模式识别的理论与方法

Pattern Recognition

裴继红

Chapter 5 (Part 1): Linear Discriminant Functions

(Sections 5.1-5.5)

引言

- 在第 3 章中，假设关于**模式类**的**内在的概率密度**是已知、或已经给出的。
  - 训练样本是用来估计这些概率密度的参数的（如ML方法，MAP方法）
- 在本章中，我们假设仅仅知道判别函数的适当的形式，依此来作为分类器训练的先验知识。
  - 基本思想与非参数估计很类似。
  - 得到的分类器——线性分类器也许不是最优的，但使用起来却非常简单。

本讲内容

- 线性判别函数和判决面
- 多类情况下的线性判别函数
- 线性分类机
- 广义线性判别函数

线性方程、线性函数回顾

1. 线性方程的常用形式
   - 斜截式、法线式、一般式、……
2. 线性方程的几何意义
   - 超平面
   - 原点到超平面的距离
3. 线性函数的几何意义
   - 点与直线（超平面）的关系
   - 点到直线（超平面）的距离

线性判别函数

- 定义

线性判别函数是将特征矢量 $\boldsymbol{x}$ 的各个分量经线性组合而成的函数。

$$g(x) = w^t x + w_0 \qquad (1)$$

这里，w 称为**权矢量**，w_0 称为**阈值**或**偏置**

- 对式（1）的**两类**分类器，判别函数使用如下的**决策规则**:

若 $\boldsymbol{g}(\boldsymbol{x}) > 0$，则判别为$\omega_1$；若 $\boldsymbol{g}(\boldsymbol{x}) < 0$，则判别为 ω_2

$\Leftrightarrow$ 若 $w^t x > -w_0$，则判别为 ω_1；否则，判别为 ω_2

特别，若 $g(x) = 0 \Rightarrow x$ 可以指派为任意一个类

两类情况的线性分类器模型

① 有 d 个输入单元的简单线性分类器，其中每一个单元对应输入矢量的一个分量值；

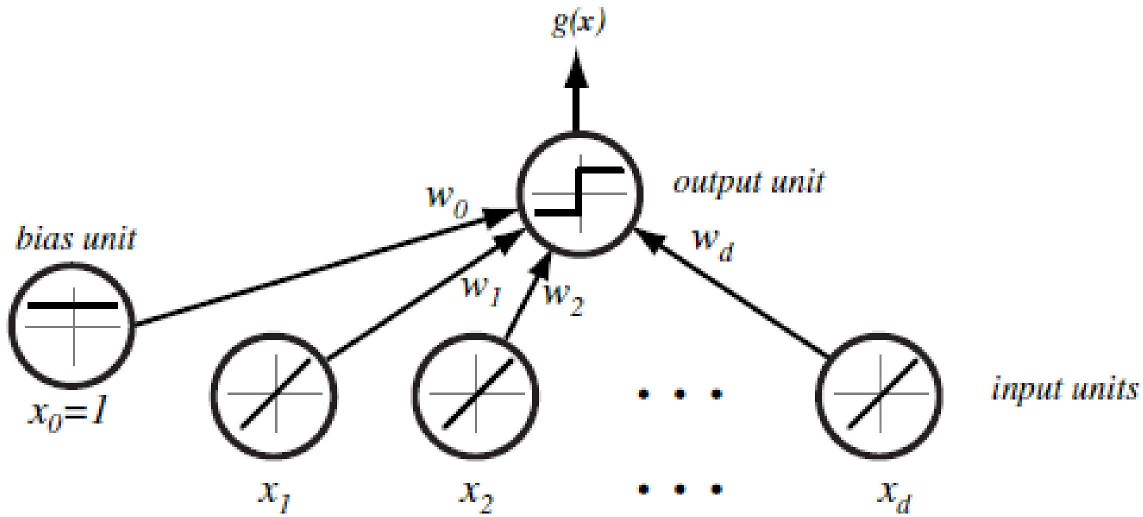
② 偏置单元的输入是常数=1.0；

③ 输入单元是线性的，输出的是其特征值的准确值；

④ 每个输入的特征值 x_i 与其对应的权值 w_i 相乘；

⑤ 输出单元的有效输入是所有这些积之和 $\sum w_i x_i$ 。

⑥ 输出单元的输出为+1 （若$\mathbf{w}^t\mathbf{x} + w_0 > 0$ ）或−1 （ $\mathbf{w}^t\mathbf{x} + w_0 <= 0$ ）



决策面、线性决策面

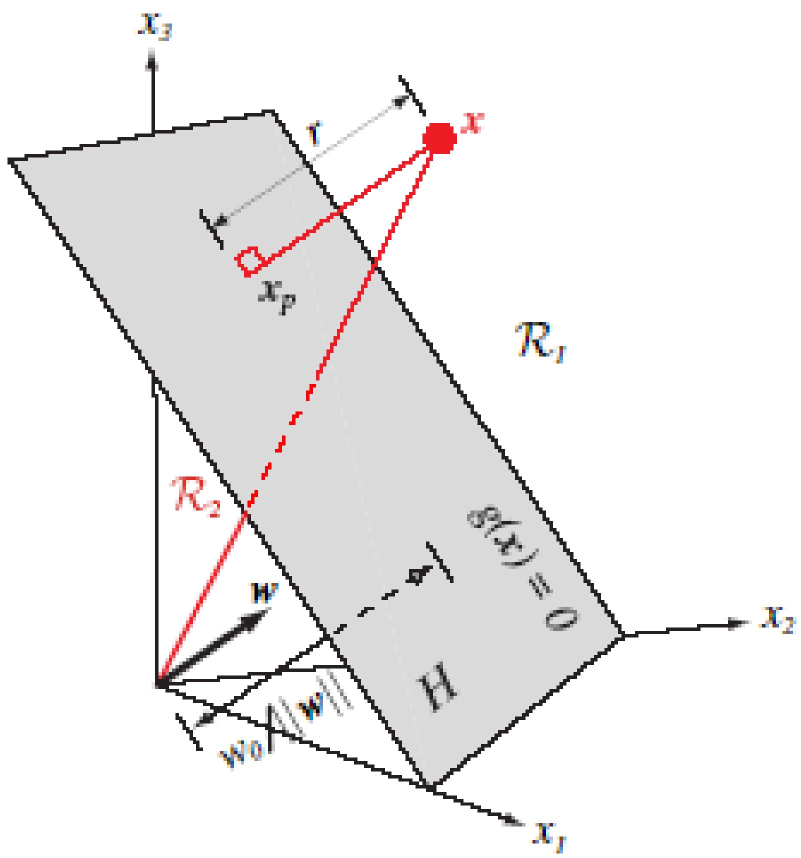
- 方程 $g(x) = 0$ 定义了一个将属于 ω_1 类的点与属于 ω_2 类的点分隔开来的决策面

- 由于 $g(x)$ 是线性函数，因此决策面 $g(x) = 0$ 是一个线性超平面

  说明：线性函数 $g(x)$ 的输出值，给出了从特征矢量 x 到该超平面的距离的代数度量

模式矢量点与判决边界：Figure5.2

- 线性决策边界***H***，边界中 $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{w}^t\boldsymbol{x} + w_0 = 0$，将特征空间划分为两个半空间 R_1 (其中 $\boldsymbol{g}(\boldsymbol{x}) > 0$)，以及 R_2 (其中 $\boldsymbol{g}(\boldsymbol{x}) < 0$).



$$r = \frac{g(x)}{\|\mathbf{w}\|} \qquad d(0,\mathrm{H}) = \frac{w_0}{\|\mathbf{w}\|}$$

决策面的法向量 w

- 超平面 *H* 将特征空间划分为两个区域：
  - 区域 R1 对应于类 ω_1
  - 区域 R2 对应于类 ω_2

- 对于在**决策面**上的两个点 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$：

$$\boldsymbol{w}^t\boldsymbol{x}_1+\boldsymbol{w}_0 = \boldsymbol{w}^t\boldsymbol{x}_2+\boldsymbol{w}_o$$

进行简单的变换：

$$\boldsymbol{w}^t(\boldsymbol{x}_1-\boldsymbol{x}_2)=0$$

$\boldsymbol{w}$与超平面中的任何矢量均正交。因此， $\boldsymbol{w}$是超平面的法向量。

代数距离度量

- 函数 $\boldsymbol{g}(\boldsymbol{x})$ 表示了 $\boldsymbol{x}$ 距决策边界的距离有多远:

点 $\boldsymbol{x}$ 可以表示为

$$x = x_p + r\frac{w}{\|w\|}$$

这里，$\boldsymbol{x}_p$ 是 $\boldsymbol{x}$ 在决策边界上的投影点， $\boldsymbol{r}$ 是到决策面边界的**代数距离**，$\boldsymbol{w}$ 与 $\boldsymbol{x}-\boldsymbol{x}_p$ 共线（平行）。由于 $\boldsymbol{g}(\boldsymbol{x}_p) = \boldsymbol{w}^t\boldsymbol{x}_p + w_0 = 0$，故有

$$\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{w}^t\boldsymbol{x} + w_0 = \boldsymbol{w}^t(\boldsymbol{x}_p + \boldsymbol{r}\,\boldsymbol{w}/\|\boldsymbol{w}\|) + w_0 = \boldsymbol{r}\|\boldsymbol{w}\|$$

这样，$\boldsymbol{g}(\boldsymbol{x})$ 的绝对值越大，则点 $\boldsymbol{x}$ 距离决策边界越远

多类情况下的线性判别函数

若样本类别数在2类以上时，有两种构造线性判别函数的方法:

① 学习并构造出划分每一个类与其它类样本的线性函数

② 学习并构造出划分每两个不同的样本类的划分函数

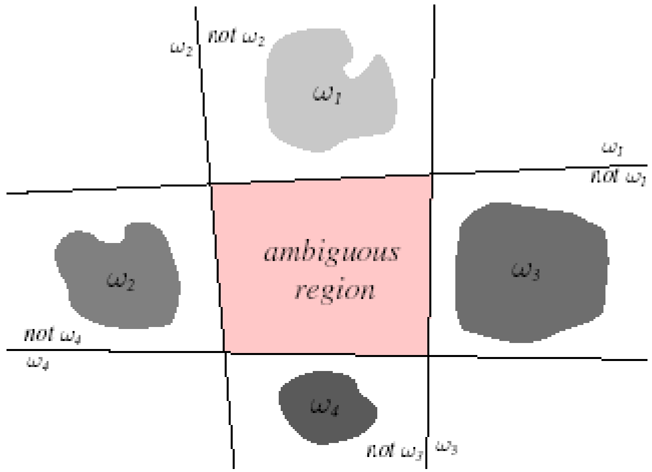
若样本类数为 c，则在第一种情况下，共有 c 个划分函数；

而在第二种情况下，共有$c(c-1)/2$个划分函数。

- 在上述两种情况下，在特征空间中都存在无法确定测试样本类型的不确定区域

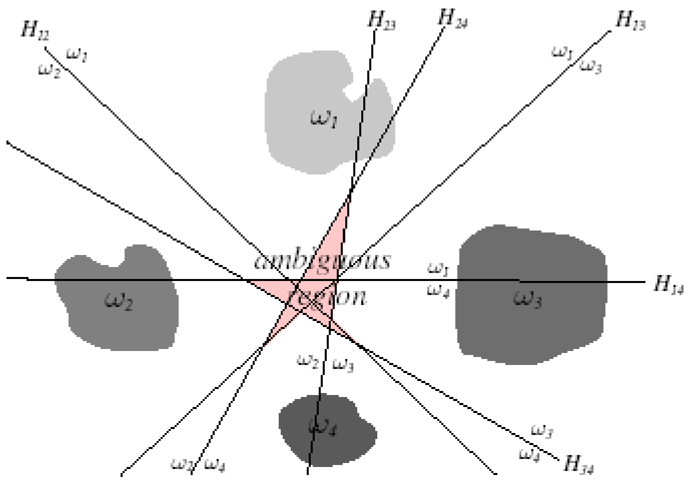
多类情况下：决策边界、特征空间的不确定区域

对于 4 类问题的线性决策边界。



上面的图表示了采用

ω_i /~ω_i 二分方法的结果；



下面的图表示了采用

ω_i /ω_j 二分方法的结果对应的决策边界为 H_{ij}。

图中的粉红色区域是类的类型不确定的区域。

线性分类机

为了避免在特征空间中出现不确定的区域，可以采用线性分类机：

定义 c 个线性函数，将取得最大值的那个函数所在的类作为 x 的类

$$g_k(\boldsymbol{x}) = \boldsymbol{w}_k^t \boldsymbol{x} + w_{k0} \quad k = 1, \ldots, c$$

即若 $g_i(\boldsymbol{x}) > g_j(\boldsymbol{x})$， $\forall j \neq i$，则将 $\boldsymbol{x}$ 归类为 ω_i 类。
我们将这种情况下的分类器称为“**线性分类机**”。

- 线性分类机将特征空间划分为 c 个决策区域，并且若 $\boldsymbol{x}$ 处在区域R_i中，$g_i(x)$ 的判决值取得最大值。 对于两个邻近的区域 R_i 和 R_j; 其划分边界是超平面H_{ij}的一部分，其定义为

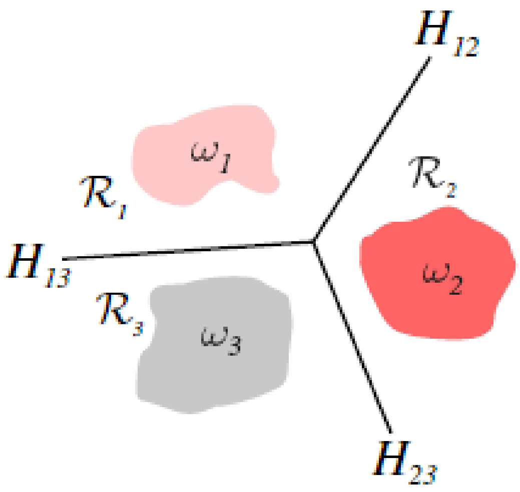
$$g_i(\boldsymbol{x}) = g_j(\boldsymbol{x})$$

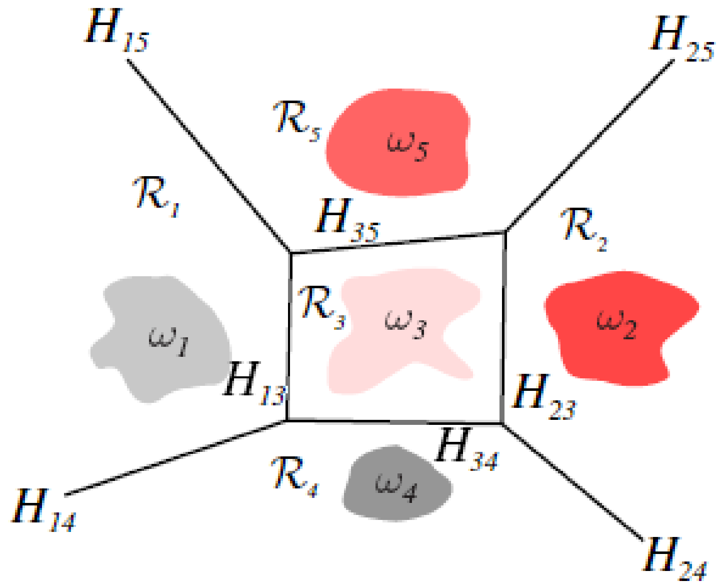
$$\Leftrightarrow \quad (\boldsymbol{w}_i - \boldsymbol{w}_j)^t\boldsymbol{x} + (w_{i0} - w_{j0}) = 0$$

$$\Leftrightarrow \quad \boldsymbol{w}_i - \boldsymbol{w}_j \text{ 是 } \boldsymbol{H}_{ij} \text{ 的法线, 且} \qquad d(\boldsymbol{x}, H_{ij}) = \frac{g_i - g_j}{\|\boldsymbol{w}_i - \boldsymbol{w}_j\|}$$

线性分类机的决策边界

- 下图是对于3类和5类的问题，由学习分类机产生的决策边界





线性分类机的特点

- 易证：

  线性分类机的决策区域是凸的，

  这一约束限制了分类器的适应性和准确性。

广义线性判决函数

- 线性判别函数 ***g*(*x*)** 可以写为:

$$g(x)=w_0+\sum_{i=0}^{d}w_i x_i$$

其中，***d*** 是特征空间的维数.

- 可以通过附加的项，获得二次型的判别函数:

$$g(x)=w_0+\sum_{i=0}^{d}w_i x_i+\sum_{i=0}^{d}\sum_{j=0}^{d}w_{ij}x_i x_j$$

- 二次判别函数通过分量之间的乘积，可以引入 ***d*(*d*-1)/2** 个系数。可以产生更复杂的曲面（二次曲面）。

广义线性判别函数

广义判别函数一般形式为：

$$g(x)=\sum_{i=0}^{d'} a_i y_i(x) \qquad \text{或} \qquad g(x)=\mathrm{a}^t\mathrm{y}$$

上述是所有 $\boldsymbol{y_i(x)}$ 函数之加权和。 $\boldsymbol{y_i(x)}$ 函数也称为$\boldsymbol{\varphi}$函数。

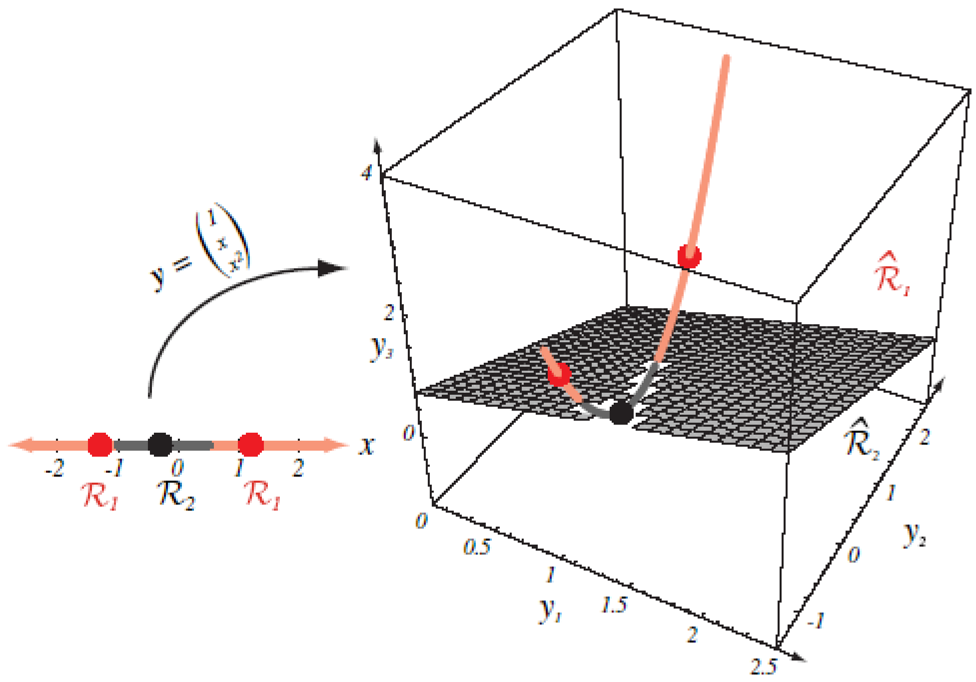
可以看出，上面的和式并不是 $\boldsymbol{x}$ 的线性函数，但却是 $\boldsymbol{y_i(x)}$ 的线性函数.

$\boldsymbol{y_i(x)}$ 函数的作用是将 $\boldsymbol{d}$-维 $\boldsymbol{x}$-空间映射到 $\boldsymbol{d'}$维的 $\boldsymbol{y}$-空间。

例如: $\boldsymbol{g(x)=a_1+a_2x+a_3x^2}$ $\qquad$ $\boldsymbol{y=(1 \quad x \quad x^2)^t}$

特征空间的非线性映射：Figure5.5

- 映射 $\mathbf{y} = (1, \mathbf{x}, \mathbf{x}^2)^t$ 将一条直线变换为3维空间中的抛物线。
- 图中，用一个平面将映射得到的 $\mathbf{y}$-空间分隔为相应的两类。而这在原始的1维 $\mathbf{x}$-空间中，无法得到单连通的决策区域



对特征空间映射的几点说明

从 $\boldsymbol{x}$-空间到 $\boldsymbol{y}$-空间的映射:

- 若 $\boldsymbol{x}$ 服从某一个确定的概率分布，则对应的分布在新的空间中一般是退化的。
  - 当新空间的维数高于原空间时，原分布在新空间中除对应曲线或曲面上的点外，其他地方均为0。
- 即使在 $\boldsymbol{y}$ 空间中简单的判决函数，在 $\boldsymbol{x}$ 空间中一般也是相当复杂的。
  - 空间的维数越高，其自由度越大（可描述的参数越多）。这样，就需要更多的学习样本。

特征空间的高维映射： Figure5.6

- 一个2维输入空间 $\boldsymbol{x}$ 通过多项式 f 映射为空间 $\boldsymbol{y}$。

这里，$(y_1, y_2, y_3) = (x_1, x_2, x_1x_2)$。

- 在变换空间中，线性判别式是一个分隔曲面的超平面。超平面正侧的点对应于类 ω_1，而下面的点对应于类 ω_2。
- 而在空间 $\boldsymbol{x}$ 中， R_1不是单连通的。