

# Homework 8

PB17000297 罗晏宸

May 3 2020

## 1

试证明对于不含冲突数据（即特征向量完全相同但标记不同）的训练集，必存在一个与训练集一致（即训练误差为 0）的决策树。

**解**

证明. 决策树中叶节点生成于样本属于同一类或者样本在当前属性集上取值相同时，对于不含冲突数据的训练集，不会出现在当前属性集上取值相同但不同标记的样本，因此决策树的叶节点均为同类样本，即从根节点到每个叶节点的路径都代表一个向量，这个决策树与训练集是一致的。  $\square$

## 2

已知正例点  $\mathbf{x}_1 = (1, 2)^T$ ,  $\mathbf{x}_2 = (2, 3)^T$ ,  $\mathbf{x}_3 = (3, 3)^T$ , 负例点  $\mathbf{x}_4 = (2, 1)^T$ ,  $\mathbf{x}_5 = (3, 2)^T$ , 试求 Hard Margin SVM 的最大间隔分离超平面和分类决策函数，并在图上画出分离超平面、间隔边界以及支持向量。

**解** 在 2 维空间中设 1 维超平面（即直线）方程为  $\boldsymbol{\omega} \cdot \mathbf{x} - b = (A, B) \cdot (x, y)^T + C = Ax + By + C = 0$ 。为避免歧义，用  $x_i$  与  $y_i$  表示点  $\mathbf{x}_1$  的分量，也即在二维空间中的坐标。平行的两个间隔边界间距为  $\frac{|C_1 - C_2|}{\sqrt{A^2 + B^2}}$ ，对

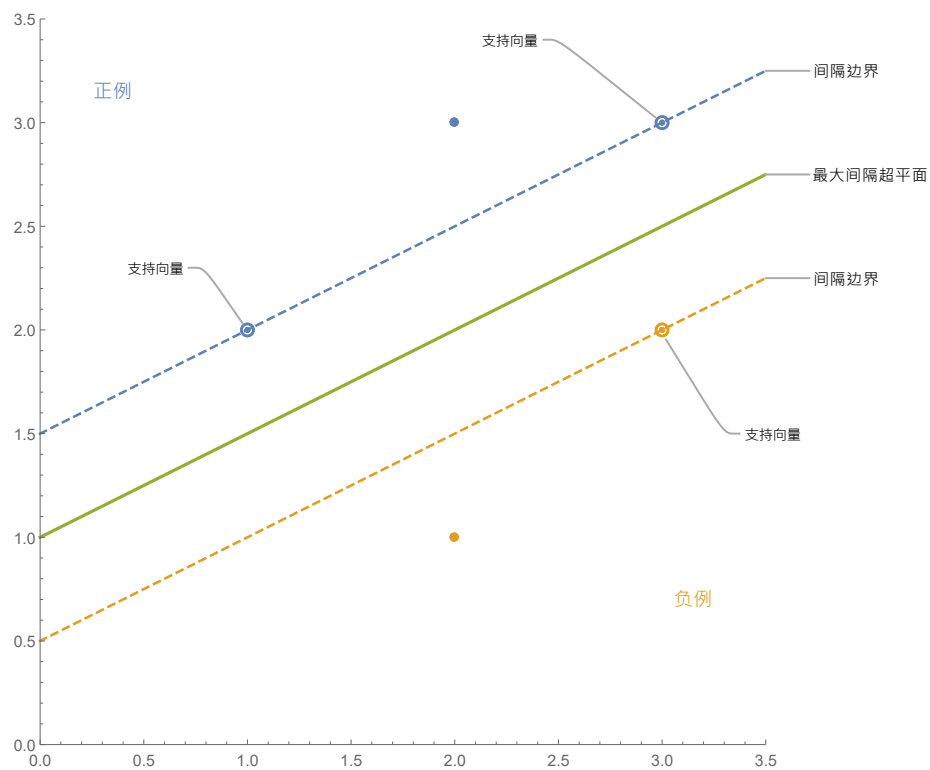


图 1: 最大间隔分离器、间隔边界与支持向量

于数据集求间距的最大值点，即求

$$\arg \max \begin{cases} \frac{|2B - A|}{\sqrt{A^2 + B^2}}, & -2 \leq \frac{A}{B} < -1, & C_1 = -3A - 3B, & C_2 = -2A - B \\ \frac{|B|}{\sqrt{A^2 + B^2}}, & -1 \leq \frac{A}{B} < -\frac{1}{2}, & C_1 = -3A - 3B, & C_2 = -3A - 2B \\ \frac{|2A|}{\sqrt{A^2 + B^2}}, & -\frac{1}{2} \leq \frac{A}{B} \leq 0, & C_1 = -A - 2B, & C_2 = -3A - 2B \end{cases}$$

得到

$$\frac{A}{B} = -\frac{1}{2}, \quad C_1 = -\frac{3B}{2}, \quad C_2 = -\frac{B}{2}$$

如图1所示，最大间隔分离超平面  $\{\mathbf{x} : \boldsymbol{\omega} \cdot \mathbf{x} - b = 0\}$ ，其中  $\boldsymbol{\omega} = (-1, 2)$ ， $b = 2$ ，即

$$\{\mathbf{x} : -x + 2y - 2 = 0\}$$

分类决策函数为

$$h(x) = \text{sign} \{ \boldsymbol{\omega} \cdot \boldsymbol{x} - b \} = \text{sign} \{ -x + 2y - 2 \}$$