Chapter 1

绪论

Chapter 2

聚类分析

2.1 距离聚类的概念

2.2 相似性测度和聚类准则

2.3 基于距离阈值的聚类算法

2.4 层次聚类法

2.5 动态聚类法

聚类过程中，聚类中心位置或个数发生变化.

两种常用的算法：

- K - 均值算法（或 C - 均值算法）
- 迭代自组织的数据分析算法

2.5.1 K - 均值算法

基于使聚类准则函数最小化。

准则函数 聚类集中每一样本点到该类中心的距离平方和。

对于第 j 个聚集类，准则函数定义为：

$$J_j = \sum_{i=1}^{N_j} ||X_i - Z_j||^2, X_i \in S_j$$

算法描述

算法讨论

聚类准则函数 J_k 与 K 的关系曲线

2.5.2 迭代自组织的数据分析算法

迭代自组织的数据分析算法也常称为 ISODATA 算法 (Iterate Selft-Organizing Data Analysis Techniques Algorithm, ISODATA).

算法特点

- 加入了试探性步骤，组成人机交互的结构；
- 可以通过类的自动合并与分裂得到较合理的类别数。

基本思路

1. 选择初始值
2. 按最邻近规则进行分类
3. 聚类后的处理：计算各类中的距离函数等指标
4. 判断结果是否符合要求，符合则结束，否则回到 2

算法描述 P31

常用指标 各指标综合考虑

1. 聚类中心之间的距离
2. 诸聚类域中样本数目
3. 诸聚类域中样本的标准差向量

Chapter 3

判别函数及几何分类法

3.1 判别函数

统计模式识别

$$\left\{\begin{array}{l} \text{聚类分析法（第二章）} \\ \text{判别函数法}\left\{\begin{array}{l}\left.\begin{array}{l}\text{线性判别函数法}\\ \text{非线性判别函数法}\end{array}\right. => xx \\ \text{统计决策方法}\end{array}\right.\end{array}\right.$$

3.1.1 判别函数

定义 直接用来对模式进行分类的准则函数。

3.2 线性判别函数

3.2.1 线性判别函数的一般形式

$$d(\mathbf{X}) = w_1x_1 + w_2x_2 + \ldots + w_nx_n + w_{n+1} = \mathbf{W}_0^T\mathbf{X}w_{n+1}$$

式中，$\mathbf{W}_0 = [w_1, w_2, \ldots, w_n]^T$ 称为权向量或参数向量；$\mathbf{X} = [x_1, x_2, \ldots, x_n]^T$ 是 n 维特征向量，又称模式向量或样本向量；w_{n+1} 是常数，称为阈值权。

3.2.2 线性判别函数的性质

两类情况

若已知两类模式 ω_1 和 ω_2

3.3 广义线性判别函数

3.4 线性判别函数的集合性质

3.5 感知器算法

3.5.1 前置概念

训练与学习

确定性分类器

感知器 感知器（preception）是一种神经网络模型，...

3.5.2 算法描述

规范化处理 将第二类样本全部乘以 (-1) 使得对于所有两类样本，判别函数的性质描述为

$$d(\mathbf{X}) = \mathbf{W}^T\mathbf{X} > 0 \tag{3.1}$$

算法步骤 P54

感知器算法就是一种赏罚过程：当分类器发生分类错误时，对分类器进行“罚”—修改权向量，使其向正确的方向转换；分类正确时，对其进行“赏”—这里表现为“不罚”，即权向量不变。

3.5.3 收敛性

经过算法的有限次迭代运算后，求出一个使训练集中所有样本都能正确分类的 W，则称算法是收敛的。可以证明感知器算法是收敛的。

3.5.4 感知器算法用于多类情况

转化成多个二类可分问题

对于 M 类模式应存在 M 个判别函数 ... P56

3.6 梯度法

3.6.1 梯度法基本原理

梯度 设函数 $f(\mathbf{Y})$ 是向量 $\mathbf{Y} = [y_1, y_2, \ldots, y_n]^T$ 的函数，则 $f(\mathbf{Y})$ 的梯度定义为：

$$\nabla f(\mathbf{Y}) = \frac{d}{d\mathbf{Y}} f(Y) = \left[\frac{\partial f}{\partial y_1}, \frac{\partial f}{\partial y_2}, \ldots, \frac{\partial f}{\partial y_n}\right]^T$$

梯度向量的最重要的性质之一：指出函数 f 在其自变量增加时，增长最快的方向。

梯度算法

P59

3.6.2 固定增量法

准则函数 $J(W,X)=\frac{1}{2}(|W^TX|-W^TX)$
该准则函数有唯一最小值 0

推导 ∇J

求 $W(k+1)$

$$W(k+1)=W(k)+\begin{cases}0, & W^T(k)X>0\\ cX, & W^T(k)X\le 0\end{cases}$$

3.7 最小平方误差算法

least Mean Square Error, LSME

原理 把对满足 $XW>0$ 的 q 求解，改为满足

$$XW=B \tag{3.2}$$

式中：
$B=[b_1,b_2,\dots,b_i,\dots,b_N]^T$ 是各分量均为正值的分量

矛盾方程组 在方程组中，当方程个数多于未知数时，通常没有精确解存在，称为矛盾方程组，一般求近似解。在模式识别中由于训练样本数 N 总是大于模式的维数 n，因此在式中方程的个数总是大于未知数 W 分量的个数，是矛盾方程组，只能求近似解。方法是求满足

$$||XW^*-B||=\text{极小} \tag{3.3}$$

的 W^*，称为最小二乘近似解，也称最优近似解。

准则函数 LMSE 算法的准则函数定义为：

$$J(W,X,B)=\frac{1}{2}||XW-B||^2 \tag{3.4}$$

LMSE 算法递推公式的推导

伪逆矩阵 $X^{=(X^TX)^{-1}X^T}$ 称为 X 的伪逆

模式类别可分性的判别

方法 每次迭代计算后检查一下 $XW(k)$ 的各分量和误差向量 $e(k)$，从而可以判断是否收敛。具体情况见 P66

3.7.1 算法特点

3.8 非线性判别函数

线性判别函数的特点 形式简单，容易学习；用于线性可分的模式类。

非线性判别函数 用于线性不可分情况，分段线性，超曲面

3.8.1 分段线性判别函数

特点

- 相对简单
- 能逼近各种形状的超曲面

一般分段线性判别函数

P70

用各类函数进行分类判决实际上是用各类选出的子类判别函数进行判决 => 判别面由子类的判别函数决定。

基于距离的分段线性判别函数

P71

3.8.2 分段线性判别函数的学习方法

已知子类划分时的学习方法

- 每个子类看成独立的类
- 应用线性可分多类情况 3 的学习算法

已知子类数目时的学习方法

未知子类数目时的学习方法

Chapter 4

基于统计决策的概率分类法

4.1 研究对象及相关概率

4.1.1 两类研究对象

- 确定性事件事物间有确定的因果关系。
- 随机事件事物间没有确定的因果关系，观察到的特征具有统计特性，是一个随机向量。

4.1.2 概率知识

略

4.2 贝叶斯决策

4.2.1 最小错误率贝叶斯决策

问题分析 在概率中，要看 X 来自于哪一类，使用后验概率 $P(\omega_i|\boldsymbol{X})$ 效果最好

利用贝叶斯公式将后验概率转化成易于得到的 $P(\boldsymbol{X}|\omega_i)$ 和 $P(\omega_i)$.

决策规则 设有 M 类模式，分类规则为:

$$\text{若} P(\omega_i|\boldsymbol{X}) = \max\{P(\omega_j|\boldsymbol{X}), \quad j = 1, 2, \ldots, M\} \tag{4.1}$$

上式称为最小错误率贝叶斯决策规则。

由于先验概率 $P(\omega_i)$ 和类概率函数 P85

对两类问题，该式相当于

4.2.2 最小风险贝叶斯决策

风险的概念

- 自动灭火系统
- 疾病判断

不同的错判造成的损失不同，因此疾病风险不同，两者紧密相连。

决策规则

P87 风险计算

决策规则:

$$\text{若} r_k(\boldsymbol{X}) = \min\{r_i(\boldsymbol{X}), i = 1, \ldots, M\}, \text{则} \boldsymbol{X} \in \omega_k \tag{4.2}$$

(0-1) 损失最小风险贝叶斯决策

4.2.3 正态分布模式的贝叶斯决策

正态分布概率模型的优点

- 物理上的合理性
- 数学上的简单性

正态分布

略

多变量正态随机向量

正态分布的最小错误率贝叶斯决策规则

当 $p(\boldsymbol{X}|\omega_i)$ 呈正态分布时，只需要知道 M 和 C 即可。

多类情况 具有 M 中模式类别的多元高斯分布函数:

两类问题 对应判别函数

不同判别面

4.3 贝叶斯分类器的错误率

4.3.1 错误率的概念

错误率 将应属于某一类的模式错分到其他类中的概率。是衡量分类器性能优劣的重要参数。

$$\text{定义为}P(e) = \int_{-\inf}^{\inf} P(e|\boldsymbol{X}) d\boldsymbol{X} \tag{4.3}$$

式中：$\boldsymbol{X} = [x_1, \ldots, x_n]^{\mathrm{T}}$; $p(e|\boldsymbol{X})$ 是 $\boldsymbol{X}$ 的条件概率

4.3.2 错误率分析

两类问题的错误率

两类的错分

错误率计算

正态分布贝叶斯决策的错误率计算

正态分布的对数似然比

4.3.3 错误率的估计

已设计好分类器时错误率的估计

1. 先验概率未知 — 随机抽样

2. 先验概率已知 — 选择性抽样

留一法

交叉法

4.4 聂曼-皮尔逊 (Neyman-Person) 决策

4.5 概率密度函数的参数估计

4.6 概率密度函数的非参数估计

4.7 后验概率密度函数分类的势函数方法