**《机器学习》课程实验报告**

|  |  |
| --- | --- |
| **实验名称:** | Classification with SVM model |
| **姓 名:** | 李若昊 |
| **学 号：** | 5120180269 |
| **教 师:** | 赵旭剑 |
| **完成时间:** | 2020-5-15 |

|  |
| --- |
| **一、实验目的和内容**  **本次试验很简单，利用libsvm工具，通过训练集中的数据训练一个模型，并且在测试集上验证准确度。** |
| **二、实验过程**  **①因为这次试验选用libsvm工具，所以要先把老师给的测试集(.data)变成libsvm能读得懂的格式，我用Excel导入train.data中的数据并且利用宏转换成相应格式：**  **然后保存成UTF-8编码的文本文件待用，且对提供的测试数据集文件也做相同操作。**  **图中每组数据冒号的前后分别代表特征的编号及数值，开头的那一列表示这个点属于第几个分类。**  **②因为是实验，所以不用easy.py全自动操作。**  **SVM涉及距离计算，所以接下来应该先调整特征值的范围，进行缩放，以免有一项数据太大，在算距离的时候主导了结果。但是我看了一下数据已经在-1~1区间了，那就跳过**  **③接下来分类，要分得好，参数就要给的好。应该用grid.py蛮力求解c和g，但是因为一些问题没有做成。我直接用默认性能较好的RBF核，默认的三折交叉验证训练模型，并用测试集测试模型：**  **没有用grid.py暴力求解参数c和g是因为这个问题尝试解决好久一直没能被解决**  **理论上要用这个方式求得最优的c和g去训练模型，分类效果会更优。**  **并且分类结果在svm-toy中应该会有直观的显示，但是…我没找出办法来运行它，但是结果返回的准确率看起来还不错：** |
| **三、实验结果** |
| **四、实验总结**  **这次试验了解了libsvm的用法，并通过实践对SVM有了更直观的了解** |