

# 分段线性距离分类器

为什么要分段？

多峰的数据分布

单峰和多峰数据怎么处理？

单峰直接分类

多峰的划分成多个单峰，即划分子类

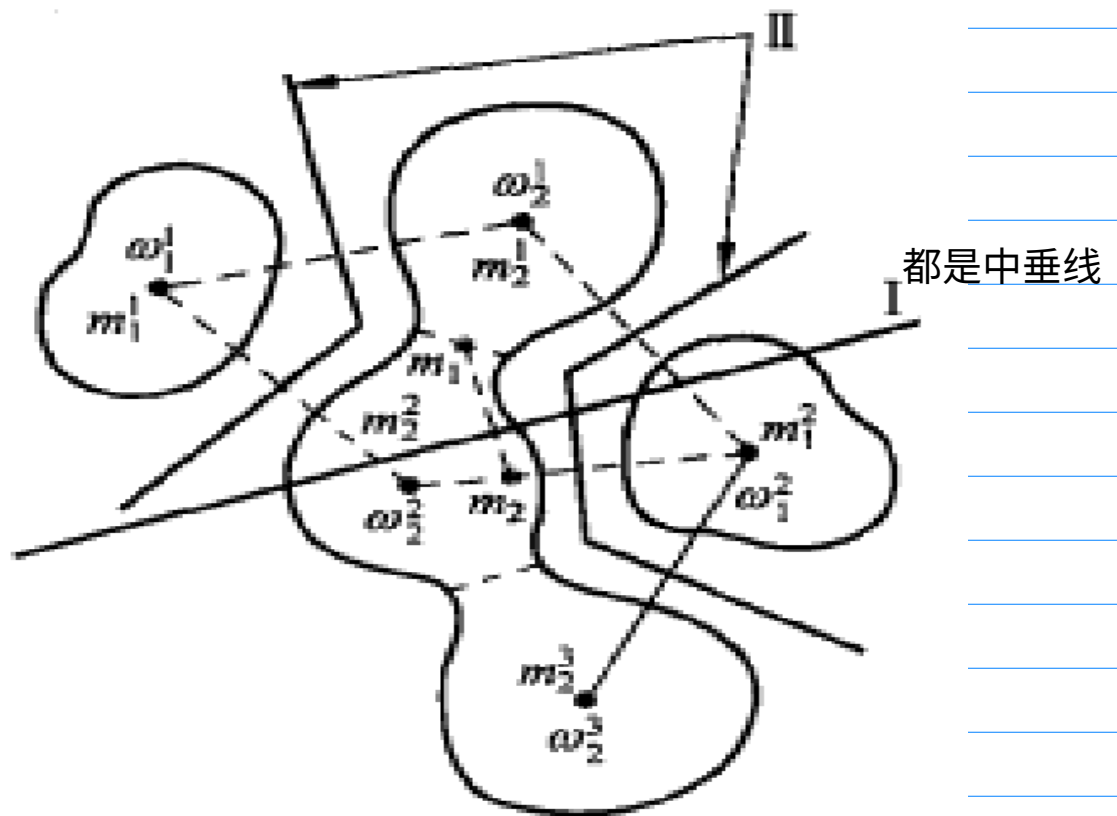


图 5-3 分段线性距离分类器示例

沿着这种思路,在各类的数据分布是多峰的情况下,我们可以把每类划分成若干个子类,使每个子类是单峰分布且尽可能在各维上对称。每个子类取均值作为模板,这样每个类就有多个模板,一个两类问题就可以用多类的最小距离分类器来解决,即:对一个待分类样本,比较它到各个子类均值的距离,把它分到距离最近的子类所属于的类。这样所得到的分类面就是由多段超平面组成的,如图5-3所示。这种分离器称作分段线性距离分类器。这种做法对多类同样适用。

$$\hat{g}_k(x) = \min_{i=1, \dots, c} g_i(x)$$

$$g_i(x) = \min_{l=1, 2, \dots, L} \|x - m_{il}^l\|$$

淘宝店铺-酷流科技 掌柜：我是雷锋的朋友

在第2章我们看到,当两类的类条件概率密度为正态分布,两类先验概率相等,而且各维特征独立且方差相等时,最小错误率贝叶斯决策就是直观的最小距离分类器:以两类各自的均值为中心点,新样本离哪类的中心点近就决策为哪一类。

$$\|x - \mu_k\|^2 = \min_{i=1, \dots, C} \|x - \mu_i\|^2$$

$$\text{If } x \in W_k$$

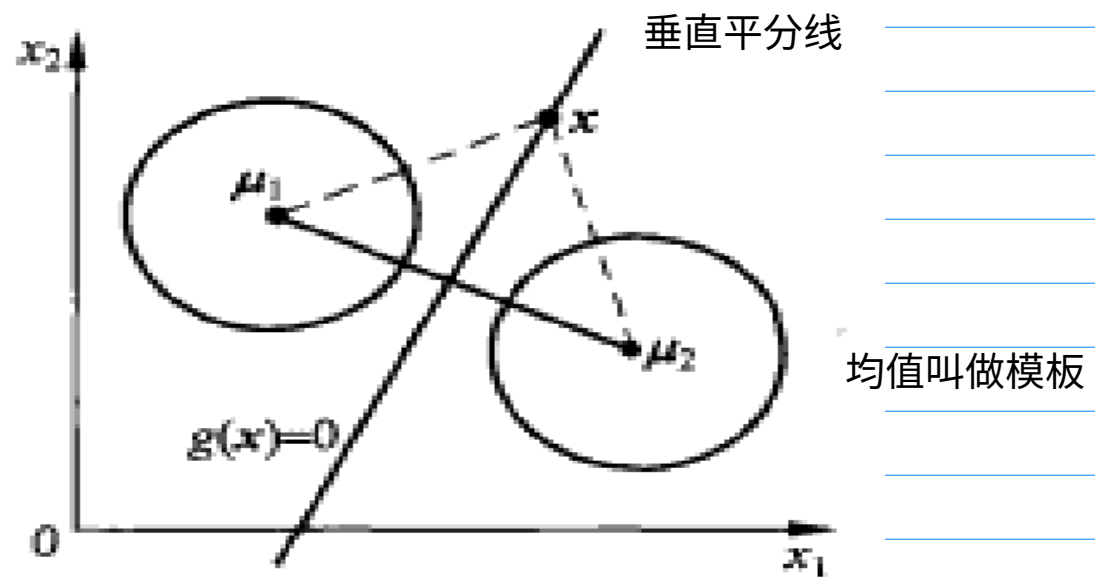


图 5-2 两类的最小距离分类器

淘宝店铺-酷流科技 掌柜：我是雷锋的朋友