**

**HUNAN UNIVERSITY**

《DM&BID》

Report X

|  |  |
| --- | --- |
| **报告名称：** | 支持向量机SVM |
| **学生姓名：** | 杨超然 |
| **学生学号：** | 202106060220 |
| **专业班级：** | 电商2102班 |
| **学 院：** | 工商管理学院 |
| **指导老师：** | 江资斌 |
| **日 期：** | 2023.4.3 |

目录

[一、SVM基本思想 3](#_Toc1918940387)

[（一）SVM分类思路 3](#_Toc721673663)

[（二）最大超平面及其优点 3](#_Toc181615785)

[二、 SVM分类的三种情况 4](#_Toc842236108)

[（一） 完全线性可分 4](#_Toc1397549779)

[（二） 广义线性可分：样本无法完全线性可分 5](#_Toc1590488414)

[（三） 线性不可分：无法被超平面线性分开 6](#_Toc1609819889)

[三、 基于python的SVM代码实现 6](#_Toc96405870)

[（一）线性可分SVM 6](#_Toc1090787252)

[1.1导入相关库 6](#_Toc1940933572)

[1.2生成模拟数据 7](#_Toc993946674)

[1.3线性可分支持向量机 7](#_Toc2133943552)

[1.4结果可视化 8](#_Toc64889917)

[（二） 广义线性可分SVM 9](#_Toc1830625990)

[2.1导入相关库 9](#_Toc332803361)

[2.2生成模拟数据 9](#_Toc1378671539)

[2.3广义线性可分支持向量机 9](#_Toc2131488490)

[2.4结果可视化 9](#_Toc1752335823)

[（三） 非线性可分SVM 10](#_Toc917442203)

[3.1导入相关库 10](#_Toc518520361)

[3.2生成模拟数据 10](#_Toc283067801)

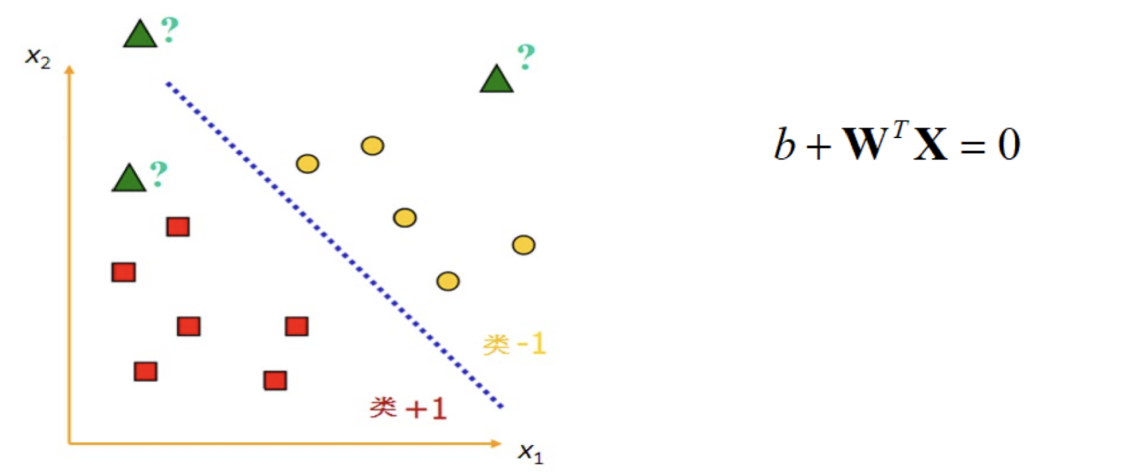
[3.3非线性SVM 11](#_Toc844253302)

[3.4结果可视化 11](#_Toc940790985)

# 一、SVM基本思想

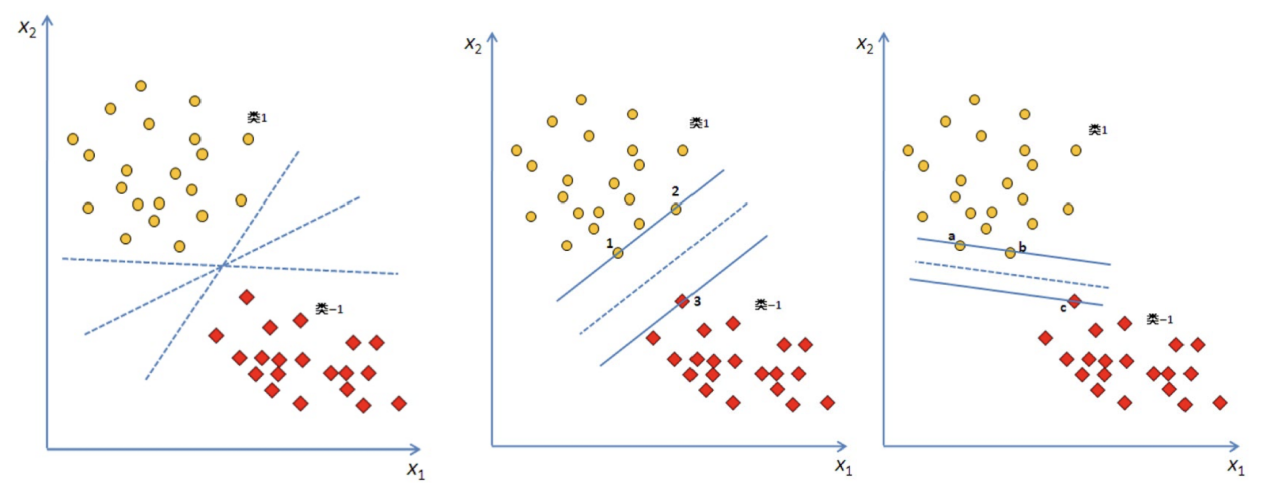
## （一）SVM分类思路

以训练样本为研究对象，在p维特征空间中**找到一个超平面，将两类数据进行有效的划分**



## （二）最大超平面及其优点

一个能将训练样本分开的平面有很多，其中**最大边界超平面**是支持向量分类的超平面，其简单的解释为**距两个类别（-1类和1类）的边界观测点最远的超平面**

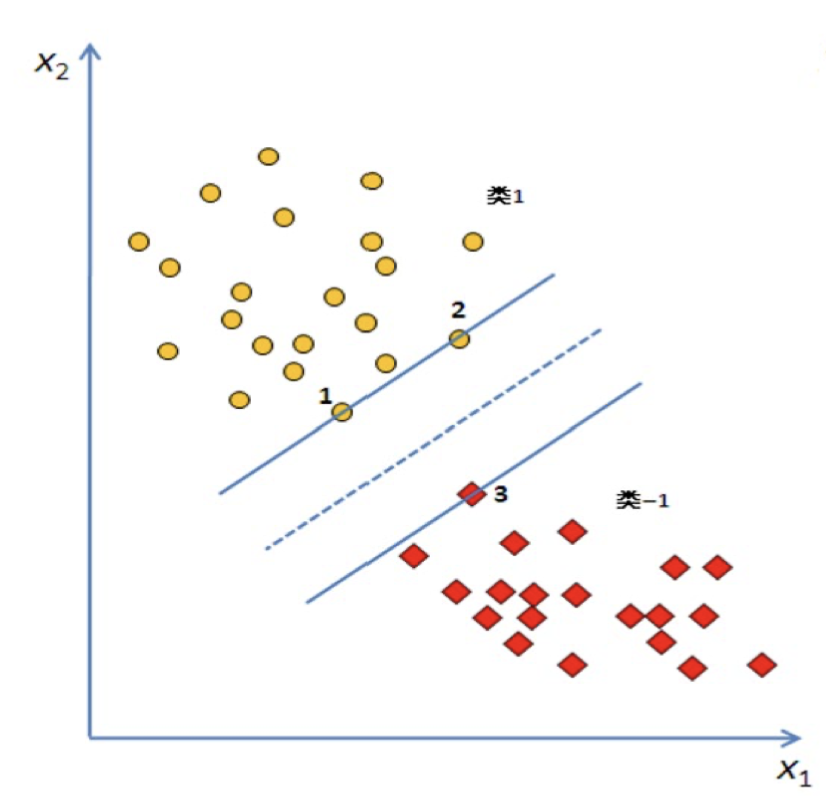


最大边界超平面的优点在于：（1）其不仅是距训练样本集边界观测点最远的，而且是距离测试样本集边界观测点最远的，这样正确预测的把握程度较高。（2）仅取决于两类别的边界观测点。

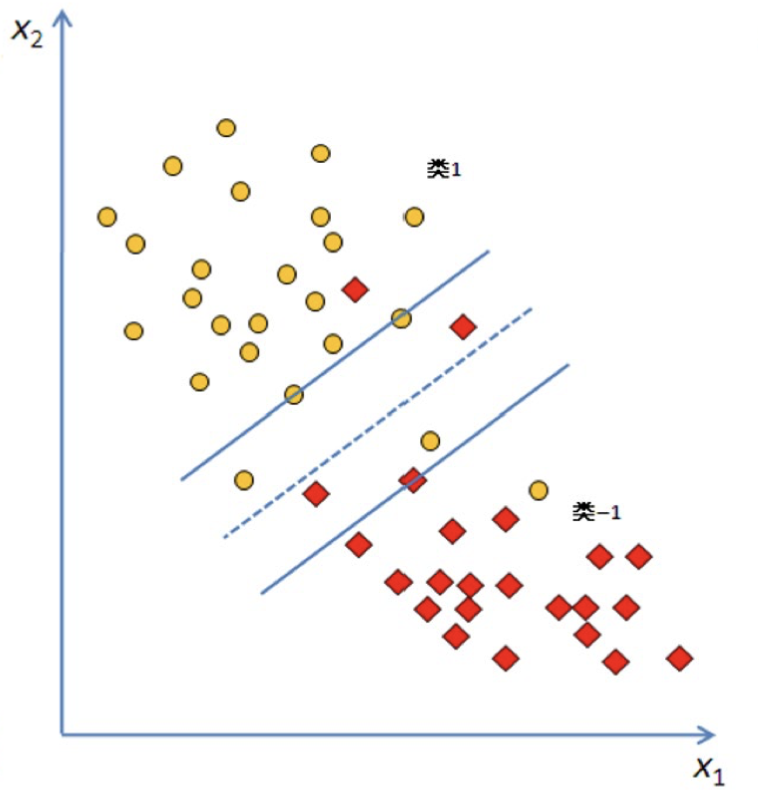
如上图的1、2、3，最大边界超平面对这些观测点的移动程度极为敏感，且仅依赖于这些极少的观测，这些观测被称为**支持向量**。

# SVM分类的三种情况

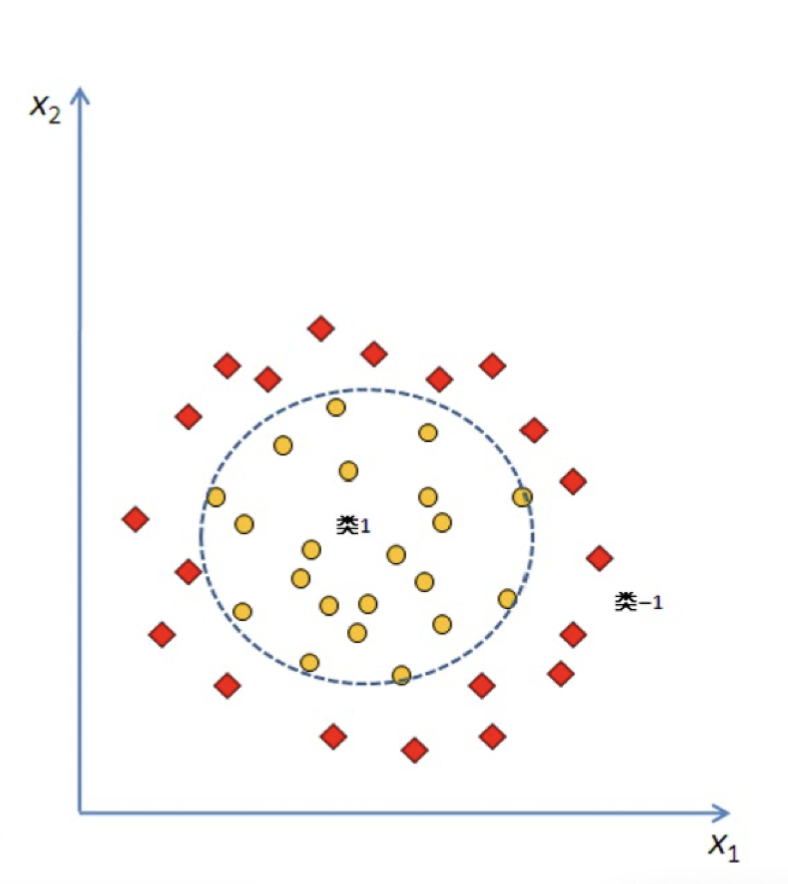
## 完全线性可分



## 广义线性可分：样本无法完全线性可分



## 线性不可分：无法被超平面线性分开



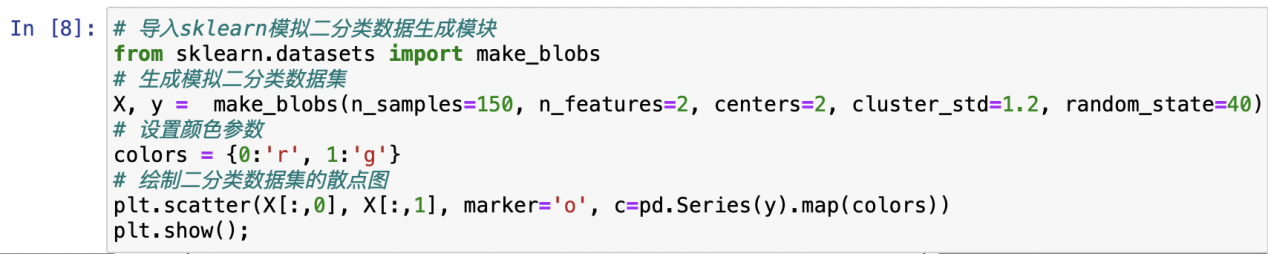
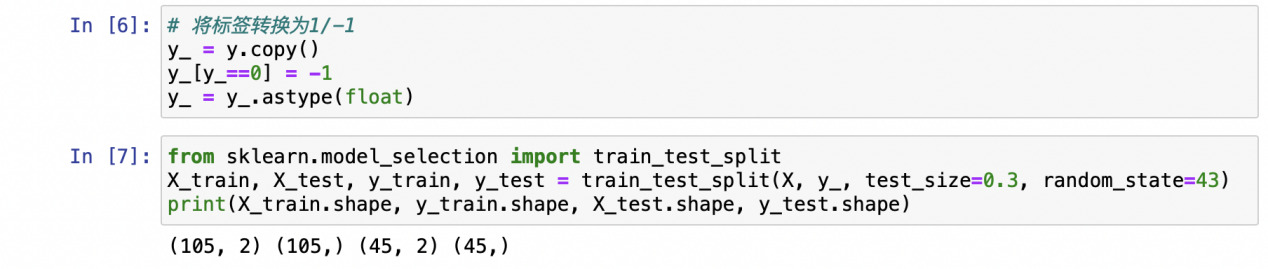
# 基于python的SVM代码实现

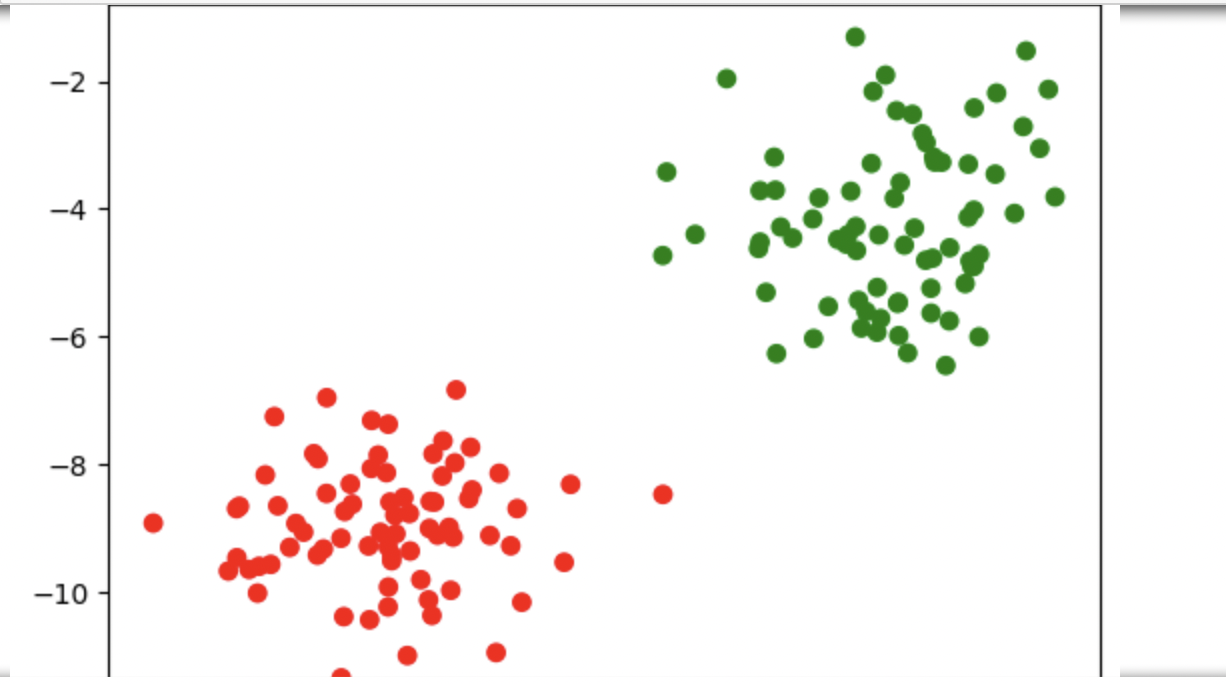
## （一）线性可分SVM

1.1导入相关库

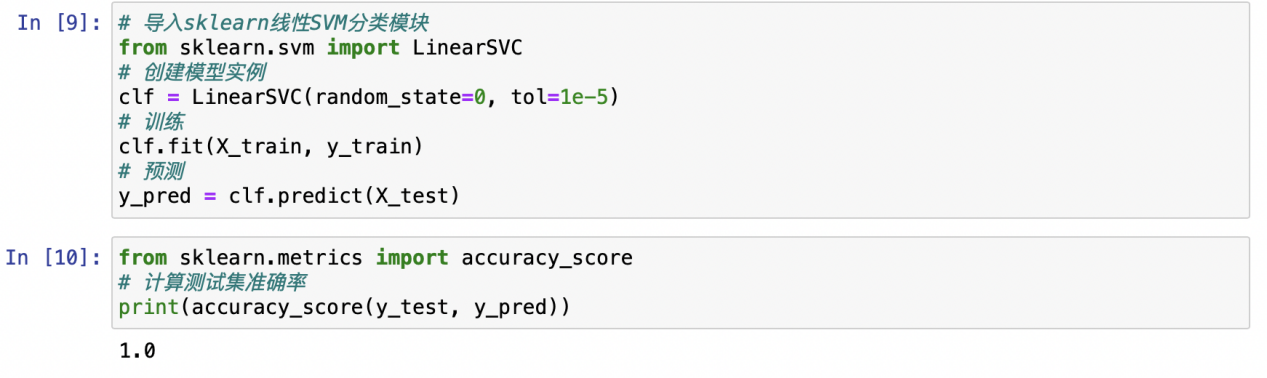


1.2生成模拟数据



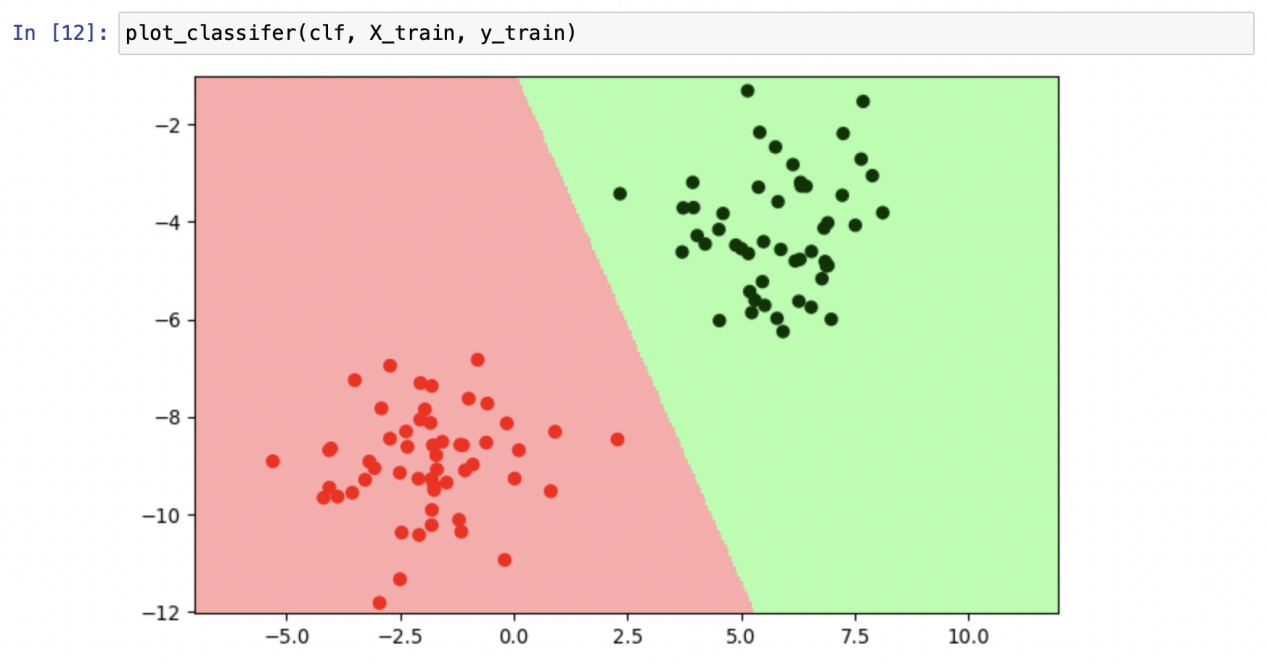


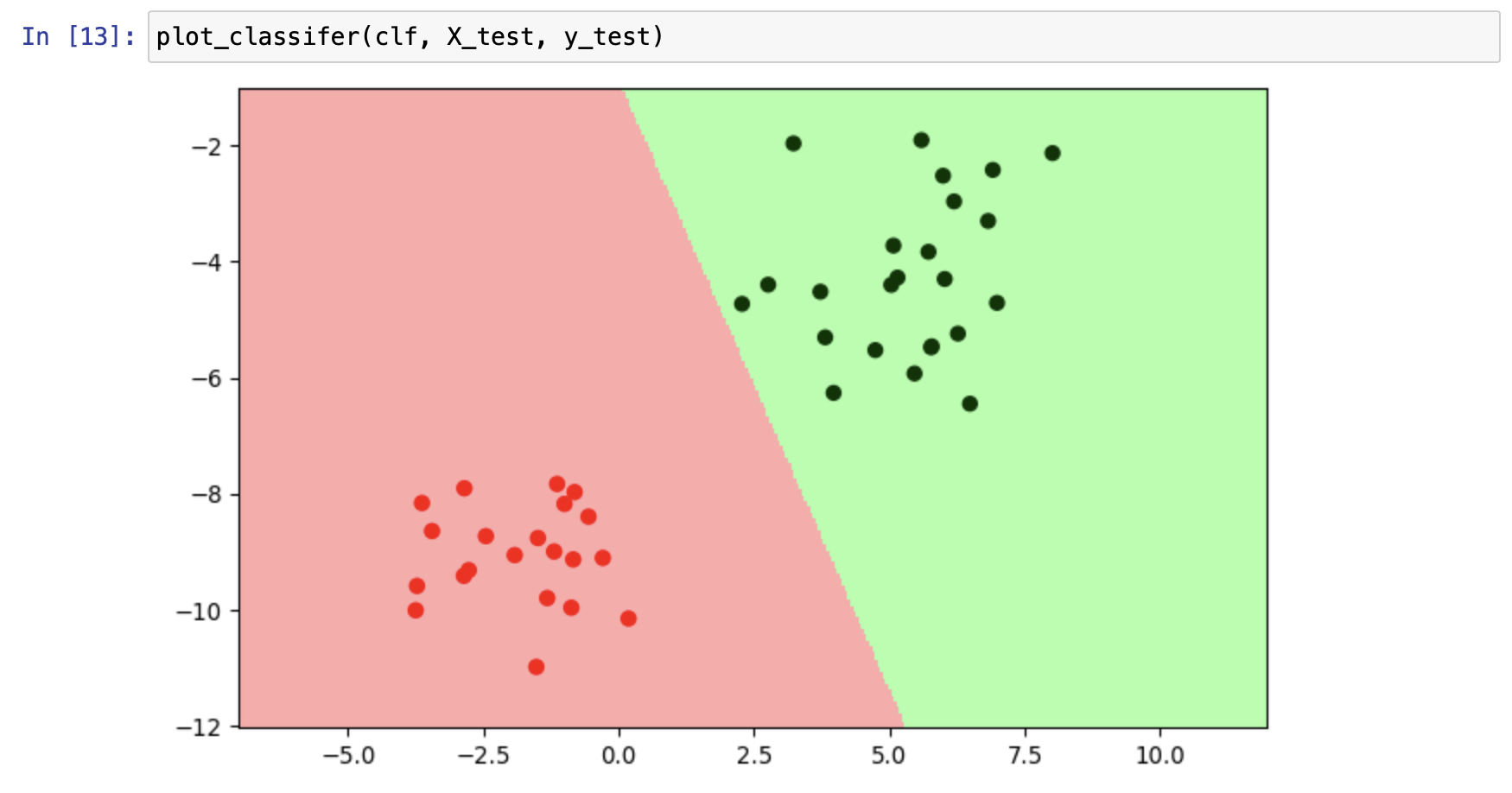
1.3线性可分支持向量机





1.4结果可视化



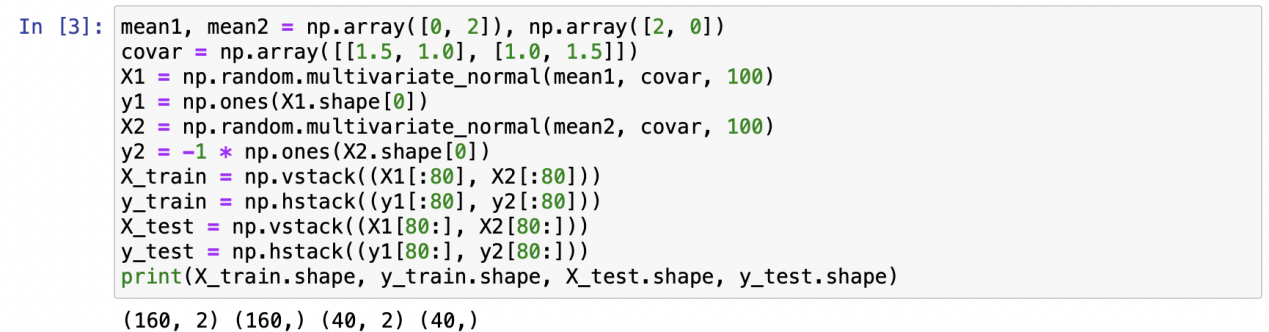


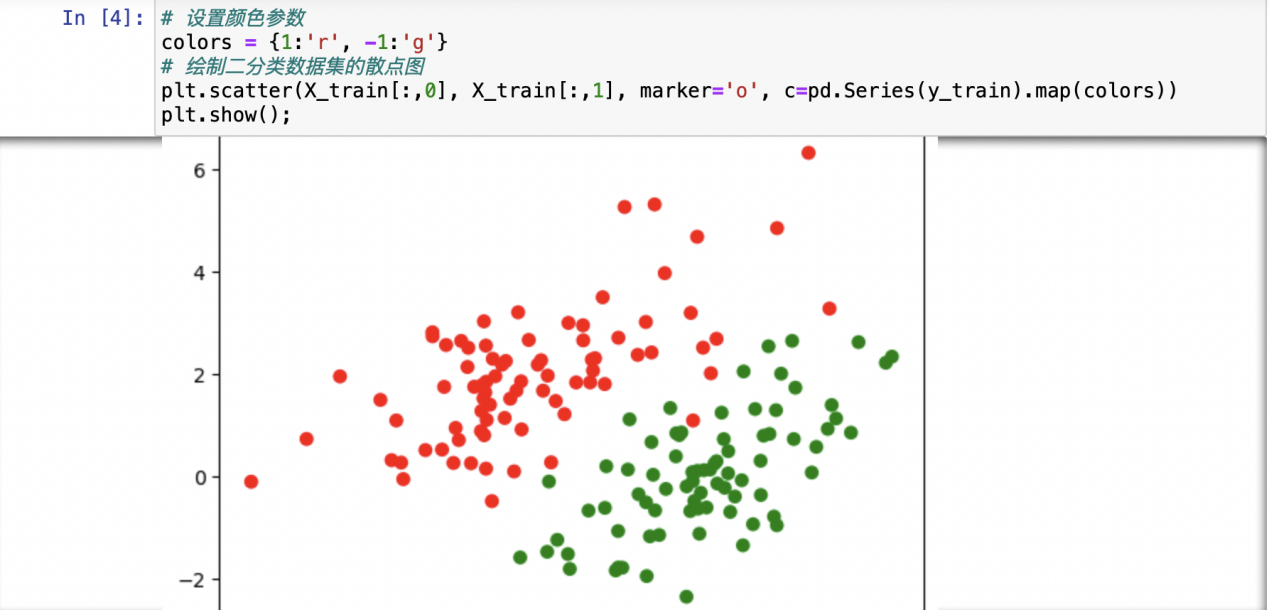
## 广义线性可分SVM

2.1导入相关库

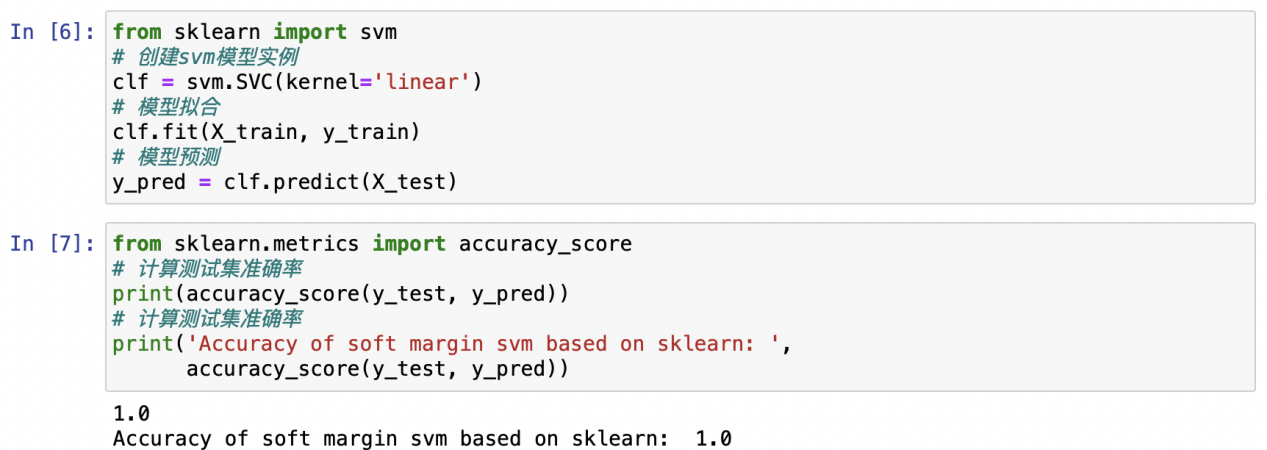
同（一）中操作

2.2生成模拟数据

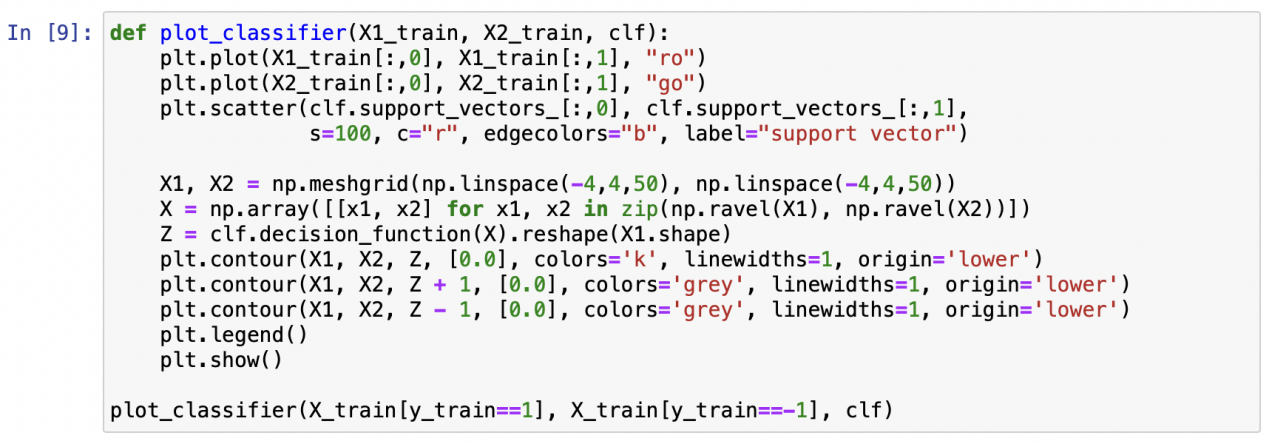


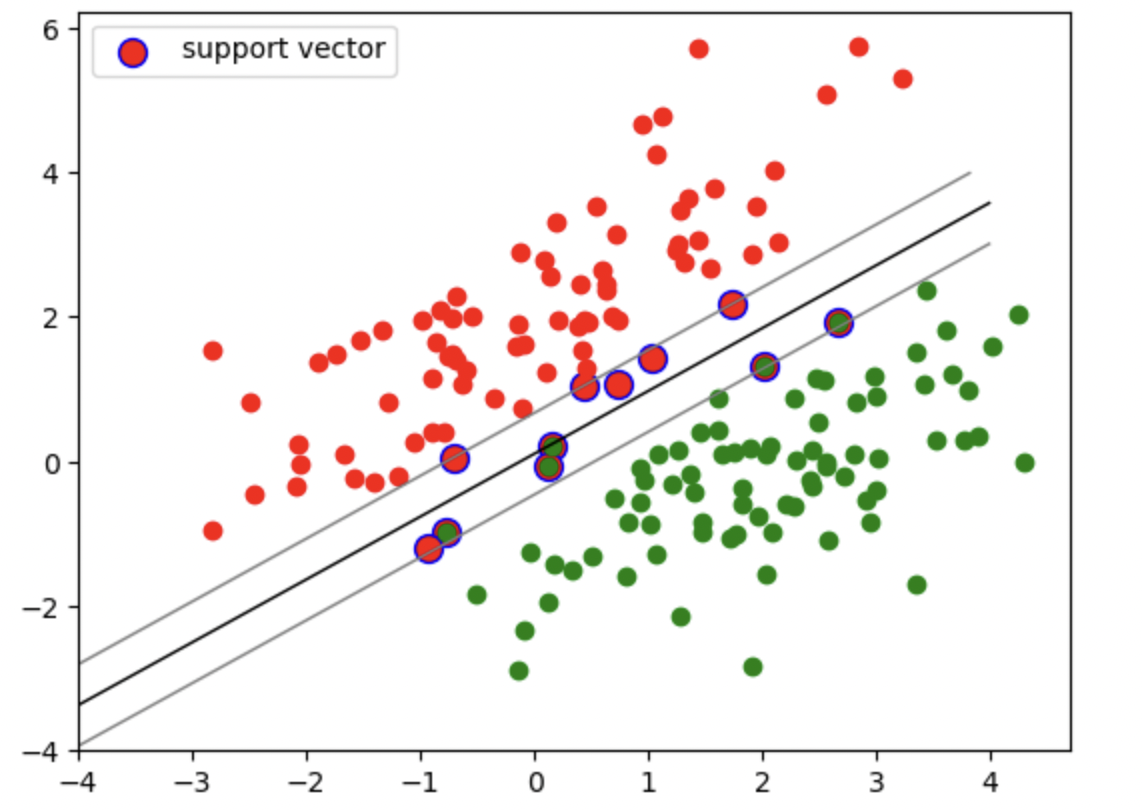


2.3广义线性可分支持向量机



2.4结果可视化





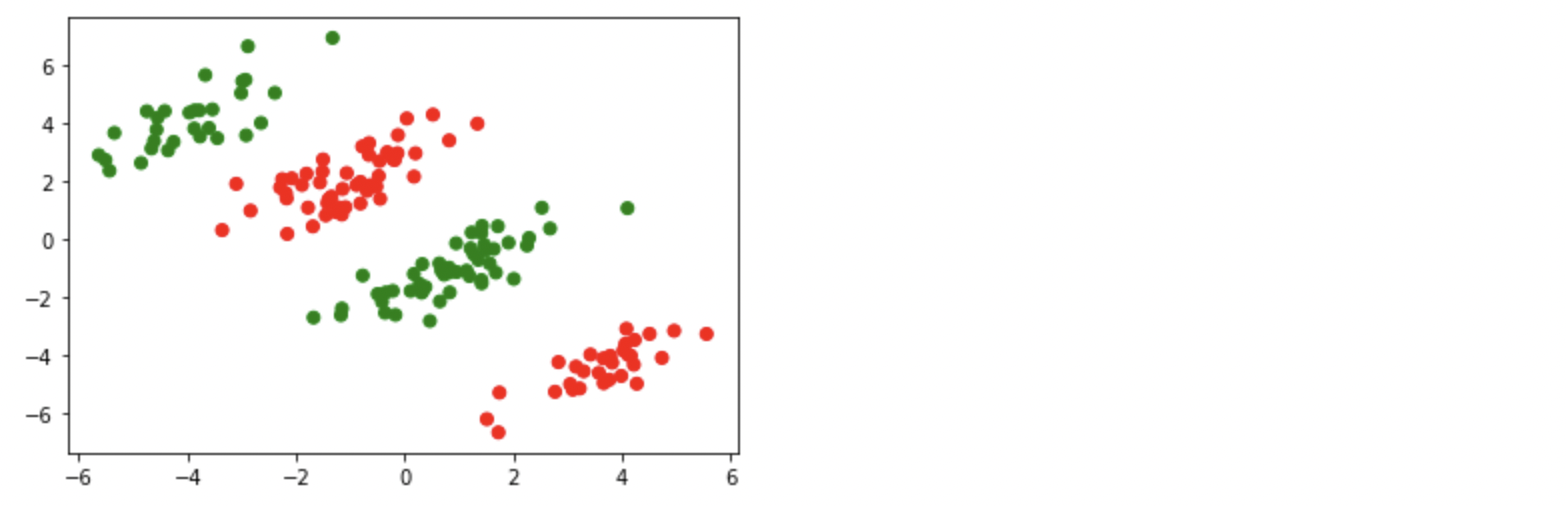
## 非线性可分SVM

3.1导入相关库

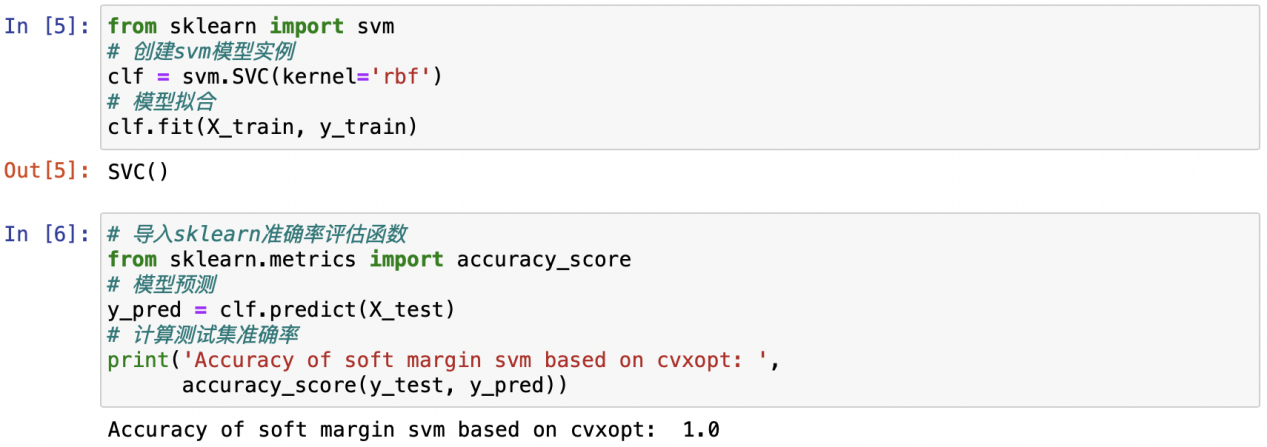
同上

3.2生成模拟数据





3.3非线性SVM



3.4结果可视化

