
作业 5

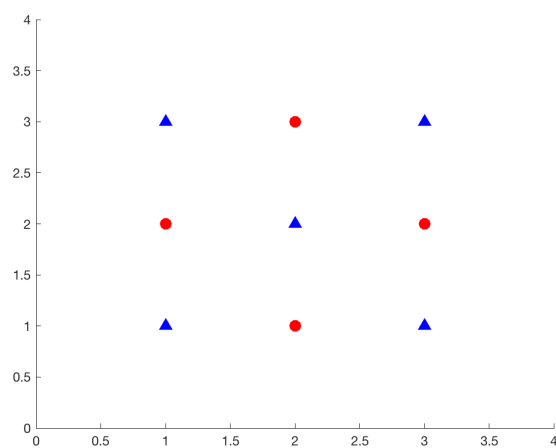
郭一隆

April 26, 2018

1 非线性分类器

1.1 问题描述

求分离 ● 和 ▲ 的函数。



1.2 求解思路

记 ▲ 为第 0 类, ● 为第 1 类。
从数据本身的分布特点考虑:

- 周期性:

两类数据点在二维平面的分布具有很强的周期性。由于数据恰好分布在格点上, 可考虑用奇偶性来分类。即对于数据 $\mathbf{x} = (x_1, x_2)$, 其分类

$$i = (x_1 + x_2) \bmod 2$$

上述分类函数只能应对坐标为整数的数据, 扩展至全平面:

$$i = (\text{round}(x_1) + \text{round}(x_2)) \bmod 2$$

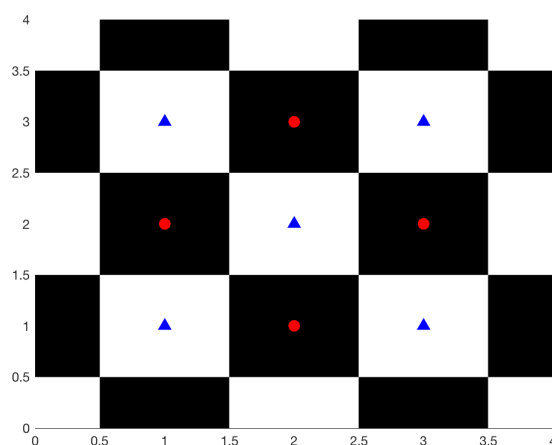


Figure 1: 取整奇偶性分类函数

- 对称性:

由于数据点全体分别关于直线 $x_1 = 2$ 和直线 $x_2 = 2$ 对称, 考虑映射

$$(x_1, x_2) \rightarrow (x'_1, x'_2) = (|x_1 - 2|, |x_2 - 2|)$$

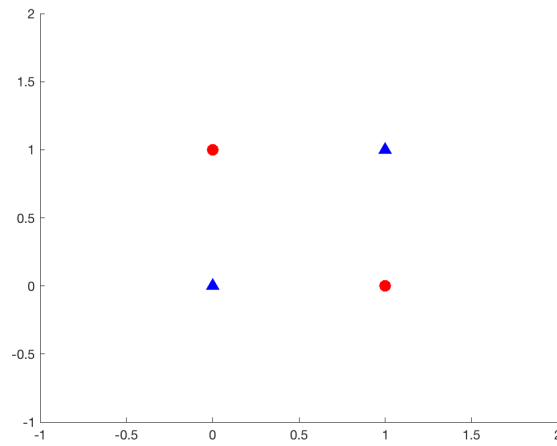


Figure 2: 经过对称折叠后的数据分布

得到的新数据分布（如图2），与课件中提到的广义线性识别函数的例子相同。可通过增加第三维度

$$x'_3 = x'_1 x'_2$$

进一步构造出高维平面

$$d(\mathbf{X}') = x'_1 + x'_2 - 2x'_3 - \frac{1}{3} = x'_1 + x'_2 - 2x'_1 x'_2 - \frac{1}{3} = 0$$

映射回原坐标系即

$$d(\mathbf{X}) = |x_1 - 2| + |x_2 - 2| - 2|x_1 - 2||x_2 - 2| - \frac{1}{3} = 0$$

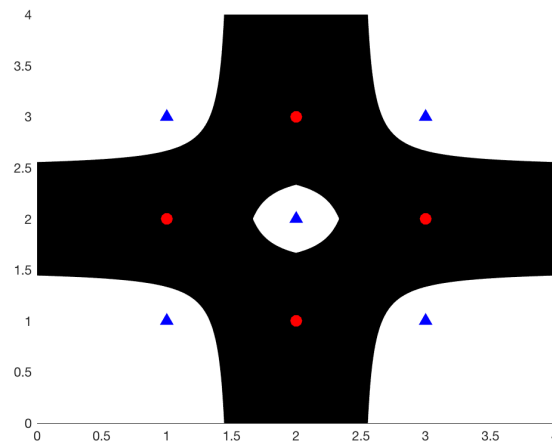


Figure 3: 对称双曲线分类函数

如图3所示，分类界面由对称的 4 支双曲线交叠而成。

- 两条直线分割：

仍考虑映射

$$(x_1, x_2) \rightarrow (x'_1, x'_2) = (|x_1 - 2|, |x_2 - 2|)$$

得到如图2所示的数据分布，可考虑用两条直线所夹区域来识别：

$$d(\mathbf{X}') = (x'_1 - x'_2)^2 - \frac{1}{4} = 0$$

即

$$d(\mathbf{X}) = (|x_1 - 2| - |x_2 - 2|)^2 - \frac{1}{4} = 0$$

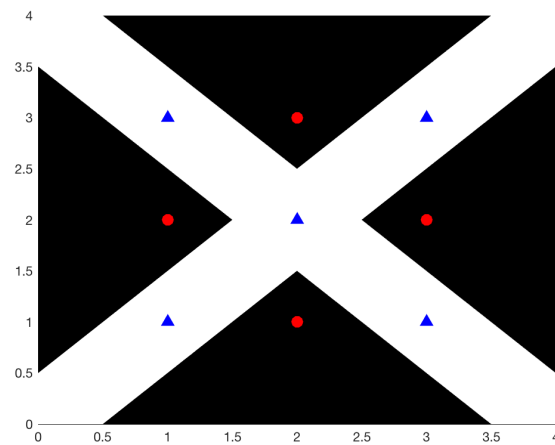


Figure 4: 多直线分类函数

2 迭代修正求权向量