第4章 线性判别函数

- 线性判别函数基本概念
- Fisher线性判别
- 感知准则函数
- 最小平方误差准则函数
- 多类问题简介

4.1 引言

- 问题的提出

$\because P(\omega_i)$和$p(x \mid \omega_i)$估计困难

$\therefore$ 考虑直接方法，利用样本直接设计分类器

线性决策面是较简单的，易于实现，因此本章主要讨论线性分类器。

4.1 引言

- 判别函数形式

$$g(x) = \omega^T x + \omega_0$$

$$g_i(x) = \omega_i^T x + \omega_{i0} \qquad i = 1, 2, \cdots, c$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \qquad \omega = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_d \end{bmatrix}$$

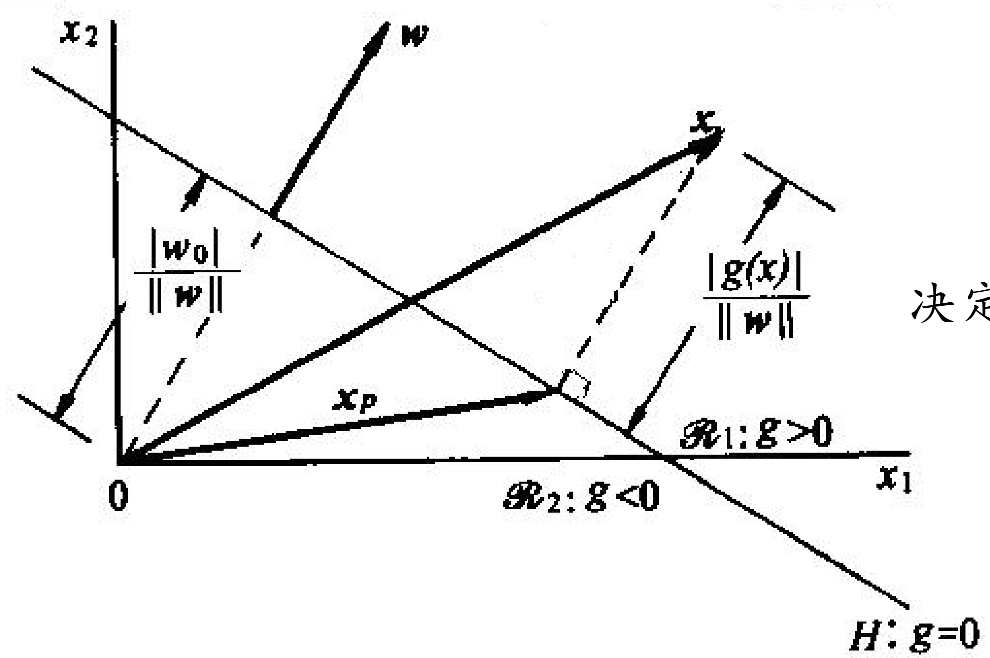
$$g_j(x) = \max g_i(x), i = 1, \cdots, c \quad x \in \omega_j$$

4.1 引言

- 两类情况

$$g(x) = g_1(x) - g_2(x)$$

$$\begin{cases} g(x) > 0 & x \in \omega_1 \\ g(x) < 0 & x \in \omega_2 \\ g(x) = 0 & \text{任意分至某一类或拒绝} \end{cases}$$



决定分类界面的方向，ω_0决定分类界面的位置

4.1 引言

- 线性分类器的设计步骤

（1）有一已知类别的样本集H

（2）确定一准则函数 J

$$\begin{cases} J\text{是}H,\omega,\omega_0\text{的函数} \\ \text{极值解对应最好的“决策”}—\omega^*,\omega_0^* \end{cases}$$

（3）利用最优化技术求出准则函数极值解 ω^*,ω_0^*

4.1 引言

- 与最优分类器的关系

　　贝叶斯分类器是在错误率或风险下为最优的分类器。线性分类器针对错误率或风险是“次优”的。但对于所采用的准则函数$J(\omega)$则是最优的。

- 线性可分性

　　已知一样本集，如果有一个线性分类器能把每个样本正确分类，则称这组样本集为线性可分的，否则称为线性不可分的。反之，如果样本集是线性可分的，则必然存在一个线性分类器能把每个样本正确分类。

4.2 Fisher线性判别

由1936年的经典论文开始最早研究线性判别函数，发明了最大似然估计方法。

共著作有

《研究者用的统计方法》

《统计方法和科学推理》

《近交的理论》

《试验设计》

《自然选择的遗传理论》

《根据孟德尔遗传方式的亲属间的相关》

R. A. Fisher爵士,1890.2.17-1962.7.29,英国

现代统计学与现代演化论的奠基者之一

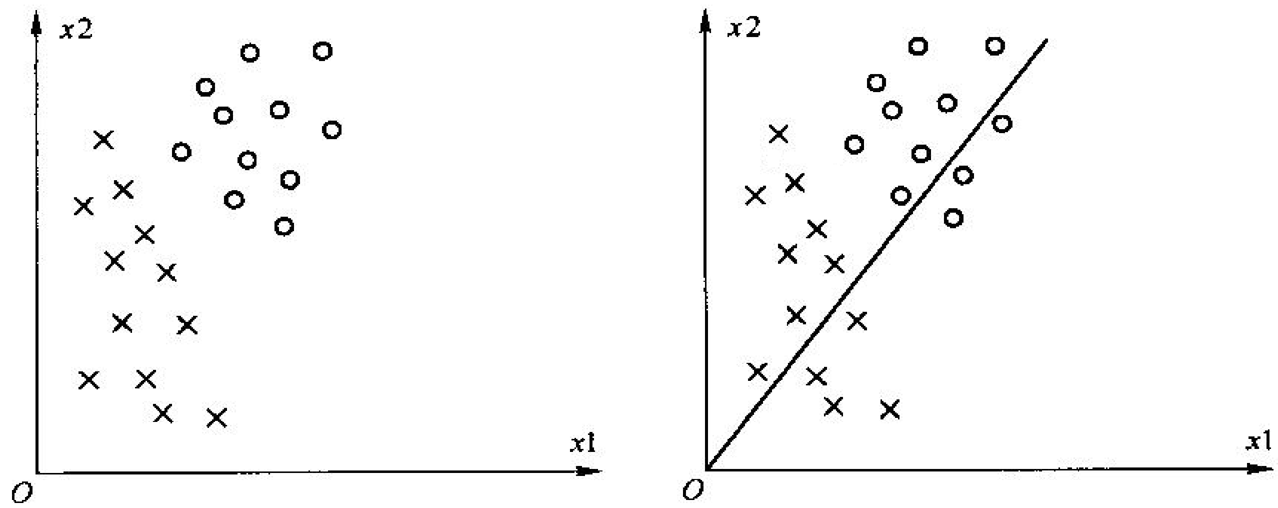
4.2 Fisher线性判别（两类）

- ## 问题的提出

在低维空间中行得通的方法在高维空间里往往失效，因此需要降低空间维数。

- ## 基本原理

将高维空间数据向某一直线上投影，如何确定该直线在特征空间中的最佳方向？



Fisher法即寻找最佳的投影方向ω，使投影后样本最佳可分

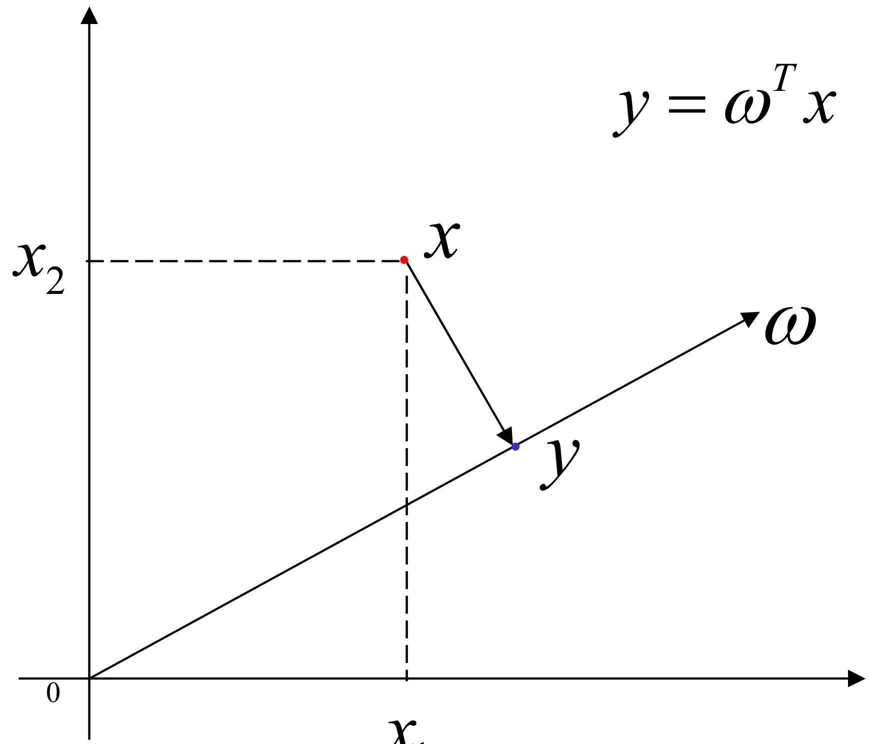
4.2 Fisher线性判别（两类）

- 投影的表示

$$y = \omega^T x$$

$$y = \omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_d x_d$$

$$\omega = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_d \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$



投影示意图

4.2 Fisher线性判别（两类）

- 基本参量定义（d维x空间）

（1）各类样本均值向量

$$m_i = \frac{1}{N_i} \sum_{x \in \omega_i} x \qquad i = 1,2$$

（2）样本类内离散度矩阵S_i和总类内离散度矩阵 S_ω

$$S_i = \sum_{x \in \omega_i} (x - m_i)(x - m_i)^T \qquad i = 1,2$$

$$S_\omega = S_1 + S_2$$ 对称半正定阵，当N>d时通常非奇异

（3）样本类间离散度矩阵

$$S_b = (m_1 - m_2)(m_1 - m_2)^T$$ 对称半正定阵

4.2 Fisher线性判别（两类）

- 补充定义—正定二次型

设有实二次型 $f = X^T AX$，若对任意向量X都有f > 0，则称矩阵A为正定矩阵；若f < 0，则称A为负定矩阵；若f≧0，则称A为半正定矩阵。

- 补充定义—非奇异阵

当|A|=0时，方阵A为奇异阵；|A| ≠0时，方阵A为非奇异阵。方阵A可逆的充要条件是|A| ≠0。

可逆阵亦称非退化阵或非奇异阵或满秩阵

不可逆的方阵也称退化阵或奇异阵或降秩阵

4.2 Fisher线性判别（两类）

- 基本参量定义（一维y空间）

（1）各类样本均值向量

$$\tilde{m}_i = \frac{1}{N_i}\sum_{y\in\omega_i} y \qquad i = 1,2$$

（2）样本类内离散度矩阵$\tilde{S}_i^2$和总类内离散度矩阵 $\tilde{S}_\omega$

$$\tilde{S}_i^2 = \sum_{y\in\omega_i}(y-\tilde{m}_i)^2 \qquad i = 1,2$$

$$\tilde{S}_\omega = \tilde{S}_1^2 + \tilde{S}_2^2$$

4.2 Fisher线性判别（两类）

- 定义Fisher准则函数

$$J_F(\omega)=\frac{(\tilde{m}_1-\tilde{m}_2)^2}{\tilde{S}_1^2+\tilde{S}_2^2}\qquad \frac{\text{越大越好}}{\text{越小越好}}$$

原则：各类样本尽可能分开，类内样本尽量密集，需求令准则函数极大的ω^*。

4.2 Fisher线性判别（两类）

- 准则函数的显式化

$$(\tilde{m}_1-\tilde{m}_2)^2=(\omega^T m_1-\omega^T m_2)^2=\omega^T(m_1-m_2)(m_1-m_2)^T\omega=\omega^T S_b\omega$$

$$\because \tilde{S}_i^2=\sum_{x\in\omega_i}(y-\tilde{m}_i)^2=\sum_{x\in\omega_i}(\omega^T x-\omega^T m_i)^2=\omega^T\left[\sum_{x\in\omega_i}(x-m_i)(x-m_i)^T\right]\omega=\omega^T S_i\omega$$

$$\therefore \tilde{S}_1^2+\tilde{S}_2^2=\omega^T(S_1+S_2)\omega=\omega^T S_\omega\omega$$

$$\therefore J_F(\omega)=\frac{\omega^T S_b\omega}{\omega^T S_\omega\omega}$$

求令准则函数极大的 ω^*

4.2 Fisher线性判别（两类）

利用$Lagrange$乘子法

求令准则函数极大的ω^*

假设$\omega^T S_\omega \omega = C \neq 0$

$$L(\omega,\lambda) = \omega^T S_b \omega - \lambda(\omega^T S_\omega \omega - C)$$

$$\frac{\partial L(\omega,\lambda)}{\partial \omega} = S_b \omega - \lambda S_\omega \omega = 0$$ 二次型对向量求导

$$S_b \omega^* = \lambda S_\omega \omega^*$$

$\because S_\omega$非奇异

$$\therefore S_\omega^{-1} S_b \omega^* = \lambda \omega^*$$

$$\because S_b = (m_1 - m_2)(m_1 - m_2)^T$$

$$\therefore \lambda \omega^* = S_\omega^{-1}(m_1 - m_2)\boxed{(m_1 - m_2)^T \omega^*}$$ 内积为常量，以R替代

$$\therefore \omega^* = \frac{R}{\lambda} S_\omega^{-1}(m_1 - m_2)$$ 要求的是方向，常量系数可去除

4.2 Fisher线性判别（两类）

- 概念补充——二次型对向量求导

$$x^T A x = [x_1 \quad \cdots \quad x_d]\begin{bmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{d1} & \cdots & a_{dd} \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} = \left[\sum_{\alpha=1}^{d} a_{\alpha 1} x_1 \quad \cdots \quad \sum_{\alpha=1}^{d} a_{\alpha d}\, x_d\right]\begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$

$$= \sum_{\beta=1}^{d}\sum_{\alpha=1}^{d} a_{\alpha\beta} x_\alpha x_\beta$$

$$\frac{\partial}{\partial x_i}\left[x^T A x\right] = \frac{\partial}{\partial x_i}\left|\sum_{\beta=1}^{d}\sum_{\alpha=1}^{d} a_{\alpha\beta} x_\alpha x_\beta\right| = 2a_{ii}x_i + \sum_{j\neq i}(a_{ij} + a_{ji})x_j$$

$$\therefore \frac{\partial}{\partial x}\left[x^T A x\right] = (A + A^T)x \qquad \text{当} A = A^T, \frac{\partial}{\partial x}\left[x^T A x\right] = 2Ax$$

4.2 Fisher线性判别（两类）

- 投影过程 $y = \omega^{*T} x$
- 一维阈值的选取

（1）当维数与样本数都很大时，可用贝叶斯分类器

（2）依据先验知识选择

$$y_0^{(1)} = \frac{\tilde{m}_1 + \tilde{m}_2}{2}$$

$$y_0^{(2)} = \frac{N_1 \tilde{m}_1 + N_2 \tilde{m}_2}{N_1 + N_2}$$

$$y_0^{(3)} = \frac{\tilde{m}_1 + \tilde{m}_2}{2} + \frac{\ln[P(\omega_1)/P(\omega_2)]}{N_1 + N_2 - 2}$$

$$y \mathop{\lessgtr}^{>}_{<} y_0 \rightarrow x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$