**课后作业三：支持向量机**

任务：基于SVM构建机器学习模型, 对Omniglot中的字符分类。

数据集：Omniglot Dataset是全语言文字数据集，包含各种语言的不同字母表，如日语的平假名和片假名，韩语的元音和辅音，最常见的拉丁字母abcd等。Omniglot Dataset共包含50个不同语言的字母表，每个字母表中包含不同的字符，共1623种字符，每种字符由20个不同的人书写。本次实验选取其中200种字符，每种字符的15张图片作为训练集，5张图片作为测试集。

**实现：**

1.导入必要的库

import csv

import numpy as np

import pandas as pd

import scipy.io

from sklearn.model\_selection import train\_test\_split #机器学习中非常重要的库，包括一些分类、回归、聚类、降维、模型选择和预处理

from sklearn.svm import SVC

2.加载数据

train\_data = scipy.io.loadmat(r'D:\dataenclorse\third\train\_data.mat')

test\_data = scipy.io.loadmat(r'D:\dataenclorse\third\test\_data.mat')

3.获取训练数据和标签，并划分训练集和测试集。从 .mat 文件中加载训练数据并展平图像数据。生成对应的标签，每个类别有 15 张图片，共有 200 个类别。

X\_train = train\_data['train']

X\_train = X\_train.reshape(-1, 28 \* 28)  # 将图像数据展平为一维数组

y\_train = np.repeat(np.arange(1, 201), 15)  # 生成标签

# 划分训练集和验证集

X\_train, X\_val, y\_train, y\_val = train\_test\_split(X\_train, y\_train, test\_size=0.2, random\_state=42)

4.初始化和训练SVM模型，创建一个线性核的SVM模型，并用训练数据进行训练。

# 初始化SVM模型

svm\_model = SVC(kernel='linear', C=1.0, random\_state=42)

# 训练SVM模型

svm\_model.fit(X\_train, y\_train)

5.用训练好的模型在验证集和测试集上进行预测。

# 在验证集上进行预测

y\_val\_pred = svm\_model.predict(X\_val)

# 在测试集上进行预测

X\_test = test\_data['test']

X\_test = X\_test.reshape(-1, 28 \* 28)  # 将测试数据展平为一维数组

y\_test\_pred = svm\_model.predict(X\_test)

L = list(range(1, 1001))

6.将测试集的预测结果保存到submission.csv中文件中

file\_path = r'D:\dataenclorse\third\submission.csv'

with open(file\_path, 'w', newline='', encoding='utf-8') as f:

        fieldnames = ['图片', '预测结果']

        f\_csv = csv.DictWriter(f, fieldnames=fieldnames)

        f\_csv.writeheader()

        for i in range(0, len(y\_test\_pred)):

            f\_csv.writerow({'图片':L[i], '预测结果':y\_test\_pred[i]})

pass

7.结果如下：

只展示部分：

