**本 科 实 验 报 告**

实验名称： 支持向量机实现

学 员： 孙源辰 学 号： 201802001067

培养类型： 无军籍本科学员 年 级： 大三

专 业：人工智能与大数据 所属学院： 计算机学院

指导教员： 刘新旺 职 称： 副教授

实 验 室： 305-205 实验日期： 2021春季学期

国防科学技术大学训练部制

《本科实验报告》填写说明

1．学员完成人才培养方案和课程标准要所要求的每个实验后，均须提交实验报告。

2．实验报告封面必须打印，报告内容可以手写或打印。

3．实验报告内容编排及打印应符合以下要求：

（1）采用A4（21cm×29.7cm）白色复印纸，单面黑字打印。上下左右各侧的页边距均为3cm；缺省文档网格：字号为小4号，中文为宋体，英文和阿拉伯数字为Times New Roman，每页30行，每行36字；页脚距边界为2.5cm，页码置于页脚、居中，采用小5号阿拉伯数字从1开始连续编排，封面不编页码。

（2）报告正文最多可设四级标题，字体均为黑体，第一级标题字号为4号，其余各级标题为小4号；标题序号第一级用“一、”、“二、”……，第二级用“（一）”、“（二）” ……，第三级用“1.”、“2.” ……，第四级用“（1）”、“（2）” ……，分别按序连续编排。

（3）正文插图、表格中的文字字号均为5号。

一、实验目的和要求

通过python实现，了解和掌握支持向量机的有关知识。

二、实验内容和原理

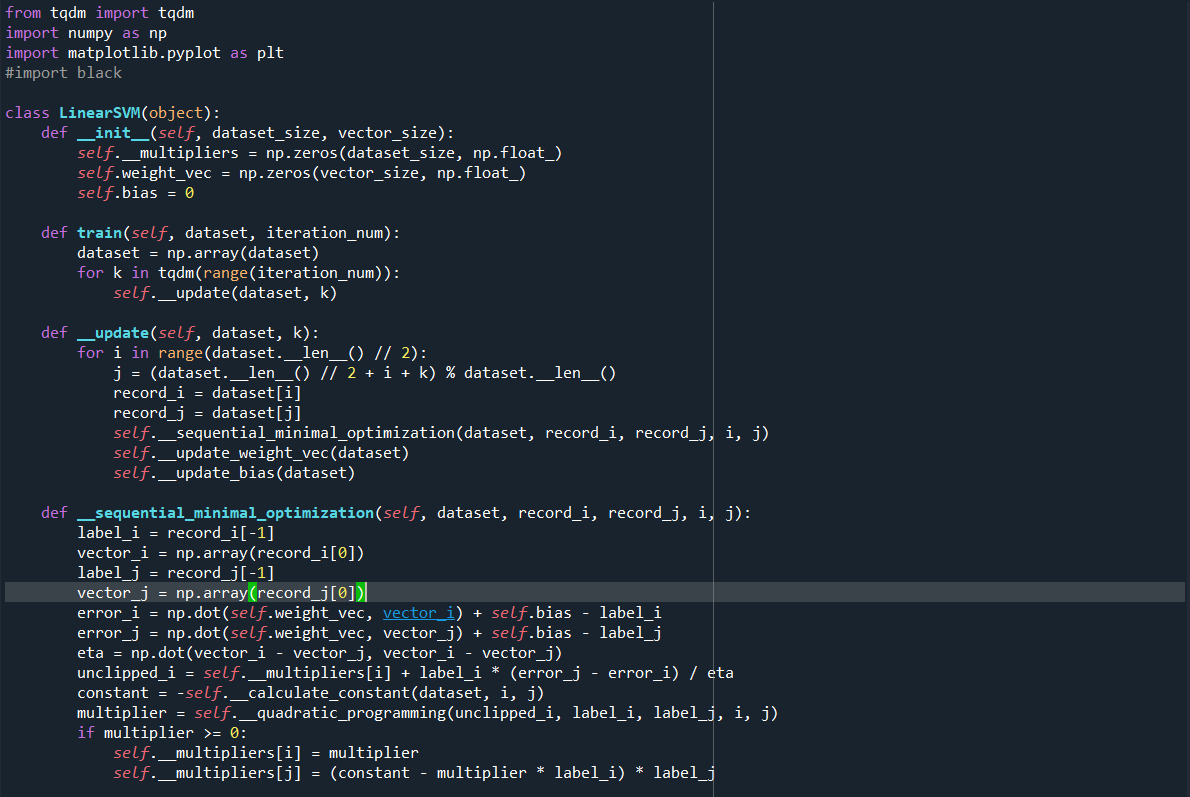
用python实现线性支持向量机并带有松弛变量；实现带有核函数的支持向量机。支持向量机是分类预测的常用手段之一，通过在超空间用距离向量训练出的超平面对数据进行划分。

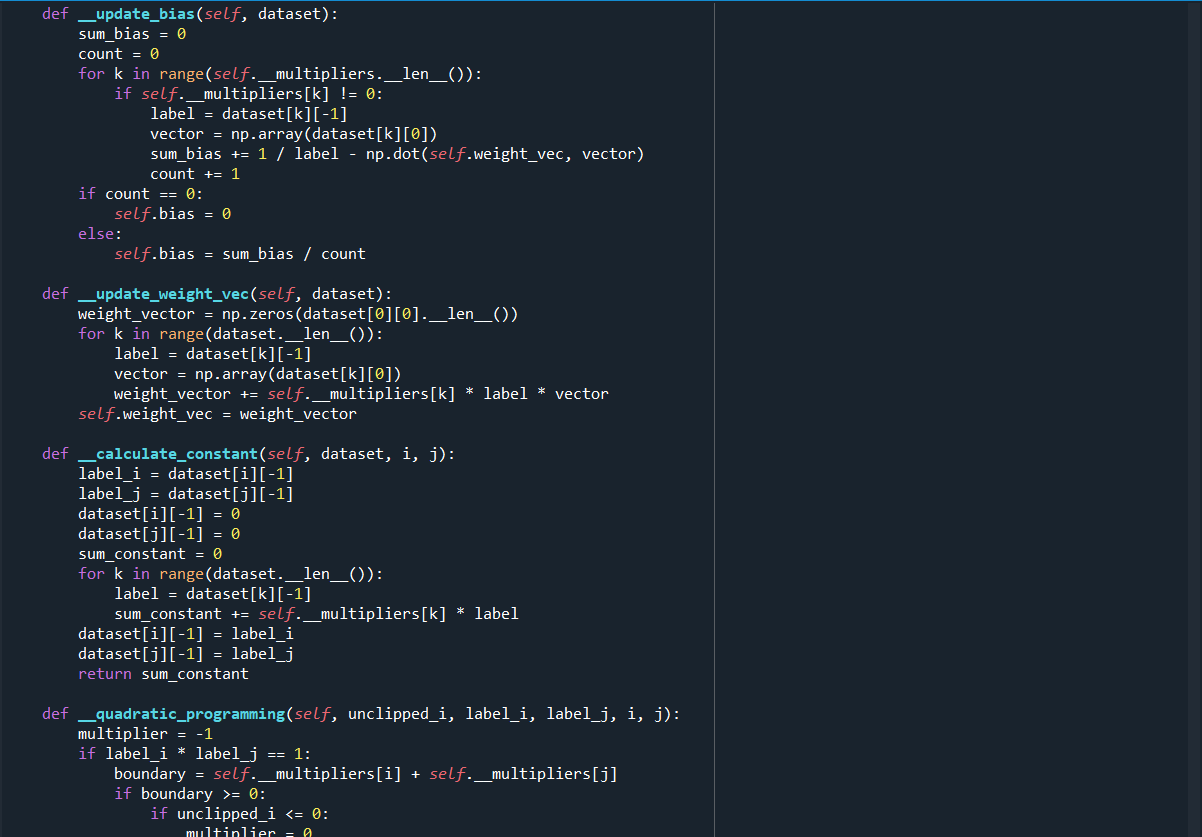
三、操作方法与实验步骤

(一).编程思路

1.线性支持向量机

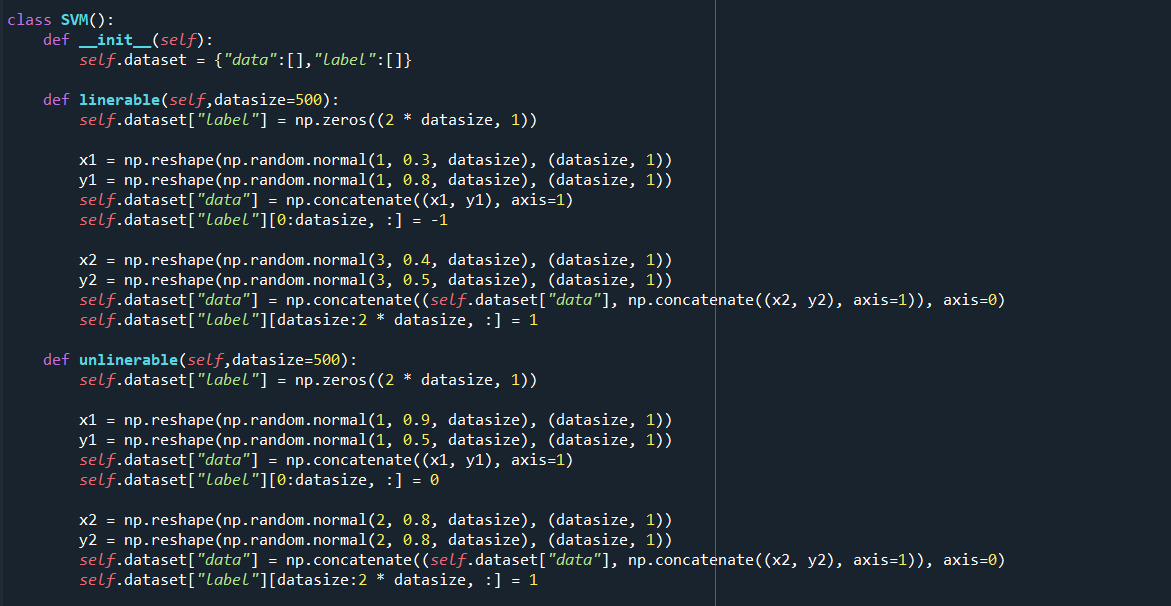
(1)向量机类定义





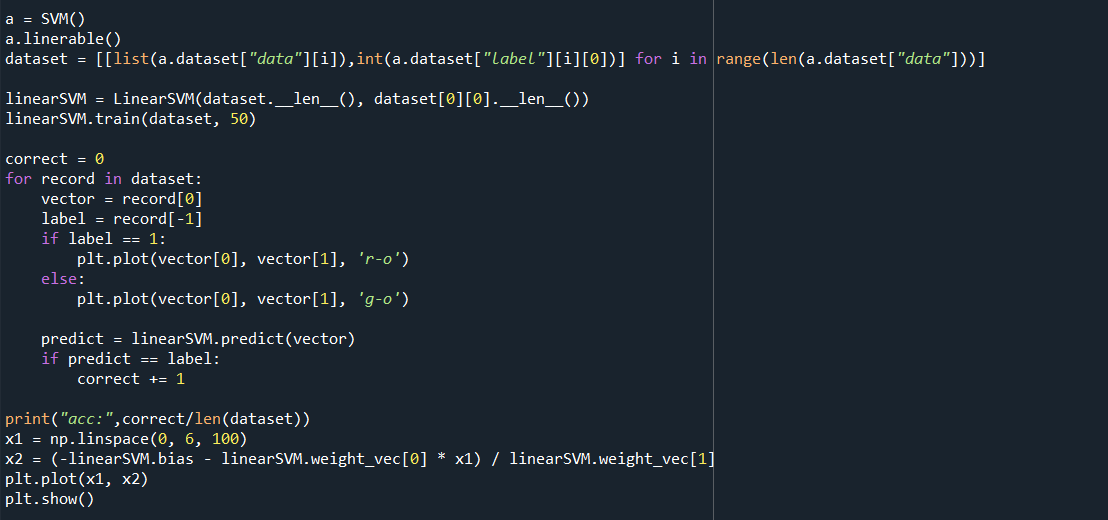


(2)数据集初始化



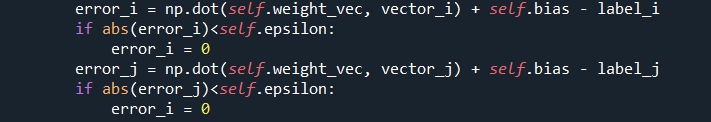
两个数据集，一个线性一个非线性，为了之后埋下伏笔

(3)测试运行部分



进行预测输出

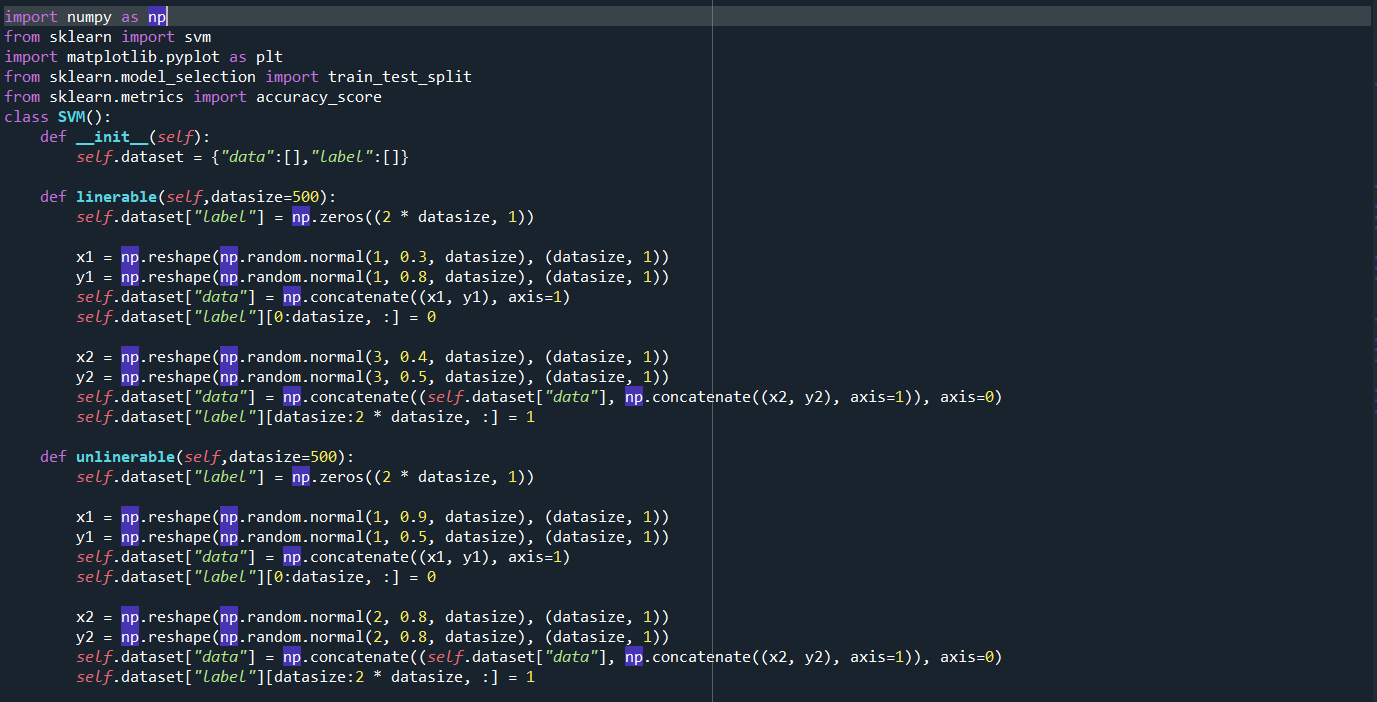
2.引入宽松变量的线性支持向量机



如图，在计算错误时若错误落在±ε里则不算错误

3.带有核函数的支持向量机

(1)数据集定义部分



和上面用的是同样的随机数生成方法

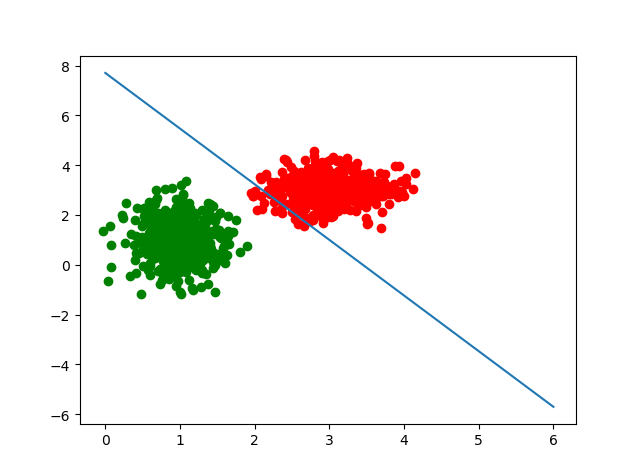
(2)训练模拟部分



通过改变核函数、参数来模拟多种情况并可视化，其中一个数据集画在一张图上，共分为正确的两类数据和划分错误的数据。

四、实验结果及分析

(一).线性支持向量机



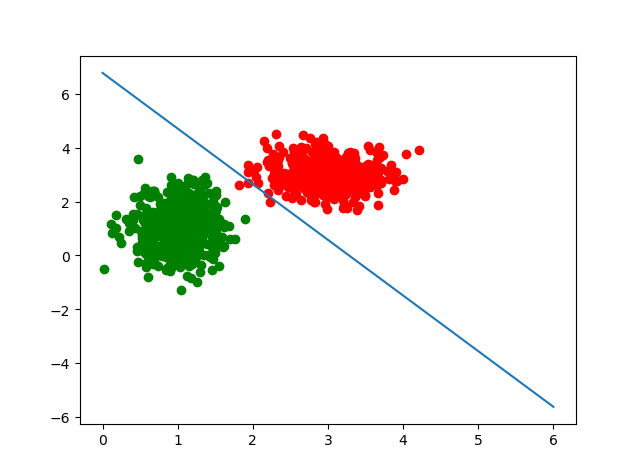
预测结果：

lineresult

这里引入了第三方进度条库tqdm，因为运行时间有点长所以用此显示运行进度。

可以看到迭代50次，预测准确率98.5%，斜率基本正确而截距略有问题。

(二).引入宽松变量的线性支持向量机



预测结果：

softlineresult

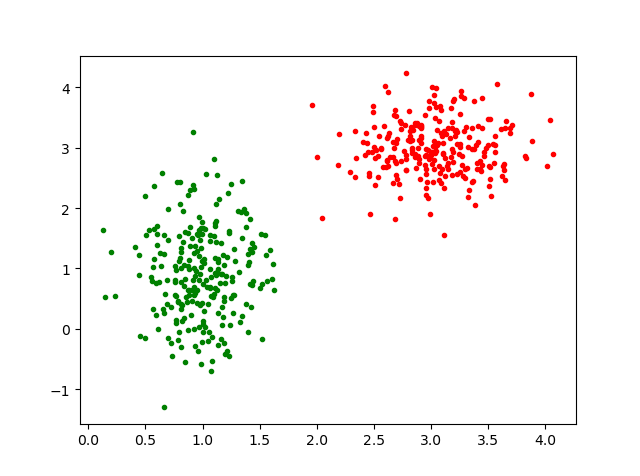
可以看到准确率有一定的提升。因为在一定程度上避免了过拟合。

(三).带有核函数的支持向量机

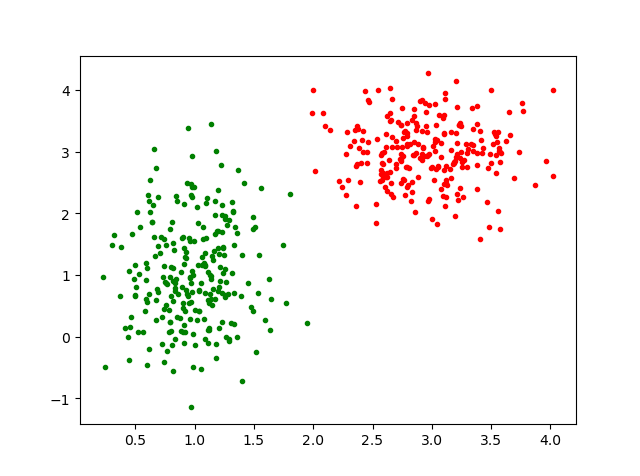
1.线性核函数（C=1，γ=100）

(1)线性可分数据集：

训练集分类结果



测试集分类结果

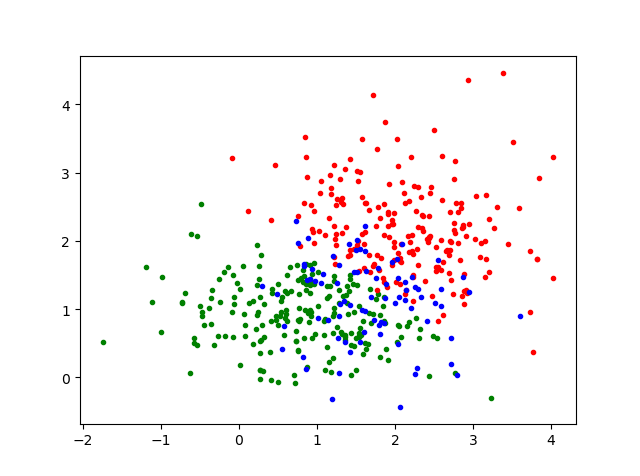


准确率：

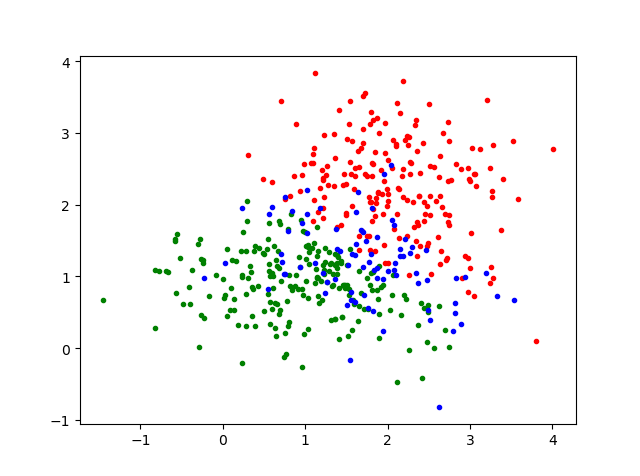
linear1-100-result

(2)线性不可分数据集：

训练集分类结果：



测试集分类结果：



准确率：

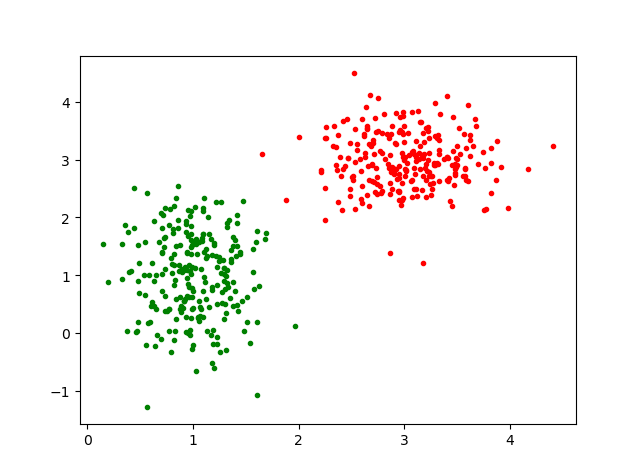
linear1-100-u-result

总结：线性核函数在线性可分数据集上的表现优异，很快便达到1.0的准确率，而对于非线性可分数据集，误差很高并集中在两类交界处。

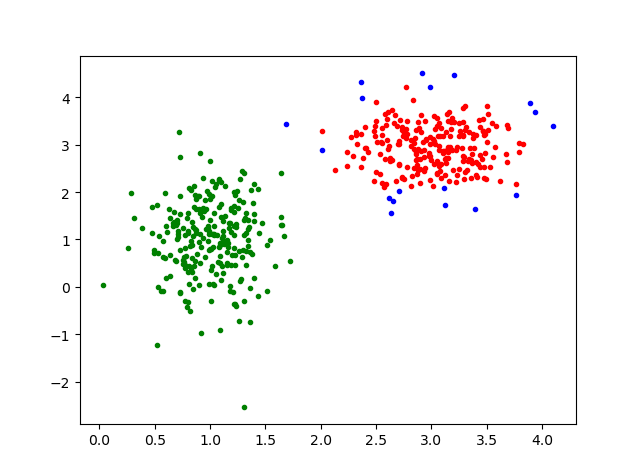
2.高斯核函数

(1)线性可分数据集（C=1，γ=100）：

训练集分类结果：



测试集分类结果：



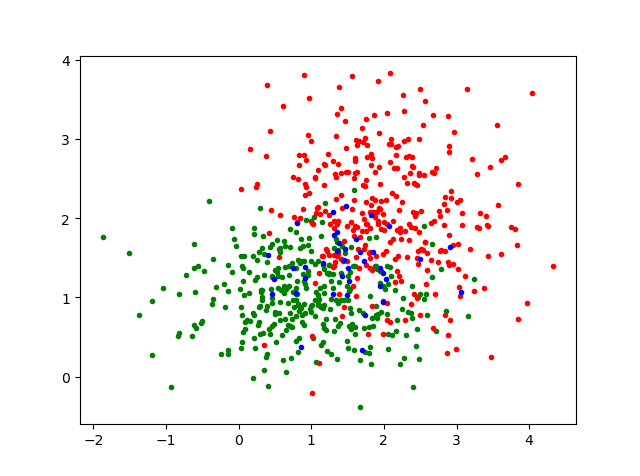
准确率：

RBF1-100-l-result

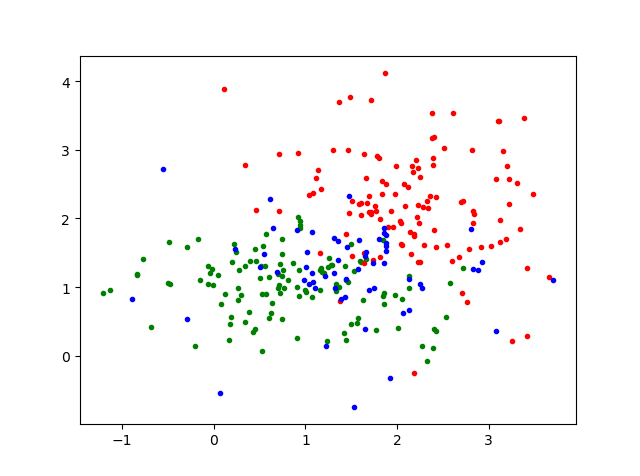
总结：对于线性可分数据集，虽然训练误差为0，但是仍在测试集的边缘数据上存在预测误差。

(2)线性不可分数据集（C=1，γ=100）：

训练集分类结果：



测试集分类结果：

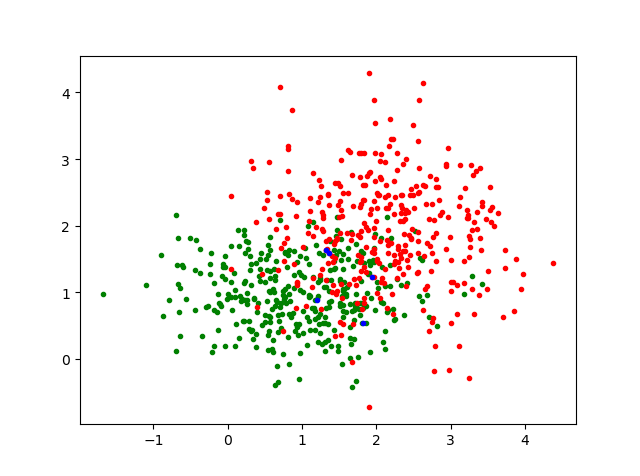


准确率：

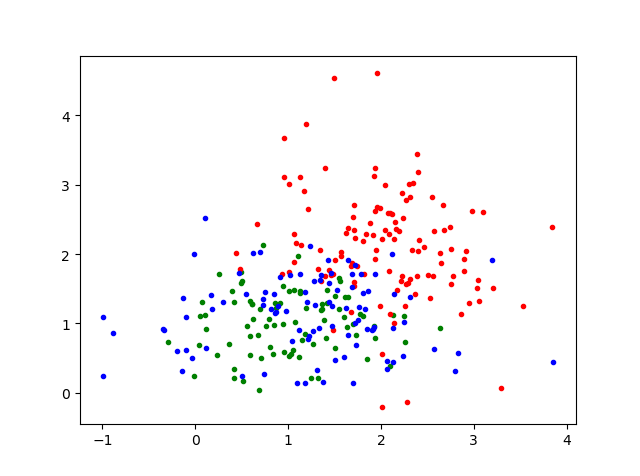
RBF1-100-result

(3)线性不可分数据集（C=1，γ=1000）：

训练集分类结果：



测试集分类结果;

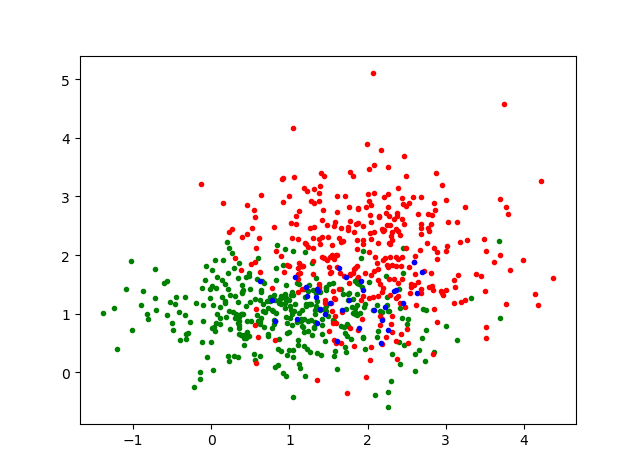


准确率：

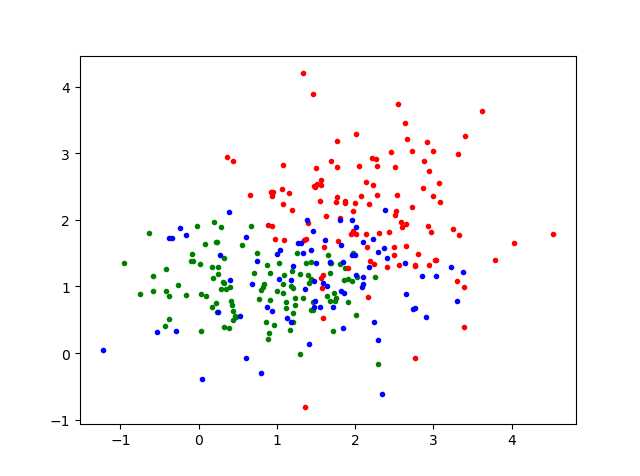
RBF1-1000-result

(4)线性不可分数据集（C=2，γ=100）：

训练集分类结果：



测试集分类结果：

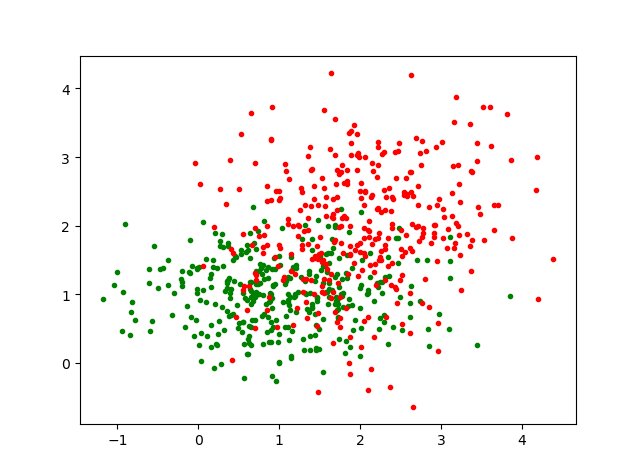


准确率：

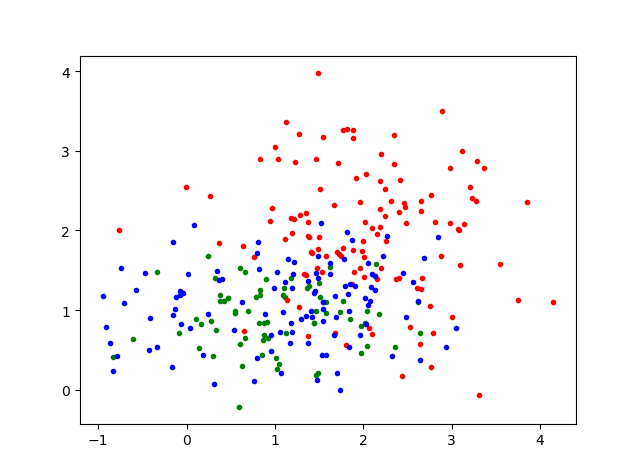
RBF2-100-result

(5)线性不可分数据集（C=2，γ=1000）：

训练集分类结果：



测试集分类结果：



准确率:

RBF2-1000-result

总结：

1. 在C相同的情况下，γ越大训练误差越小，但极容易过拟合使测试误差增大
2. 在γ相同的情况下，C越大训练误差越小，但极容易过拟合使测试误差增大
3. 对于线性可分数据集，线性核函准确率明显优于高斯核函数，同时过拟合风险更小
4. 对与线性不可分数据集，高斯核函数准确率明显高于线性核函数，但是有较高的过拟合风险