说明文档及相关截图

### 一、概述

采用python进行代码编写，在代码中我借助了LibSVM工具包，实现了（支持向量机）SVM对数据集的二分类。

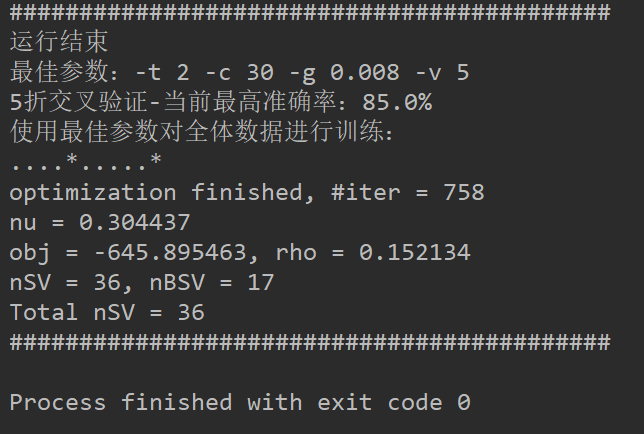
其中，我尝试调节参数c以及使用不同的核函数（线性核函数和径向基核函数，也尝试了多项式核函数，但效果并不出色），并调节了核函数参数g。

同时，我也利用了libsvm的k折交叉验证，进行了5折交叉验证，手动调参粗略得到85%的准确率得分。

### 二、运行方法

代码用到的工具包有libsvm.svmutil（用于使用其SVM相关方法）、pandas（用于读取数据集）、numpy（用于矩阵运算）。Python含有这两个包即可正常运行。

### 三、运行结果



**图 1-SVM\_libsvm.py运行结果图**

**说明：**对全体数据的训练得出的一系列输出中，主要参数解释如下

#ite：为迭代次数；

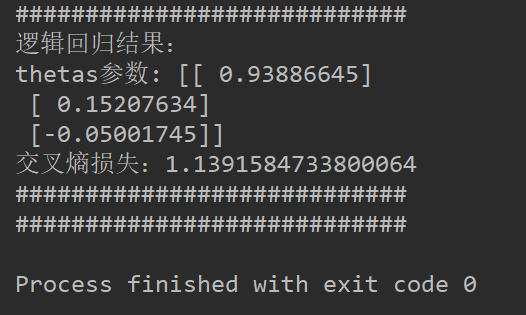
obj：二次规划的最小值；

rho：决策函数常数项；

nSV：支持向量数（所有满足一定约束条件的支持向量）；

nBSV：边界上的支持向量数；

Total nSV：总的支持向量数。



**图 2-****logistic.py运行结果图**

### 四、比较

**SVM和****Logistic回归的比较**

二者都能够实现对数据的二分类，如果不考虑核函数，那么这两者都属于线性分类器，且都是一种预测模型，对数据进行训练后再对测试数据给出判断。

不同的是，Logistic回归模型考虑的是全局情况，每个样本点都会对其有影响；而SVM模型则更注重局部——边界线附近的样本点，也就是更加关注支持向量。Logistic回归的目的大致为寻找一条边界将不同的类别分隔开来，而SVM则是保证正确分类的情况下让样本到边界的距离尽可能远。

此外，核函数也是SVM的一个重要特点，使用核函数可以让SVM分类器将低维不可分数据映射到高维可分空间中进行处理，从而可以轻松处理那些线性不可分的数据。