模式识别

第一章 概论

任课教师：柳欣老师

email：starxliu@163.com

课件内容综合来自于互联网

课程介绍

- 人工智能……
- 人类的心智活动：感知觉（包括模式识别）、注意、记忆、语言、思维、意识等，研究领域为认知科学（**Cognitive Science**），主要从脑与神经科学、心理学的角度
- 模式识别（**Pattern Recognition**）是研究用数学方法使计算机自动识别事物的一门科学，其目的是用机器完成类似于人类智能通过视觉、听觉等感官去识别外界环境所进行的工作，典型应用：语音识别、图像识别等

课程介绍

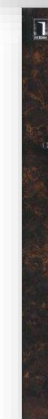
- 模式识别学科形成于**50~60**年代，但其中统计模式识别方法中的线性判别函数等内容**30**年代已提出
- 模式识别是一门理论与应用并重的技术科学，也是一个高度交叉的学科，应用领域日益扩大，现有的理论与方法还有不足
- 本课程主要介绍统计模式识别的基本概念和理论，具有代表性的一些方法**/**算法及其应用。要求了解模式识别的基本概念，掌握基本原理和基本方法

课程介绍

- 预备知识：
  - **线性代数**：向量、矩阵的基本运算、逆、行列式、特征值、特征向量等；
  - **概率论与数理统计**：概率（先验、条件）、概率密度、随机变量和分布、数学期望、全概率和贝叶斯公式、正态分布、参数估计、假设检验
  - **高等数学、最优化方法、信息论、程序设计基础**
- 相关学科
  - 机器学习
  - 人工智能
  - 图像处理
  - 计算机视觉

教材和参考书目

- **教材：**
  - **《模式识别》清华大学出版社**
    - **（第2版），边肇祺、张学工等著，2000年**
    - **（第3版），张学工等著，2010年**
- **参考书：**
  - 《模式分类》（第2版），R.O.Duda等著，李宏东等译，机械工业出版社，2003年
  - 《模式识别》（第4版），Sergios Theodoridis著，李晶皎译，电子工业出版社，2016年
- **网络课程MOOC**
  - Machine Learning，Stanford，Andrew Ng
    coursera 網易公开课
  - Machine Learning Crash Course
    Google



课程要求

- 讲课＋讨论＋实验
- 考核：
  - 平时成绩（到课、作业、实验报告）30%
  - 期末考试　70%
  - *课程内容讲授或讨论：个人、实验、小组

学习要点

- 重点掌握模式识别的基本概念，基本方法和算法原理。
- 注重理论与实践紧密结合，注意如何将所学知识运用到实际应用之中
- 为研究新的模式识别的理论和方法打下基础
- 不要被繁琐的数学推导吓倒
- 基本要求：完成课程学习，通过考试，获得学分。
- 提高：能够将所学知识和内容用于课题研究，解决实际问题。
- 飞跃：通过模式识别的学习，改进思维方式，为将来的工作打好基础，终身受益。

题外话

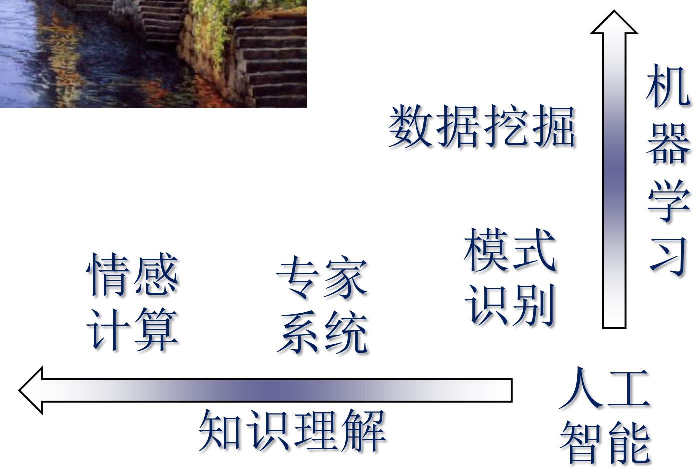
- 基本：完成课程学习，通过考试，获得学分。
- 提高：能够将所学知识和内容用于课题研究，解决实际问题。
- 飞跃：通过模式识别的学习，改进思维方式，为将来的工作打好基础，终身受益。

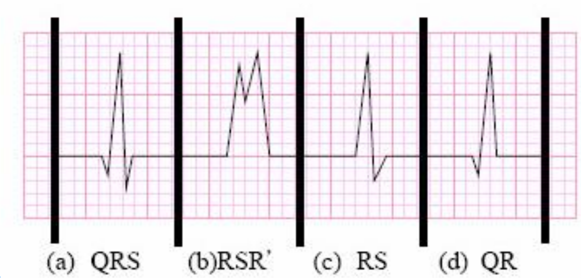
1.1 模式识别的基本概念

- 人对事物的认识
  - 按分类来进行认识
  - 认识的不同层次

generation of image captions

- **模式识别（pattern recognition）**
  - 指用计算机实现人的**模式识别**能力
  - 确定一个样本的类别属性（模式类）的过程
  - 样本（特征空间）到类别的映射
  - 把样本根据其特征归类，又称“模式分类”(pattern classification)





1.1 模式识别的基本概念

- Pattern recognition is the study of how machines can observe the environment, learn to distinguish patterns of interest from their background, and make sound and reasonable decisions about the categories of the patterns. (Anil K. Jain)

- 模式识别（pattern recognition）
- 模式识别是指对表征事物或现象的各种形式的(数值的、文字的和逻辑关系的)信息进行处理和分析，以对事物或现象进行描述、辨认、分类和解释的过程，是信息科学和人工智能的重要组成部分。

1.1 模式识别的基本概念

什么是模式？

- 广义地说，存在于时间和空间中可观察的物体，如果我们可以区别它们是否相同或是否相似，都可以称之为模式。
- 模式所指的不是事物本身，而是从事物获得的信息，因此，模式往往表现为具有时间和空间分布的信息。
- 模式的直观特性:
  - 可观察性
  - 可区分性
  - 相似性

1.1 模式识别的基本概念

什么是识别？

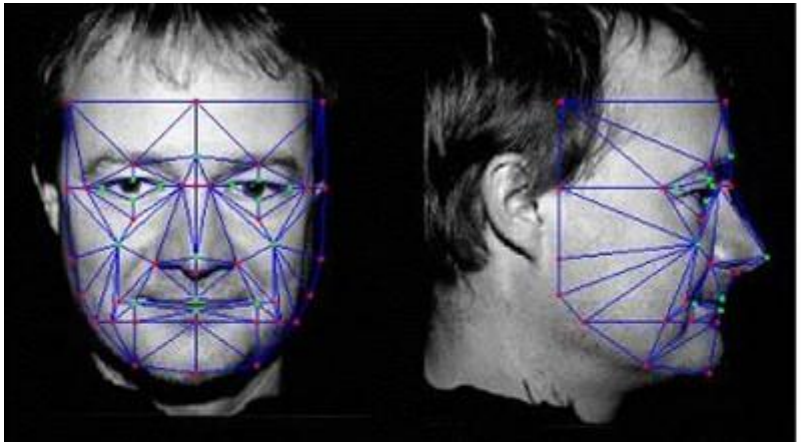
- 把具体的样本归类到某一个模式，可以叫做模式的识别（或分类）。

- 识别是时时刻刻都在发生的；
- 识别（Recognition)是再认知的过程；

- 识别行为：
  - 识别具体事物；
  - 识别抽象事物。

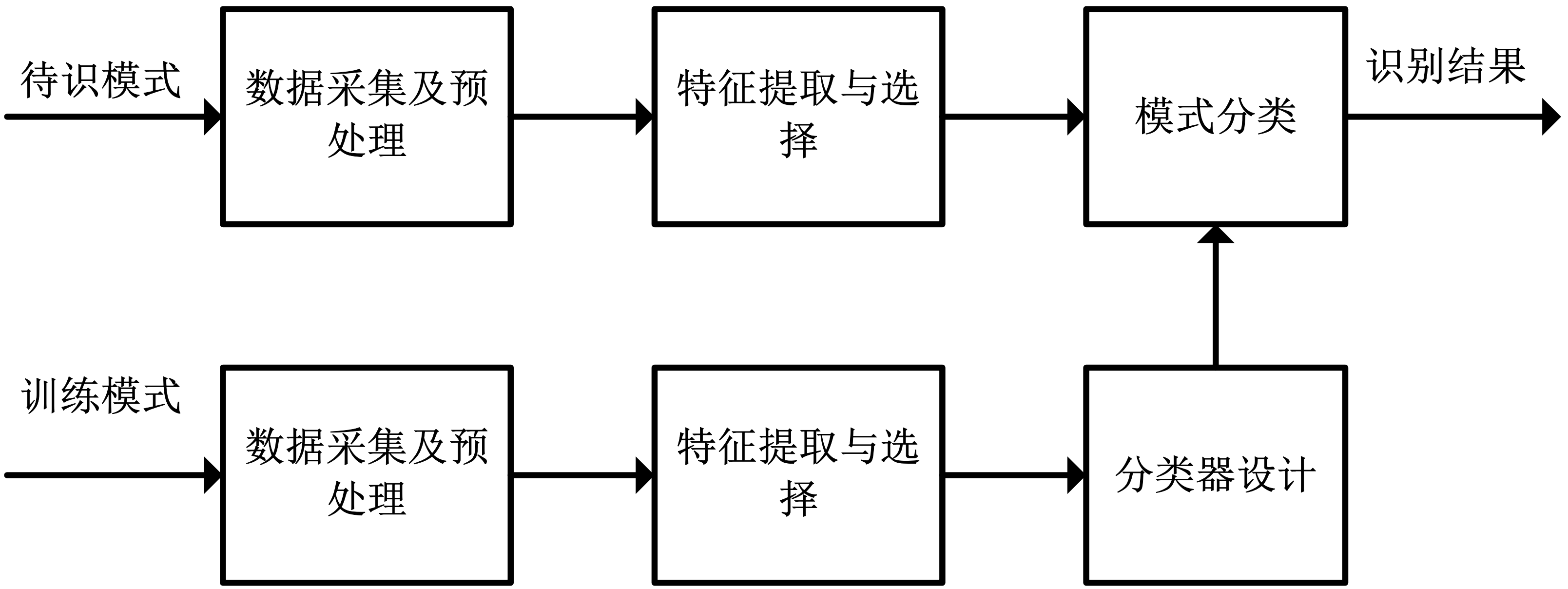
1.1 模式识别的基本概念

- **样本、特征、模式、类**
  - 样本(sample)，研究**对象**的个体；样本集
  - 特征(feature)，表征样本的**观测**
  - 模式(pattern)，法也；特征的组合（时、空）
  - 类别(class)，具有某些共同特性(模式)样本的集合
  - 已知/未知样本，类别标号已知/未知



1.2 模式识别系统

- 模式识别系统组成与过程



1.2 模式识别系统

- 空间类型

对象空间

↓

模式空间

↓

特征空间

↓

类型空间

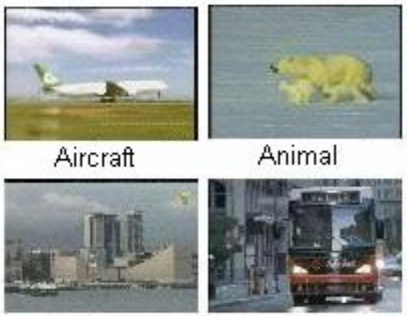
模式采集： 从客观世界（对象空间）到模式空间的过程称为模式采集。

特征提取和特征选择： 由模式空间到特征空间的变换和选择。

类型判别： 特征空间到类型空间所作的操作。

1.2 模式识别系统

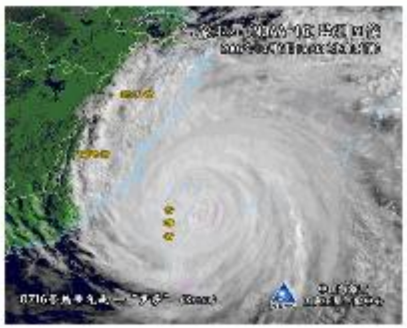
- 数据获取：模式的数字化表达



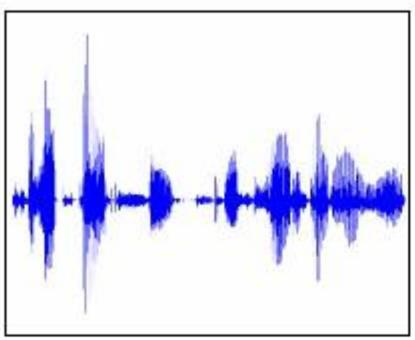
图片

电视

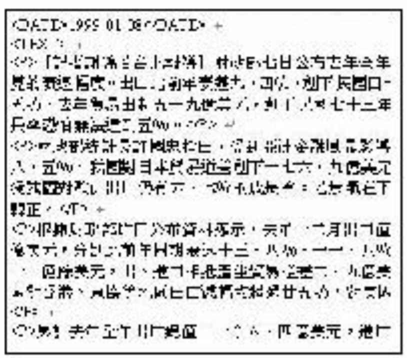
视频监控



遥感图像



语音



文本

网络数据

医学图像

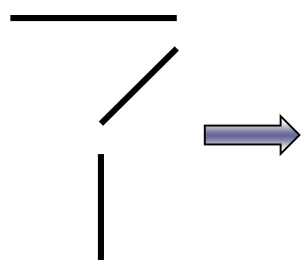
1.2 模式识别系统

- **预处理：去噪、复原、强化**
- **特征提取和选择：**
  - 特征是指用于描述模式性质（特性）的一种定量的概念。（**苹果和桔子的大小、颜色、味道**）
  - **特征提取**是指采用映射（或变换）实现由模式测量空间向特征空间的转变或者将特征空间的维数从高维变成低维。（**遥感图像等**）
  - **特征选择**，从一组中挑选最有效的特征以降低维数。

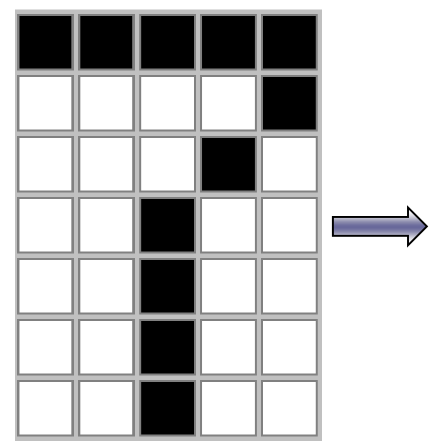
特征提取和选择更依赖于具体问题和领域，需要相应的领域的专业知识

1.2 模式识别系统

- **特征向量表示法**
  - 印刷体数字图像往往用一个N×M的数组表示。如果N＝5，M＝7，则一个数字就用5×7共35个网格是黑是白来表示。如令是黑为“1”，是白为“0”，那么一个数字就可用35维的二进制向量表示。这就是典型的特征向量表示法
- **结构表示法**
  - 对对象所包含的成分进行分析



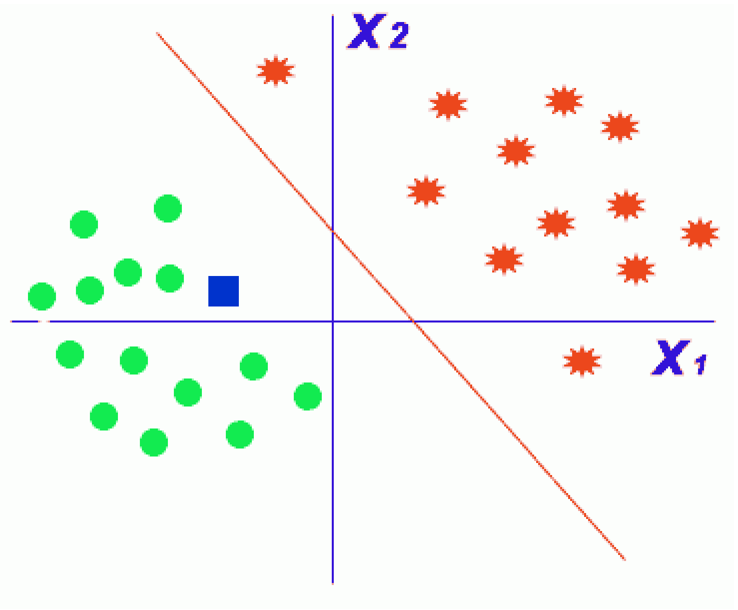
$(1 \to 3 \to 2)$



$$\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

1.2 模式识别系统

- 分类决策
  - **训练**是依据特征空间的分布，决定分类器的具体参数。一般说来采用什么样式的决策分界由设计者决定，如可用直线、折线或曲线作为类别的分界线。分界线的类型可由设计者直接确定，也可通过训练过程产生，但是这些分界线的具体参数则利用训练样本经训练过程确定
  - **分类决策**是对待分类样本进行分类决策的过程



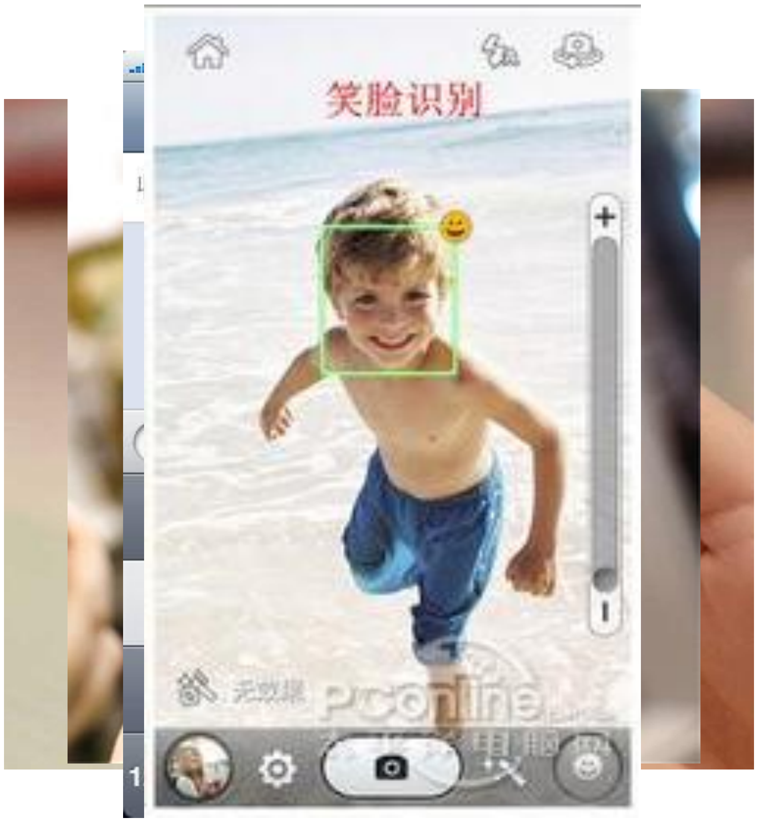
- 在许多应用领域中模式识别并不一定作为独立的环节存在

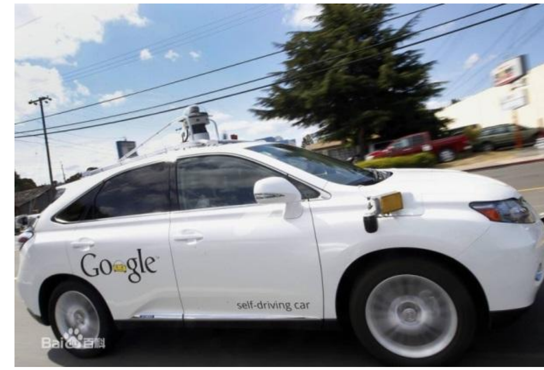
1.2 模式识别系统

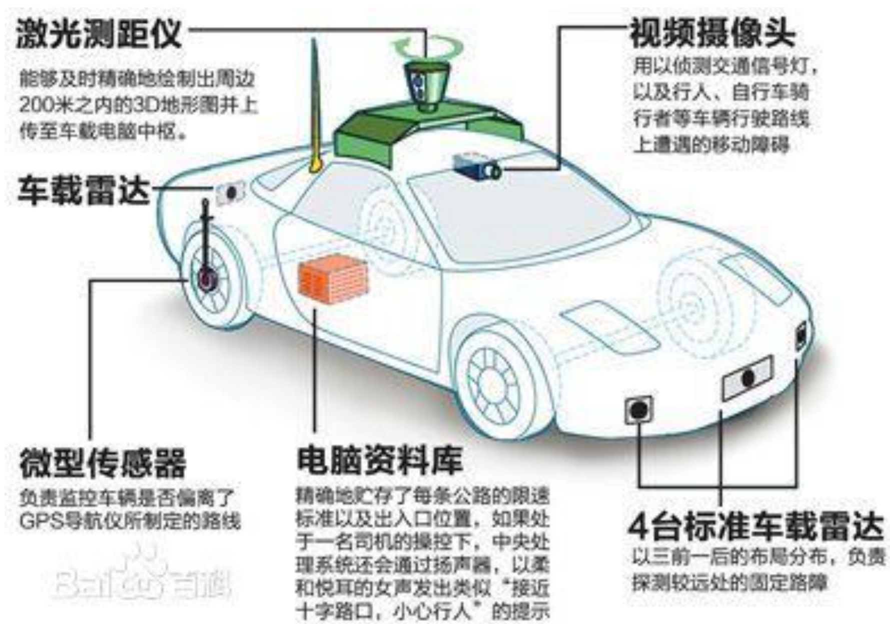
- 实例：车牌识别

1.2 模式识别系统

- 我们身边的模式识别？

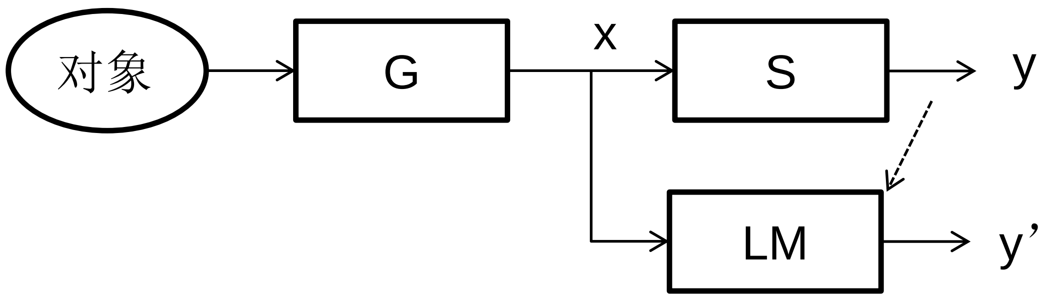




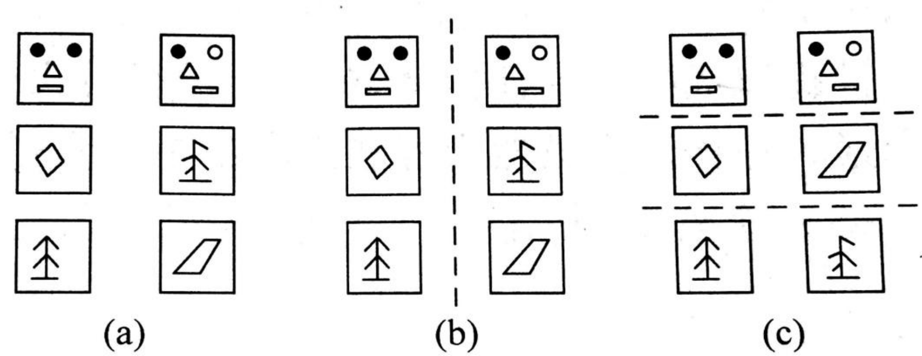


1.3 模式识别的基本方法

- 基于知识：人工智能、句法模式识别
- 基于数据：统计模式识别



- 监督模式识别
- 非监督模式识别（聚类分析）

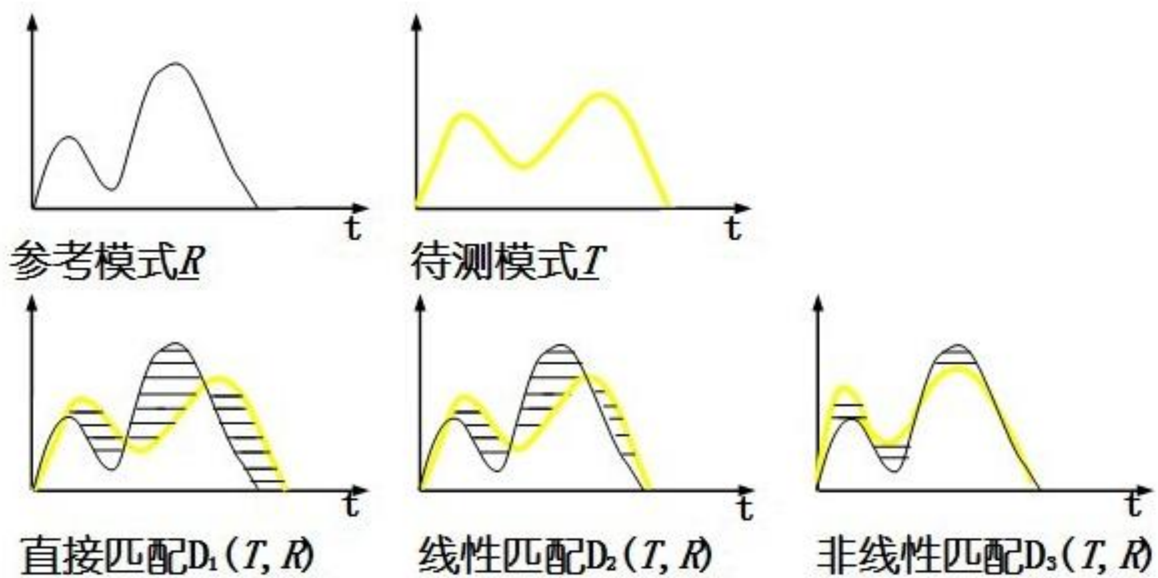


1.3 模式识别的基本方法

- 模版匹配法(templete matching)
- 统计模式识别(statistical pattern recognition): 1950s-
- 结构模式识别(句法方法) (structural pattern recognition) 1970s-
- 模糊模式识别(fuzzy) : 1980s-
- 神经网络方法(neural network): 1980s-
- 支持向量机(核方法) : (Support Vector Machine) 1990s-
- 深度学习(Deep learning)：2010s-

1.3 模式识别的基本方法

- 模版匹配法(templete matching)
  - 首先对每个类别建立一个或多个模版，输入样本和数据库中每个类别的模版进行比较,例如求相关或距离，根据相似性(相关性或距离)大小进行决策。
  - 优点:直接,简单；
  - 缺点:适应性差



1.3 模式识别的基本方法

- **统计模式识别(statistical pattern recognition):**
  - 理论基础：概率论，数理统计
  - 主要方法：线性、非线性分类、Bayes决策、聚类分析
  - 优点：

    1）比较成熟

    2）能考虑干扰噪声等影响

    3）识别模式基元能力强
  - 缺点：

    1）对结构复杂的模式抽取特征困难

    2）不能反映模式的结构特征，难以描述模式的性质

    3）难以从整体角度考虑识别问题

1.3 模式识别的基本方法

- **结构模式识别(句法方法)**
  - 美籍华人付京荪(Purdue Univ.)提出句法结构模式识别（50年代 Noam Chomsky 基于文法的形式语言理论）
  - 采用比较简单的子模式（基元）来描述复杂模式，观察对象表达为一个由基元组成的句子，基元构成模式所遵循的规则即为文法（句法）
  - 模式描述方法：　符号串，树，图
  - 模式判定：　用一个文法表示一个类，m类就有m个文法，然后判定未知模式遵循哪一个文法
  - 在学习过程中，确定基元与基元之间的关系，推断出生成景物的方法。
  - 判决过程中，首先提取基元，识别基元之间的连接关系，使用推断的文法规则做句法分析
  - 实际中，往往与统计相结合，统计方法完成基元识别

1.3 模式识别的基本方法

- 理论基础：形式语言，自动机技术
- 主要方法：自动机技术、CYK剖析算法、Early算法、转移图法
- 优点：
  1）识别方便，可以从简单的基元开始，由简至繁。
  2）能反映模式的结构特征，能描述模式的性质。
  3）对图象畸变的抗干扰能力较强。
- 缺点：
  当存在干扰及噪声时，抽取特征基元困难，且易失误。

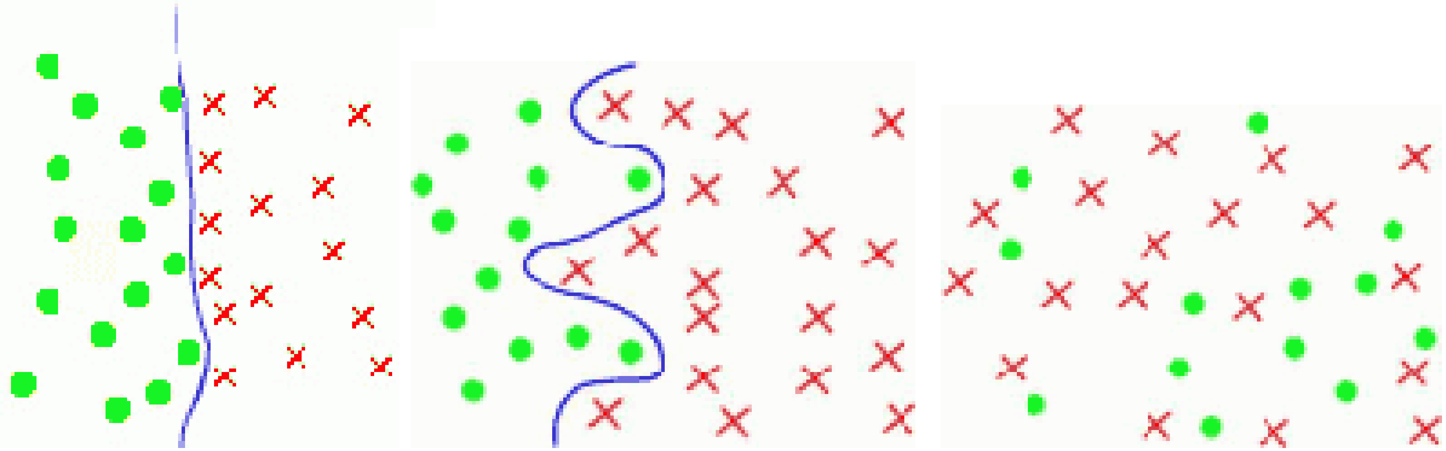
1.3 模式识别的基本方法

- **模糊模式识别（fuzzy）：**
  - 模式判定：

    是一种集合运算。用隶属度将模糊集合划分为若干子集， m类就有m个子集，然后根据择近原则分类。
  - 理论基础：模糊数学
  - 主要方法：模糊统计法、二元对比排序法、推理法、模糊集运算规则、模糊矩阵
  - 优点：由于隶属度函数作为样本与模板间相似程度的度量，故往往能反映整体的与主体的特征，从而允许样本有相当程度的干扰与畸变。
  - 缺点：准确合理的隶属度函数往往难以建立，故限制了它的应用

1.4 关于模式识别的一些基本问题

- 模式类的紧致性



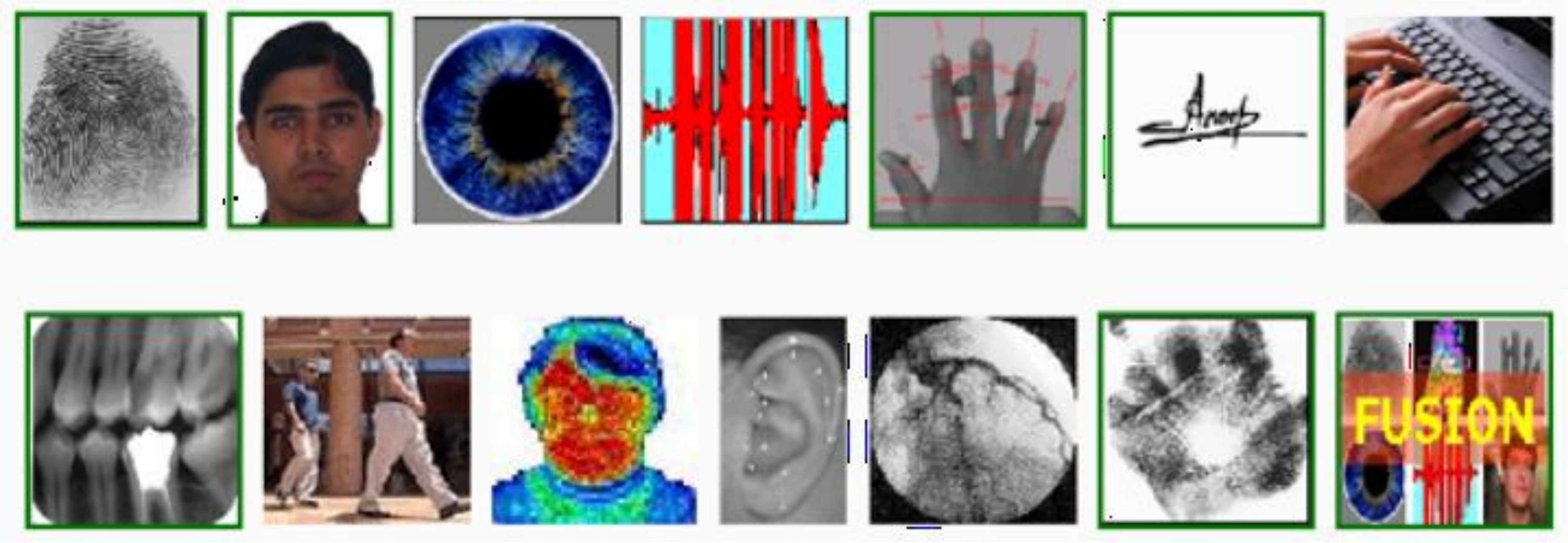
- 分类都是带有主观性的行为，常缺乏纯客观的分类标准。
- 靠哪些特征决定相似并进行分类，取决于行为的目的和方法

1.5 模式识别的应用和研究组织

- 语音识别
- 说话人识别、生物特征识别
- 字符与文字识别：印刷、手写、联机
- 安全监控（身份识别、视频监控、交通监控）
- 空间探测与环境资源检测，遥感图像
- 生物信息学，医学图像
- 文本图像分析，文本分类
- 工业自动化：零部件/物品分类
- 数据挖掘、多媒体数据检索（文档、图像、视频、音乐检索 ）
- ……

1.5 模式识别的应用和研究组织

- 生物特征识别

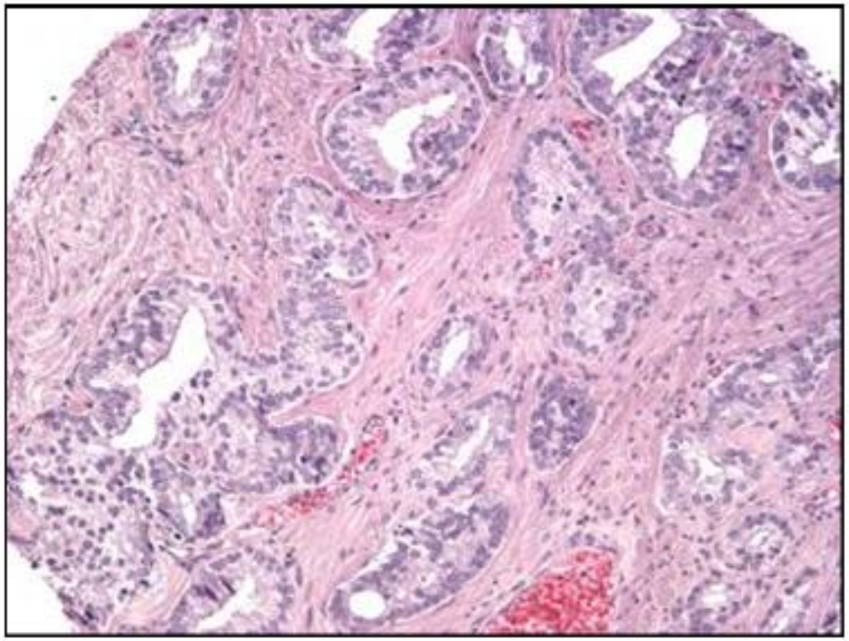


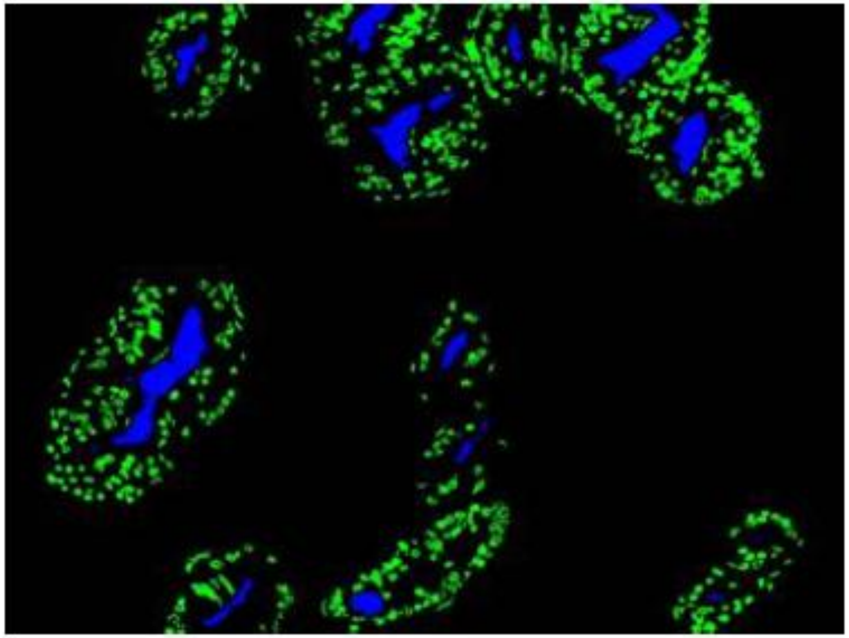
1.5 模式识别的应用和研究组织

- 遥感图像地表分类

1.5 模式识别的应用和研究组织

- 医学图像分析





Cancer detection and grading using microscopic tissue data.

1.5 模式识别的应用和研究组织

其它

- 工业用途：产品质量检验，设备故障检测，智能机器人的感知系统；
- 商业用途：钱币的自动识伪，信函的自动分拣，电话信息查询，声控拨号；
- 医学用途：对心电、脑电、CT等信号进行处理和识别
- 安全领域：生理特征鉴别(Biometrics)，网上电子商务的身份确认，对公安对象的刑侦和鉴别；
- 军事领域：巡航导弹景物识别，战斗单元的敌我识别；
- 办公自动化：文字识别技术和声音识别技术；
- 数据挖掘：数据分析；网络应用：文本分类。

1.5 模式识别的应用和研究组织

- 国际组织
  - IAPR，1973年 IEEE发起了第一次关于模式识别的国际会议“ICPR”，成立了国际模式识别协会---“IAPR”，每2年召开一次国际学术会议
  - IEEE Computer Society: TC on PAMI (Pattern Analysis and Machine Intelligence) 1977年 IEEE的计算机学会成立了模式分析与机器智能（PAMI）委员会，每2年召开一次模式识别与图象处理学术会议
- 国内组织
  - 中国自动化学会:模式识别与机器智能(PRMI)专业委员会,1981年成立,IAPR成员组织
  - 中国计算机学会:人工智能与模式识别专业委员会
  - 中国人工智能学会

1.5 模式识别的应用和研究组织

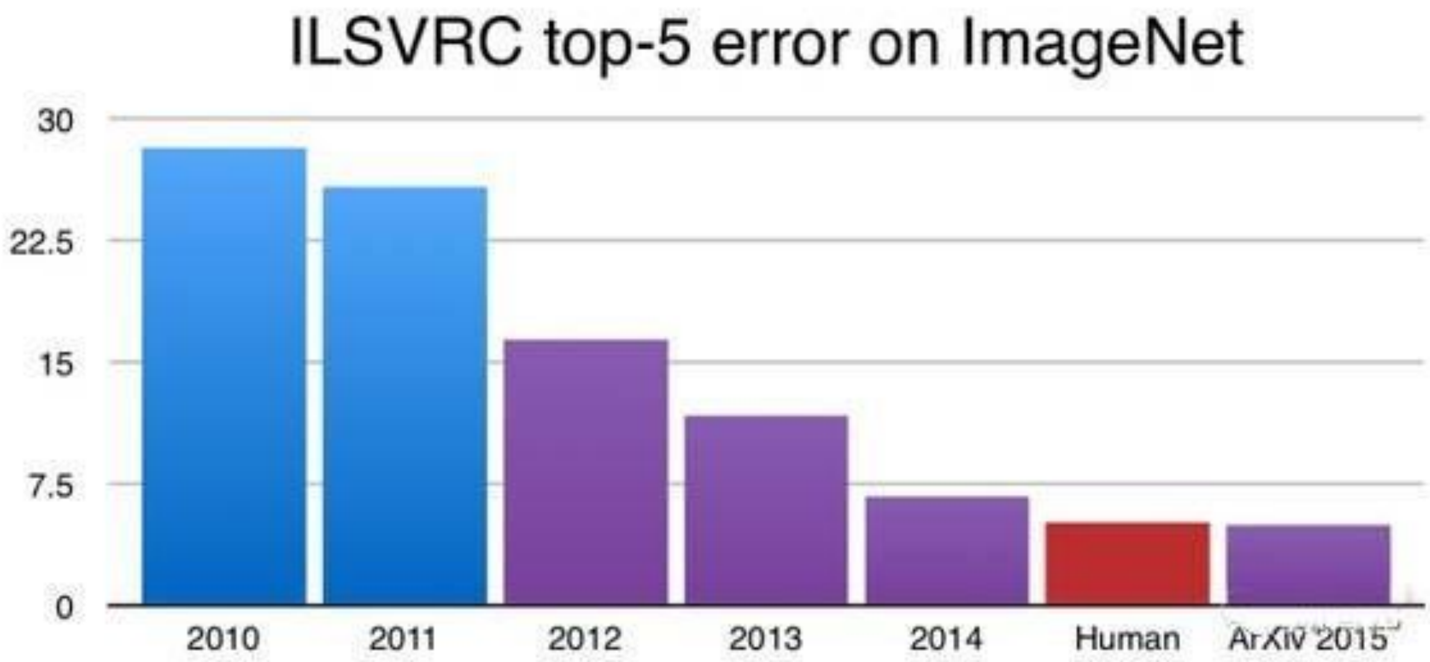
- 主要期刊
  - IEEE Trans. on PAMI,1978-,IEEE Computer Society
  - Pattern Recognition,1968-,PR Society, Elsevier
  - Pattern Recognition Letter,1980-,IAPR, Elsevier
  - Int. Journal of PR and AI, 1988- (World Scientific)
  - Pattern Analysis and Applications, 1997- (Springer)
  - Int. J. Document Analysis & Recognition, 1998-
  - 模式识别与人工智能
  - 中国图像与图形学学报

1.5 模式识别的应用和研究组织

- 主要会议，系列性国际会议
  - ICPR:2年一次
  - ICCV: 2年一次, IEEE International Conference on Computer Vision
  - CVPR:每年一次, IEEE Conference on Computer Vision and Pattern Recognition
  - ICDAR: 2年一次, 规模International Conference on Document Analysis and Recognition
  - ICIP - IEEE International Conference on Image Processing

1.5 模式识别的应用和研究组织

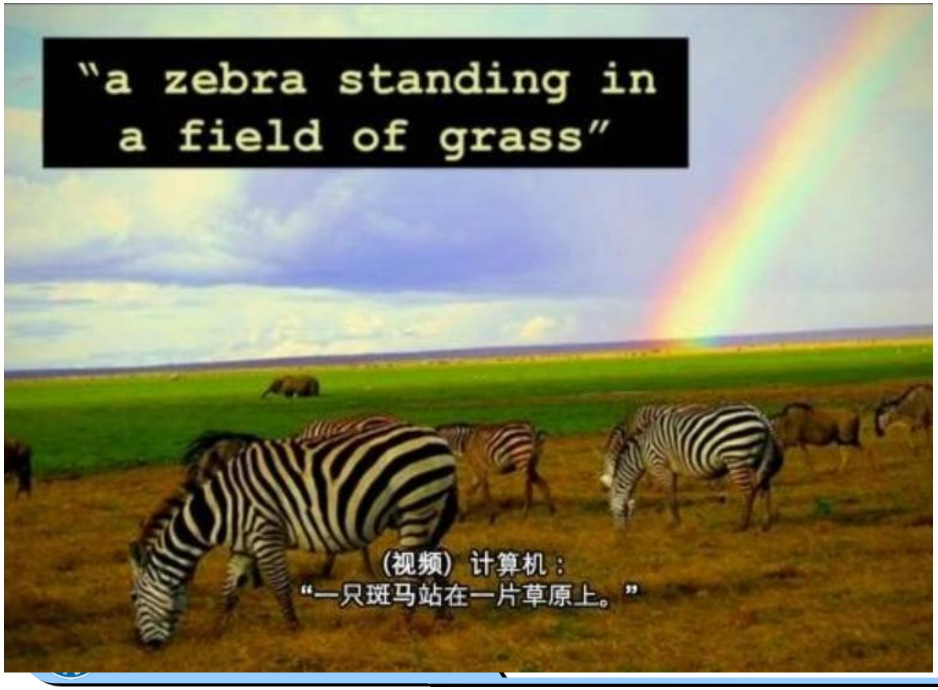
- **图像识别准确率的指数级增长**
  - 2015年的 ImageNet 挑战赛，在图像识别准确率上，机器首次超过了人类（2015年3.5%，2016年2.99%左右）



- **ImageNet：2009年建立，含有1500万张照片， 涵盖了22000种物品（李飞飞，**斯坦福大学人工智能与视觉实验室主任，谷歌机器学习部门负责人）

2012年，
Google Brain，
Andrew Ng
Jeff Dean
“Cat”

- 图像理解
  - Labeling
  - Caption



1.2 特征矢量和特征空间

特征矢量：

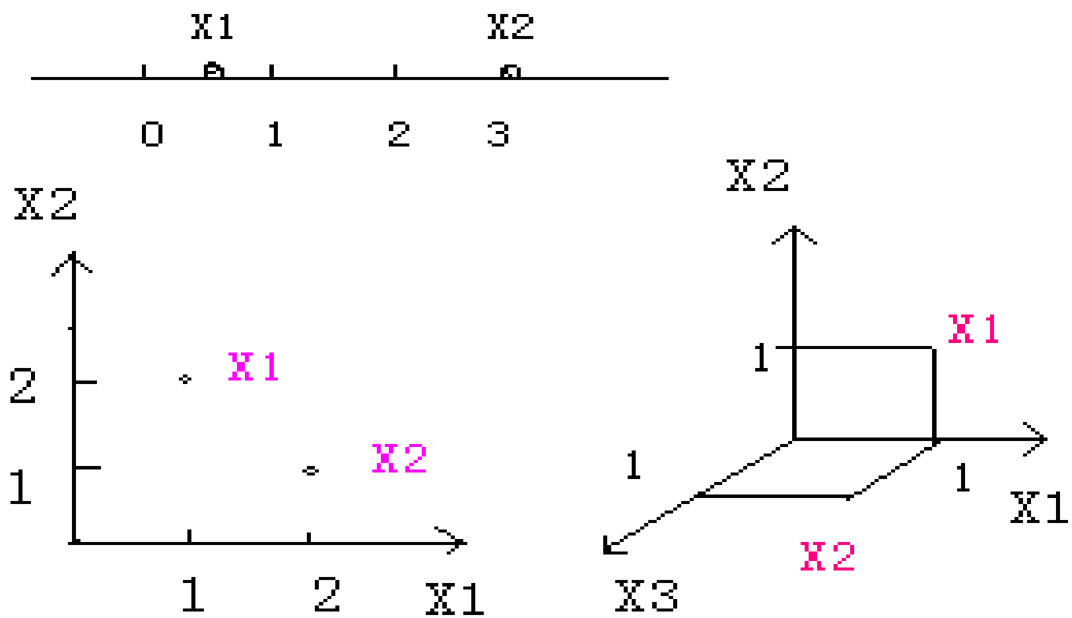
设一个研究对象的n 个特征量测量值分别为 $x_1, x_2, \ldots, x_n$，我们将它们作为一个整体来考虑，让它们构成一个n 维特征矢量$\vec{x}$ 。

特征空间：

各种不同取值的特征矢量的全体构成了n维特征空间。

注：特征矢量就是特征空间中的一个点。

1.2 特征矢量和特征空间



1.3 随机矢量的描述

随机矢量：

在模式识别过程中，要对许多具体对象进行测量，以获得许多次观测值。

每次观测值不一定相同，所以对许多对象而言，各个特征分量都是随机变量，即许多对象的特征向量在n维空间中呈随机性分布，称为随机矢量。

1.3 随机矢量的描述

(一)随机矢量的分布函数:

设 $\vec{X}=(X_1,X_2,\cdots,X_n)'$ 为随机矢量,

$\vec{x}=(x_1,x_2,\cdots,x_n)'$为确定性矢量。

随机矢量的联合概率分布函数定义为:

$$F(x_1,x_2,\cdots,x_n)=P(X_1\le x_1,X_2\le x_2,\cdots,X_n\le x_n)$$

$$F(\vec{x})=P(\vec{X}\le\vec{x})$$

式中$P(\cdot)$表示括号中事件同时发生的概率。

1.3　随机矢量的描述

(一)随机矢量的分布函数：

随机矢量$\vec{X}$ 的联合概率密度函数定义为：

$$p(x_1, x_2, \cdots, x_n) \triangleq p(\vec{x})$$
$$= \partial^n F(x_1, x_2, \cdots, x_n) / \partial x_1 \partial x_2 \cdots \partial x_n$$

1.3 随机矢量的描述

(一)随机矢量的分布函数:

设集合由c类模式组成，第i类为 ω_i ，

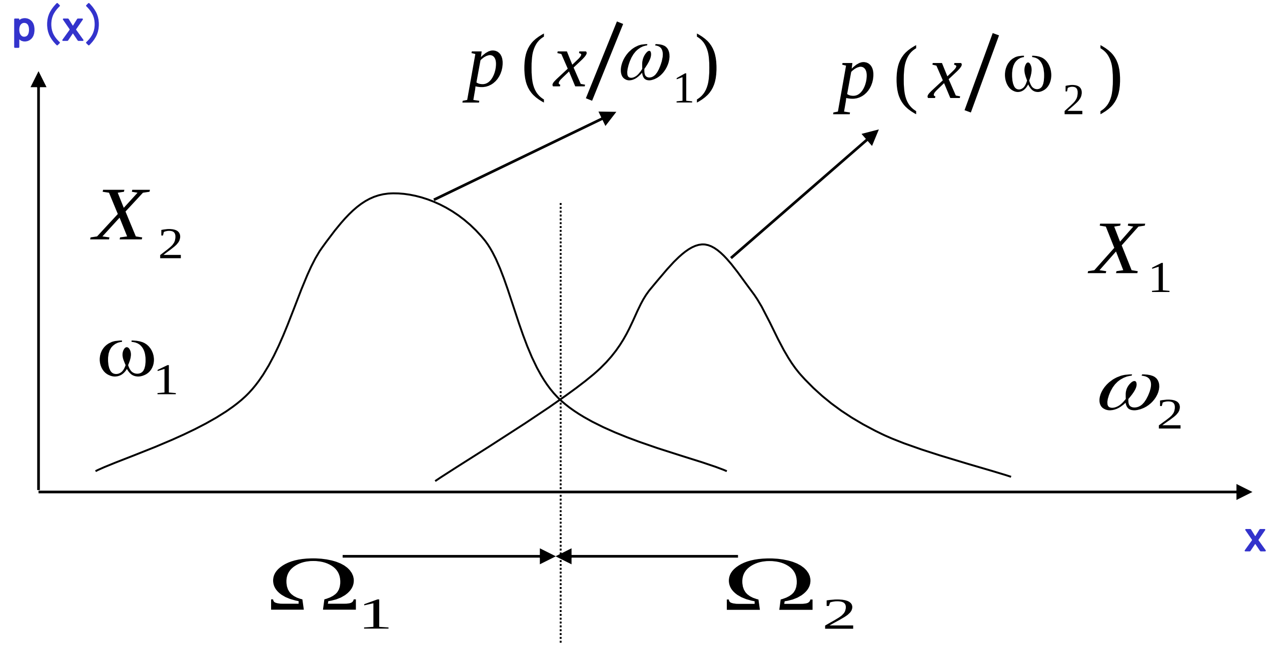
ω_i 类模式特征矢量有其自己的分布函数和密度函数。

ω_i 类的模式特征矢量的分布函数及密度函数分别定义为

$$F(\vec{x}|\omega_i) = P(\vec{X} \le \vec{x}|\omega_i)$$

$$p(\vec{x} \mid \omega_i) = \partial^n F(x_1, x_2, \cdots, x_n \mid \omega_i) / \partial x_1 \partial x_2 \cdots \partial x_n$$

1. 3　随机矢量的描述



1.3 随机矢量的描述

(二) 随机矢量的数字特征:

(1) 均值矢量(期望矢量)

n 维随机矢量 $\vec{X}$ 的数学期望 $\vec{\mu}$ 定义为:

$$\vec{\mu} = \mathrm{E}[\vec{X}] = \bar{\vec{X}} = \begin{pmatrix} \mathrm{E}[X_1] \\ \mathrm{E}[X_2] \\ \cdots \\ \mathrm{E}[X_n] \end{pmatrix} = \int_{X^n} \vec{x} p(\vec{x}) \mathrm{d}\vec{x}$$

随机变量X的数学期望（或称均值）记作E(X)，它描述了随机变量的取值中心。随机变量$(X - E(X))^2$的数学期望称为X的方差，记作σ^2，而σ称为X的均方差（标准差）。它们描述了随机变量的可能取值与均值的偏差的疏密程度。

若X是连续型随机变量，其分布密度为p(x)，则（当积分绝对收敛时）

$$m = E(X) = \int_{-\infty}^{\infty} xp(x)dx$$

$$\sigma^2 = E\{(X-m)^2\} = \int_{-\infty}^{\infty} (x-m)^2 p(x)dx$$

若X是离散型随机变量，其可能取值为x_k，k=1,2,...，且$P(X=x_k) = p_k$，则（当级数是绝对收敛时）

$$m = E(X) = \sum_{k=1}^{\infty} x_k p_k$$

$$D(X) = \sum_{k=1}^{\infty} (x_k - m)^2 p_k$$

1.3 随机矢量的描述

(二)随机矢量的数字特征：

其中，$\vec{\mu}$ 的分量：

$$\mu_i = \mathrm{E}[X_i] = \int_{-\infty}^{+\infty} x_i p(x_i)\mathrm{d}x_i =$$

$$\int_{-\infty}^{+\infty} \cdots \int x_i p(x_1, x_2, \ldots, x_n)\mathrm{d}x_1\mathrm{d}x_2 \ldots \mathrm{d}x_n = \bar{X}_i$$

式中 $p(x_i)$ 是 $\vec{X}$ 的第 i 个分量的边缘密度。随机矢量 $\vec{X}$ 的均值矢量 $\vec{\mu}$ 的各分量是相应的各随机分量的均值。

1.3 随机矢量的描述

(二)随机矢量的数字特征：

(2) 条件期望

在模式识别中，经常以类别ω_i 作为条件，在这种情况下随机矢量$\vec{x}$ 的条件期望矢量定义为

$$\vec{\mu}_{\omega_i} = \mathrm{E}[\vec{X} \mid \omega_i] = \int_{X^n} \vec{x} p(\vec{x} \mid \omega_i) \mathrm{d}\vec{x}$$

1. 3　随机矢量的描述

(二)随机矢量的数字特征：

(3) 协方差矩阵

随机矢量$\vec{X}$ 的自协方差矩阵表征各分量围绕其均值的散布情况及各分量间的相关关系，其定义为：

$$\Sigma = \mathrm{E}[(\vec{X} - \bar{\vec{X}})(\vec{X} - \bar{\vec{X}})']$$

$$= \int_{X^n} (\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})' p(\vec{x}) \mathrm{d}\vec{x} = (\sigma_{ij}^2)_{n \times n}$$

1.3　随机矢量的描述

(二) 随机矢量的数字特征：

(3) 协方差矩阵

式中σ_{ij}^2是$\vec{X}$的第i个分量与第j个分量的协方差，当 $i=j$ 时，σ_{ij}^2 便是X_i的方差。

$$\sigma_{ij}^2 = \mathrm{E}[(X_i - \overline{X}_i)(X_j - \overline{X}_j)]$$

$$= \iint_{-\infty}^{+\infty} (x_i - \mu_i)(x_j - \mu_j)\, p(x_i, x_j)\, \mathrm{d}x_i \mathrm{d}x_j$$

Definition [edit]

Throughout this article, boldfaced unsubscripted **X** and **Y** are used to refer to random vectors, and unboldfaced subscripted X_i and Y_i are used to refer to random scalars.

If the entries in the column vector

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$

are random variables, each with finite variance, then the covariance matrix Σ is the matrix whose (i, j) entry is the covariance

$$\Sigma_{ij} = \mathrm{cov}(X_i, X_j) = \mathrm{E}\begin{bmatrix}(X_i - \mu_i)(X_j - \mu_j)\end{bmatrix}$$

where

$$\mu_i = \mathrm{E}(X_i)$$

is the expected value of the ith entry in the vector **X**. In other words,

$$\Sigma = \begin{bmatrix} \mathrm{E}[(X_1 - \mu_1)(X_1 - \mu_1)] & \mathrm{E}[(X_1 - \mu_1)(X_2 - \mu_2)] & \cdots & \mathrm{E}[(X_1 - \mu_1)(X_n - \mu_n)] \\ \mathrm{E}[(X_2 - \mu_2)(X_1 - \mu_1)] & \mathrm{E}[(X_2 - \mu_2)(X_2 - \mu_2)] & \cdots & \mathrm{E}[(X_2 - \mu_2)(X_n - \mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n - \mu_n)(X_1 - \mu_1)] & \mathrm{E}[(X_n - \mu_n)(X_2 - \mu_2)] & \cdots & \mathrm{E}[(X_n - \mu_n)(X_n - \mu_n)] \end{bmatrix}.$$

Scatter matrix

From Wikipedia, the free encyclopedia

For the notion in quantum mechanics, see scattering matrix.

In multivariate statistics and probability theory, the **scatter matrix** is a statistic that is used to make estimates of the covariance matrix, for instance of the multivariate normal distribution.

Definition [edit]

Given n samples of m-dimensional data, represented as the m-by-n matrix, $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$, the sample mean is

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j$$

where $\mathbf{x}_j$ is the jth column of X.

The **scatter matrix** is the m-by-m positive semi-definite matrix

$$S = \sum_{j=1}^{n} (\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})^T = \sum_{j=1}^{n} (\mathbf{x}_j - \overline{\mathbf{x}}) \otimes (\mathbf{x}_j - \overline{\mathbf{x}}) = \left(\sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j^T\right) - n\overline{\mathbf{x}}\overline{\mathbf{x}}^T$$

where T denotes matrix transpose, and multiplication is with regards to the outer product. The scatter matrix may be expressed more succinctly as

$$S = X\, C_n\, X^T$$

where C_n is the n-by-n centering matrix.

1.3 随机矢量的描述

(二)随机矢量的数字特征:

(4) 自相关矩阵

随机矢量$\vec{X}$的自相关矩阵定义为

$$R = \mathrm{E}[\vec{X}\vec{X}']$$

由定义可知,$\vec{X}$的协方差矩阵和自相关矩阵间的关系是

$$\Sigma = R - \overline{\vec{X}}\,\overline{\vec{X}}' = R - \vec{\mu}\vec{\mu}'$$

1.3 随机矢量的描述

(二)随机矢量的数字特征:

(4) 相关系数

$$r_{ij} = \sigma_{ij}^2 / (\sigma_{ii}\sigma_{jj})$$

由布尼亚科夫斯基不等式知: $\left|\sigma_{ij}^2\right| \le \sigma_{ii}\sigma_{jj}$

$$-1 \le r_{ij} \le 1$$

相关系数矩阵定义为 : $r = (r_{ij})_{n\times n}$

1.3　随机矢量的描述

(二)随机变量、随机矢量间的统计关系

(1) 不相关

随机矢量 $\vec{X}$ 的第i个分量X_i 和第j个分量X_j，若有

$$\sigma_{ij}^2 = E[(X_i - \overline{X}_i)(X_j - \overline{X}_j)] = 0, \quad (i \neq j)$$

则称它们不相关。这等价于

$$E(X_i X_j) = E[X_i]E[X_j]$$

1. 3　随机矢量的描述

（二）随机变量、随机矢量间的统计关系

(1) 不相关

随机矢量$\vec{X}$和$\vec{Y}$不相关的充要条件是互协方差矩阵：

$$\mathrm{cov}(\vec{X},\vec{Y}) = \phi$$

亦即

$$E(\vec{X}\vec{Y}') = E[\vec{X}]E[\vec{Y}']$$

1.3 随机矢量的描述

(二) 随机变量、随机矢量间的统计关系

(2) 正交

随机矢量 $\vec{X}$ 和 $\vec{Y}$ 若满足

$$\mathrm{E}[\vec{X}\vec{Y}] = 0$$

则称 $\vec{X}$ 和 $\vec{Y}$ 正交。

1.3　随机矢量的描述

(二) 随机变量、随机矢量间的统计关系

(3) 独立

随机矢量 $\vec{X}$ 和 $\vec{Y}$ 的联合概率密度函数

$p(\vec{x},\vec{y})$ 若满足

$$p(\vec{x},\vec{y}) = p(\vec{x})\,p(\vec{y})$$

则称 $\vec{X}$ 和 $\vec{Y}$ 独立。

独立→不相关
不相关≠独立

1.4　正态分布

(1)一维随机变量的正态分布

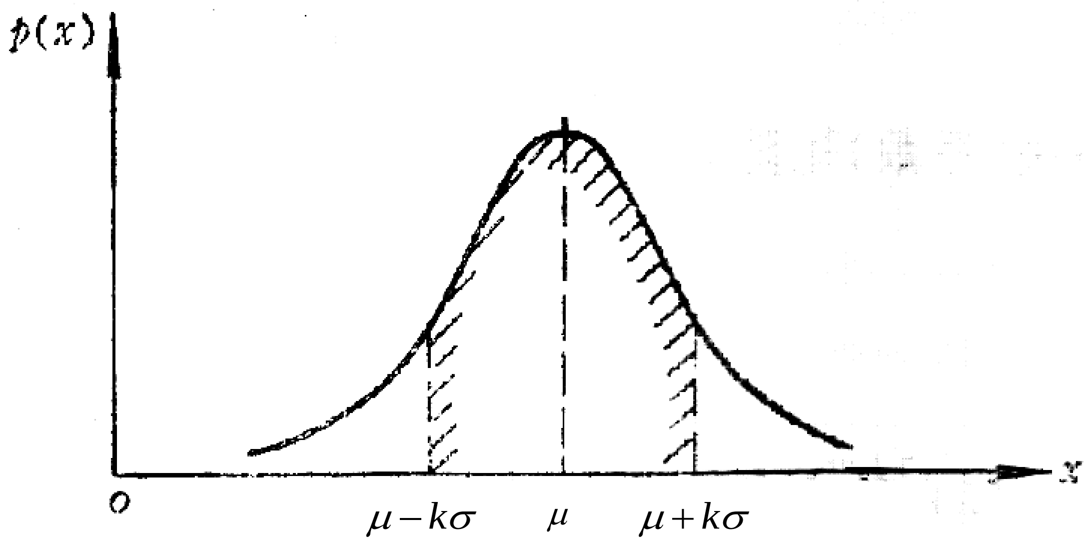
正态分布的一维随机变量X的概率密度函数

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

式中，μ 为数学期望，σ^2 为方差。

1.4　正态分布

（1）一维随机变量的正态分布



1.4　正态分布

（1）一维随机变量的正态分布

正态分布概率密度函数由两个参数μ 和σ^2就可以完全确定。

$$\mu = E[X] = \int_{-\infty}^{\infty} xp(x)dx$$

$$\sigma^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx$$

1.4　正态分布

（2）随机矢量的正态分布

正态分布随机矢量 $\vec{X}=(X_1, X_2, \cdots, X_n)'$

的概率密度函数定义为：

$$p(x_1, x_2, \cdots, x_n)=p(\vec{x})=\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp[-\frac{1}{2}(\vec{x}-\vec{\mu})'\Sigma^{-1}(\vec{x}-\vec{\mu})]$$

$$p(x_1, x_2, \cdots, x_n)=p(\vec{x})=\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp[-\frac{1}{2}(\vec{x}-\vec{\mu})'\Sigma^{-1}(\vec{x}-\vec{\mu})]$$

1.4 正态分布

（2）随机矢量的正态分布

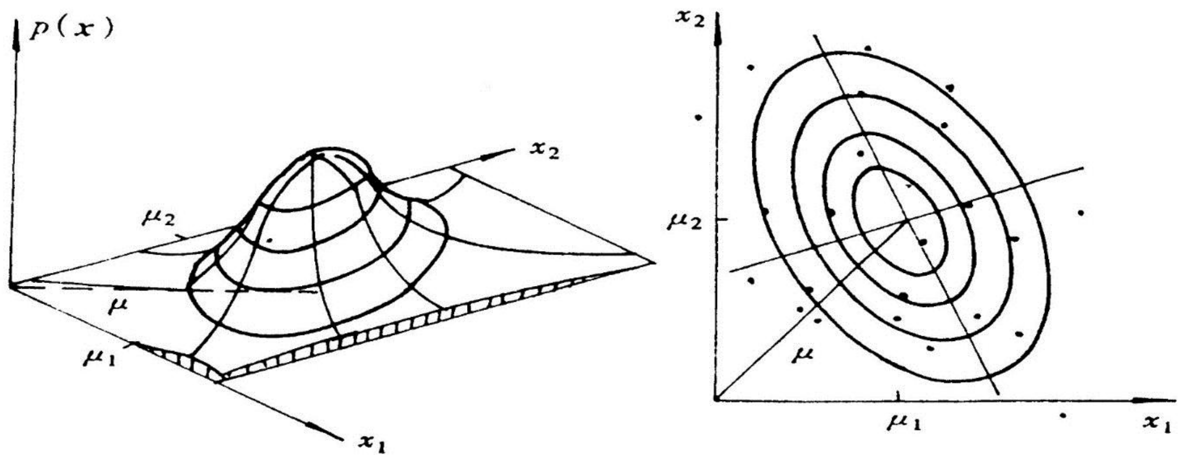
式中，$\vec{\mu} = E[\vec{X}] = (E[X_1], E[X_2], \cdots, E[X_n])'$

为 X 的数学期望矢量；Σ 为 X 的协方差矩阵，

$$\Sigma = E[(\vec{X} - \vec{\mu})(\vec{X} - \vec{\mu})'] = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2n}^2 \\ \vdots & \vdots & & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma_{nn}^2 \end{pmatrix}$$

1.4 正态分布

（2）二维随机变量的正态分布



(a) 二维正态分布概密函数　　(b) 二维正态分布等概密点轨迹

1.4 正态分布

（3）正态分布的随机矢量的性质

- 分布函数完全由 $\vec{\mu}$ 和 Σ 确定
- 等概率密度点的轨迹为一超椭球面
- 对正态分布，不相关等价于独立
- 其边缘密度和条件密度仍然是正态分布
- 正态分布随机矢量的线性变换仍为正态随机矢量