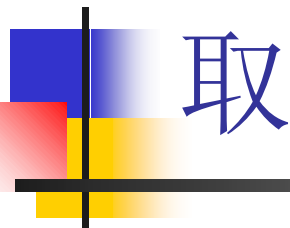


关于特征的选择与提取特征提取





学习指南

- 前面章节讨论的问题
 - 主要讨论模式识别的重要概念
 - 贝叶斯分类器、线性分类器、非线性分类器、KNN等
 - 设计分类器时的讨论问题
 - 在 d 维特征空间已经确定的前提下，讨论分类方法与分类器设计
 - 即：对要分类的样本怎样描述这个问题是已经确定的



学习指南

- 前面章节讨论的问题
 - 例如对苹果与梨的划分问题
 - 我们使用尺寸、重量和颜色三种度量来描述
 - 这种度量方法是已经确定好的
 - 在这种条件下研究用线性分类器好还是非线性分类器好，以及这些分类器的其它设计问题



学习指南

- 本章讨论的问题
 - 对已有的特征空间进行改造,着重于研究对样本究竟用什么样的度量方法更好
 - 譬如用三种度量来描述苹果与梨
 - 那么是否运用这三种度量是最有效的呢?
 - 颜色:
 - 这一个指标对区分红苹果与梨很有效
 - 区分黄苹果与梨就会困难得多
即, 这个指标就不很有效了



学习指南

- 本章讨论的问题
 - 使用什么样的特征描述事物，也就是说使用什么样的特征空间是个很重要的问题
 - 由于对特征空间进行改造目的在于提高其某方面的性能，因此又称特征的优化问题



学习指南

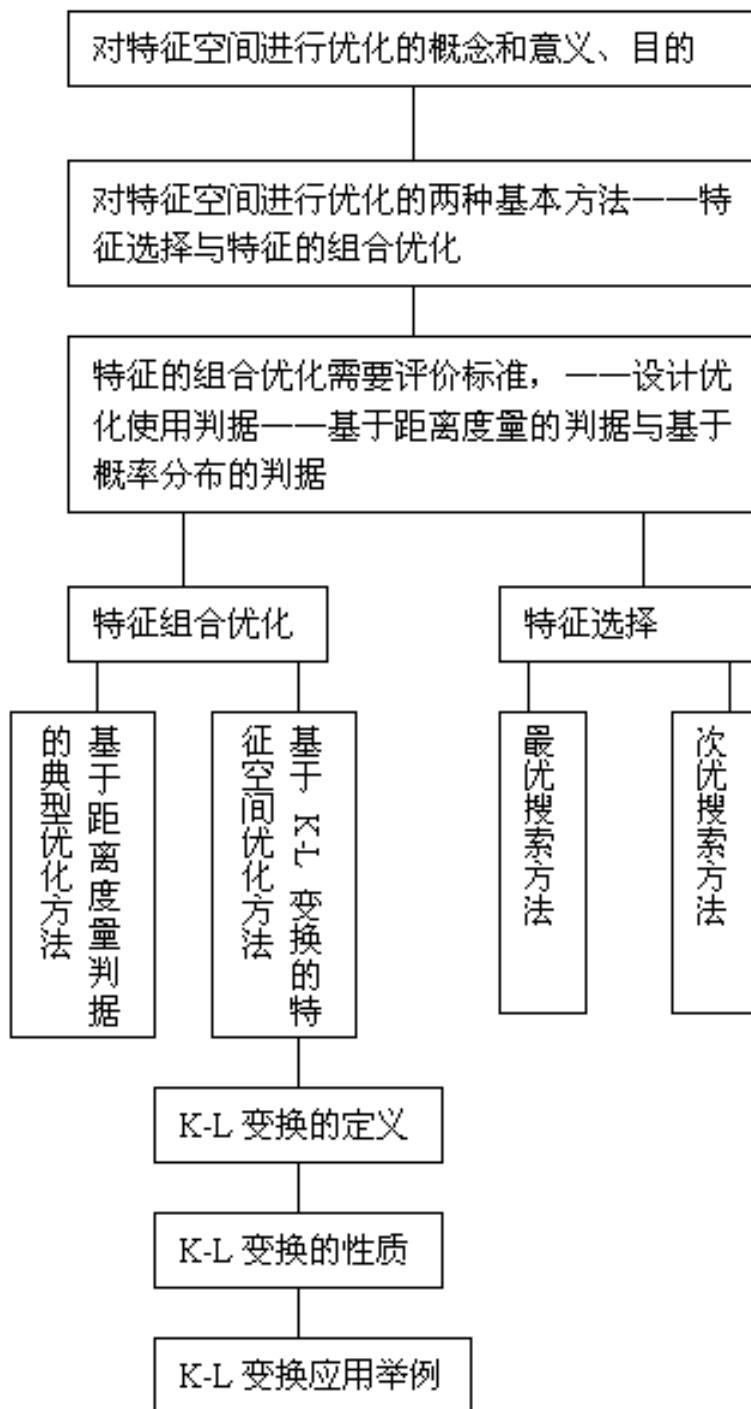
- 降维主要有两种途径
 - 对特征空间的改造、优化、主要的目的是降维，即把维数高的特征空间改成维数低的特征空间，降维主要有两种途径
 - 特征的选择：一种是删选掉一些次要的特征
 - 问题在于如何确定特征的重要性，以及如何删选
 - 特征的提取：另一种方法是使用变换的手段，在这里主要限定在线性变换的方法上，通过变换来实现降维



本章重点

- 1. 弄清对特征空间进行优化的含义
- 2. 对特征空间进行优化的两种基本方法——特征选择与特征的组合优化
- 3. 对特征空间进行优化的一些常用判据
- 4. 利用线段变换进行特征空间优化的基本方法，

本章知识点

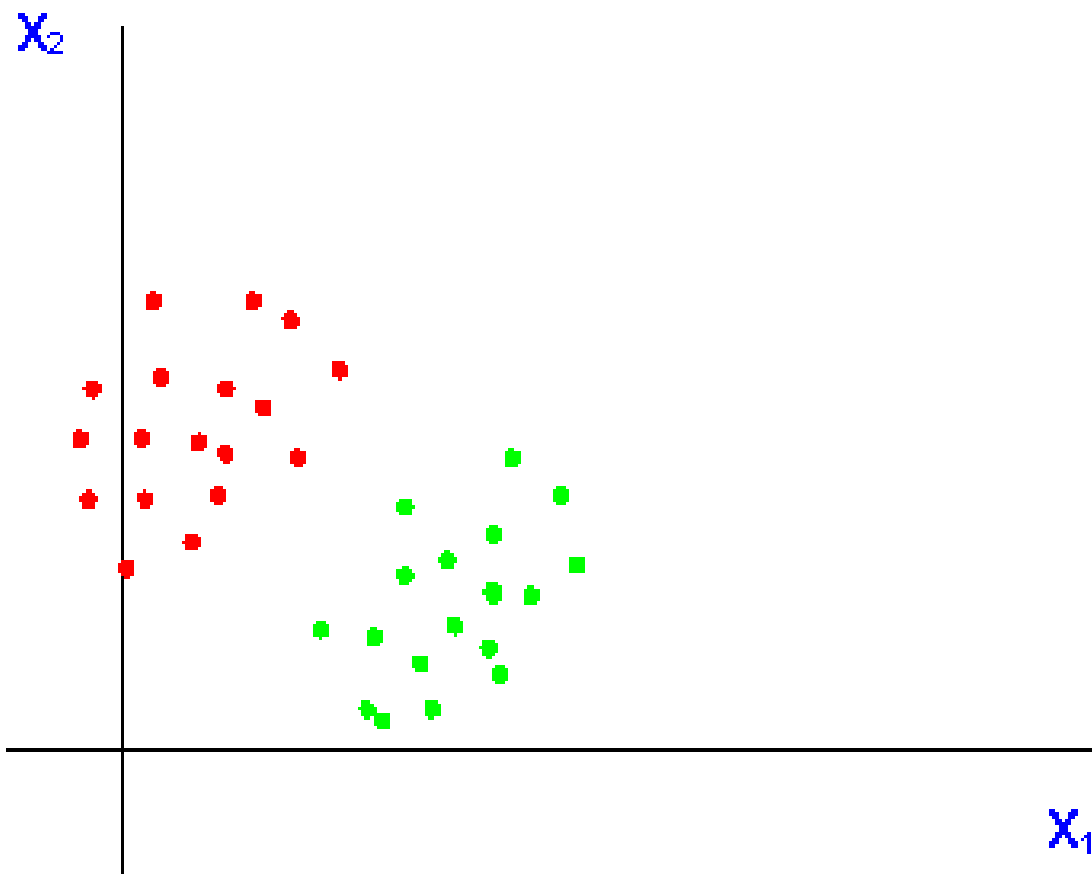




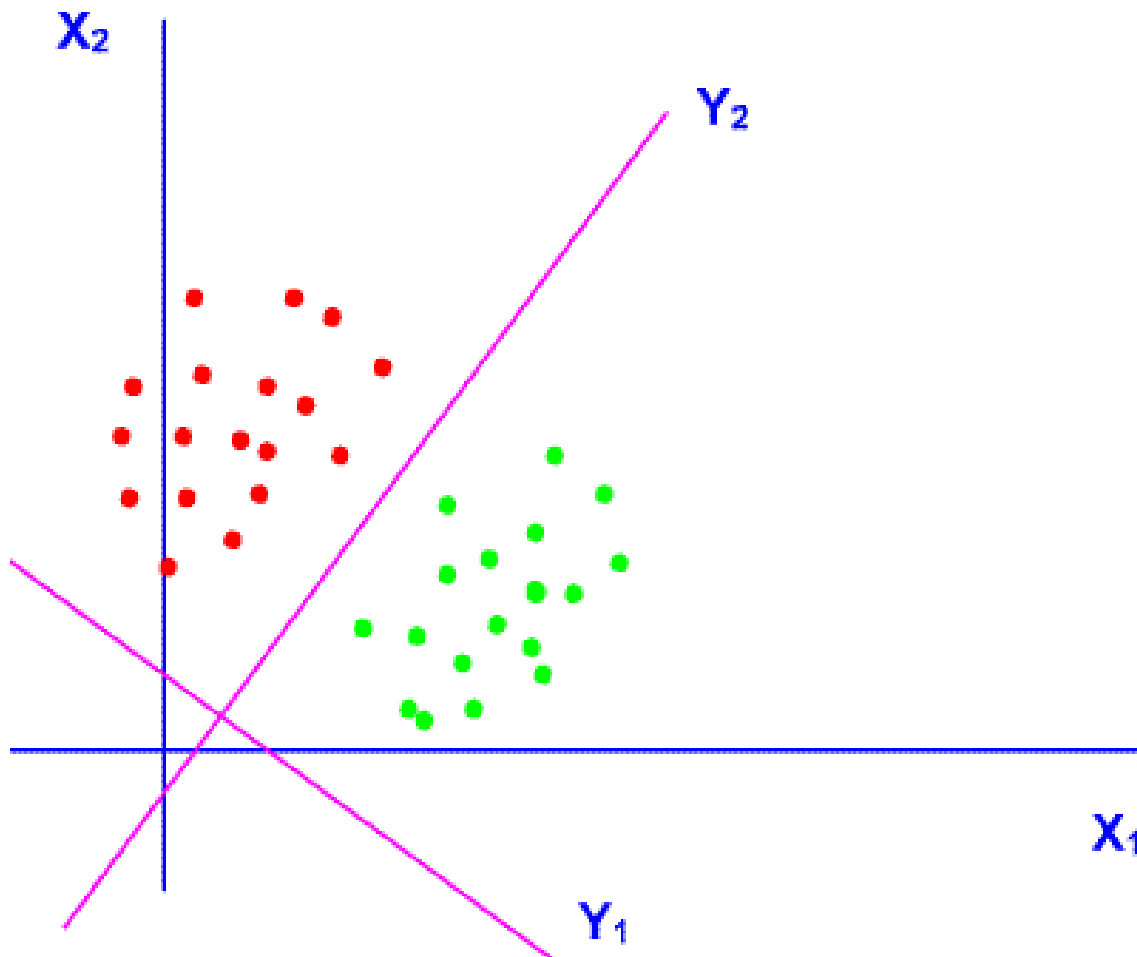
8.1 基本概念

- 已经讨论的问题
 - 是在 d 维特征空间已经确定的前提下进行的
 - 讨论的分类器设计问题是: 选择什么准则、使用什么方法, 将已确定的 d 维特征空间划分成决策域的问题
- 另一个重要问题
 - 如何确定合适的特征空间
 - 如何描述每个样本和每个模式

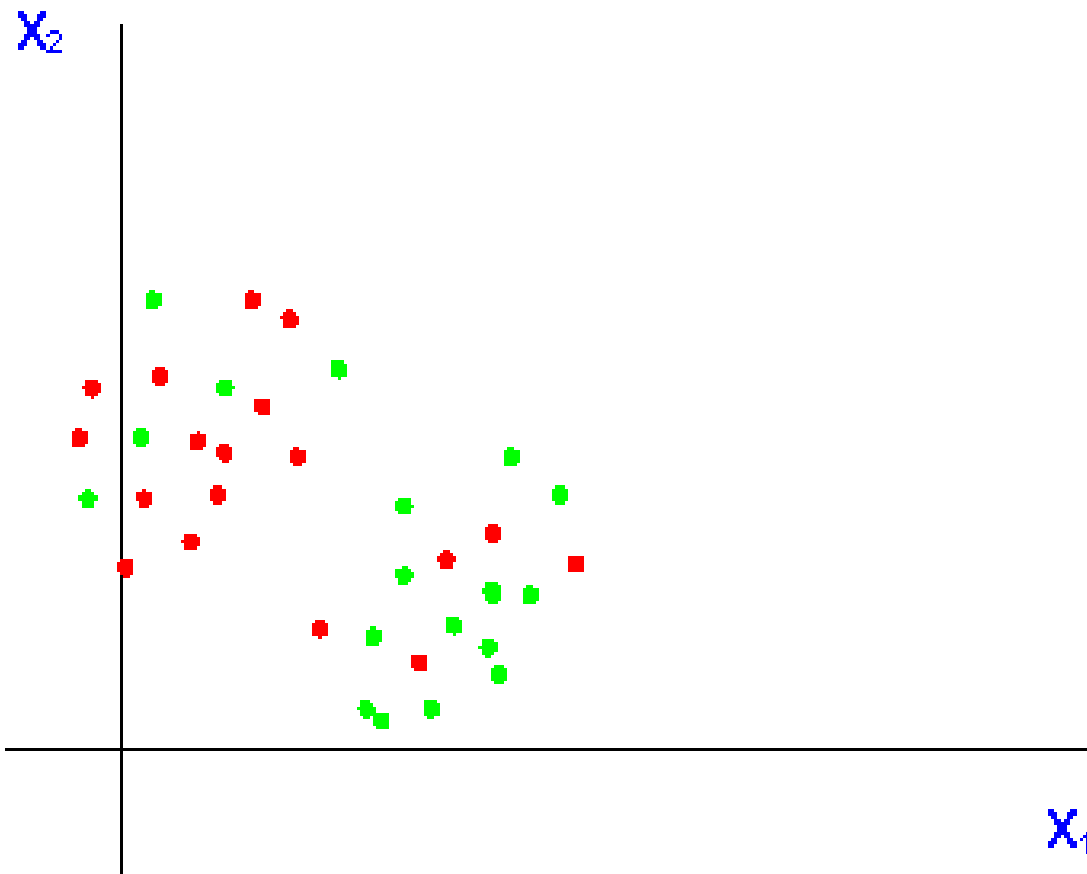
基本概念



基本概念



基本概念





基本概念

■ 问题的重要性

- 如果所选用的特征空间能使同类物体分布具有紧致性，即各类样本能分布在该特征空间中彼此分割开的区域内，这就为分类器设计成功提供良好的基础。
- 反之，如果不同类别的样本在该特征空间中混杂在一起，再好的设计方法也无法提高分类器的准确性。



基本概念

- 三个不同的层次描述
 - 1.物理量的获取与转换
 - 指用什么样的传感器获取电信号
 - 摄取景物则要用摄像机
 - 文字与数字识别首先要用扫描仪等设备
 - 手写体文字所用传感器与印刷体文字也很可能不同
 - 这些都属于物理量的获取，并且已转换成电信号，为计算机分析打下了基础
 - 对从传感器中得到的信号，可称之为原始信息
 - 它要经过加工、处理才能得到对模式分类更加有用的信号



基本概念

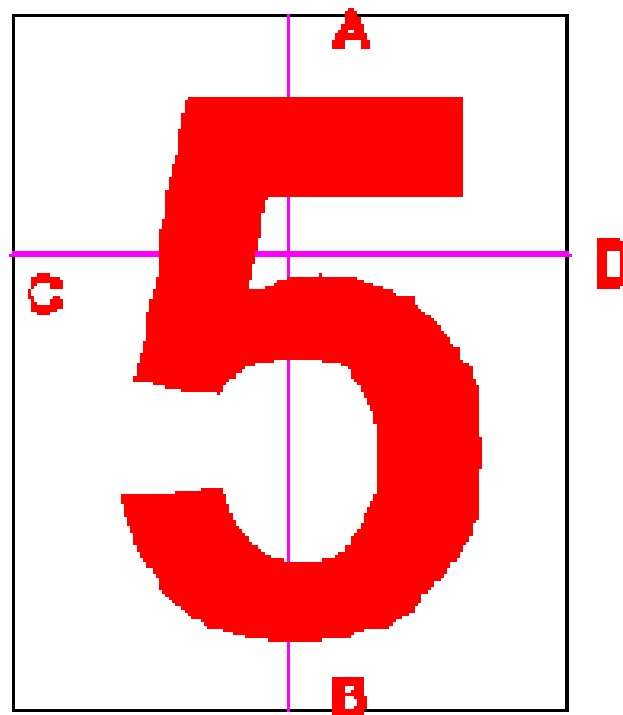
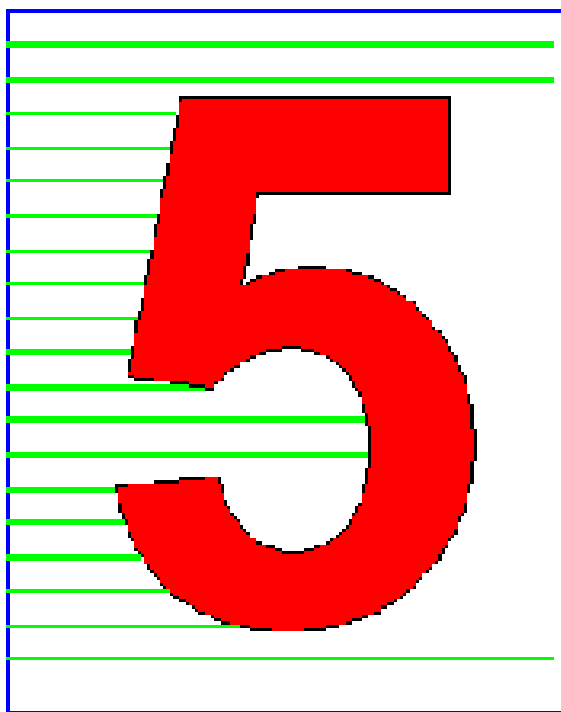
- 三个不同的层次描述

- 2. 描述事物方法的选择与设计

- 在得到了原始信息之后，要对它进一步加工，以获取对分类最有效的信息
 - 这部分信息必须对原始信息进行加工，而设计所要信息的形式是十分关键的
 - 例如:对阿拉伯数字的识别可以提出各种不同的想法
 - * 分析从框架的左边框到数字之间的距离变化反映了不同数字的不同形状，这可以用来作为数字分类的依据
 - * 强调分析不同截面的信号，如在框架的若干部位沿不同方向截取截面分析从背景到字，以及从字到背景转换的情况

基本概念

阿拉伯数字的识别可以提出各种不同的想法





基本概念

- 三个不同的层次描述
 - 2. 描述事物方法的选择与设计
 - 设计对事物的描述方法是充分发挥设计者智慧的过程
 - 这个层次的工作往往因事物而易，与设计者本人的知识结构也有关
 - 这是一个目前还无法自动进行的过程
 - 这个层次的工作是最关键的，但因为太缺乏共性，也不是本章讨论的内容



基本概念

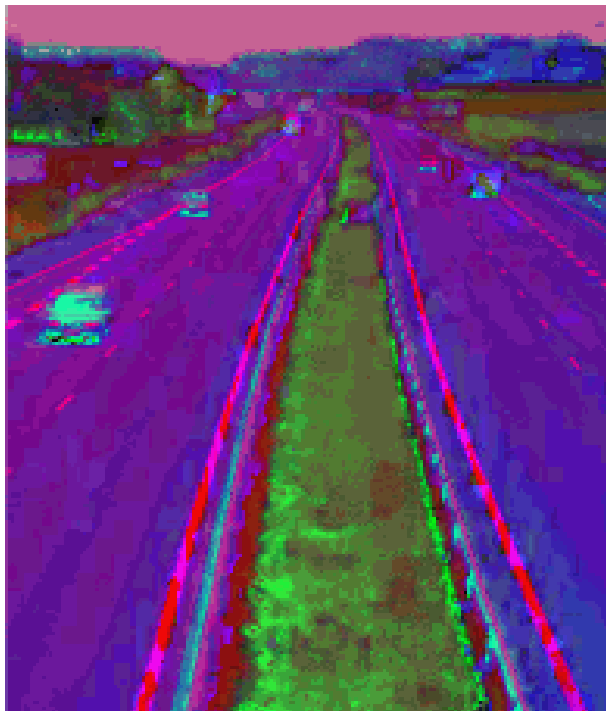
- 三个不同的层次描述
 - 3.特征空间的优化
 - 这个层次的工作发生在已有了特征的描述方法之后，也就是已有了一个初始的特征空间，如何对它进行改造与优化的问题
 - 所谓优化是要求既降低特征的维数，又能提高分类器的性能，这就是本章着重讨论的问题
 - 这里讨论的是一些不同模式识别系统中具有共性的问题，**特征选择与特征提取**



基本概念

- 例 用RGB颜色空间和HSI颜色空间
 - RGB和HSI是两种常用的颜色空间，
 - 虽然它们描述颜色的范围是一样的，
 - 也有确定的转换关系，
 - 但是用这两种不同的特征描述图像，
 - 对以后的识别工作会有很大影响

基本概念





基本概念

- 特征选择和特征提取
 - 对特征空间进行优化有两种基本方法
 - 一种为特征选择，
 - 一种为特征提取
 - 两者区别
 - 特征选择: 删掉部分特征
 - 特征提取: 通过一种映射，也就是说新的每一个特征是原有特征的一个函数



基本概念

■ 特征选择和特征提取

■ 两者区别

假设已有**D**维特征向量空间, $y = \{y_1, y_2, \dots, y_D\}$

特征选择:

样本由**d**维的特征向量描述: $x = \{x_1, x_2, \dots, x_d\}$ $d < D$

由于**X**只是**Y**的一个子集, 因此每个分量 x_i 必然能在原特征集中找到其对应的描述量 $x_i = y_j$



基本概念

■ 特征选择和特征提取

■ 两者区别

假设已有D维特征向量空间， $y = (y_1, y_2, \dots, y_D)$

特征提取：

找到一个映射关系： $A: Y \rightarrow X$

使新样本特征描述维数比原维数降低。其中每个分量 x_i 是原特征向量各分量的函数，即

$$x_i = x_i(y_1, y_2, \dots, y_D)$$



基本概念

■ 特征选择和特征提取

■ 两者区别

- 因此这两种降维的基本方法是不同的。
- 在实际应用中可将两者结合起来使用，比如先进特征提取，然后再进一步选择其中一部分，或反过来。



8.2 类别可分离性判据

- 为什么需要类别可分离性判据
 - 特征选择与特征提取的任务是求出一组对分类最有效的特征
 - 所谓有效是指在特征维数减少到同等水平时，其分类性能最佳
 - 因此需要有定量分析比较的方法，判断所得到的特征维数及所使用特征是否对分类最有利
 - 这种用以定量检验分类性能的准则称为**类别可分离性判据**



类别可分离性判据

- 为什么需要类别可分离性判据
 - 一般说来分类器最基本的性能评估是其分类的错误率
 - 如果能用反映错误率大小的准则，在理论上是最合适的
 - 对错误率的计算是极其复杂的，以至于很难构筑直接基于错误率的判据
 - 为此人们设法从另一些更直观的方法出发，设计出一些准则，用来检验不同的特征组合对分类性能好坏的影响，甚至用来导出特征选择与特征提取的方法
- 这些准则就是类别可分离性判据



类别可分离性判据

- 类别可分离性判据的种类
 - 基于距离度量的可分性判据
 - 基于概率分布的可分性判据等



8.2.1 基于距离的可分性判据

- 基于距离的可分性判据
 - 基于距离度量是人们常用来进行分类的重要依据
 - 一般情况下同类物体在特征空间呈聚类状态，即从总体上说同类物体各样本由于具有共性
 - 因此类内样本间距离应比跨类样本间距离小
 - Fisher准则正是以使类间距离尽可能大同时又保持类内距离较小这一种原理为基础的
 - 同样在特征选择与特征提取中也使用类似的原理，这一类被称为基于距离的可分性判据。



基于距离的可分性判据

- 基于距离的可分性判据

- 在讨论Fisher准则时曾用过两个描述离散度的矩阵

一个是类间离散矩阵 S_b $S_b = (\bar{x}_1 - \bar{x}_2)(\bar{x}_1 - \bar{x}_2)^T$

另一个是类内离散度矩阵 S_w

$$S_w = S_1 + S_2$$

$$S_i = \sum_{X \in X_i} (X - \bar{x}_i)(X - \bar{x}_i)^T, i=1, 2$$



基于距离的可分性判据

- 基于距离的可分性判据
 - 如果推广至c类别情况

$$S_b = \sum_{i=1}^c P_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T$$

$$S_w = \sum_{i=1}^c P_i E_i[(X - \bar{x}_i)(X - \bar{x}_i)^T]$$

其中 \bar{x} 为所有样本的总均值向量， P_i 表示各类别的先验概率， E_i 表示i类的期望符号。



基于距离的可分性判据

- 1 计算特征向量间平均距离的判据

$$J_d(X) = \text{tr}(S_r + S_b)$$

该式可写成

$$J_d(X) = \frac{1}{2} \sum_{i=1}^c P_i \sum_{j=1}^c P_j \frac{1}{N_i N_j} \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} \mathcal{D}(X_k^{(i)}, X_l^{(j)})$$

在欧氏距离情况下有

$$\mathcal{D}(X_k^{(i)}, X_l^{(j)}) = (X_k^{(i)} - X_l^{(j)})^T (X_k^{(i)} - X_l^{(j)})$$



基于距离的可分性判据

- 1 计算特征向量间平均距离的判据

利用均值向量 \bar{x}_i 与总均值向量 \bar{x} ，有

$$\bar{x}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} X_k^{(i)} \quad \bar{x} = \sum_{i=1}^c P_i \bar{x}_i$$

$$J_d(X) = \sum_{i=1}^c P_i \left[\frac{1}{N_i} \sum_{k=1}^{N_i} (X_k^{(i)} - \bar{x}_i)^T (X_k^{(i)} - \bar{x}_i) + (\bar{x}_i - \bar{x})^T (\bar{x}_i - \bar{x}) \right]$$

$$J_d(X) = \text{tr}(S_w + S_b)$$

基于距离的可分性判据

- 2 考虑类内类间欧氏距离的其它判据
 - 判据 $J_d(X)$ 是计算特征向量的总平均距离
 - 以下一些判据则基于使类间离散度尽量大，类内离散度尽量小的考虑而提出：

$$J_2(X) = \text{tr}(S_W^{-1} S_b) \quad J_3(X) = \ln \left[\frac{|S_b|}{|S_W|} \right]$$

$$J_4(X) = \frac{\text{tr}(S_b)}{\text{tr}(S_W)} \quad J_5(X) = \frac{|S_W + S_b|}{S_W}$$



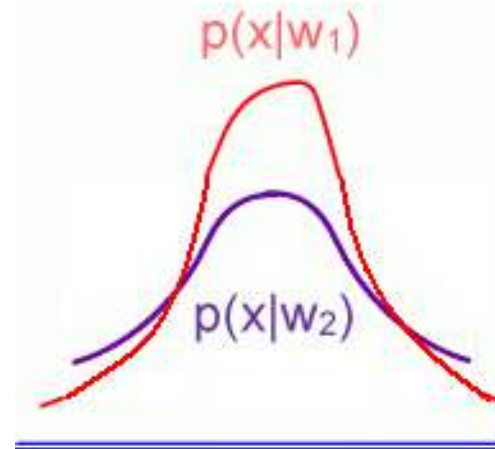
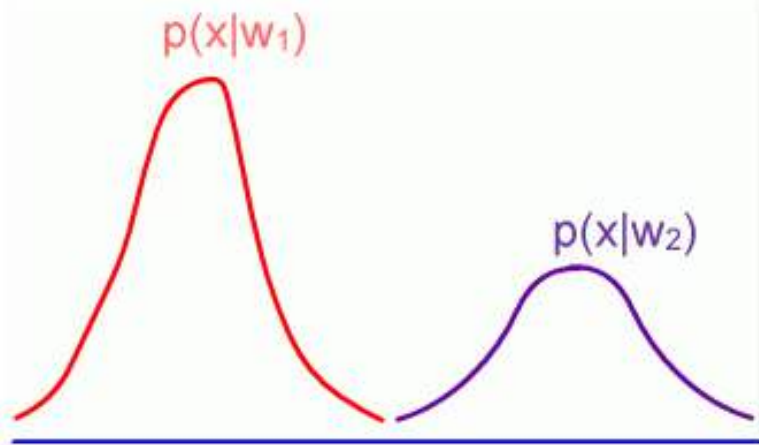
8.2.2 基于概率分布的可分性判据

- 基于概率分布的可分性判据
 - 上一节讨论的是样本在特征空间的分布距离作为特征提取的依据。
 - 该种原理直观，计算简便。
 - 但是这种原理没有考虑概率分布，因此当不同类样本中有部分在特征空间中交迭分布时，简单地按距离划分，无法表明与错误概率之间的联系。
 - 基于概率分布的可分性判据则依据如下观察到的现象。

8.2.2 基于概率分布的可分性判据

- 基于概率分布的可分性判据

观察现象:不考虑各类的先验概率, 或假设两类样本的先验概率相等





基于概率分布的可分性判据

■ 基于概率分布的可分性判据

观察现象

- 如果两类条件概率分布互不交迭，即对 $p(X|\omega_2) \neq 0$ 处都有 $p(X|\omega_1) = 0$ ，则这两类就完全可分；
- 另一种极端情况是对所有 X 都有 $p(X|\omega_1) = p(X|\omega_2)$ ，则两类就完全不可分。



基于概率分布的可分性判据

■ 基于概率分布的可分性判据

- 因此人们设计出与概率分布交迭程度有关的距离度量方法，这些距离 J_p 有以下几个共同点：

- 1. J_p 是非负，即 $J_p \geq 0$
- 2. 当两类完全不交迭时 J_p 达到其最大值
- 3. 当两类分布密度相同时， $J_p = 0$

这种函数的一般式可表示为：

$$J(\cdot) = \int g[p(x | \omega_1), p(x | \omega_2), p_1, p_2] dx$$



基于概率分布的可分性判据

- 一些常用的概率距离度量
 - Bhattacharyya距离和Chernoff界限

Bhattacharyya距离的定义用下式表示

$$J_B = -\ln \int [p(X | \omega_1)p(X | \omega_2)]^{1/2} dx$$

Chernoff界限的定义与其相似，为

$$J_B = -\ln \int [p^s(X | \omega_1)p^{1-s}(X | \omega_2)] dx$$

因此 J_B 是 J_C 的一个特例



基于概率分布的可分性判据

- 一些常用的概率距离度量
 - 散度

另一种常用的基于概率距离度量的判据是利用似然比或对数似然比。

对两类问题，其对数似然比为
$$l_{ij} = \ln \frac{p(X | \omega_i)}{p(X | \omega_j)}$$

如果对某个 X , $p(X | \omega_i) = p(X | \omega_j)$ 则 $l_{ij} = 0$

反之若两者差异越大，则 l_{ij} 的绝对值也大。



基于概率分布的可分性判据

- 一些常用的概率距离度量
 - 散度

以上只是对某一X值而言，为了对整个特征空间概率分布的差异程度作出评价，将对 ω_i 类及对 ω_j 的可分性信息分别定义为

$$I_{ij} = E[l_{ij}(X)] = \int_{\mathcal{X}} p(X | \omega_i) \ln \frac{p(X | \omega_i)}{p(X | \omega_j)} dX$$

$$I_{ji} = E[l_{ji}(X)] = \int_{\mathcal{X}} p(X | \omega_j) \ln \frac{p(X | \omega_j)}{p(X | \omega_i)} dX$$



基于概率分布的可分性判据

- 一些常用的概率距离度量
 - 散度

而总的平均可分信息则可表示成

$$J_D = I_{ij} + I_{ji} = \int_{\mathcal{X}} [p(X | \omega_i) - p(X | \omega_j)] \ln \frac{p(X | \omega_i)}{p(X | \omega_j)} dX$$

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/868123024102006057>