



模式识别与机器学习

Pattern Recognition and Machine Learning



主讲教师

- 李 映 教授
- 办公室：计算机学院204房间
- Email: lybyp@nwpu.edu.cn
- Tel (wechat) : 13891433893

- 魏巍 副教授
- 办公室：计算机学院103房间
- Email: weiweinwpu@nwpu.edu.cn
- Tel: 13772538134



参考书籍

- 李航，统计学习方法，清华大学出版社
- 周志华，机器学习，清华大学出版社
- 李映，模式识别与机器学习，讲义，2016
- **Christopher M. Bishop, Pattern Recognition and Machine Learning, 2006**



参考文献

CCF推荐的相关A或B类期刊及会议

- IEEE Transaction on Pattern Analysis and Machine Intelligence (TPAMI)
- International Journal of Computer Vision (IJCV)
- CVPR(International Conference on Computer Vision and Pattern Recognition)
- ICML(International conference on Machine Learning)
- Pattern Recognition
- AAI、IJCAI、ICCV、TIP、ECCV等



内容安排

- 模式识别和机器学习概述
- 模式识别和机器学习的研究内容
- 模式识别和机器学习的发展动向
- 模式识别和机器学习的基本方法



目 录

- 第一章 绪论
 - 引言
 - 模式识别的基本概念
 - 模式识别系统
 - 机器学习的主要方法
 - 随机变量及分布



目 录

- 第二章 贝叶斯统计决策理论
 - 引言
 - 最小错误率判决规则
 - 最小风险判决规则
 - 最大似然比判决规则
 - Neyman-Pearson判决规则
 - 最小最大判决规则
 - 分类器设计
 - 正态分布时的统计决策
 - 小结



目 录

- 第三章 概率密度函数的估计
 - 引言
 - 最大似然估计
 - 贝叶斯估计
 - **EM**估计方法
 - 非参数估计方法
 - 小结



目 录

- 第四章 线性回归分析
 - 多元线性回归分析
 - 脊回归方法
 - 核回归方法
 - 小结



目 录

- 第五章 线性判别函数
 - 引言
 - 线性判别函数和决策面
 - 广义线性判别函数
 - 感知器
 - 最小平方误差判别
 - 小结



目 录

- 第六章 其他分类方法
 - 邻近法
 - 决策树与随机森林
 - 罗杰斯特回归
 - **Boosting**方法
 - 小结



目 录

- 第七章 无监督学习和聚类

- 引言
- 基于模型的方法
- 混合模型的估计
- 动态聚类算法
- 分层聚类方法
- 谱聚类方法
- 模糊聚类方法
- 小结



目 录

- 第八章 人工神经网络
 - 引言
 - 多层神经网络
 - 自组织映射神经网络
 - 深度学习
 - 小结



目 录

- 第九章 核方法和支持向量机
 - 引言
 - 核学习机
 - 支持向量分类机
 - 支持向量回归机
 - 小结



目 录

- 第十章 特征选择与提取
 - 引言
 - 特征的评价准则
 - 特征选择的方法
 - 非线性特征提取方法
 - 多维尺度法
 - 小结



第一章

绪论



研究目的

- 模式识别和机器学习是一门交叉学科，涉及：
 - **数学**：概率论、线性代数、模糊数学、离散数学
 - **计算机**：计算机理论、自动机理论
 - **信息科学**：信号处理、图象处理、形式语言与专家系统
- 提高计算机的感知能力
- 建立智能化的人机接口(HCI)



目 录

- 引言
- 模式识别的基本概念
- 模式识别系统
- 机器学习的主要方法
- 随机变量及分布

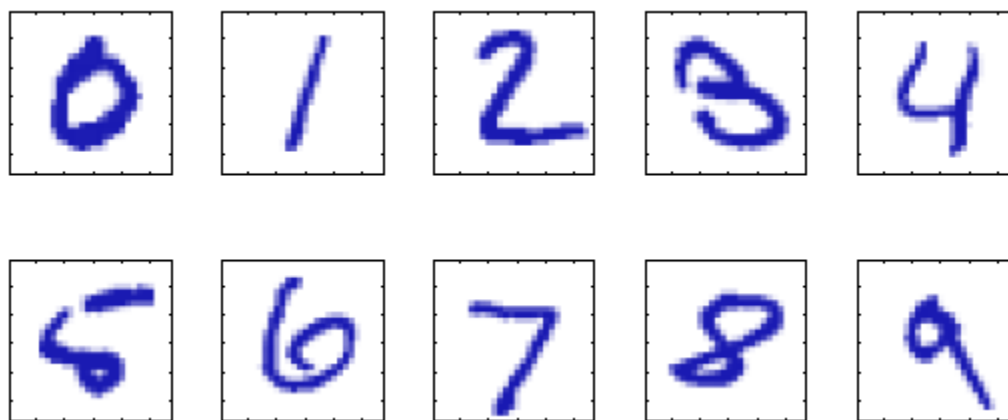


1.1 引言

- 寻找数据中的模式

例：手写数字的识别

如下图所示，每个数字相当于一个 28×28 像素的图像，因此可以表示为由784个数字组成的矢量。我们的目标是建立一个机器，将这样一个向量作为输入，并将产生的辨别数字0,...,9作为输出。





手写数字的识别

- 学习阶段

利用一个大小的 N 个数字的集合 (训练样本), 来自适应地调整模型的参数。这些在训练集中的数字分类被认为是预先知道的。

- 测试阶段

运行机器学习算法的结果可以看作是一个函数，此函数以一个新的数字图像 x 作输入，以一个输出向量 y 作输出。一旦模型训练完成，它就可以用来辨别新的图像。



1.2 模式识别的基本概念

1.2.1 模式和模式识别

- 例1：医生给一个病人看病
- 测量病人的体温和血压，化验血沉，询问临床表现；
- 通过综合分析，抓住主要病症；
- 医生运用自己的知识，根据主要病症，作出正确的诊断。



分析

- 用模式识别的观点进一步分析：
- 样本：医院里的众多患者，每个患者都是一个样本。
- 单一样本：请医生给出诊断的某一个患者，就是众多患者中的一个样本。
- 样本的测量值：患者的体温、血压等测量值。
- 模式：样本的各测量值的综合。
- 模式样本：具有某种模式的样本。



分析

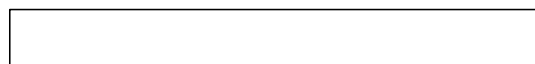
- **模式采集**：获取某样本的各测量值的过程。
- **样本特征**：患者的主要病症。
- **特征提取、特征选择**：模式样本各测量值经过综合分析找出主要病症。
- **分类判决**：医生运用自己的知识作出诊断。
- **判决准则、判决规则**：医生的知识。
- **判决结果**：把患者区别成某种疾病的患者。
也就是把样本（患者）区别为相应类型（疾病）。



举例2

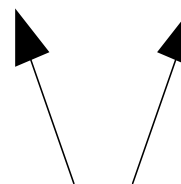
- 例2：选取做家具的木材。

型材



单一样本

包括 松木、桦木



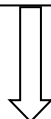
木材的类型

怎样区分它的类型？

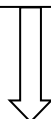


解答

观察：颜色、花纹、亮度、密度（样本测量值），
得到样本模式



综合分析：
提取、选择主要特征



根据主要区别：区分各单一样本
（松木、桦木）



模式和模式识别

- 基本概念
- 模式(pattern)是取自世界有限部分的单一样本的被测量值得综合。

模式识别(Pattern Recognition)就是试图确定一个样本的类型的属性，也就是把某一个样本归属于多个类型中的某一个类型。



1.2.2 模式空间、特征空间和类型空间

- 从模式识别的技术途径来说

模式空间→特征空间→类型空间



模式空间、特征空间和类型空间

(1) 模式空间

所有的样本观测数据构成**模式空间**。

其维数与所选择的样本和观测方法有关，也与特定的应用有关，一般来说是很大的，但是有限值。在模式空间，每个模式都是一个点，点的位置由该模式在各维上的观测数据来确定。



模式空间、特征空间和类型空间

(2) 特征空间

获得最能揭示样本属性的观测量作为主要特征，这些主要特征就构成了**特征空间**。

其维数大大压缩了，每个样本在特征空间也是一个点，点的位置由该样本的各特征值来确定。

模式空间 $\xrightarrow{\text{特征提取 (选择)}}$ 特征空间



模式空间、特征空间和类型空间

(3) 类型空间

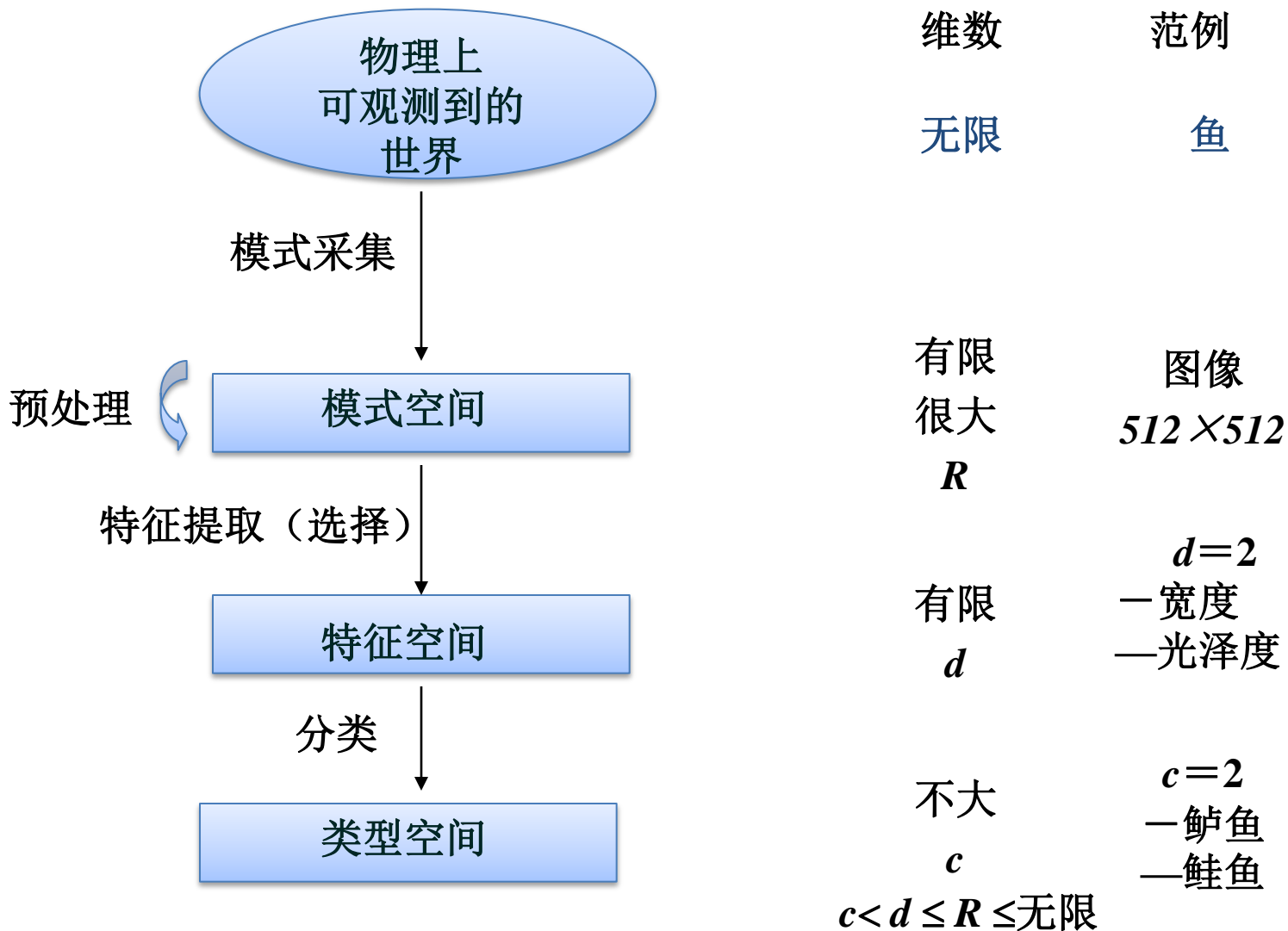
根据适当的判决规则，把特征空间里的样本区分成不同的类型，从而把特征空间塑成了**类型空间**。

其维数与类型的数目相等。





模式识别过程





1.2.3 预处理

- 预处理的功能包括：
 - 清除或减少模式采集中的噪声及其它干扰，**提高信噪比**
 - 消除或减少数据图像的模糊及几何失真，**提高清晰度**
 - **转变模式的结构**，以便后续处理（如非线性模式转为线性模式）



预处理

➤ 预处理的方法包括：

- 滤波
- 变换
- 编码
- 标准化
- A/D转换



预处理----- A/D转换

- 为了便于计算机处理，往往需要将模拟量转化为数字量，也就是进行A/D转换。在此过程中必须考虑2个问题：采样间隔与量化等级。
- 采样间隔（采样频率）：表示单位时间内（秒），要求多少个采样值。
- 量化级：表示每个采样值要有多少个量化级，才能满足要求。



1.2.4 特征提取/选择

- 通过特征抽取，可把在高维向量空间中的向量映射到低维特征空间中。**特征提取就是找出模式的充分表示。**
- **特征提取的分类准则：**在该准则下，选择对分类贡献较大的特征，删除贡献甚微的特征。



特征抽取方法

- 通过**样本的统计学习**得到
 - 常见的有**人工神经网络(ANN)**、**支撑向量机(SVM)**等。
- 通过**某种变换**得到
 - **傅立叶变换**
 - **分形维数**



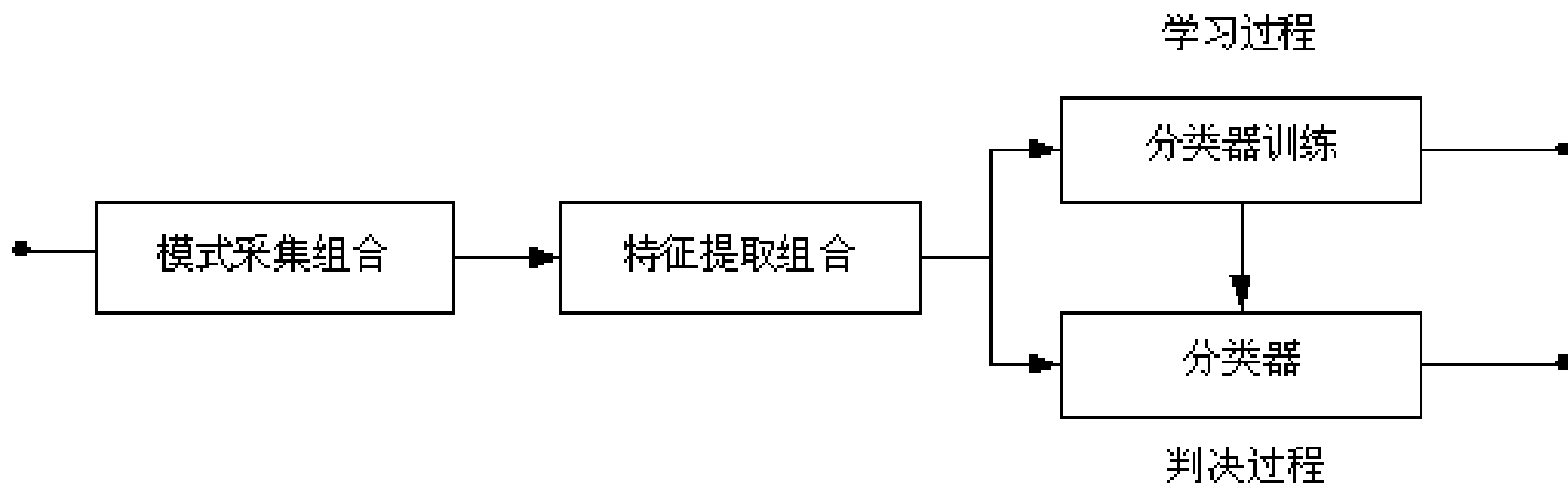
1.2.5 分类

- 分类目标：
 - (1) 把特征空间划分成类型空间。
 - (2) 把未知类别属性的样本确定为类型空间的某一个类型。
 - (3) 在给定条件下，可以否定样本属于某种类型。
- 实际分类过程中，对于预先给定的条件，分类中出现错误是不可避免的。因此，分类过程只能以某种错误率来完成。



1.3 模式识别系统

- 一个模式识别系统应该完成模式采集、特征提取/选择、分类等功能。系统框图如下：



模式识别系统框图



模式采集组合

- **目的：** 完成模式的采集。
- **方法：** 根据处理对象的不同，可以选用不同的传感器、测量装置、图像录取输入装置等。采集之后，还要进行滤波、消除模糊、减少噪声、纠正几何失真等预处理操作。



特征提取组合

- **目的：**实现由模式空间到特征空间的转变，有效压缩维数。
- 一般来说，特征提取组合应该是在一定分类准则下的最佳或次佳变换器，或是实现某特征选择算法的装置。



分类器

- **目的：**用来实现对未知类别属性样本的分类判决。
- 为了设计分类器，首先要确定对分类错误率的要求，选用适当的判决规则。但是为了使分类器能有效地进行分类判决，还必须对它进行训练。也就是，分类器首先要进行学习。



分类器学习

- 分类判决规则常常是样本各特征的函数，训练过程就是要**确定函数的所有权因子**。这个过程是一个输入、修正、再输入、再修正，不断反复的过程，直到分类错误率不大于给定值为止。
- 分类器完成训练之后，根据已经确定的判决规则，对未知类别属性的样本进行分类。此时，分类器就具有自动识别的能力。



实例1

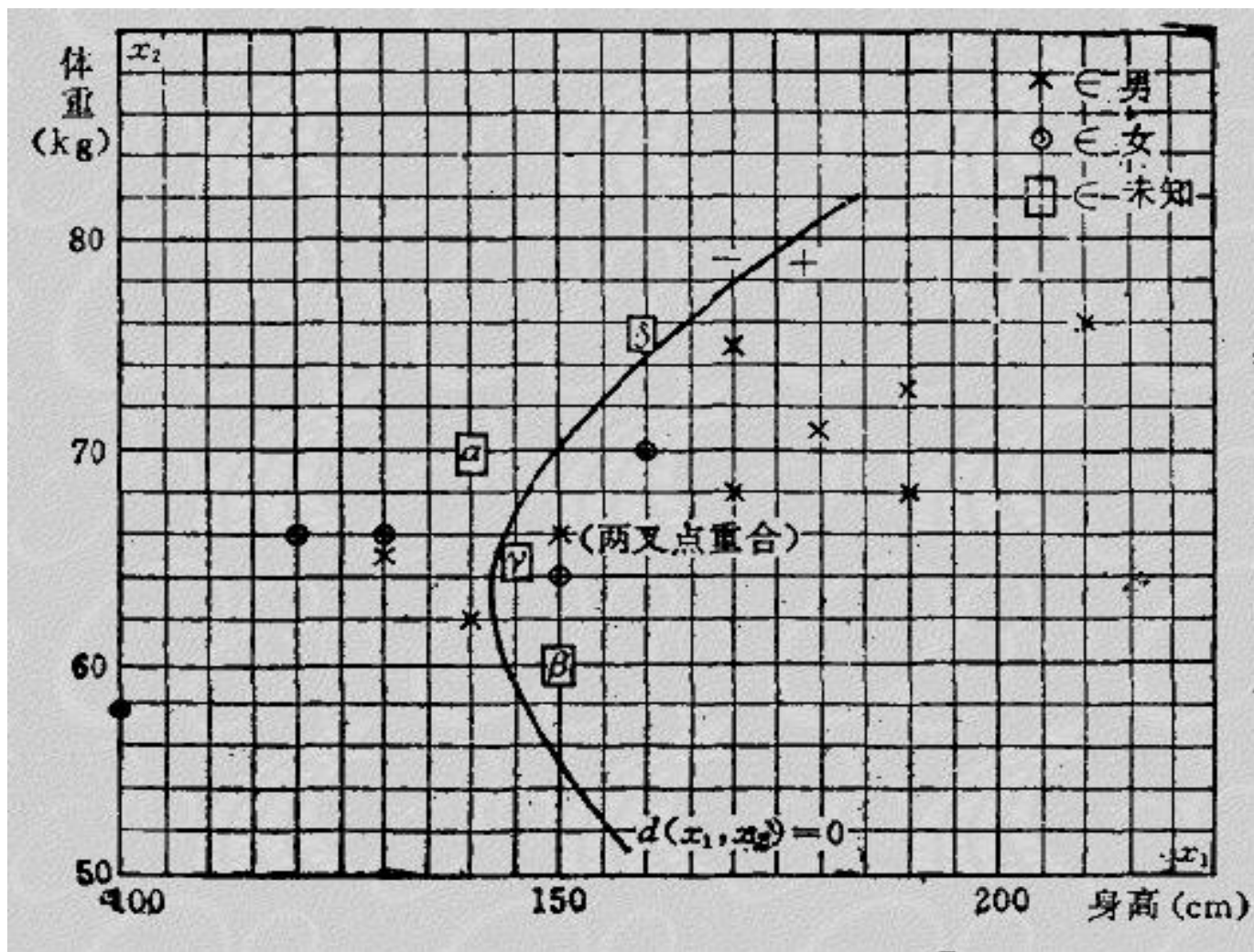
- 例1: 男女19人进行体检，测量身高和体重，但事后发现4人忘了写性别，试问这4人是男是女？

序号	身高cm	体重kg	性别	序号	身高cm	体重kg	性别
1	170	68	男	11	140	62	男
2	130	66	女	12	150	64	女
3	180	71	男	13	120	66	女
4	190	73	男	14	150	66	男
5	160	70	女	15	130	65	男
6	150	66	男	16	140	70	$\alpha?$
7	190	68	男	17	150	60	$\beta?$
8	210	76	男	18	145	65	$\gamma?$
9	100	58	女	19	160	75	$\delta?$
10	170	75	男				



解答

- 试验样本是人，分为男、女两个类别。二维的主要特征是身高、体重，构成二维特征空间。已知**15**人的性别，可以作为训练样本，根据其值确定他们在特征空间的位置。如图所示。



- 特征空间里的样本分布图



解答

图中，男性集中于右上方，女性集中于左下方，这就是聚类性质。

采用数理统计方法，可在两个性别之间描绘一条曲线，它是特征 x_1 (身高)、 x_2 (体重) 的函数，表示为 $d(x_1, x_2) = 0$ 。

可以确定判决规则：

$$\begin{cases} d(x_1, x_2) > 0, & \mathbf{x} \in \text{男} \\ d(x_1, x_2) < 0, & \mathbf{x} \in \text{女} \end{cases}$$



解答

其中, $\mathbf{x} = (x_1, x_2)$ 是向量, 称为**模式向量**。

$d(x_1, x_2) = 0$ 描绘的曲线称为分界线。

- 现考察**16--19**号体检者, 由身高、体重确定在上图中的位置。显然, **16、19**在负线一侧, 判定他们为**女性**。**17、18**位于正线一侧, 判为**男性**。上述判决方法使分类错误率最小。



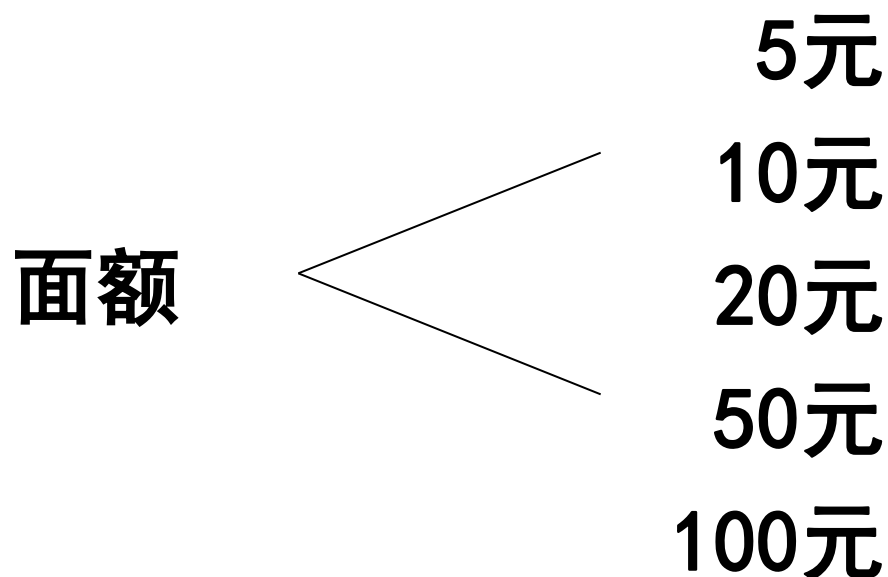
实例2：纸币识别器





实例2：纸币识别器

纸币识别器对纸币按面额进行分类





实例2：纸币识别器

	长度 (mm)	宽度 (mm)
5元	136	63
10元	141	70
20元	146	70
50元	151	70
100元	156	77



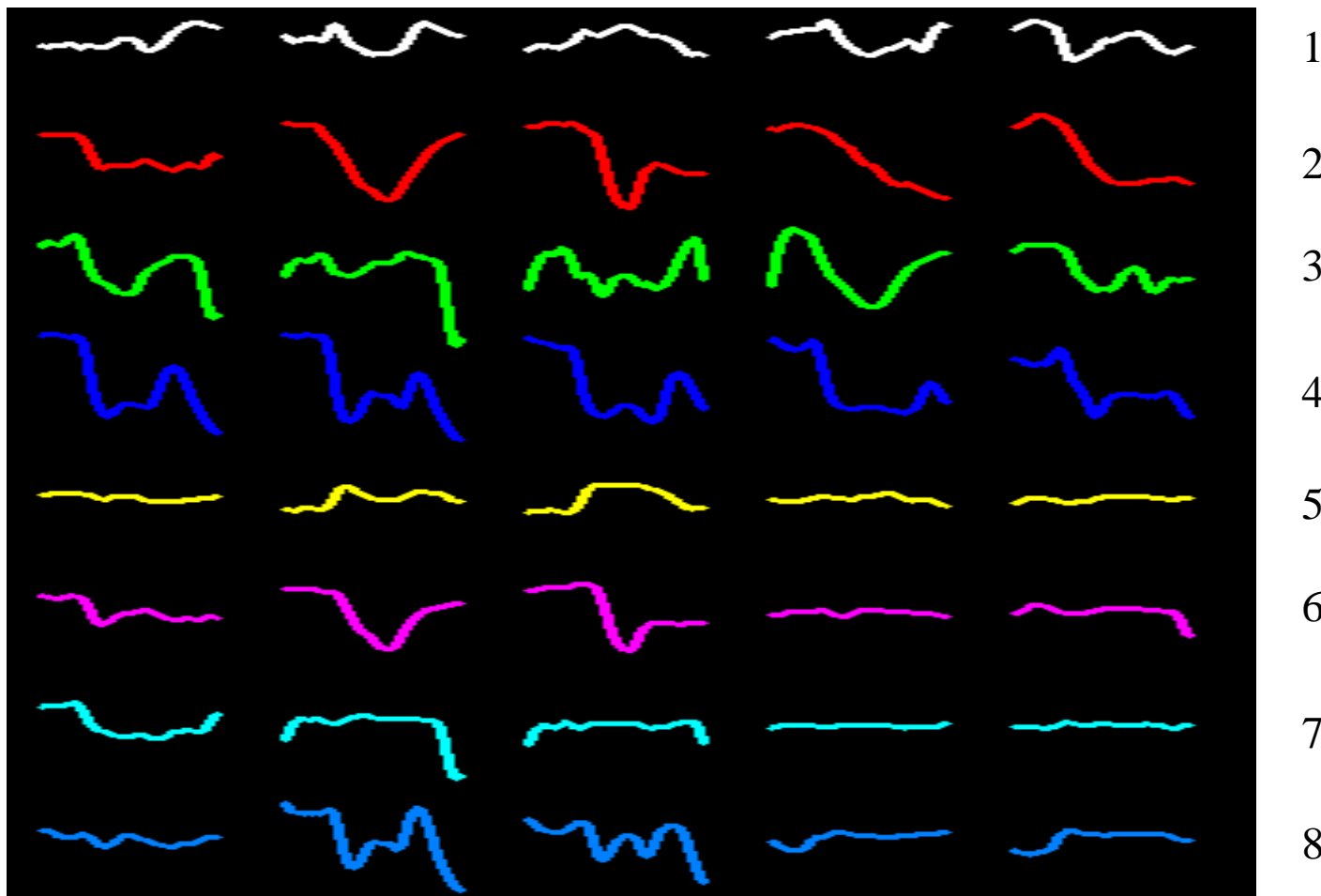
实例2：纸币识别器

	磁性	金属条位置(大约)
5元	有	54/82
10元	有	54/87
20元	有	57/89
50元	有	60/91
100元	有	63/93



5元 10元 20元 50元 100元

反射光波形图





实例2：纸币识别器

- **数据采集、特征提取：**

长度、宽度、磁性、磁性的位置，光反射亮度、光透射亮度等等

- **特征选择：**

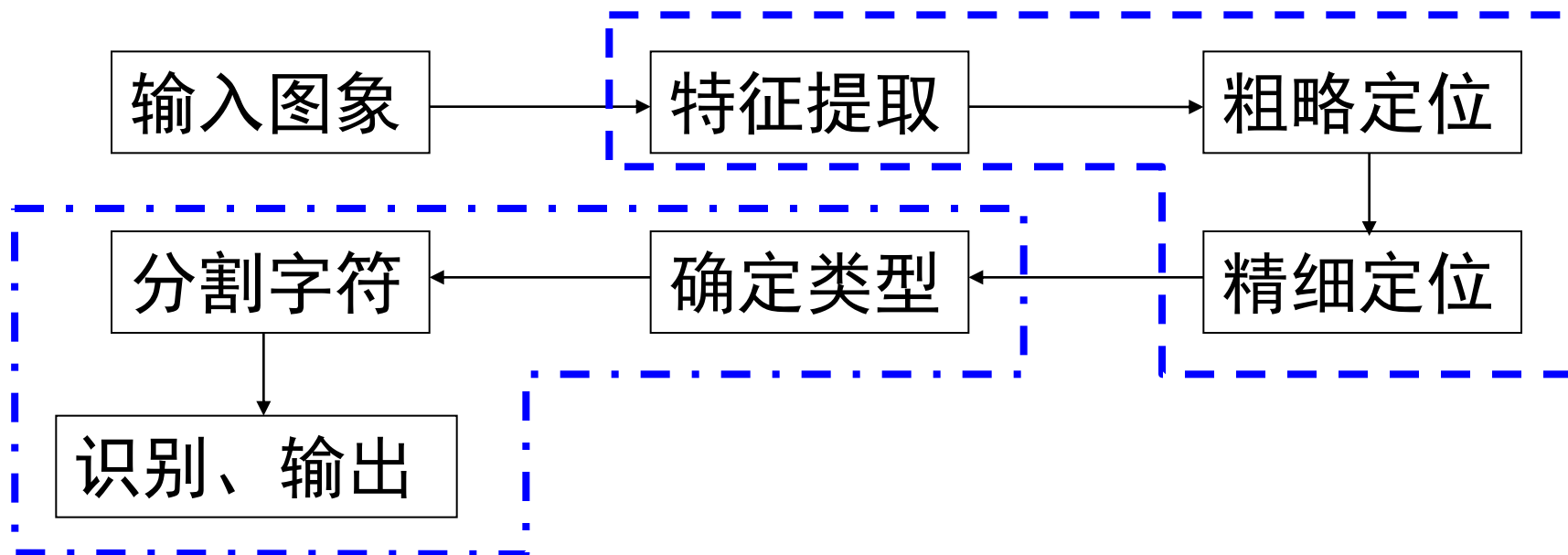
长度、磁性及位置、反射亮度

- **分类识别：**

确定纸币的面额及真伪

实例3：汽车车牌识别

- 从摄像头获取包含车牌的彩色图象
- 车牌定位和获取
- 字符分割和识别



试验结果（车牌提取）



原图



边缘图

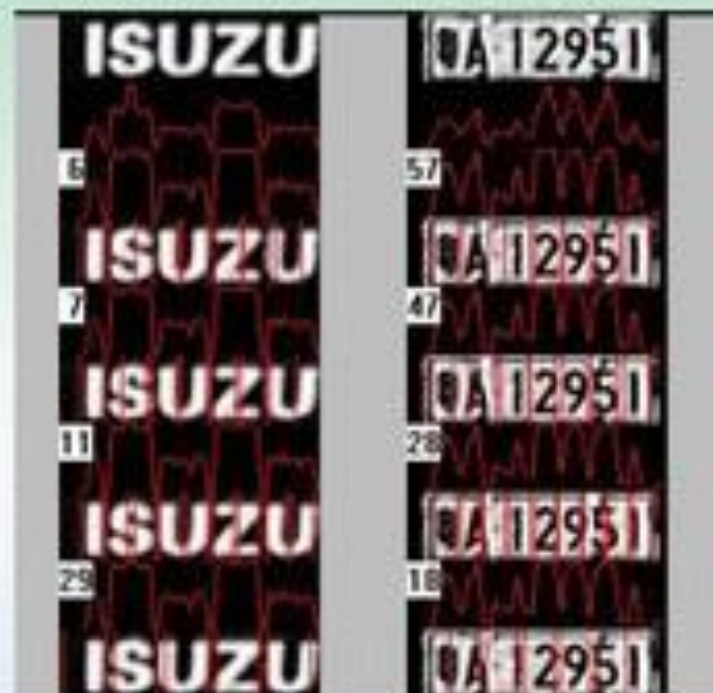


粗略定位



精细定位

试验结果（字符识别）





- 英文字母中的I和O一般避而不用，以免和数字中的1和0混淆；
- 车牌大小为440×140mm；
- 汉字字体为黑体，数字和字母字体为国家公安部门开发的特殊字体。



实例4：图像检索

查询图像：





检索结果



☐ 好 ☒ 中 ☐
差



☐ 好 ☒ 中 ☐
差



☐ 好 ☒ 中 ☐
差



☐ 好 ☒ 中 ☐
差



☐ 好 ☒ 中 ☐
差



☐ 好 ☒ 中 ☐
差



☐ 好 ☒ 中 ☐
差



☐ 好 ☒ 中 ☐
差



☐ 好 ☒ 中 ☐
差



☐ 好 ☒ 中 ☐
差



☐ 好 ☒ 中 ☐
差



☐ 好 ☒ 中 ☐
差

[返回](#)



检索结果



[返回](#) 60



检索结果



[返回](#)



1.4 机器学习的主要方法

- 机器学习

机器学习是近**20**多年兴起的一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。

机器学习算法是一类从数据中自动分析获得规律，并利用规律对未知数据进行预测的算法。



1.4.1 监督学习

➤ 监督学习

- **监督学习**从给定的训练数据集中学习出一个函数，当新的数据到来时，可以根据这个函数预测结果。**训练集中的目标是由人标注的**。常见的监督学习算法包括**回归分析**和**分类**。
- **回归分析**：函数的输出可以是一个连续的值
- **分类**：预测一个分类标签



分类器

➤ 目前最广泛被使用的分类器有：

- 人工神经网络
- 支持向量机
- 最近邻居法
- 高斯混合模型
- 朴素贝叶斯方法
- 决策树
- 径向基函数



1.4.2 无监督学习

➤ 无监督学习

- 无监督学习在学习时并不知道其分类结果是否正确，亦不知道何种学习是正确的。
其特点是仅提供输入样本，而它会自动从这些样本中找出其潜在类别规则。当学习完毕并经测试后，也可以将之应用到新的案例上。
- 常见的无监督学习算法有聚类。如k-means算法。



1.4.3 半监督学习

➤ 半监督学习

- 半监督学习(**Semi-supervised Learning**)是模式识别和机器学习领域研究的重点问题，是监督学习与无监督学习相结合的一种学习方法。它主要考虑如何利用少量的标注样本和大量的未标注样本进行训练和分类的问题。
- 半监督学习对于减少标注代价，提高学习机器性能具有非常重大的实际意义。



1.4.3 半监督学习

- 半监督学习算法有：
- 基于概率的算法
 - 在现有监督算法基础上作修改的方法
 - 依赖于聚类假设的方法
 - 基于多视图的方法
 - 基于图的方法



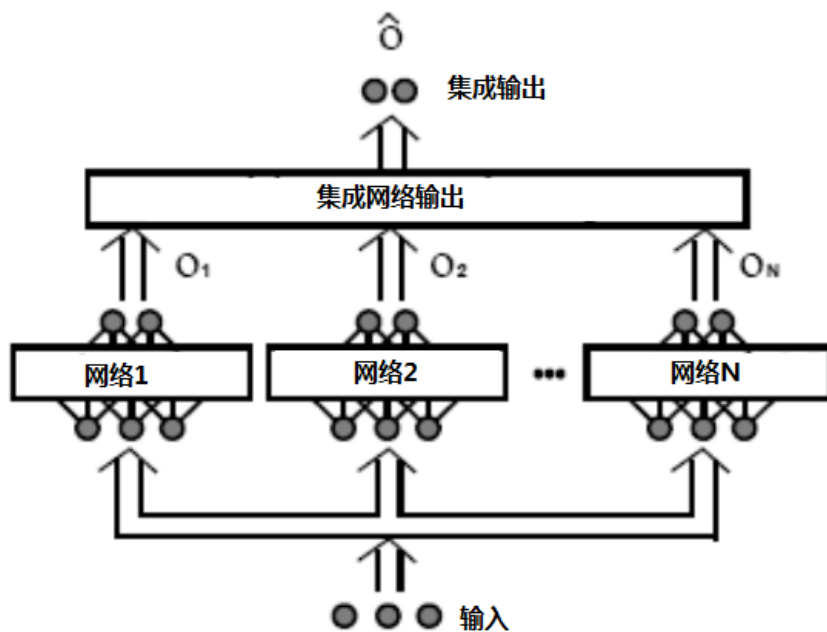
1.4.4 集成学习

➤ 集成学习

- 集成学习（**Ensemble Learning**）的思路是在对新的实例进行分类的时候，把若干个单个分类器集成起来，通过对多个分类器的分类结果进行某种组合来决定最终的分类，以取得比单个分类器更好的性能。如果把单个分类器比作一个决策者的话，集成学习的方法就相当于**多个决策者共同进行一项决策**。

集成学习

- 集成分类器包括N个单一的分类器，对于同样的输入分别给出各自的输出，这些输出通过整合以得到集成分类器整体的输出结果作为最终分类。



集成学习示意图



1.4.5 增强学习

➤ 增强学习

- **增强学习 (Q-learning)** 通过观察来学习做成如何的动作。每个动作都会对环境有所影响，学习对象根据观察到的周围环境的反馈来做出判断。
- **例：机器人对弈**
- 施教者可在游戏胜利时给出正回报，在游戏失败时给出负回报，其他时候为零回报。



1.5 随机变量及分布

- 一般地，如果A为某个随机事件，则一定可以通过如下示性函数使它与数值发生联系：

$$X = \begin{cases} 1, & \text{如果A发生} \\ 0, & \text{如果A不发生} \end{cases}$$



随机变量

- **随机变量**：试验结果能用一个数 X 来表示，这个数 X 是随着试验的结果不同而变化的，也即它是样本点的一个函数，这种量以后称为随机变量。
- **离散型随机变量**：试验结果 X 所可能取的值为有限个或至多可列个，我们能将其可能的结果一一列举出来，这种类型的随机变量称为离散型随机变量。



随机变量

- 连续型随机变量：与离散型随机变量不同，一些随机现象所出现的试验结果**不止取可列个值**，**例如**测量误差、分子运动速度、候车时的等待时间、降水量、风速、洪峰值等等，这时用来描述试验结果的随机变量还是样本点的函数，但是这随机变量能取某个区间 $[c, d]$ 或 $(-\infty, +\infty)$ 的一切值。此时，这种随机变量为连续型随机变量。



1.5.1 分布函数与参数

- 称

$$F(x) = P\{X(\omega) < x\}, -\infty < x < +\infty$$

为随机变量 X 的分布函数。

- 简记为 $X \sim F(x)$



概率分布

- 对于离散型随机变量，设 $\{x_i\}$ 为离散型随机变量的所有可能取值，而 $P(x_i)$ 是 X 取 x_i 的概率。

$\{p(x_i), i = 1, 2, 3 \cdots\}$ 称为随机变量 X 的**概率分布**，它应该满足下面关系：

$$p(x_i) \geq 0, i = 1, 2, 3 \cdots$$

$$\sum_{i=1}^{\infty} p(x_i) = 1$$



概率分布

- 对于离散型随机变量，有了概率分布，可以通过下式求得分布函数：

$$F(x) = P\{X < x\} = \sum_{x_k < x} p(x_k)$$



概率密度函数

- 对于连续型随机变量，这种随机变量可取某个区间 $[c, d]$ 或 $(-\infty, +\infty)$ 中的一切值，而且其分布函数 $F(x)$ 是绝对连续函数，即存在可积函数 $p(x)$ ，使

$$F(x) = \int_{-\infty}^x p(y) dy$$

- 其中， $p(y)$ 称为 X 的概率密度函数。



概率密度函数

- 概率密度函数 $p(x)$ 满足:

$$p(x) = F'(x)$$

$$p(x) \geq 0$$

$$\int_{-\infty}^{+\infty} p(x) dx = 1$$



常见的随机变量分布

➤ 离散型:

- 伯努利分布
- 二项分布
- 泊松分布

➤ 连续型:

- 均匀分布
- 正态分布



伯努利分布

- 在一次试验中，事件A出现的概率为 p ，不出现的概率为 $q=1-p$ ，若以 β 记事件A出现的次数，则 β 仅取0,1两值，相应的概率分布为

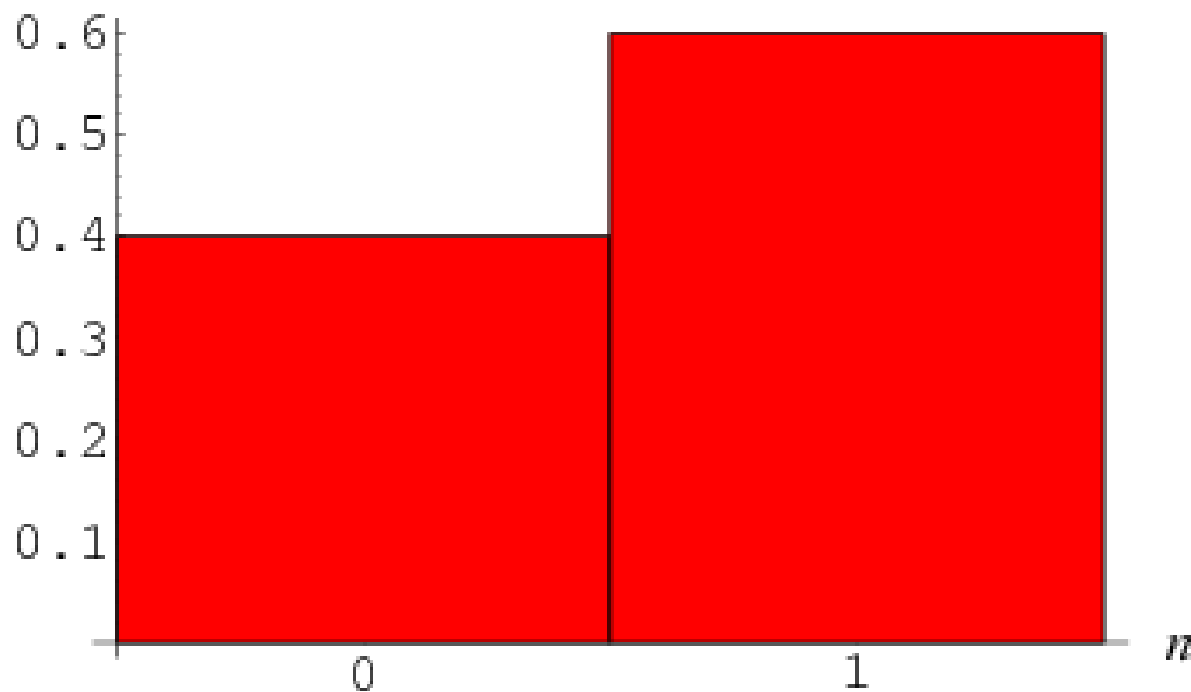
$$b_k = p\{\beta = k\} = p^k q^{1-k}, k = 0, 1$$

- 这个分布称为伯努利分布，或两点分布。



伯努利分布

$P(n)$ for $p = 0.6$



$p=0.6$ 时的伯努利分布



二项分布

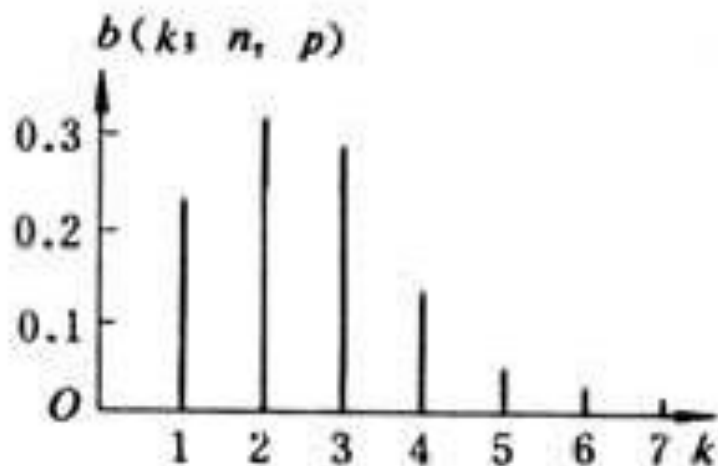
- 在n重伯努利试验中，若以 X 记事件A出现的次数，则它是一个随机变量， X 可能取的值为 $0, 1, 2, \dots, n$ ，其对应的概率由二项分布给出：

$$b(k; n, p) = p\{X = k\} = \binom{n}{k} p^k q^{n-k}, k = 0, 1, 2, \dots, n$$

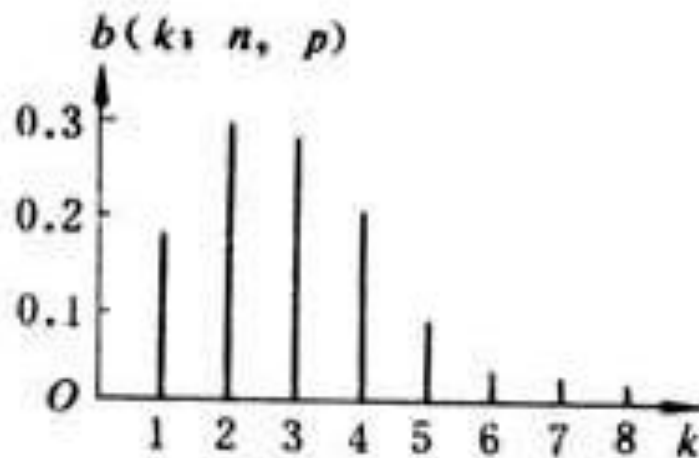
- 简记为 $X \sim B(n, p)$



二项分布



$$B(7, \frac{1}{3})$$



$$B(8, \frac{1}{3})$$

二项分布 $B(n, p)$ ($n=7, 8; p=1/3$)

二项分布概率密度函数示意图



泊松分布

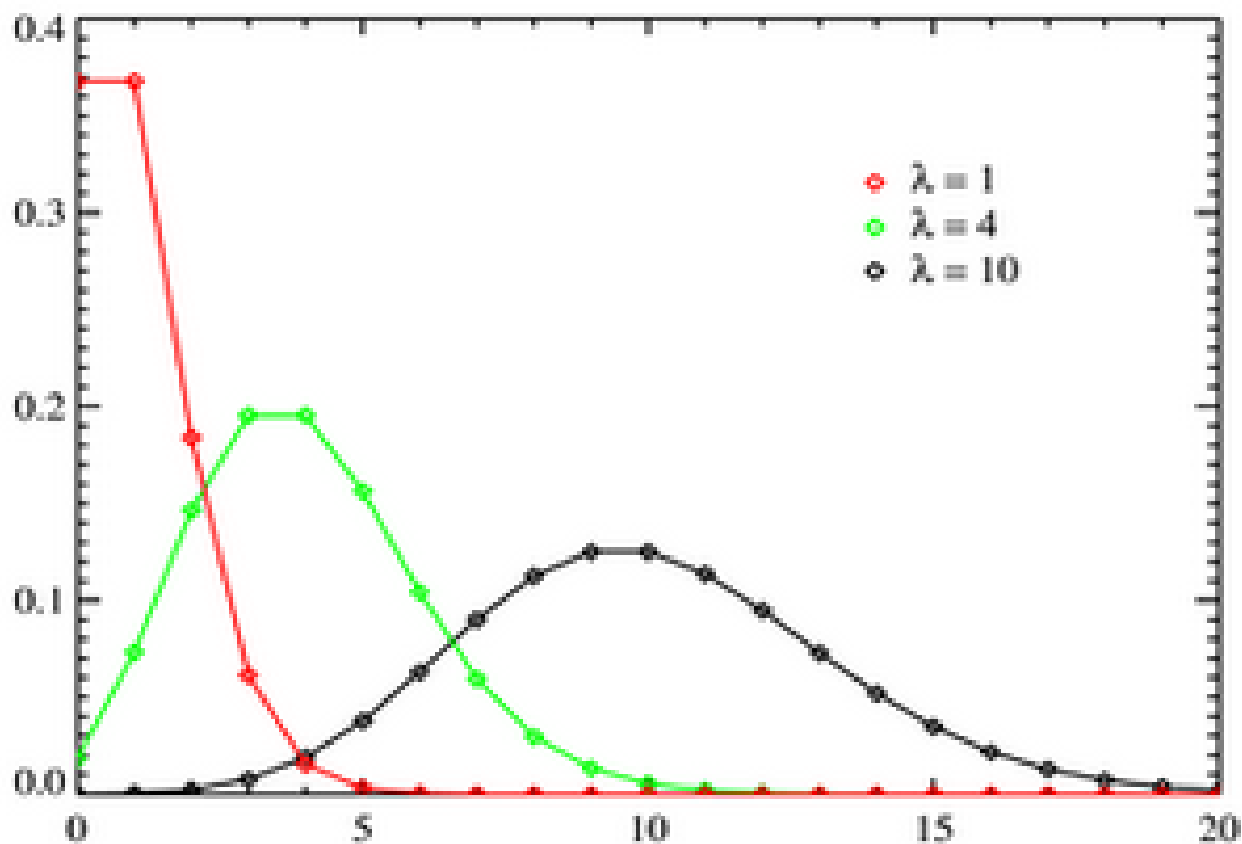
- 若随机变量 X 可取一切非负整数值，且

$$p\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, \dots$$

- 其中， $\lambda > 0$ 则称 X 服从泊松分布。
- 简记为 $X \sim p(\lambda)$



泊松分布



泊松分布概率密度函数示意图



均匀分布

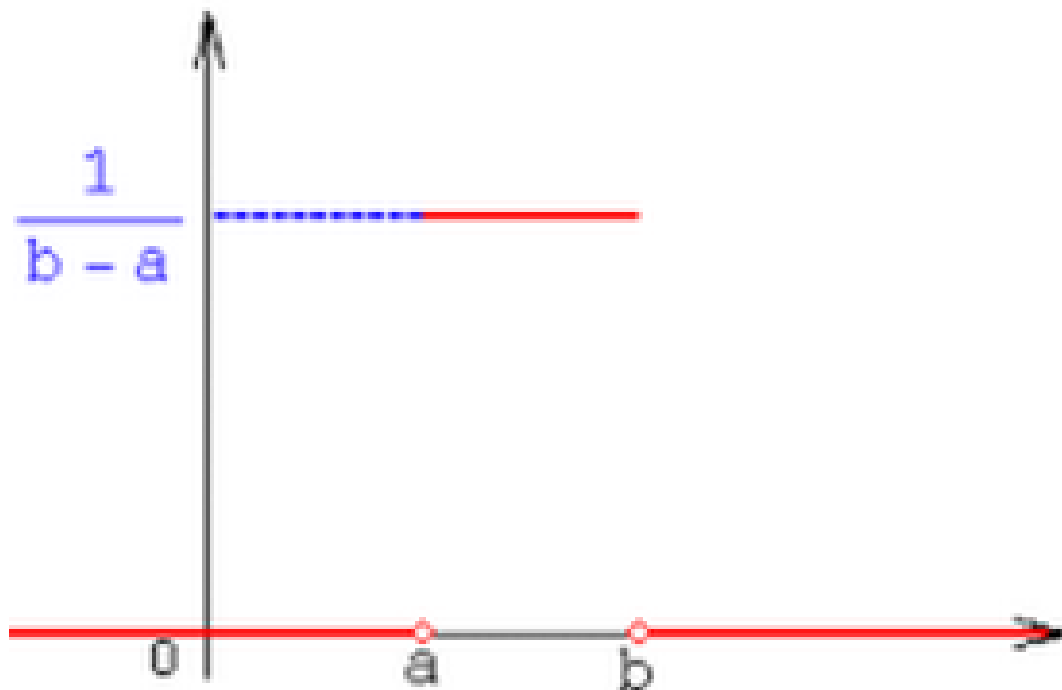
- 若 a, b 为有限数，由下列密度函数定义的分
布称为 $[a, b]$ 上的**均匀分布**：

$$p(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & x < a \text{ 或 } x > b \end{cases}$$

- 若随机变量 X 服从上的均匀分布，则 X 在 $[a, b]$
中取值落在某一区域内的概率与这个区间的
测度成正比。



均匀分布

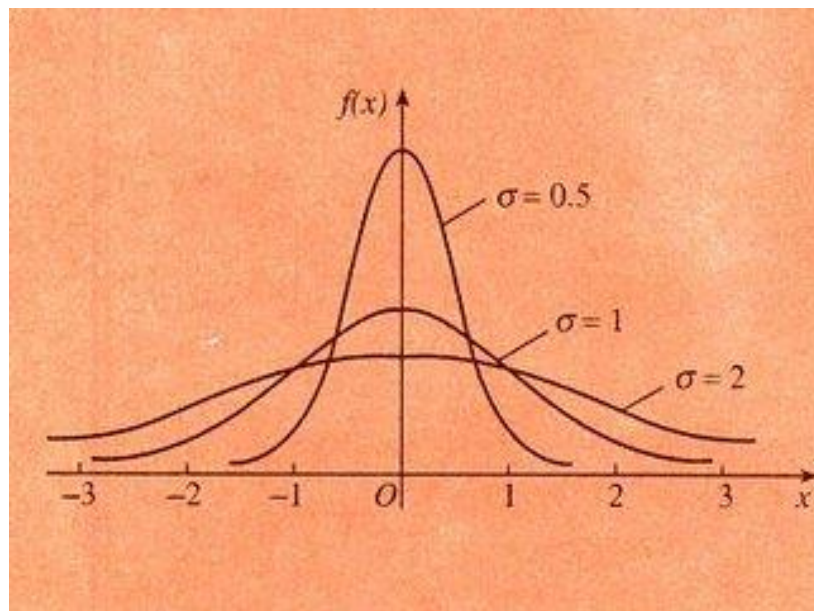


均匀分布概率密度函数示意图



1.5.2 正态分布及其性质

- 单变量正态分布
- 多元正态分布





单变量正态分布

- 定义: $\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]$
- 其中: μ 为随机变量 x 的期望, 即平均值;
 σ^2 为 x 的方差, σ 为均方差, 或标准差。

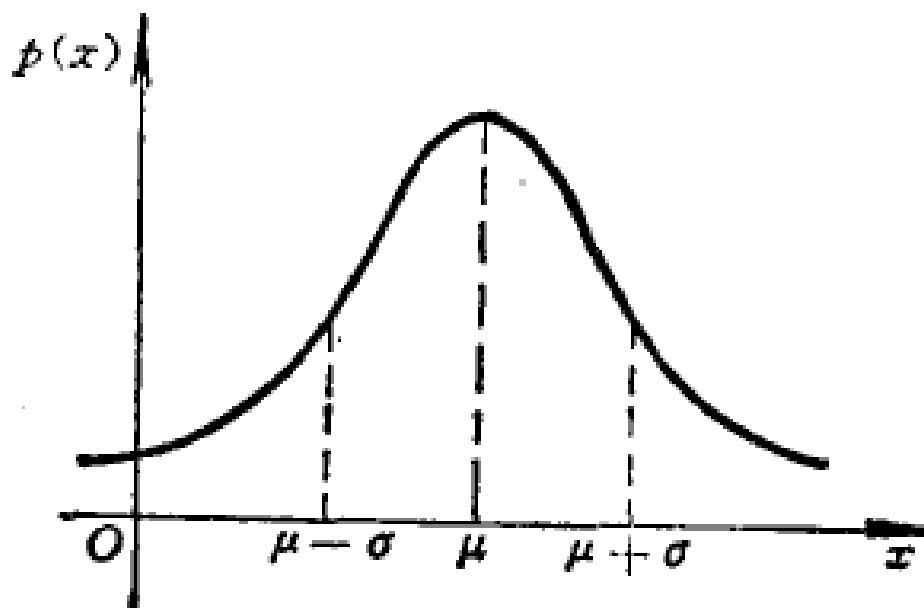
$$\mu = E(x) = \int_{-\infty}^{\infty} x \cdot \rho(x) dx$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot \rho(x) dx$$



单变量正态分布

- 概率密度函数的一般图形如下：



一维概率密度函数



单变量正态分布

- $\rho(x)$ 具有以下性质：

$$\begin{cases} \rho(x) \geq 0, (-\infty < x < \infty) \\ \int_{-\infty}^{\infty} \rho(x) dx = 1 \end{cases}$$

- 从 $\rho(x)$ 的图形上可以看出，只要有两个参数就可以完全确定其曲线。为了简单，常记 $\rho(x)$ 为 $N(\mu, \sigma^2)$ 。



多元正态分布

- **定义:**
$$\rho(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right]$$
- 其中: $\mathbf{x} = [x_1, x_2, \dots, x_d]^T$ 为 d 维随机向量, 对于 d 维随机向量 \mathbf{x} , 它的均值向量 μ 是 d 维的, Σ 是 $d \times d$ 维协方差矩阵, Σ^{-1} 是 Σ 的逆矩阵, $|\Sigma|$ 为 Σ 的行列式。
 μ 、 Σ 分别是向量 \mathbf{x} 和矩阵 $(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T$ 的期望



多元正态分布

- 具体说：若 x_i 是 x 的第 i 个分量， μ_i 是 μ 的第 i 个分量， σ_{ij}^2 是 Σ 的第 i 、 j 个元素。

$$\mu_i = E[x_i] = \int x_i \rho(x) dx = \int_{-\infty}^{\infty} x_i \rho(x_i) dx_i$$

- 其中 $\rho(x_i)$ 为边缘分布：

$$\rho(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \rho(x) dx_1 dx_2 \cdots dx_d$$



多元正态分布的性质

(1) 参数 μ 和 Σ 对分布的决定性:

$\rho(\mathbf{x})$ 可由 μ 、 Σ 完全确定

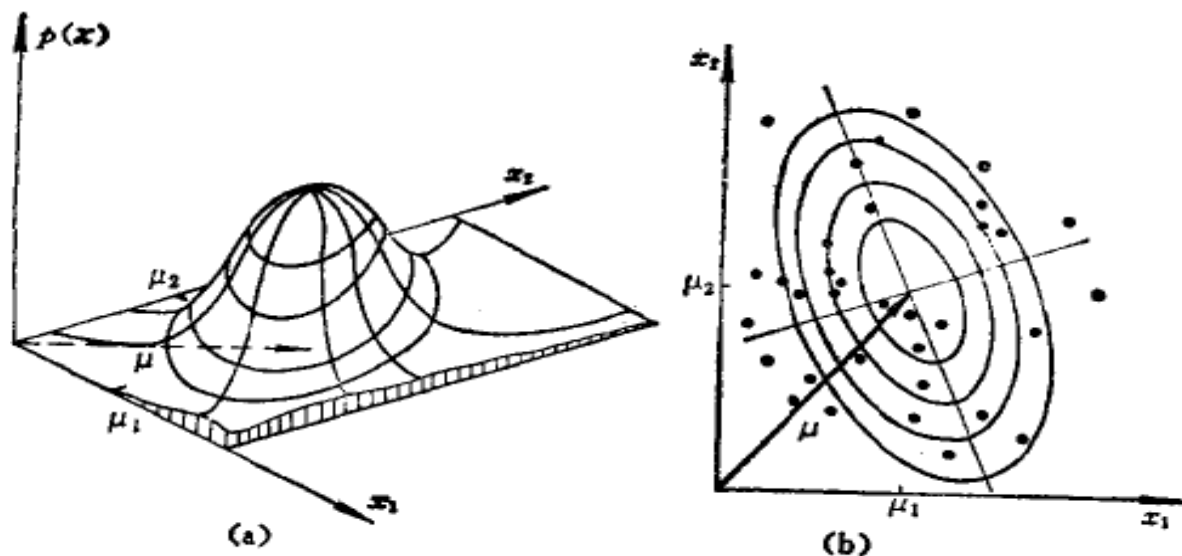
(2) 等密度点的轨迹为一超椭球面:

由 $\rho(\mathbf{x})$ 的定义公式可知, 当右边指数项为常数时, 密度 $\rho(\mathbf{x})$ 的值不变, 所以等密度点满足:

$$(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) = \text{constant}$$

多元正态分布的性质

可以证明，上式的解是一个超椭球面，其主轴方向取决于 Σ 的本征向量（特征向量），主轴的长度与相应的本征值成正比。如下图所示：





多元正态分布的性质

(3) 不相关性等价于独立性:

对服从正态分布的两个分量互不相关, 则它们之间一定独立。

(4) 边缘分布与条件分布的等价性:

不难证明正态随机向量的边缘分布与条件分布仍服从正态分布。

(5) 线性变换的正态性:

对于多元随机向量的线性变换, 仍为多元正态分布的随机向量。



多元正态分布的性质

(6) 线性组合的正态性

若 \mathbf{x} 为多元正态随机向量，则线性组合
 $y = \mathbf{a}^T \mathbf{x}$ 是一维的正态随机变量：

$$\rho(y) \sim N(\mathbf{a}^T \boldsymbol{\mu}, \mathbf{a}^T \boldsymbol{\Sigma} \mathbf{a})$$

其中， \mathbf{a} 与 \mathbf{x} 同维。



1.5.3 混合分布模型

- 上面提到的分布都相对简单，而实际中，很多数据并不完美地适应某个分布。此时，我们常用多个简单模型的线性组合来刻画数据。高斯混合模型使我们最常使用的模型，下面我们着重介绍混合高斯模型的基本概念，它的参数求解方法常用EM算法，我们将在后续章节作介绍。



单高斯分布模型（GSM）

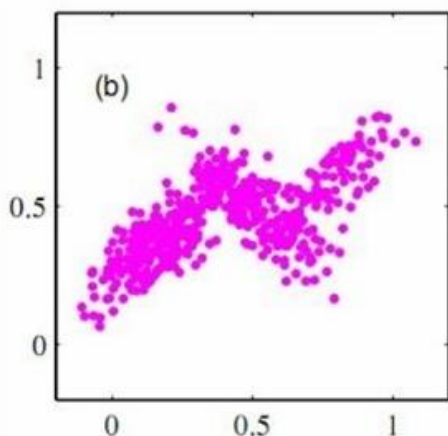
- 多维变量 \mathbf{x} 服从高斯分布时，它的概率密度函数PDF为：

$$N(\mathbf{x}; \mu, \Sigma) = \frac{1}{\sqrt{2\pi} |\Sigma|} \exp\left[-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right]$$

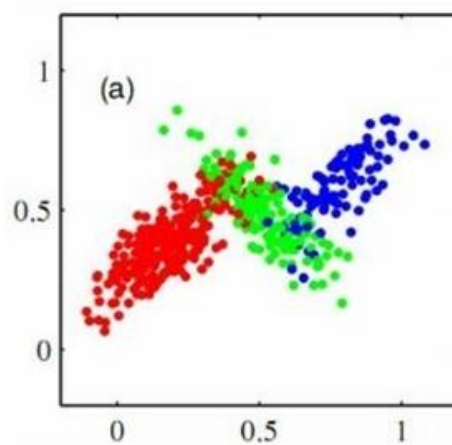
- 从几何上讲，单高斯分布模型在二维空间应该近似于椭圆，在三维空间上近似于椭球。遗憾的是在很多分类问题中，属于同一类别的样本点并不满足“椭圆”分布的特性。这就引入了高斯混合模型。

高斯混合模型

- 如下图所示，（a）中的数据显然不成“椭圆”形状，因此不能用单一的高斯模型去刻画，而（b）可以将数据分割为3个部分，每个部分都近似成“椭圆”形状，可以用高斯模型刻画，因此整个数据可以用高斯混合模型来刻画。



(a)



(b)



高斯混合模型 (GMM)

- GMM认为数据是从几个GSM中生成出来的，即：

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k N(\mathbf{x} | \mu_k, \Sigma_k)$$

- K需要事先确定好，就像K-means中的K一样。 π_k 是权值因子。其中的任意一个高斯分布 $N(\mathbf{x} | \mu_k, \Sigma_k)$ 称为这个模型的一个组分。



小 结

- 引言
- 模式识别的基本概念
- 模式识别系统
- 机器学习的主要方法
- 随机变量及分布