**6.2 试使用LIBSVM，在西瓜数据集3.0α上分别用线性核和高斯核训练一个SVM，并比较其支持向量的差别。**

**①问题描述：**

本题旨在使用由国立台湾大学学者开发的LIBSVM框使用数据集进行训练，分别将SVM使用线性核和高斯核进行训练，比较其最后计算出的支持向量的差别，并比较基于训练集进行划分的图像的区别。

**②实现过程：**

a.使用FormatDatalibsvm.xls文件将原始训练集和测试集转变为LIBSVM格式的数据集，其中任选11个数据作为训练集，6个数据作为测试集。

b.使用svm\_read\_problem()方法分别读取LIBSVM格式的训练集和测试集。

c.使用svm\_train()方法训练SVM模型，在该方法中可确定所使用的内核的类型，分别使用线性核和高斯核，其中cost均设定为100。

d.使用svm\_save\_model()方法保存已训练好的SVM模型。

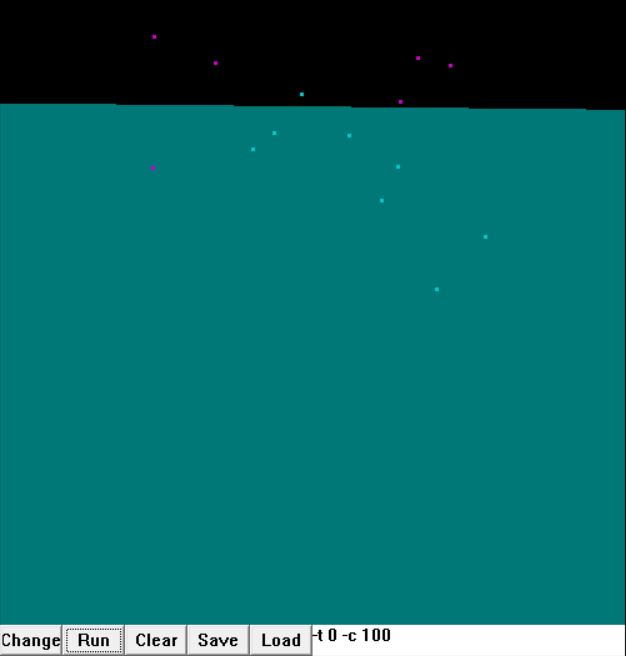
e.使用svm\_predict()方法引入测试集根据已训练好的模型进行预测，并计算出预测的准确率。

f.在保存完成的SVM模型中可提取出支持向量。

g.使用svm-toy.exe导入训练集数据生成模型划分结果的图像。

**③实现结果：**

线性核训练集划分结果图：



线性核支持向量：

100 1:0.556 2:0.215

72.982448712921567 1:0.403 2:0.237

100 1:0.481 2:0.149

100 1:0.437 2:0.211

-100 1:0.243 2:0.267

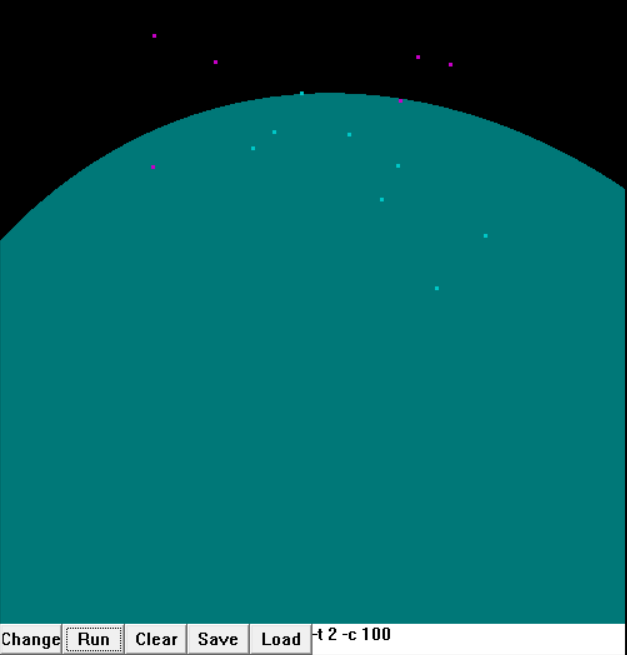
-94.70098479301096 1:0.343 2:0.099

-100 1:0.639 2:0.161

-78.281463919910607 1:0.719 2:0.103

线性核准确率：66.6667％

高斯核训练集划分结果图：



高斯核支持向量：

100 1:0.556 2:0.215

65.396727296159042 1:0.403 2:0.237

100 1:0.481 2:0.149

100 1:0.437 2:0.211

-100 1:0.243 2:0.267

-91.732482393981925 1:0.343 2:0.099

-100 1:0.639 2:0.161

-73.664244902177145 1:0.719 2:0.103

高斯核准确率：66.6667％

**④结果分析：**

在已有数据集基础上使用线性核和使用高斯核训练出的SVM模型中，两者得到的支持向量的总数都为8，但支持向量的具体内容不一致，差别仅出现在支持向量的权值上，而支持向量对应的属性值则完全相同。使用线性核绘制出的图像中，正反例区域的分界线为一条直线，使用高斯核绘制出的图像中，正反例区域的分界线为一条曲线。进一步地可以通过调整高斯核的gamma值使高斯核SVM训练模型更加贴合本数据集。