人工智能程序设计实验报告（五）

线性分类器程序设计

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 姓名 | 桑浩翔 | 学号 | E42114027 | 成绩 |  |

|  |
| --- |
| **1.实验目标：**  1）能理解线性分类模型；  2）能利用SciKit-Learn编写、训练和评估线性分类模型；  **2.实验任务与要求：**  1）使用sklearn.datasets.fetch\_openml函数获取MNIST手写体数字图像数据集，该数据集包含了7万张0-9是个数字的手写体图像。  2）把该样本集按照6:1的比例随机分为一个训练集和一个测试集。  3）从训练样本中随机采样32张图像，使用matplotlib子图方式绘制到一幅图像中，每张字符上方输出该字符的label。  4）使用训练集拟合一个Softmax Regression模型，对数字图像进行分类。  5）统计模型的训练集准确率与测试集准确率。  6）输出训练集上的混淆矩阵，并以图像的方式保存为文件，插入到本文档的结果中。  7）随机选择32个测试图像，使用你训练的模型进行预测，然后用matplotlib把每一张图像及其预测结果绘制出来，请采用子图方式绘制，并把绘制结果保存为图像文件，然后插入到本文档的结果中。  8）提交内容包括：  A.本实验报告，请把实验报告文件名中的########替换为你的学号，XXX替换为你的姓名。  B.源代码。源文件命名规则为AIP-5-########.py，其中########替换为你的学号。  例如学号为E02018007的同学，提交的源代码文件命名为：AIP-5-E02018007.PY。 |
| **3.把绘制的图像保存为文件，并插入到下方。（因语法错误不能生成可执行文件的，无运行结果，不粘贴截图）**  **3) 的图（随机采样32张，还没训练）**    **正在训练：**    **5) 的正确率和6) 的混淆矩阵**    **7) 图片、我的预测结果和真实值** |
| **4.把你完成的源文件插入到下方。（插入方法：先把光标置于本段文字的下一段落的开头，然后在菜单栏中选中插入->对象->文件中的文字，在弹出的对话框中选择你编写的源文件）**  from sklearn.datasets import fetch\_openml  from sklearn.model\_selection import train\_test\_split  import matplotlib.pyplot as plt  from sklearn.linear\_model import SGDClassifier    mnist = fetch\_openml('mnist\_784', as\_frame = False)  x, y = mnist['data'], mnist['target']  x\_train, x\_test, y\_train, y\_test = train\_test\_split(mnist['data'], mnist['target'], test\_size = 1 / 7)  x\_draw, \_, y\_draw, \_ = train\_test\_split(x\_train, y\_train, train\_size = 32)  for i in range(32):  plt.subplot(4, 8, i + 1)  plt.imshow([x\_draw[i][j:j+28] for j in range(0, 784, 28)])  plt.title("truth = %d" % int(y\_draw[i]))  plt.axis('off')  plt.show()  clf = SGDClassifier(loss = 'log', verbose = 2)  clf.fit(x\_train, y\_train)  acc1 = clf.score(x\_train, y\_train)  acc2 = clf.score(x\_test, y\_test)  print(acc1, acc2)  from sklearn.metrics import confusion\_matrix  cmat = confusion\_matrix(y\_train, clf.predict(x\_train))  print(cmat)  x\_draw, \_, y\_draw, \_ = train\_test\_split(x\_test, y\_test, train\_size = 32)  y\_drawp = clf.predict(x\_draw)  for i in range(32):  plt.subplot(4, 8, i + 1)  plt.imshow([x\_draw[i][j:j+28] for j in range(0, 784, 28)])  plt.title("truth = %d\nMy predict = %d" % (int(y\_draw[i]), int(y\_drawp[i])))  plt.axis('off')  plt.show() |

参考资料：

Python实验五资料.pdf

关于mnist的相关信息（主要是该向量的长度和规格）  
<https://www.openml.org/search?type=data&sort=runs&id=554&status=active>

将一个一维向量转化为二维矩阵  
<https://stackoverflow.com/questions/14681609/create-a-2d-list-out-of-1d-list>

把子图的坐标轴关掉  
<https://stackoverflow.com/questions/25862026/turn-off-axes-in-subplots>

序列解包时有不想要的元素  
<https://blog.csdn.net/qdPython/article/details/117364209>

Softmax Regression的含义  
<https://www.youtube.com/watch?v=LLux1SW--oM>

关于SGDClassifier的进一步解释

<http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html>