**深 圳 大 学 实 验 报 告**

**课 程 名 称： 数据挖掘导论**

**实验项目名称： 支持向量机分类器实验**

**学 院： 计算机与软件学院**

**专 业： 计算机科学与技术**

**指 导 教 师： 李俊杰**

**报告人： 吴汇杰 学号： 2015150006 班级： 高性能班**

**实 验 时 间： 2017年5月8日**

**实验报告提交时间： 2017年5月22日**

**教务处制**

**一、 实验目标：**

（1）理解支持向量机

（2）掌握利用SMO进行优化的具体方法步骤

（3）掌握利用核函数对数据进行空间转换

（4）了解常用分类器的性能特点

**二、实验环境：**

1. Python编辑环境：Jupyter Notebook （或者其他Python环境均可）

**三、实验内容与步骤**

1. 使用SMO算法并比较不同δ值下的算法性能，最后填写如下列表

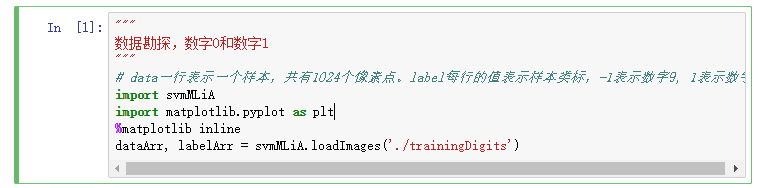
|  |  |  |  |
| --- | --- | --- | --- |
| 内核，设置 | 训练准确率 | 测试错误率 | 支持向量数 |
| RBF, 0.1 | 100% | 52.1505% | 402 |
| RBF, 5 | 100% | 3.2258% | 402 |
| RBF, 10 | 100% | 1.0753% | 120 |
| RBF, 50 | 98.7562% | 2.6882% | 40 |
| RBF, 100 | 99.5025% | 1.6129% | 36 |
| Linear | 99.92537% | 1.0753% | 33 |

2）使用sklearn进行再一次实验

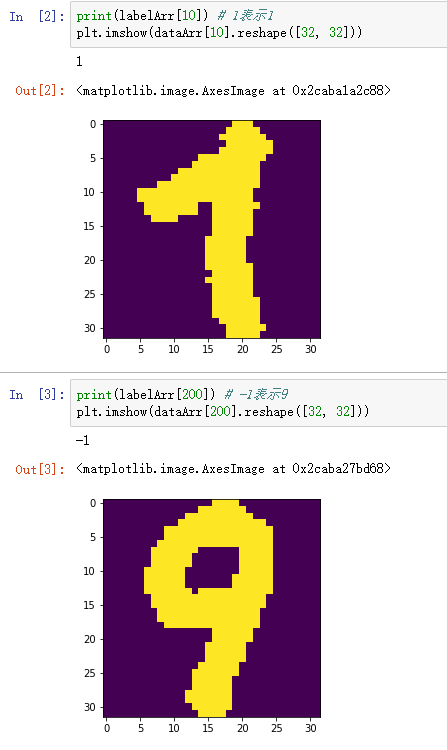
**四、实验结果**

①.使用文件中写好的SMO进行我们的实验

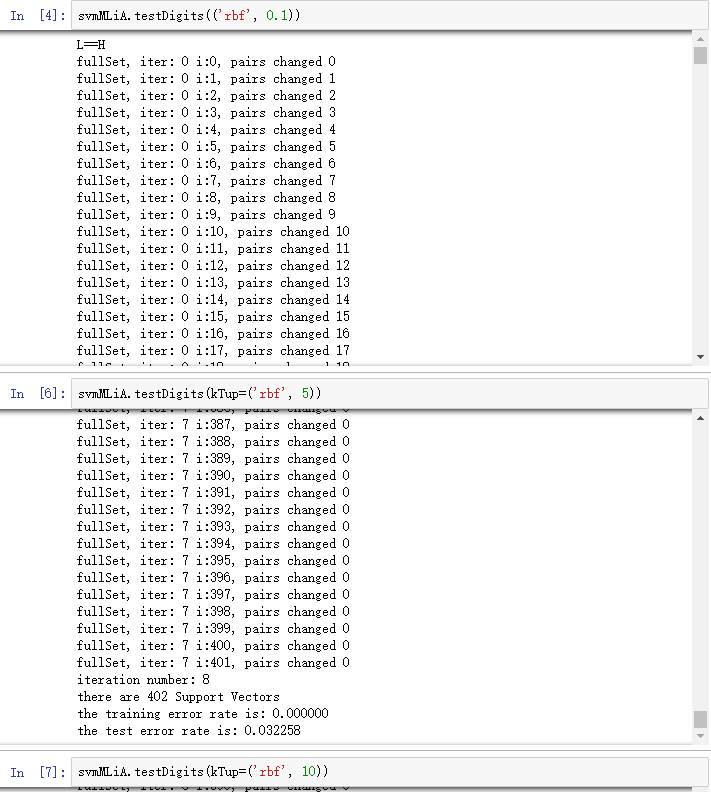
⒈先进行需要的包的导入，以及数据集的导入，代码如下图所示。



⒉分别显示数字1和数字9的一个例子，如下图所示。



⒊直接调用测试函数得到结果，如下图所示。



②．使用sklearn包进行实验

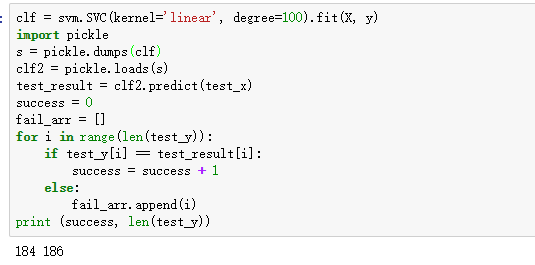
⒈导入需要使用的包以及图片内联，如下图所示。



⒉分别导入训练数据和测试数据，如下图所示。



⒊使用linear的kernel对我们的值进行训练，然后测试得到错误的仅有两个，如下图所示。



⒋因为我们的错误率比较低，所以调出我们识别错误的图片进行查看，如下图所示。



可以见到，上面的两幅图，就算让人做识别也很难识别的出来，毕竟没有一个统一的标准。

而机器可以做到如此程度，也多亏了sklearn中写的svm算法，这样才能有如此识别成功率。

当然，内置方法里面实现的线性得出的错误率其实是和使用sklearn是一样的，充分证明了，这玩意只有性能比较差，其他算法本身还是没有精简的。

**五、实验总结与体会**

**总结：**

**支持向量机是一种很好的分类方法，但他所需的运算时间还是比较长，需要进行优化的点很多，而且需要掌握很多数学相关知识才能深入理解它。**

|  |
| --- |
| **指导教师批阅意见：**  **成绩评定：**  指导教师签字：    年 月 日 |
| 备注： |