

模式识别

上机实验指导书

二零一七年四月

目 录

实验课概况.....	1
实验一、Bayes 分类器设计.....	2
实验二、基于 Fisher 准则线性分类器设计.....	4
实验三、基于感知函数准则线性分类器设计.....	7
实验四、近邻法分类器设计.....	9

实验课概况

课程名称：模式识别

实验学时：12

壹、实验的性质、任务和基本要求

（壹）实验课的性质

《模式识别》实验课是一门非独立的实验课，是同学对模式识别理论内容进行充分的理解的基础上，根据相应的原理，设计实验内容，完成实验任务，是理论知识实践化的方式，利于学生更好的吸收，领悟模式识别的原理与应用，培养学生的动手实践的能力。

（貳）实验课的基本要求

- 1、理解模式识别的基本概念
- 2、掌握各种算法的流程，以及相应的优缺点。
- 3、会使用相应的模式识别分类器等算法处理实验问题。

式、实验的分配情况

序号	实验内容	学时	选作
1	Bayes 分类器算法	3	四选二
2	Fisher 线性分类器设计	3	
3	感知器设计	3	
4	近邻法	3	

实验一、Bayes 分类器设计

1.1 实验类型：

基础型：Bayes 分类器设计

1.2 实验目的：

本实验旨在让同学对模式识别有一个初步的理解，能够根据自己的设计对贝叶斯决策理论算法有一个深刻地认识，理解二类分类器的设计原理。

1.3 实验原理：

最小风险贝叶斯决策可按下列步骤进行：

(1) 在已知 $P(\omega_i)$, $P(X|\omega_i)$, $i=1, \dots, c$ 及给出待识别的 X 的情况下，根据贝叶斯公式计算出后验概率：

$$P(\omega_i|X) = \frac{P(X|\omega_i)P(\omega_i)}{\sum_{j=1}^c P(X|\omega_j)P(\omega_j)} \quad j=1, \dots, c$$

(2) 利用计算出的后验概率及决策表，按下面的公式计算出采取 $a_i, i=1, \dots, a$ 的条件风险

$$R(a_i|X) = \sum_{j=1}^c \lambda(a_i, \omega_j) P(\omega_j|X), i=1, 2, \dots, a$$

(3)对(2)中得到的a个条件风险值 $R(a_i|X)$, $i=1,\cdots,a$ 进行比较,找出使其条件风险最

小的决策 a_k ,则 a_k 就是最小风险贝叶斯决策。

1.4 实验内容：

假定某个局部区域细胞识别中正常(ω_1)和非正常(ω_2)两类先验概率分别为

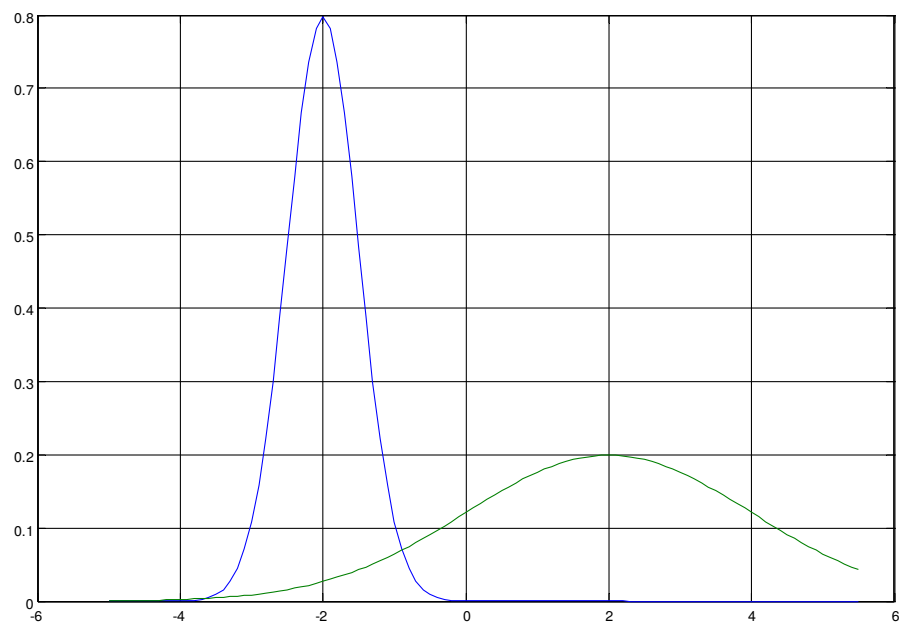
正常状态: $P(\omega_1)=0.9$;

异常状态: $P(\omega_2)=0.1$ 。

现有一系列待观察的细胞,其观察值为 x :

-3.9847	-3.5549	-1.2401	-0.9780	-0.7932	-2.8531
-2.7605	-3.7287	-3.5414	-2.2692	-3.4549	-3.0752
-3.9934	2.8792	-0.9780	0.7932	1.1882	3.0682
-1.5799	-1.4885	-0.7431	-0.4221	-1.1186	4.2532

$p(x|\omega_1)$ 和 $p(x|\omega_2)$ 类条件概率分布正态分布分别为 $(-2, 0.25)$ 和 $(2,4)$,试对观察的结果进行分类。



1.5 实验要求：

- 1) 完成分类器的设计，要求程序相应语句有说明文字。
- 2) 如果是最小风险贝叶斯决策，决策表如下：

最小风险贝叶斯决策表：

状态 决策	ω_1	ω_2
α_1	0	6
α_2	1	0

请重新设计程序，画出相应的后验概率的分布曲线和分类结果,并比较两个结果

实验二、基于 **Fisher** 准则线性分类器设计

2.1 实验类型：

线性分类器设计 (Fisher 准则)

2.2 实验目的：

本实验旨在让同学进一步了解分类器的设计概念，能够根据自己的设计对线性分类器有更深刻地认识，理解 Fisher 准则方法确定最佳线性分界面方法的原理。

2.3 实验原理：

线性判别函数的一般形式可表示成

$$g(X) = W^T X + w_0 \quad \text{其中}$$

$$X = \begin{pmatrix} x_1 \\ \cdots \\ x_d \end{pmatrix} \quad W = \begin{pmatrix} w_1 \\ w_2 \\ \cdots \\ w_d \end{pmatrix}$$

根据 Fisher 选择投影方向 W 的原则，即使原样本向量在该方向上的投影能兼顾类间分布尽可能分开，类内样本投影尽可能密集的要求，用以评价投影方向 W 的函数为：

$$J_F(W) = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{\tilde{S}_1^2 + \tilde{S}_2^2}$$

$$W^* = S_W^{-1}(m_1 - m_2)$$

上面的公式是使用 Fisher 准则求最佳法线向量的解，该式比较重要。另外，该式这种形式的运算，我们称为线性变换，其中 $m_1 - m_2$ 式一个向量， S_W^{-1} 是 S_W 的逆矩阵，如 $m_1 - m_2$ 是 d 维， S_W 和 S_W^{-1} 都是 d×d 维，得到的 W^* 也是一个 d 维的向量。

向量 W^* 就是使 Fisher 准则函数 $J_F(W)$ 达极大值的解，也就是按 Fisher 准则将 d 维 X 空间投影到一维 Y 空间的最佳投影方向，该向量 W^* 的各分量值是对原 d 维特征向量求加权求和的权值。

以上讨论了线性判别函数加权向量 W 的确定方法，并讨论了使 Fisher 准则函数极大的 d 维向量 W^* 的计算方法，但是判别函数中的另一项 W_0 尚未确定，一般可采用以下几种方法确定 W_0 如

$$W_0 = -\frac{\tilde{m}_1 + \tilde{m}_2}{2}$$

或者
$$W_0 = -\frac{N_1 \tilde{m}_1 + N_2 \tilde{m}_2}{N_1 + N_2} = \tilde{m}$$

或当 $p(\omega)_1$ 与 $p(\omega)_2$ 已知时可用

$$W_0 = \left[\frac{\tilde{m}_1 + \tilde{m}_2}{2} - \frac{\ln[p(\omega_1) / p(\omega_2)]}{N_1 + N_2 - 2} \right]$$

.....

当 W_0 确定之后，则可按以下规则分类，

$$W^T X > -w_0 \rightarrow X \in \omega_1$$

$$W^T X < -w_0 \rightarrow X \in \omega_2$$

使用 Fisher 准则方法确定最佳线性分界面的方法是一个著名的方法，尽管提出该方法的时间比较早，仍见有人使用。

2.4 实验内容：

利用 Fisher 准则对自行建立的样本或应用下面数据求出投影变换向量。

假设已经获得两类二维的模式样本： $\omega_1: \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix} \right\}$ $\omega_2: \left\{ \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \begin{pmatrix} 6 \\ 4 \end{pmatrix}, \begin{pmatrix} 6 \\ 6 \end{pmatrix}, \begin{pmatrix} 4 \\ 6 \end{pmatrix} \right\}$
 两类均服从正态分布，且先验概率相等。试用 Fisher 准则求出投影变换向量（权向量）。

2.5 实验要求：

请把数据作为样本，根据 Fisher 选择投影方向 W 的原则，使原样本向量在该方向上的投影能兼顾类间分布尽可能分开，类内样本投影尽可能密集的要求，完成 Fisher 线性分类器的设计，程序的语句要求有注释。

实验三、基于感知函数准则线性分类器设计

3.1 实验类型：

线性分类器设计（感知函数准则）

3.2 实验目的：

本实验旨在让同学理解感知准则函数的原理，通过软件编程模拟线性分类器，理解感知函数准则的确定过程，掌握梯度下降算法求增广权向量，进一步深刻认识线性分类器。

3.3 实验原理：

感知准则函数是五十年代由 Rosenblatt 提出的一种自学习判别函数生成方法，由于 Rosenblatt 企图将其用于脑模型感知器，因此被称为感知准则函数。其特点是随意确定的判别函数初始值，在对样本分类训练过程中逐步修正直至最终确定。

感知准则函数利用梯度下降算法求增广权向量的做法，可简单叙述为：任意给定向量初始值 $\bar{a}(1)$ ，第 $k+1$ 次迭代时的权向量 $\bar{a}(k+1)$ 等于第 k 次的权向量 $\bar{a}(k)$ 加上被错分类的所有样本之和与 ρ_k 的乘积。可以证明，对于线性可分的样本集，经过有限次修正，一定可以找到一个解向量 \bar{a} ，即算法能在有限步内收敛。其收敛速度的快慢取决于初始权向量 $\bar{a}(1)$ 和系

数 ρ_k 。

3.4 实验内容

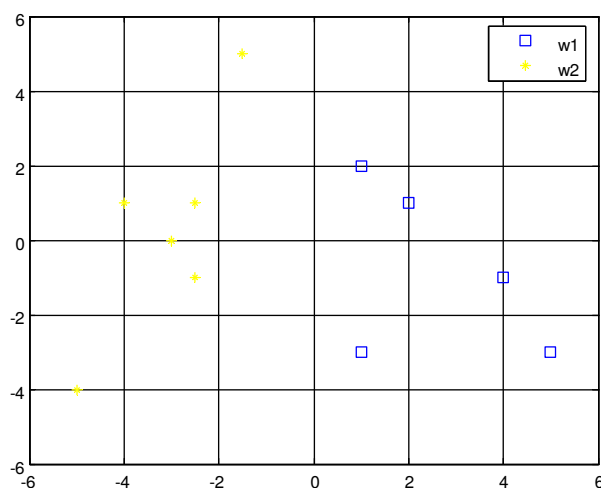
利用感知器算法对自行建立样本或应用下面数据求取权向量。

已知有两个样本空间 w1 和 w2，这些点对应的横纵坐标的分布情况是：

$x1=[1,2,4,1,5]; y1=[2,1,-1,-3,-3];$

$x2=[-2.5,-2.5,-1.5,-4,-5,-3]; y2=[1,-1,5,1,-4,0];$

在二维空间样本分布图形如下所示：



3.5 实验任务：

完成感知准则函数确定判决权向量的程序设计。

实验四、近邻法分类器设计

4.1 实验类型：

近邻法分类器设计

4.2 实验目的：

本实验旨在让同学理解近邻法的原理，通过软件编程分段线性分类器的极端情况，理解 k-近邻法邻的设计过程。

4.3 实验原理：

最近邻法可以扩展成找测试样本的 k 个最近样本作决策依据的方法。其基本规则是，在所有 N 个样本中找到与测试样本的 k 个最近邻者，其中各类别所占个数表示成 $k_i, i = 1, \dots, c$ 则决策规划是：

$$\text{如果 } k_j(X) = \max_i k_i(X), i = 1, \dots, c$$

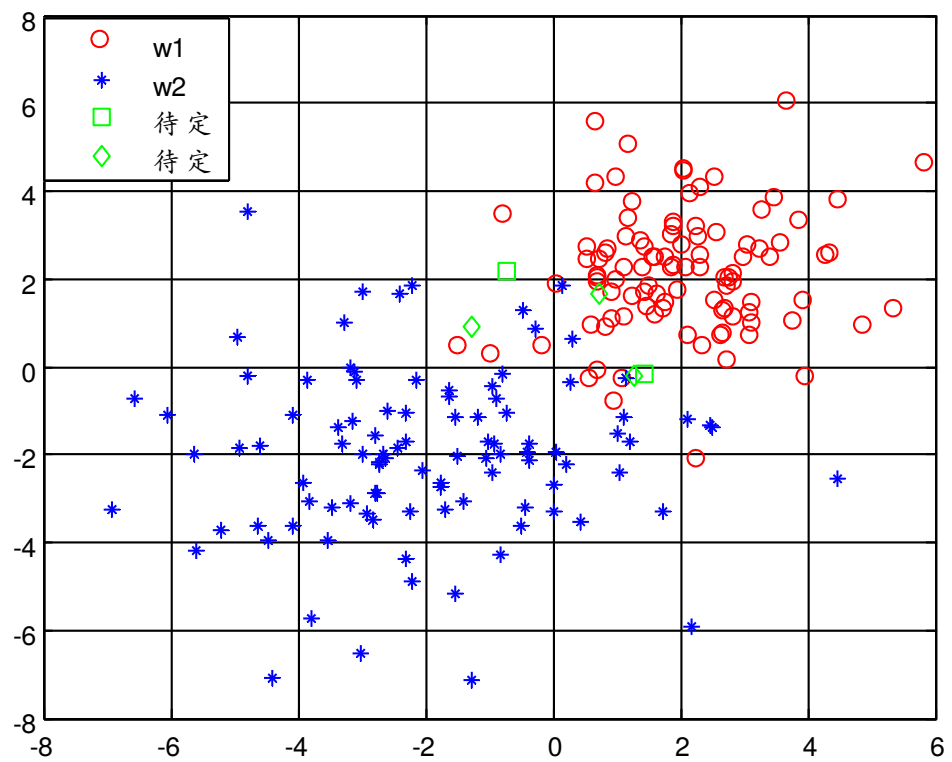
则决策 $X \in \omega_j$ (3-63)

k 近邻一般采用 k 为奇数，跟投票表决一样，避免因两种票数相等而难以决策。

4.4 实验内容

有两个类别，x，y，样本的分布规律服从正态分布，其均值和方差分别为 (2, 2)，(-2, 4)，每个类别里面分别有样本 100 个（请随机生成该两类样本，例如下图中的红色和蓝色所示）。使用 k-近邻法判断下列样本的分类情况（对应图中绿色所示）。

(-0.7303, 2.1624)，(1.4445, -0.1649)，(-1.2587, 0.9187)，(1.2617, -0.2086)，
(0.7302, 1.6587)



4.5 实验要求：

- 1、编写程序确定样本的分类情况。
- 2、分析 k 值的不同对分类的情况是否有影响。