的参数，汇报不同参数对实验结果的影响，并尝试对结果进行分析，哪些参数对实验结果影响比较大，原因是什么。
- 2) 使用肿瘤数据进行**肿瘤分类**的实验，汇报训练集以及测试集的准确率。调节参数，汇报不同的参数对实验结果的影响，并尝试对结果进行分析，哪些参数对实验结果影响比较大，原因是什么。

4. 实验内容

- 1) 准备工作

安装好对应的实验环境，简单了解 Numpy、Scikit-learn 等函数库的使用。

- 2) 实验材料（已上传到智慧树作业考试实验 1-SVM）
 - “iris.data”：鸢尾花数据
 - “iris_cls.py”：svm 进行鸢尾花分类完整代码

- “tumor.data”: 肿瘤数据

3) 实验一:

对**鸢尾花分类**实验更改不同参数, 查看不同的参数对分类结果的影响(主要修改的参数为 C, kernel, decision_function_shape)。

备注: 注意在调节参数时保证划分训练和测试数据集种子点(seed)的保持相同。

4) 实验二:

更改代码为对肿瘤数据进行**肿瘤分类**的实验。调节参数, 汇报不同的参数对实验结果的影响。

5. 实验方式

每位同学独立上机编程实验, 由助教进行实验指导。

6. 参考内容

Numpy 相关: <https://www.runoob.com/numpy/numpy-tutorial.html>

Numpy 包安装方法: 使用命令 pip install numpy

Scikit-learn 包: <https://scikit-learn.org/stable/>

Scikit-learn 函数库安装方法: 使用命令 pip install scikit-learn

SVM 函数: <https://scikit-learn.org/stable/modules/svm.html#svm-classification>

7. 实验报告要求

- 1) 现已给出 SVM 对鸢尾花分类算法实现, 要求对给出的算法做到理解, 并调试不同参数, 查看不同参数对实验结果的影响, 将不同参数对应的训练集和测试集准确率通过表格记录, 并尝试对结果进行分析, 哪些参数对实验结果影响比较大, 原因是什么。
- 2) 更改代码为对肿瘤数据的进行分类的算法实现, 并提交代码, 调试不同参数, 查看不同参数对实验结果的影响, 将不同参数对应的训练集和测试集准确率通过表格记录, 并尝试对结果进行分析, 哪些参数对实验结果影响比较大, 原因是什么。

8. 实验结果提交

- 1) 将代码和实验报告打包为“实验 1-学号-姓名.zip”, 其中实验报告的命名

是“实验 1-实验报告-学号-姓名.docx”；

2) Zip 文件通过智慧树作业页面进行提交。