

机器学习

第4次实验：使用神经网络识别手写体数字

实验目的

01

会用Python创建神经网络模型（多层感知机）

02

使用神经网络模型识别手写体数字

03

会评价神经网络模型



实验内容

使用sklearn.neural_network.MLPClassifier类实现手写数字图片识别

- ✓ MLPClassifier是一个有监督学习算法，MPL又称多层感知机，除了输入和输出层，中间可以有多个隐含层。
- ✓ sklearn.externals.joblib.dump方法用于保存模型参数
- ✓ sklearn.externals.joblib.load方法用于从文件中装载模型参数构造一个MLPClassifier对象
- ✓ score方法用于估算正确率

训练数据: digits_training.csv

测试数据: digits_testing.csv

第1列是类别，其他列是特征属性

实验内容

1

载入训练数据、分出特征属性和类别，对特征属性标准化，显示读入数据的行数

```
xTrain = trainData[:,1:NTrain]  
yTrain = trainData[:,0]
```

标准化函数

```
def normalizeData(X):  
    return (X - X.mean())/X.max()
```

实验内容

2

构建多层感知机，训练模型

`psolver='lbfgs', alpha=1e-5, hidden_layer_sizes=(48, 24)`

3

保存训练模型名字为 “mlp_classifier_model1.m”

4

载入测试数据，分出特征属性和类别，对特征属性标准化，显示读入数据的行数

5

使用模型对测试集进行预测，显示预测错误数据的数目和预测准确率

执行结果图例

```
1. 载入训练数据.....
2. 标准化训练数据.....
装载训练数据： 5000 条
3. 构建神经网络.....
4. 训练模型.....
5. 训练完毕，保存模型.....
6. 模型保存完毕，执行测试.....
7. 载入测试数据.....
8. 标准化测试数据.....
装载测试数据： 500 条
9. 使用模型预测.....
预测完毕。错误： 36 条
测试数据正确率： 0.928
10. 评价模型.....
模型内建的正确率估计： 0.928
```