4.8 讨论

线性判别函数是形式最简单的判别函数，但因为它具有计算简单、在一定条件下能够实现最优分类的性质，因此在实际中得到了广泛的应用。很多情况下，虽然所研究的问题并不是线性可分的，但因为受到维数高、样本数有限等实际情况的限制，仍经常使用线性分类器。这不但是在容许一定的错误率时的一种"有限合理性"的选择，而且实际上，在这些条件下，即使采用更复杂的分类器设计方法也往往并不能得到更好的结果，有时甚至结果更差。

在后面两章我们将要看到，在线性分类器基础上，用分段线性分类器可以实现更复杂的分类面。而本章介绍的感知器准则下的线性分类器，实际上就是人工神经网络中最典型的多层感知器的原型(参见第 11 章)。在"4.1.2 小节"中提出的广义线性判别函数，虽然在这里因为维数灾难问题没有得到进一步推广，但在第 13 章将看到，在统计学习理论中从全新的角度对广义线性分类器进行了研究，巧妙地克服了维数灾难问题，进而发展出了最新的模式识别方法——支持向量机，成为解决有限样本情况下非线性分类问题的有效手段。

习　　题

4.1 (1) 指出从 x 到超平面 $g(\boldsymbol{x})=\boldsymbol{w}^T\boldsymbol{x}+w_0=0$ 的距离

$$r = \frac{|g(\boldsymbol{x})|}{\|\boldsymbol{w}\|}$$

是在 $g(\boldsymbol{x}_q)=0$ 的约束条件下，使 $\|\boldsymbol{x}-\boldsymbol{x}_q\|^2$ 达到极小的解；

(2) 指出在超平面上的投影是

$$\boldsymbol{x}_p = \boldsymbol{x} - \frac{g(\boldsymbol{x})}{\|\boldsymbol{w}\|^2}\boldsymbol{w}$$

4.2 设有一维空间二次判别函数

$$g(x) = 5 + 7x + 9x^2$$

(1) 试映射成广义齐次线性判别函数；

(2) 总结把高次函数映射成齐次线性函数的方法。

4.3 (1)通过映射，把一维二次判别函数

$$g(x) = a_1 + a_2x + a_3x^2$$

映射成三维广义齐次线性判别函数；

(2) 若 $\boldsymbol{x}$ 在一维空间具有分布密度 $p(x)$，说明三维空间中的分布密度退化成只在一条曲线上有值，且曲线上值无穷大。

4.4 对于二维线性判别函数

$$g(\boldsymbol{x}) = x_1 + 2x_2 - 2$$

(1) 将判别函数写成 $g(\boldsymbol{x})=\boldsymbol{w}^T\boldsymbol{x}+w_0$ 的形式，并画出 $g(\boldsymbol{x})=0$ 的几何图形；

(2) 映射成广义齐次线性判别函数

$$g(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{y}$$

(3) 指出上述 X 空间实际是 Y 空间的一个子空间，且 $\boldsymbol{a}^T\boldsymbol{y}=0$ 对 X 子空间的划分与原空间中 $\boldsymbol{w}^T\boldsymbol{x}+w_0=0$ 对原 X 空间的划分相同，并在图上表示出来。

4.5 指出在 Fisher 线性判别中，$\boldsymbol{w}$ 的比例因子对 Fisher 判别结果无影响的原因。

4.6 证明两向量外积组成的矩阵一般是奇异的。

4.7 已知 Fisher 准则函数为

$$J(\boldsymbol{w})=\frac{\boldsymbol{w}^T S_B \boldsymbol{w}}{\boldsymbol{w}^T S_W \boldsymbol{w}}$$

(1) 说明是否一般的分数函数都可以用 Lagrange 乘子法求极值解；

(2) 分析 $J(\boldsymbol{w})$ 可用 Lagrange 乘子法求解的条件。

4.8 证明在正态等协差条件下，Fisher 线性判别等价于贝叶斯判别。

4.9 证明

(1)引入余量 b 以后的解区($\boldsymbol{a}^T\boldsymbol{y}_i\geqslant b$)位于原来的解区($\boldsymbol{a}^T\boldsymbol{y}_i>0$)之中；

(2)与原解区边界之间的距离为 $b/\|\boldsymbol{y}_i\|$。

4.10 证明，在几何上，感知准则函数正比于被错分类样本到决策面的距离之和。

4.11 分析在可变增量法中，余量 b 如何选择较为合理。

4.12 写出 Widrow-Hoff 法程序框图。

4.13 (1)证明矩阵恒等式

$$(A+\boldsymbol{x}\boldsymbol{x}^T)^{-1}=A^{-1}-\frac{A^{-1}\boldsymbol{x}\boldsymbol{x}^T A^{-1}}{1+\boldsymbol{x}^T A^{-1}\boldsymbol{x}}$$

(2)利用上式结果证明式(4-98)。

4.14 考虑准则函数

$$J(\boldsymbol{a})=\sum_{\boldsymbol{y}\in\mathscr{Y}(\boldsymbol{a})}(\boldsymbol{a}^T\boldsymbol{y}-b)^2$$

其中 $\mathscr{Y}(\boldsymbol{a})$ 是使 $\boldsymbol{a}^T\boldsymbol{y}\leqslant b$ 的样本集合。设 $\boldsymbol{y}_1$ 是 $\mathscr{Y}(\boldsymbol{a}_k)$ 中的唯一样本，则 $J(\boldsymbol{a})$ 的梯度为 $\nabla J(\boldsymbol{a}_k)=2(\boldsymbol{a}_k^T\boldsymbol{y}_1-b)\boldsymbol{y}_1$，二阶偏导数矩阵为 $D=2\boldsymbol{y}_1\boldsymbol{y}_1^T$。

据此证明，若最优步长选择为

$$\rho_k=\frac{\|\nabla J(\boldsymbol{a})\|^2}{\nabla J^T(\boldsymbol{a})D\nabla J(\boldsymbol{a})}$$

时，梯度下降法的迭代公式为

$$\boldsymbol{a}_{k+1}=\boldsymbol{a}_k+\frac{b-\boldsymbol{a}_k^T\boldsymbol{y}_1}{\|\boldsymbol{y}_1\|^2}\boldsymbol{y}_1$$

4.15 证明：当取

$$b=\Big[\underbrace{\frac{N}{N_1},\cdots,\frac{N}{N_1}}_{N_1\text{个}},\underbrace{\frac{N}{N_2},\cdots,\frac{N}{N_2}}_{N_2\text{个}}\Big]^T$$

时，MSE 解等价于 Fisher 解。

4.16 证明：

(1) 式(4-113)表示的向量 $\boldsymbol{y}-\frac{\boldsymbol{a}^T\boldsymbol{y}}{\|\boldsymbol{w}\|^2}\begin{bmatrix}0\\ \boldsymbol{w}\end{bmatrix}$ 表示 $\boldsymbol{y}$ 到 X 空间中超平面上的投影。

(2) 该投影正交于 X 空间的超平面。

4.17 在多类问题中，如果一组样本可被一线性机器全部正确分类，则称这组样本是线性可分的。对任意 ω_i 类，如果能用一个超平面把标以 ω_i 类的样本同所有其他样本分开，则称这组样本为总体线性可分的。举例说明，总体线性可分必定线性可分，但反之不然。

4.18 设有一组样本。若存在 $c(c-1)/2$ 个超平面 H_{ij}，使 H_{ij} 能把属于 ω_i 类的样本同属于 ω_j 类的样本分开，则称这组样本是成对线性可分的。举例说明，成对线性可分的样本不一定是线性可分的。