

Machine Learning 2019

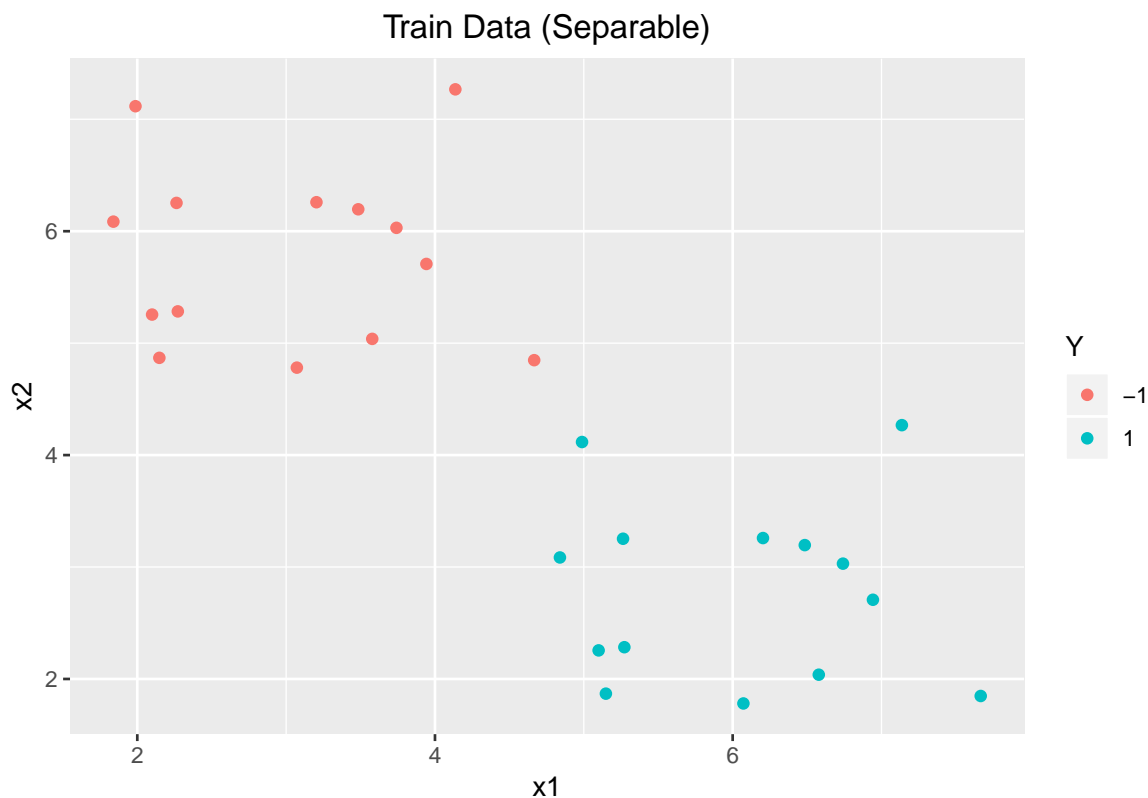
袁欣

2019 年 4 月 11 日

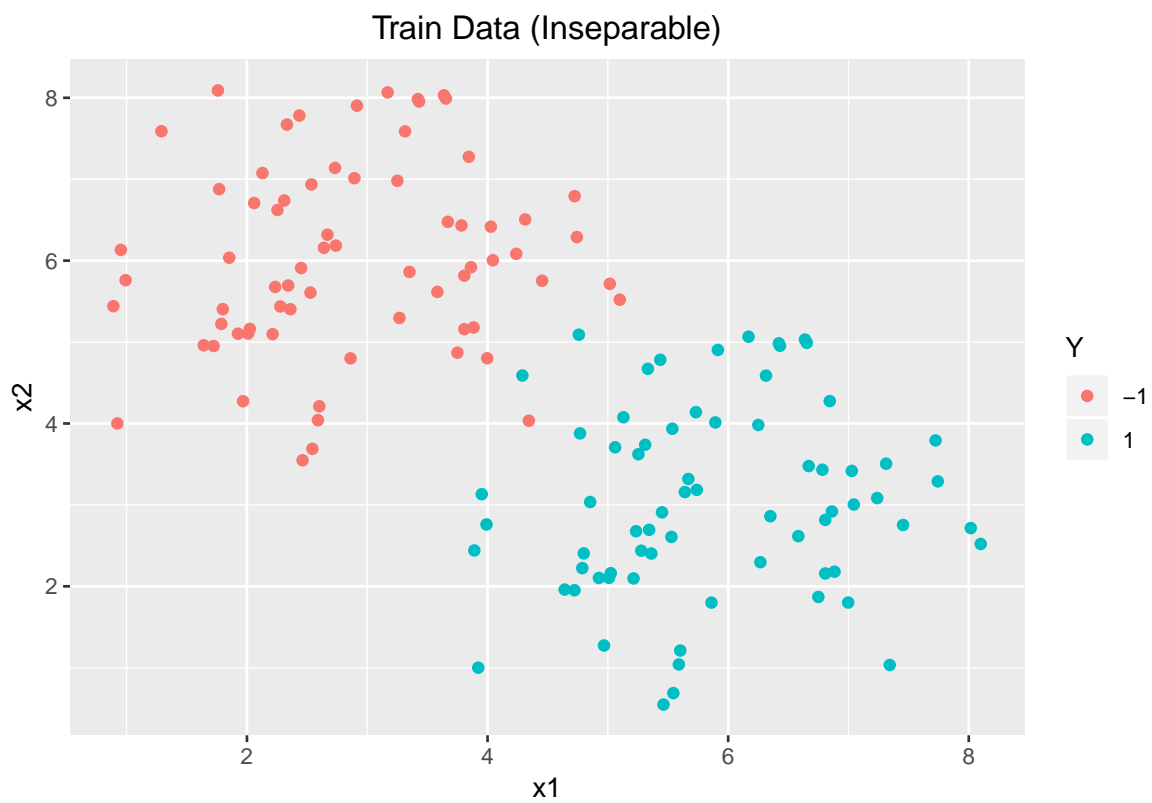
1 支持向量机

- 构建数据集

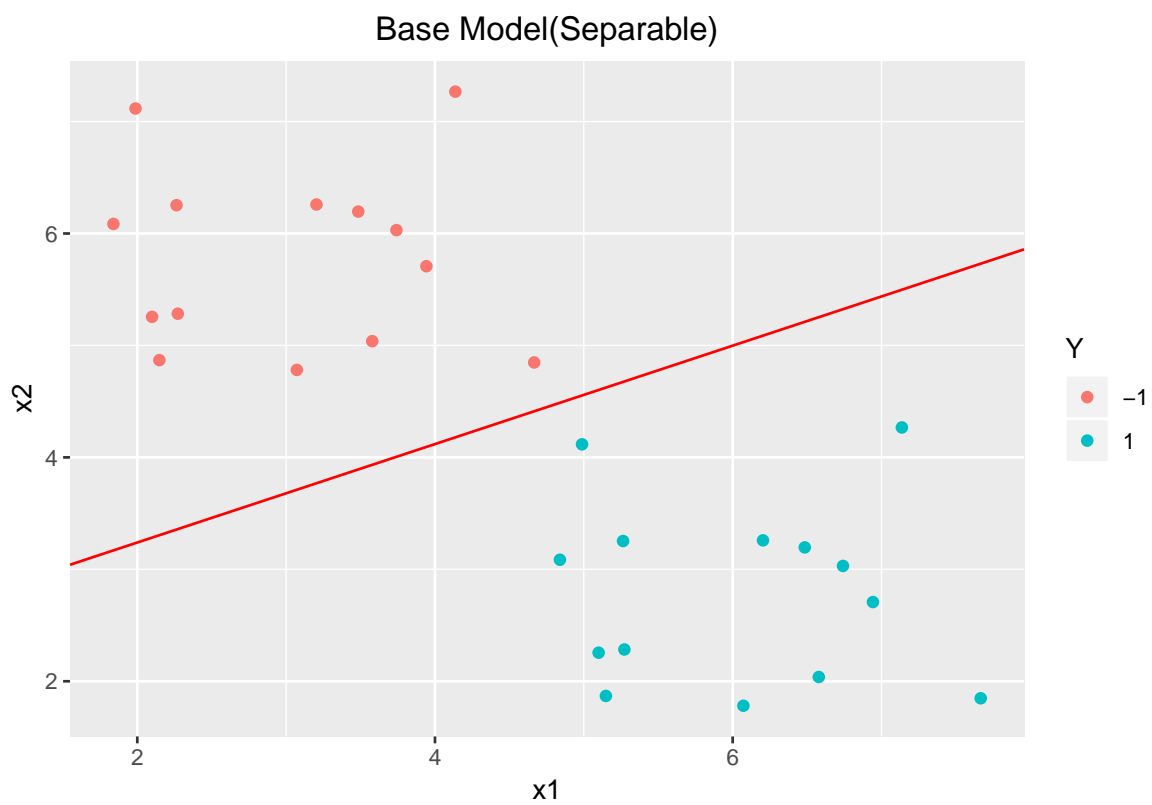
随机生成服从二元正态分布的 40 个样本，其中 20 个样本均值为 $[3, 6]^T$ ，协方差阵为单位矩阵，设置标签为-1；20 个样本均值为 $[6, 3]^T$ ，协方差阵仍为单位矩阵，设置标签为1（确保两类样本线性可分）。然后按照 7: 3 的比例将样本分为训练集与测试集，并画出训练集图形。



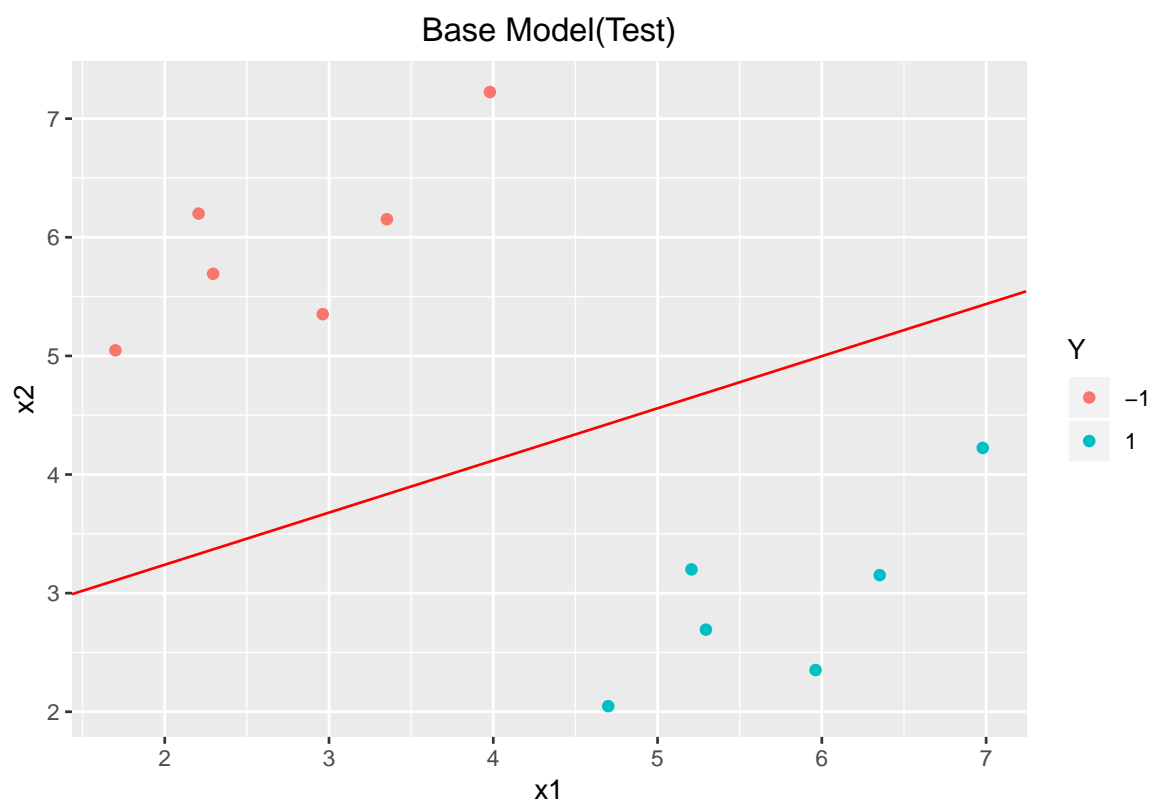
随机生成服从二元正态分布的 200 个样本，其中 100 个样本均值为 $[3, 6]^T$ ，协方差阵为单位矩阵，设置标签为-1；100 个样本均值为 $[6, 3]^T$ ，协方差阵仍为单位矩阵，设置标签为-1（确保两类样本线性不可分）。然后按照 7: 3 的比例将样本分为训练集与测试集，并画出训练集图形。



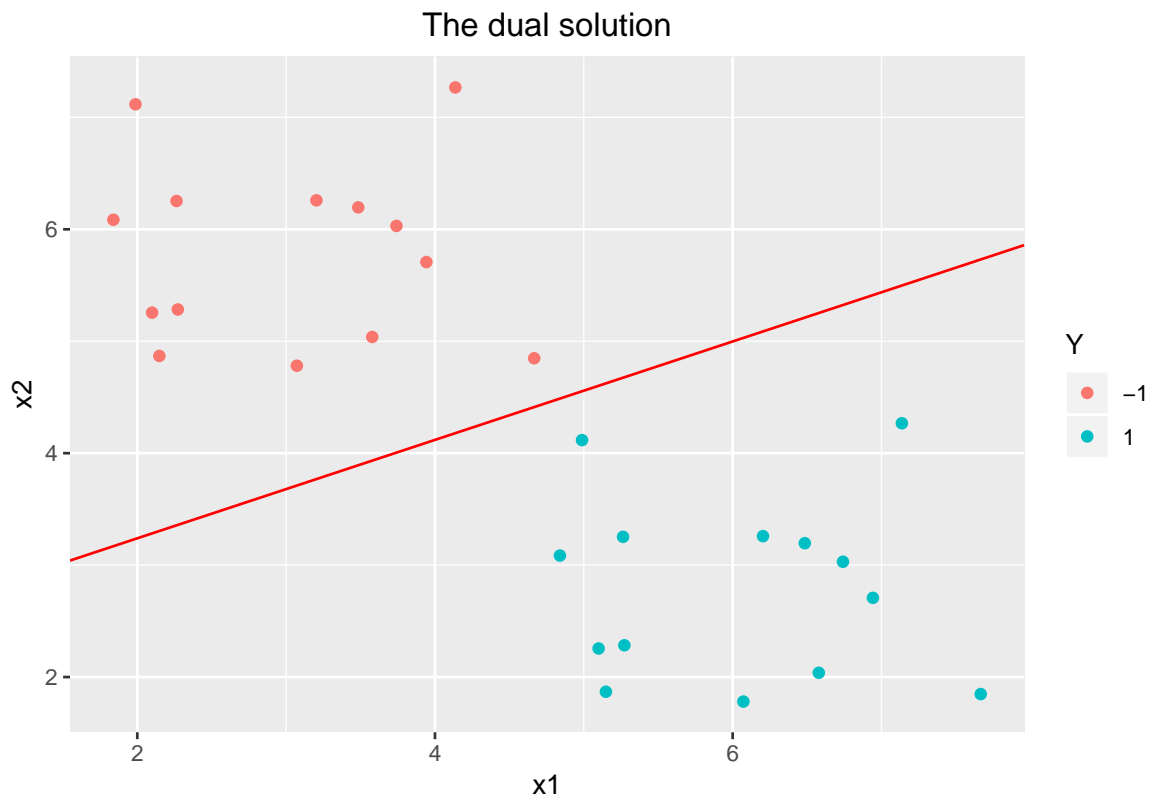
- 基本形式求解



- 测试集检验

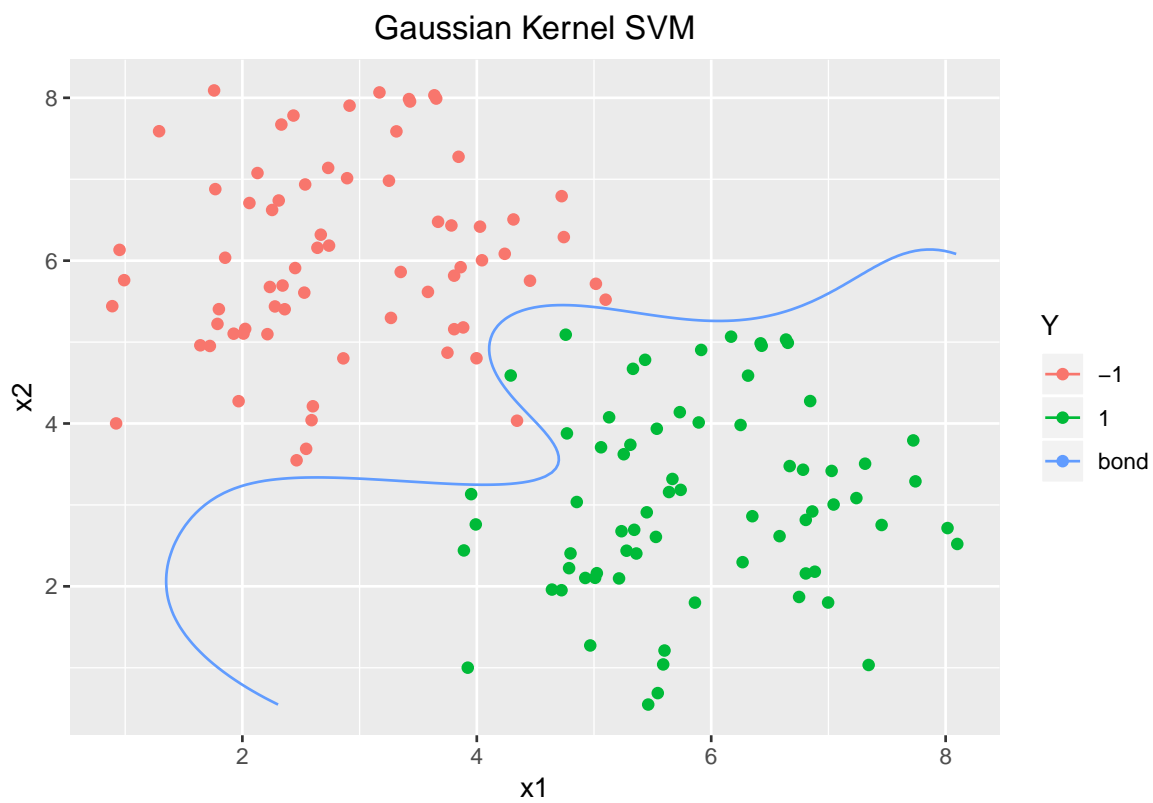


- 对偶问题求解

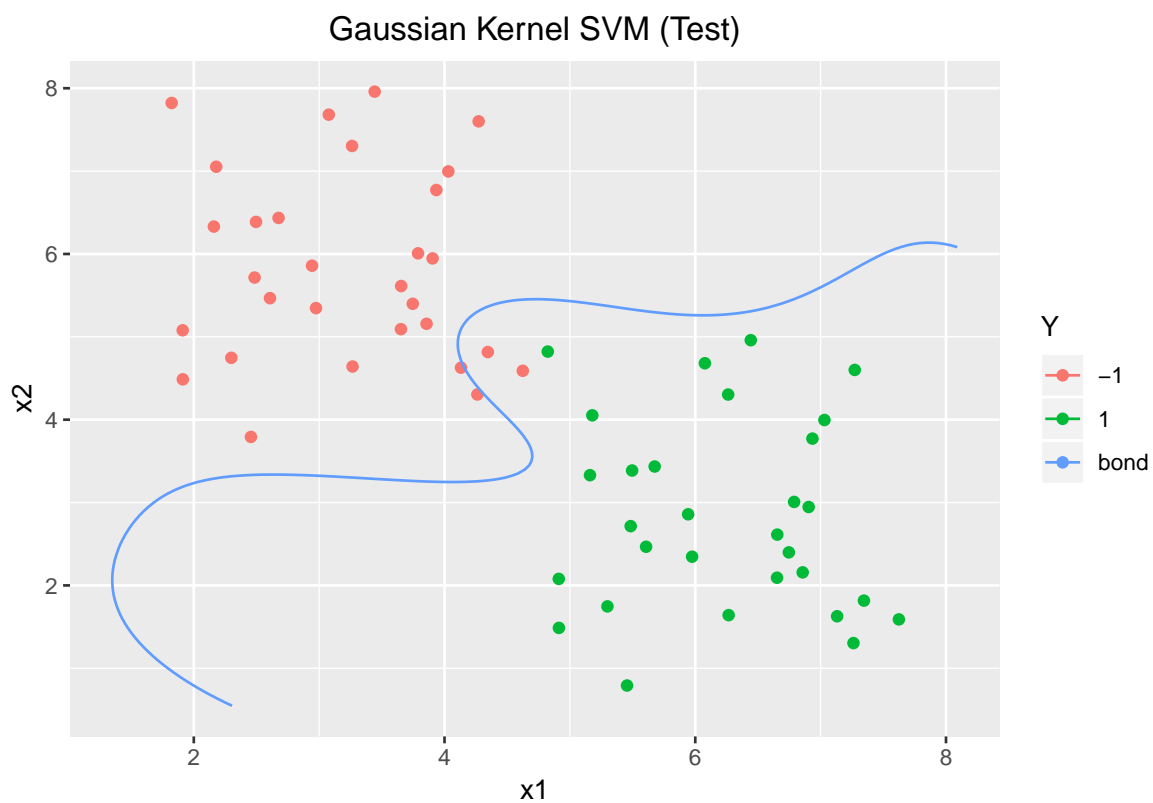


- 核函数

当原始样本空间不存在能正确划分两类样本的超平面时，我们需要将样本从原始空间映射到一个更高维的特征空间，使得样本在这个特征空间内线性可分。

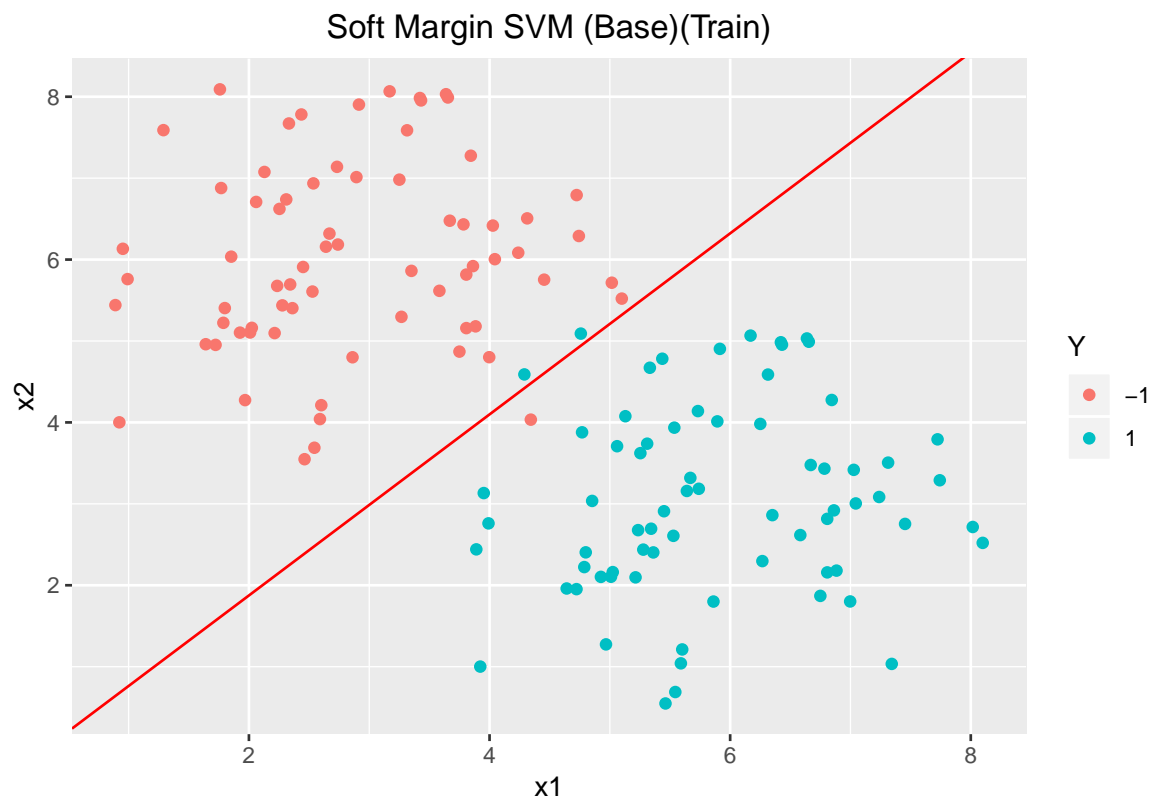


- 测试集检验

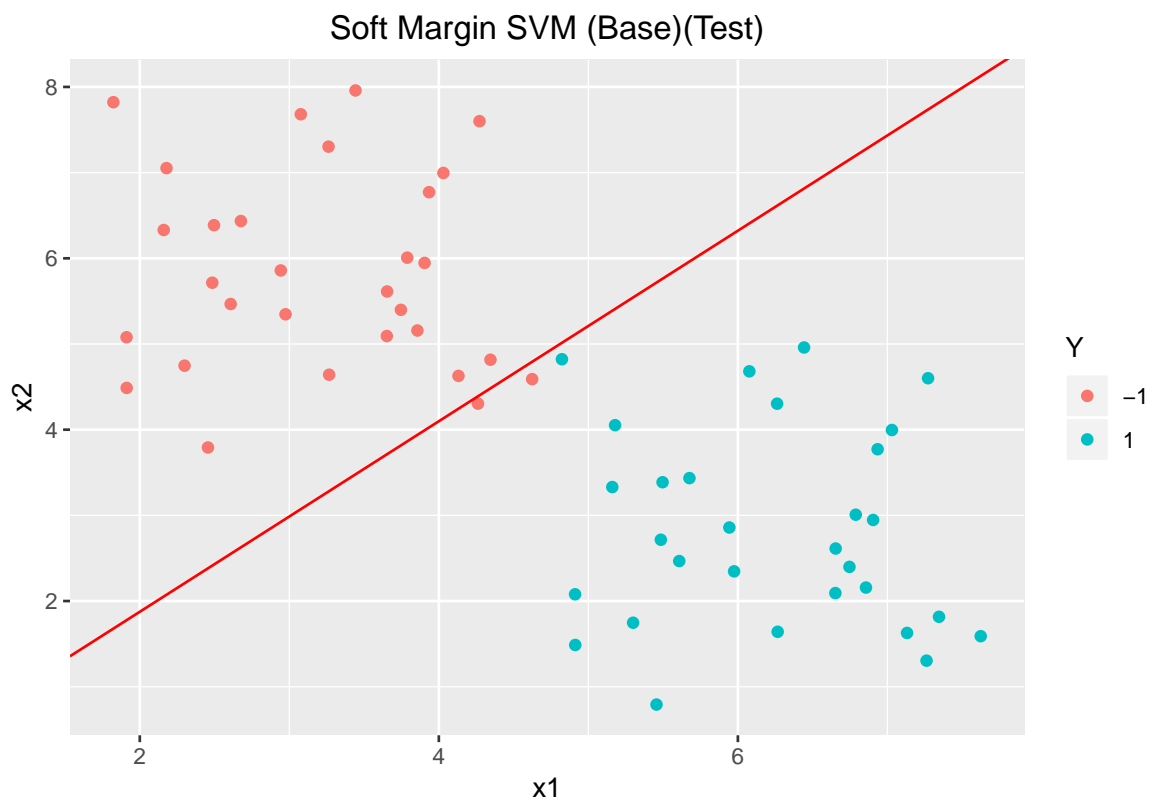


- 软间隔与正则化

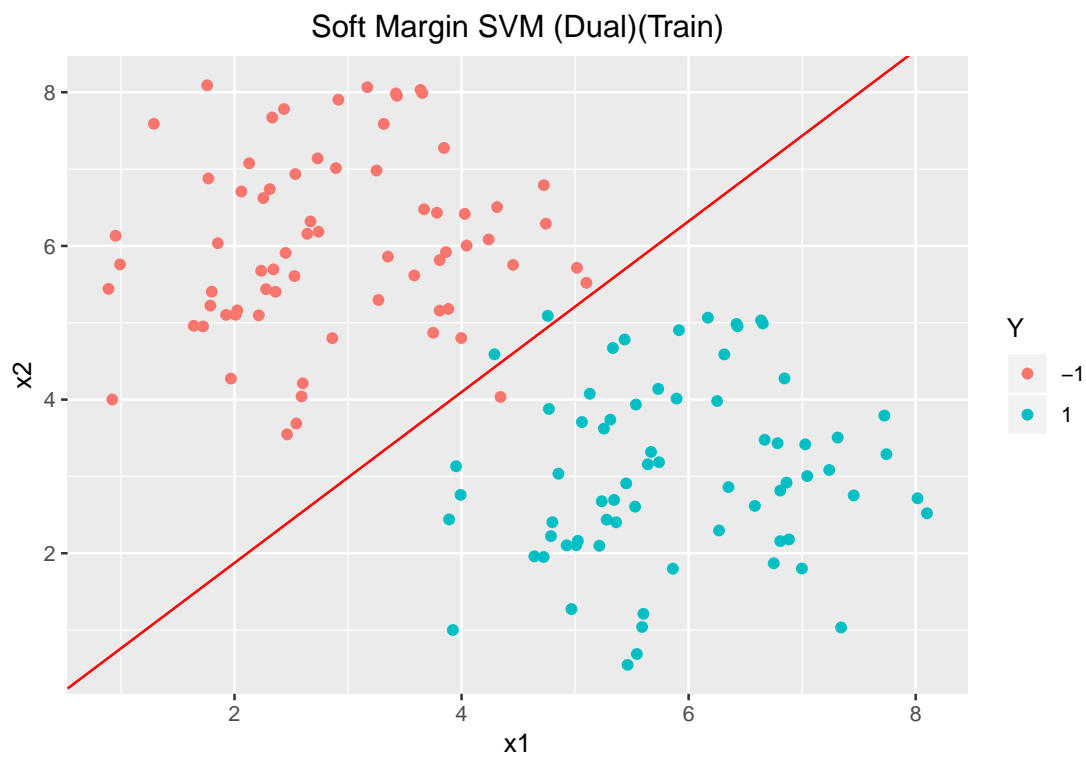
当样本空间线性不可分时，我们除了引入核函数的方法以外，还可以利用软间隔与正则化。软间隔（soft margin）的思想是允许支持向量机在一些样本上出错。



- 测试集检验

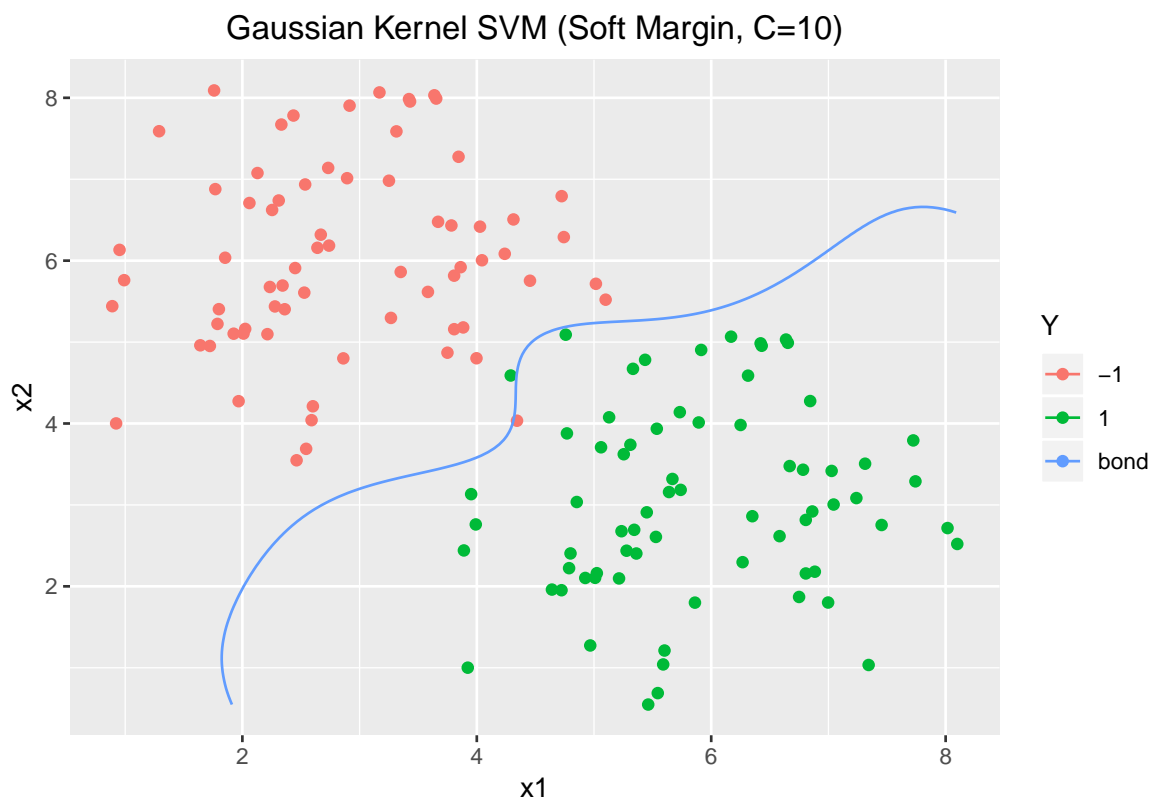


- 对偶问题求解



- 编程实现

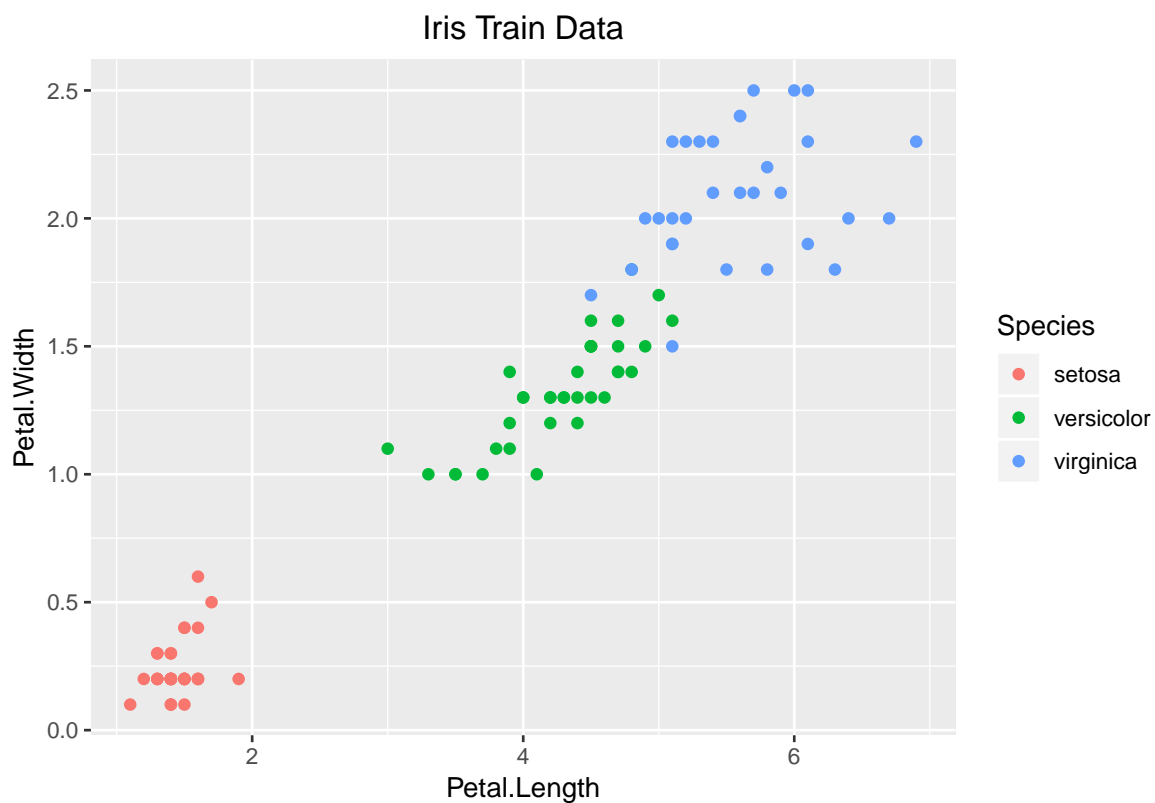
当然基于高斯核函数的支持向量机也可以加入软间隔降低模型过拟合的可能。



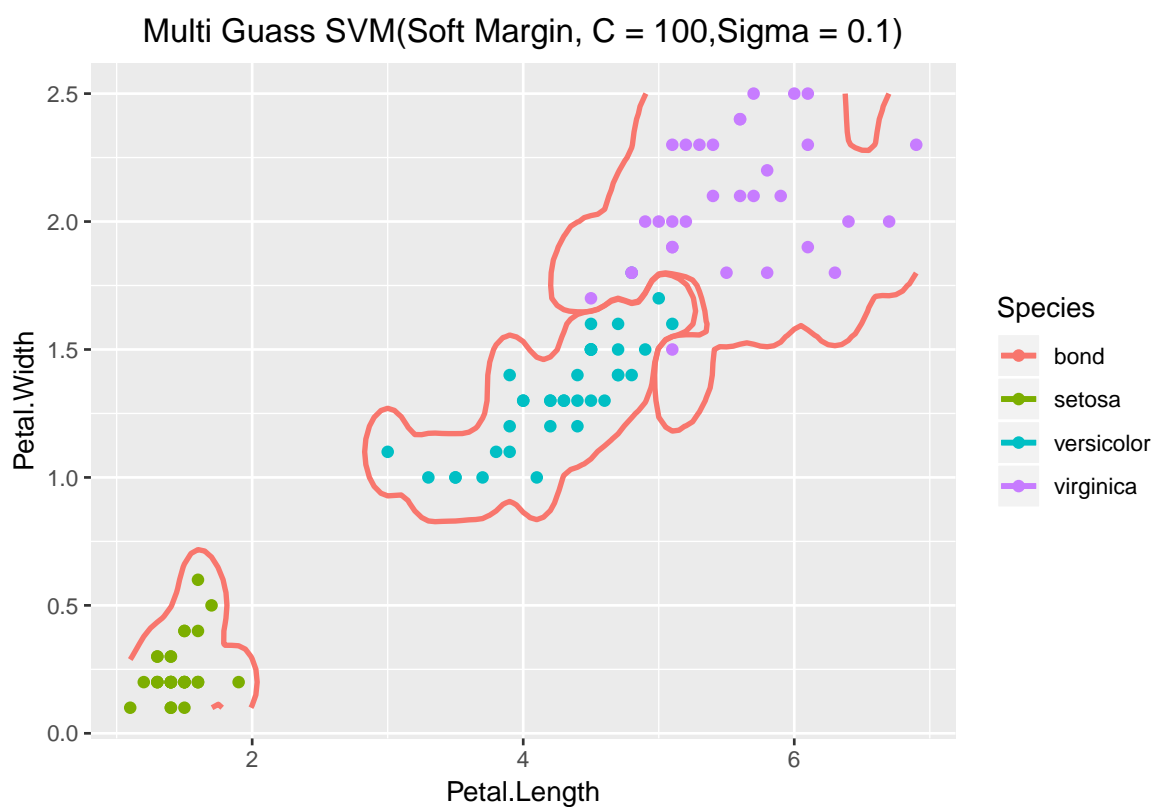
- 支持向量多分类

我们利用 OvR 训练 N 个分类器解决支持向量机多分类问题。OvR 每次将一个类的样本作为正例、所有其他类的样本作为反例来进行训练，训练完成后分别计算各分类器的 $f(x)$ ，将样本划分到 $f(x)$ 取最大值的类别。

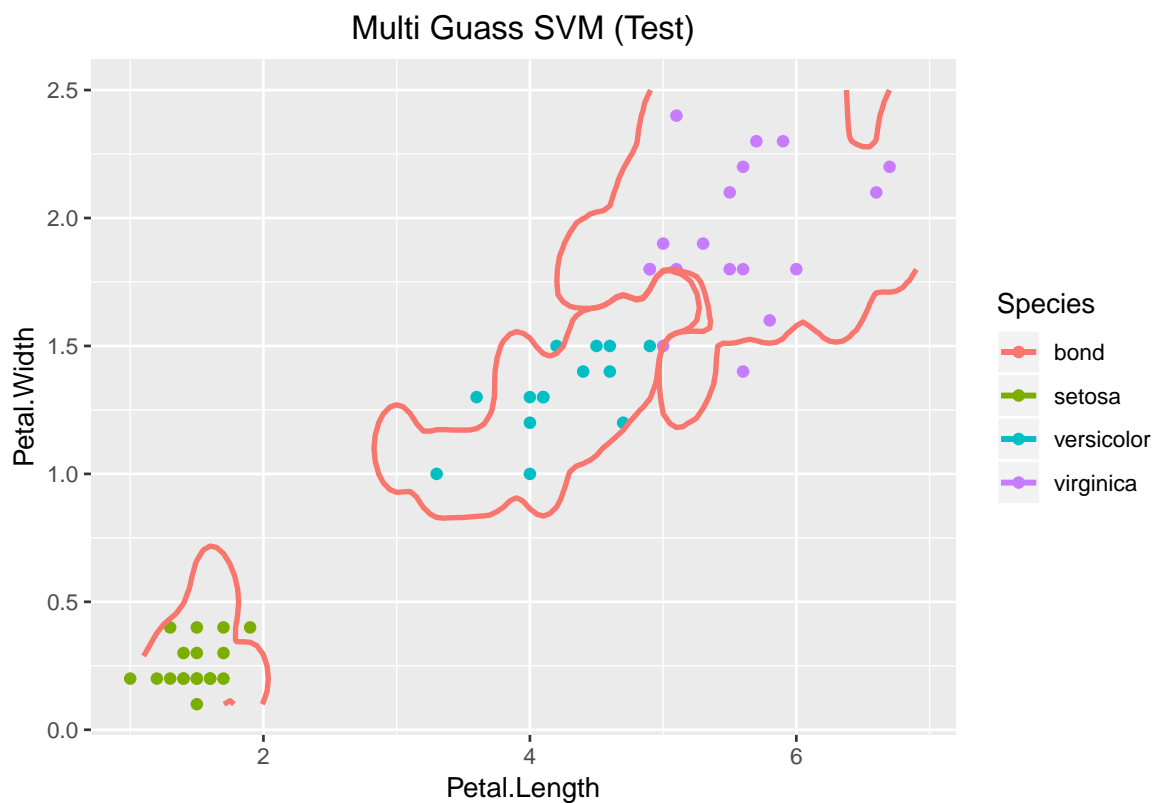
- Iris 数据集



- 编程实现

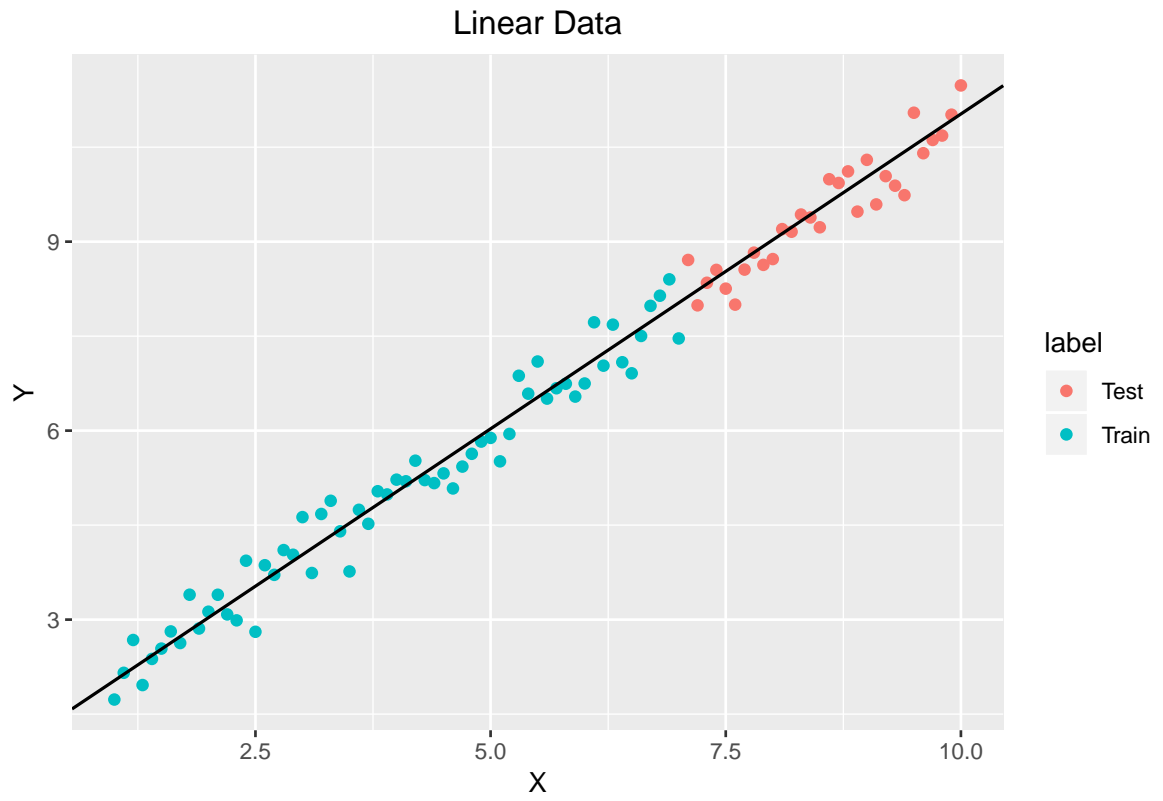


- 测试集检验



- 支持向量机回归

支持向量回归 (Support Vector Regression, 简称 SVR) 与传统模型不同的是, 它假设我们能容忍 $f(x)$ 与 y 之间最多有 ϵ 的偏差, 即仅当 $f(x)$ 与 y 之间的绝对偏差大于 ϵ 时才计算损失。



- 预测
- 注：高斯核画图

利用等高线 0 表示分类边界。