

python第五次实验相关资料

线性分类器程序设计

分类，即给定一个输入的集合，分类器致力于预测每一个类别的概率

1. Scikit-Learn

Scikit-Learn（也称为sklearn）是针对Python编程语言的免费软件机器学习库，它具有各种分类，回归和聚类算法，包括支持向量机，随机森林，梯度提升，k均值和DBSCAN，并且旨在与Python数值科学库NumPy和SciPy联合使用。

2. mnist数据集

mnist是一个手写体数字的图片数据集。训练集一共包含了 60000 张图像和标签，而测试集一共包含了 10000 张图像和标签。标签取值范围[0-9]分别表示对应的图像中手写体数字是什么

3. 算法主要作用

预测手写体数字图像的类别：给出一张手写体数字图像，要求算法尽可能正确预测该图像对应的数字是多少

实验资料

1. 打开命令行界面，分别安装scikit-learn工具库

```
pip install scikit-learn
如果因网速问题无法下载上述两个工具库，可以使用以下命令
pip install -i https://pypi.tuna.tsinghua.edu.cn/simple scikit-learn
```

2. 本次实验所用的一些方法

```
#获取MNIST手写体数字图像数据集
from sklearn.datasets import fetch_openml
mnist = fetch_openml('mnist_784', as_frame = False)
# mnist['data']对应手写体图像数据 mnist['target']对应标签数据
X,y = mnist['data'],mnist['target']

#使用SGD优化器拟合分类模型
from sklearn.linear_model import SGDClassifier
clf = SGDClassifier(loss = 'log',verbose=2)

#统计模型的训练集准确率与测试集准确率
# x对应训练集或测试集数据，y对应训练集或测试集标签
acc = clf.score(x,y)
```

```
#计算混淆矩阵, 两个参数分别表示标签和算法预测的值
from sklearn.metrics import confusion_matrix
cmat = confusion_matrix(y_train, train_pred)
```