

# 第1章：绪论

刘成林([liucl@nlpr.ia.ac.cn](mailto:liucl@nlpr.ia.ac.cn))

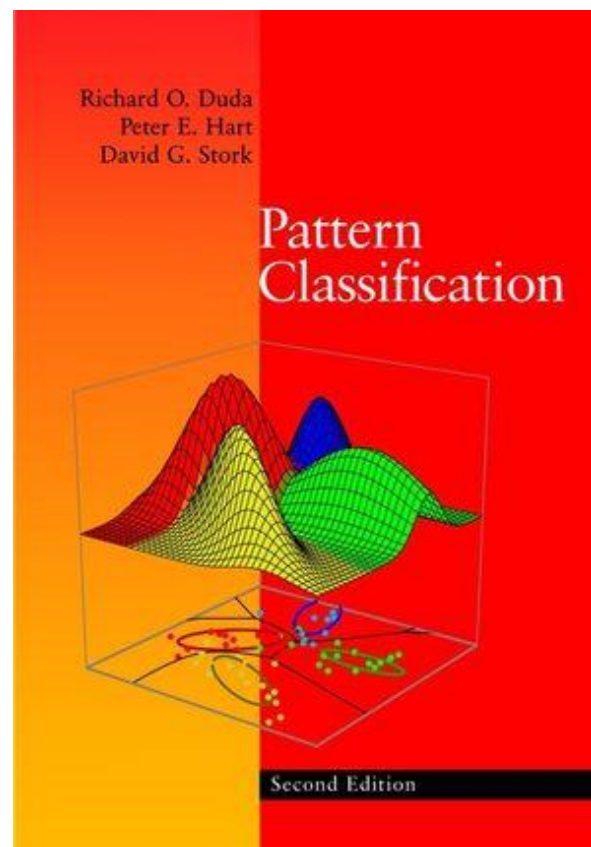
2015年9月16日

助教：杨学行([xhyang@nlpr.ia.ac.cn](mailto:xhyang@nlpr.ia.ac.cn))

吴一超([yichao.wu@nlpr.ia.ac.cn](mailto:yichao.wu@nlpr.ia.ac.cn))

# 教材

- Richard O. Duda, Peter E. Hart, David G. Stork,  
[Pattern Classification](#), 2nd Edition, John Wiley, 2001.  
（国内）模式分类，英文影印版
- 其他补充参考资料



# 提 纲

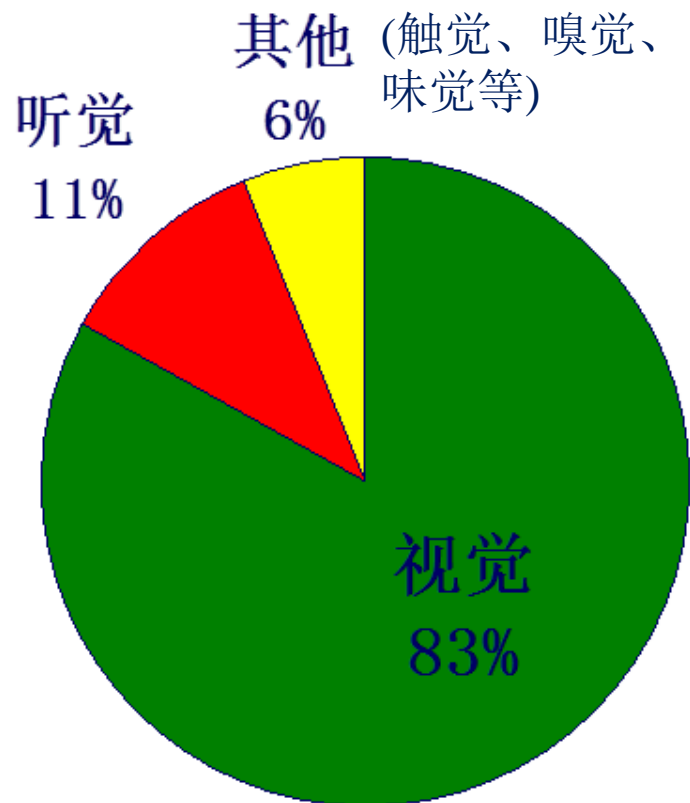
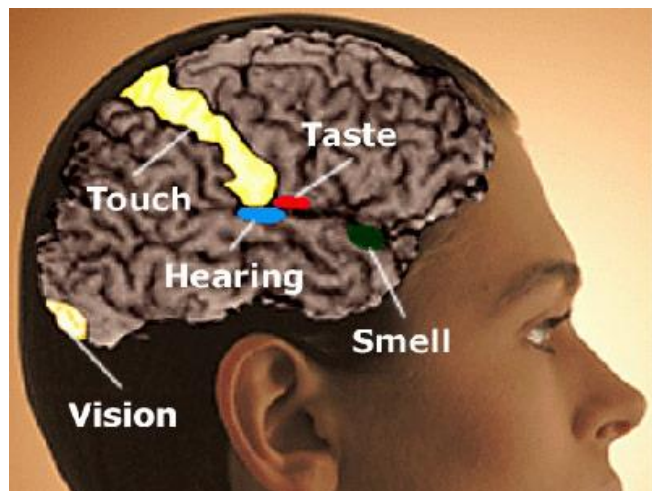
- 什么是模式识别
- 模式识别形式化
- 模式识别系统流程
- 模式分类器设计
- 模式识别方法分类
- 本课程内容体系

# 什么是模式识别

- 信息感知
- 计算机模式识别
- 模式识别应用
- 相关问题和领域

# 感知：从环境获取信息

人和动物通过感知从周围环境获取信息。感知就是模式识别过程。



# 计算机模式识别

- 模式识别：使计算机模仿人的感知能力，从感知数据中提取信息（判别物体和行为、现象）的过程



ROCKETS

11

姚明

ROCKETS

11



# 模式识别的意义

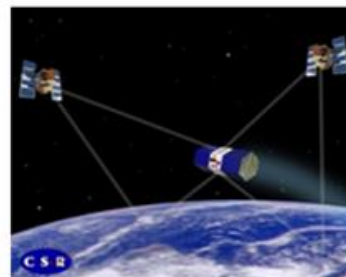
数字化感知数据：来源丰富、数量巨大



视觉监控



机器人感知



卫星遥感



Google眼镜



RGB-D成像



各种文档



电视、网络视频



互联网、物联网



中国科学院大学

University of Chinese Academy of Sciences



# 模式识别应用

- 安全监控（身份识别、行为监控、交通监控）
- 空间探测与环境资源监测（卫星/航空遥感图像）
- 智能人机交互（表情、手势、声音、符号）
- 机器人环境感知（视听触觉）
- 人类健康（医学图像、体测数据）
- 工业应用（零部件/物品分类、损伤检测）
- 文档数字化（历史书籍报纸、档案、手稿、标牌等）
- 网络搜索、信息提取和过滤（文本、图像、视频、音频、多媒体）
- 舆情分析（互联网、大数据）



# 应用：生物特征识别



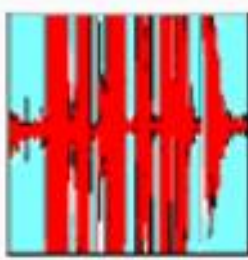
指纹



脸相



虹膜



声纹



指静脉



签名



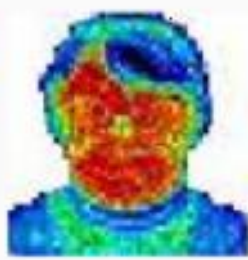
击键



骨骼



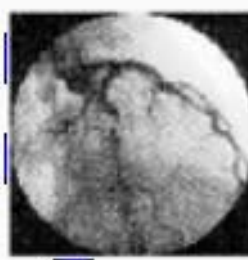
步态



红外成像



耳廓



视网膜



掌纹



多模态融合

有些生物特征（如虹膜、指静脉）精度高，但是需要客户配合。

有些（如签名、步态）精度相对较低，但是不需要配合，有适合其应用的场合。

# 应用：视频监控



2008年北京奥运会区域  
人群监控



北京城铁13号线人物入侵检测和跟踪，  
曾辅助抓获盗割电缆者



# 应用：遥感图像地表分类



原始图像(伪彩色)

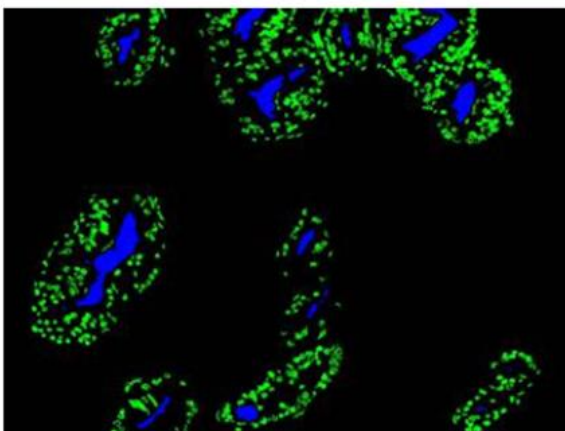
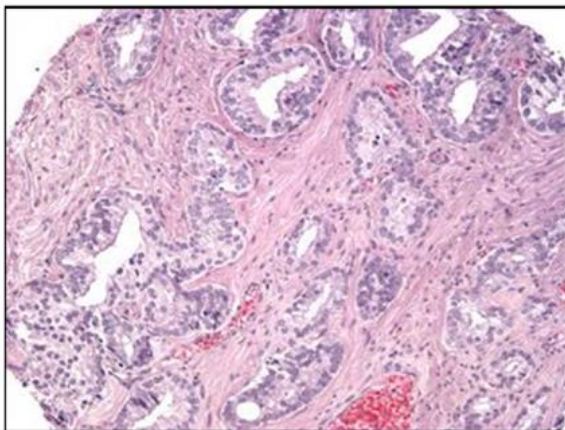


SVM

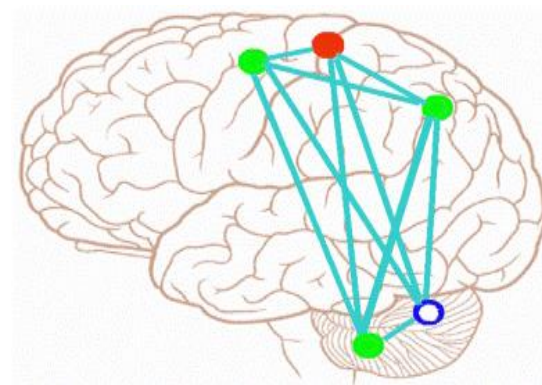
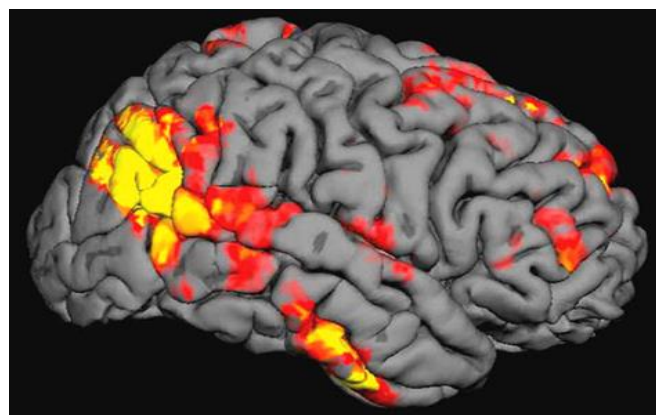


图模型

# 应用：医学图像分析



显微组织图像癌变检测



脑功能网络分析



# 应用：车牌识别



(b)



(c)



	<b>P7577</b>	P7577
	<b>G4402</b>	G4402
	<b>DU3403</b>	DU3403
	<b>GG 4025</b>	GG4025
	<b>CX0166</b>	CX0166

S.-L. Chang, et al., [Automatic license plate recognition](#), *IEEE T-ITS*, 2004.

# 应用：信函分拣



富士見市関沢

19329.9

富士見市関沢 2-10-3  
あや 夏容室

請求書在中  
三島建設株式会社 御中  
問計大井町武蔵 1574C.G  
△ 問 郡 大井町 武蔵野 1.258

3 5 0 0 2  
0711 9 8-12 鶴ヶ島市五味ヶ谷  
鶴ヶ島市五味 17146.4 99  
滝島三枝子様

# 模式识别相关问题

- 数据预处理
  - 视频、图像、信号处理等
- 模式分割
  - 检测定位、背景分离、模式-模式分离
- 运动分析
  - 目标跟踪、运动模式描述
- 模式描述与分类
  - 特征提取/选择、模式分类、聚类、机器学习
- 模式识别应用研究
  - 针对具体应用的方法与系统

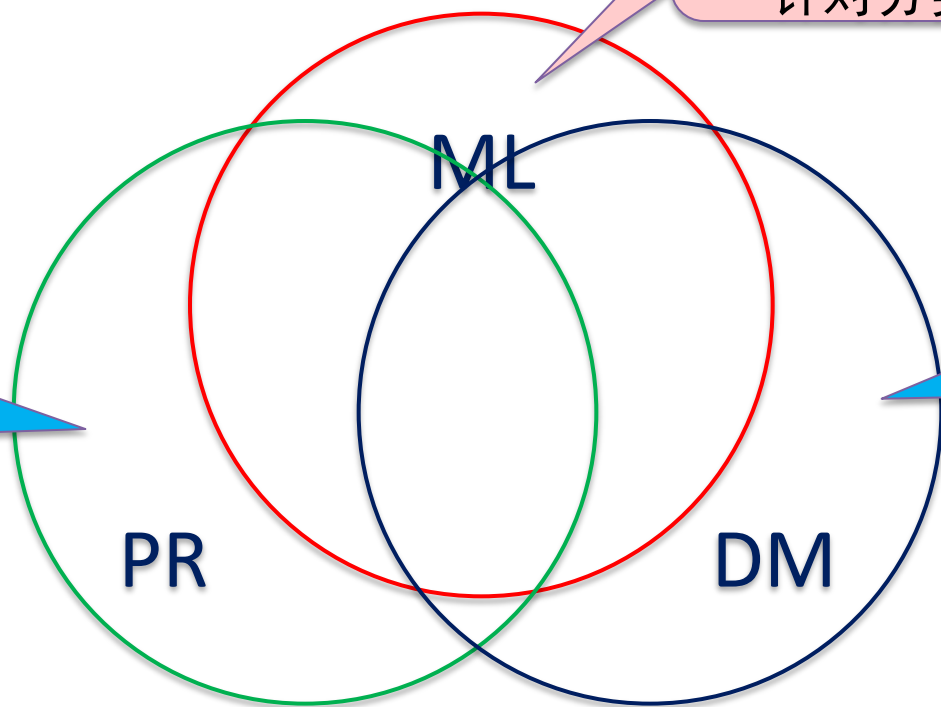


# 相关领域：模式识别-机器学习-数据挖掘

机器学习：从数据/经验中获取知识/规则/模型/参数的过程

主要研究理论算法，大部分针对分类

主要针对感知数据，面向应用



主要针对非感知和混合数据

ML/PR/DM: 基本上都在做分类

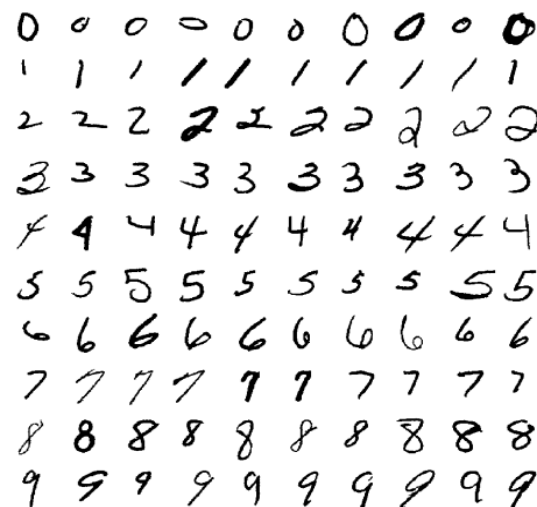
# 模式识别形式化

- 模式和模式识别
- 模式和分类器表示
- 一个例子

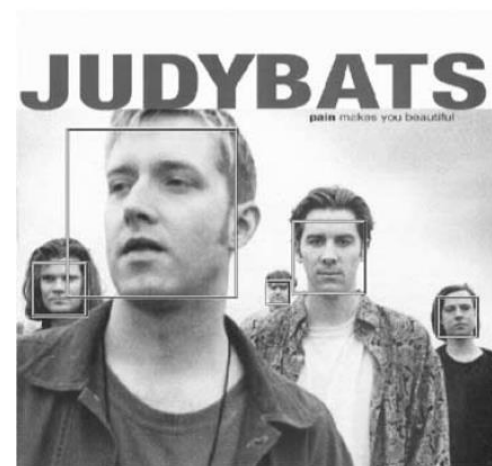
# 模式和模式识别

- 模式的两个层次
  - 样本(Sample, instance, example)
  - 类别(Class, category)

例如：100个样本、10个类别



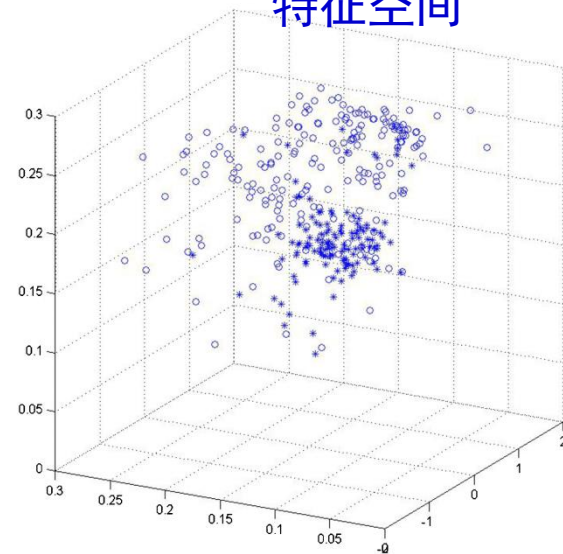
- 模式识别核心技术：模式分类
  - 检测：2-class (binary)
  - 判别：2-class, multi-class
  - 分类器设计：机器学习
  - 相关问题：特征提取、特征选择



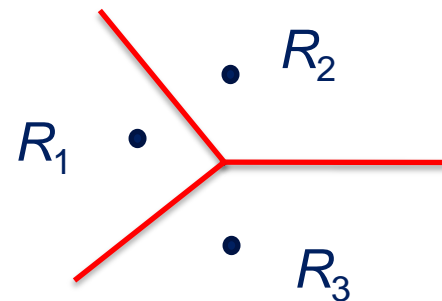
# 模式表示：两个方面

- 识别对象表示：**特征**
  - 特征度量： $\mathbf{x}=[x_1, x_2, \dots, x_n]^T$
  - 问题：特征提取、特征选择

特征空间



- 分类器表示
  - 类别模型： $M_i=M(\mathbf{x},\theta_i)$
  - 判别函数： $y_i=f(\mathbf{x},\mathbf{w}_i)$   
(Discriminant function)
- 识别（分类）
  - 距离度量（相似度）： $\min_i d(\mathbf{x}, M_i)$
  - 决策区域： $R_i = \arg \max_i f(\mathbf{x}, \mathbf{w}_i)$

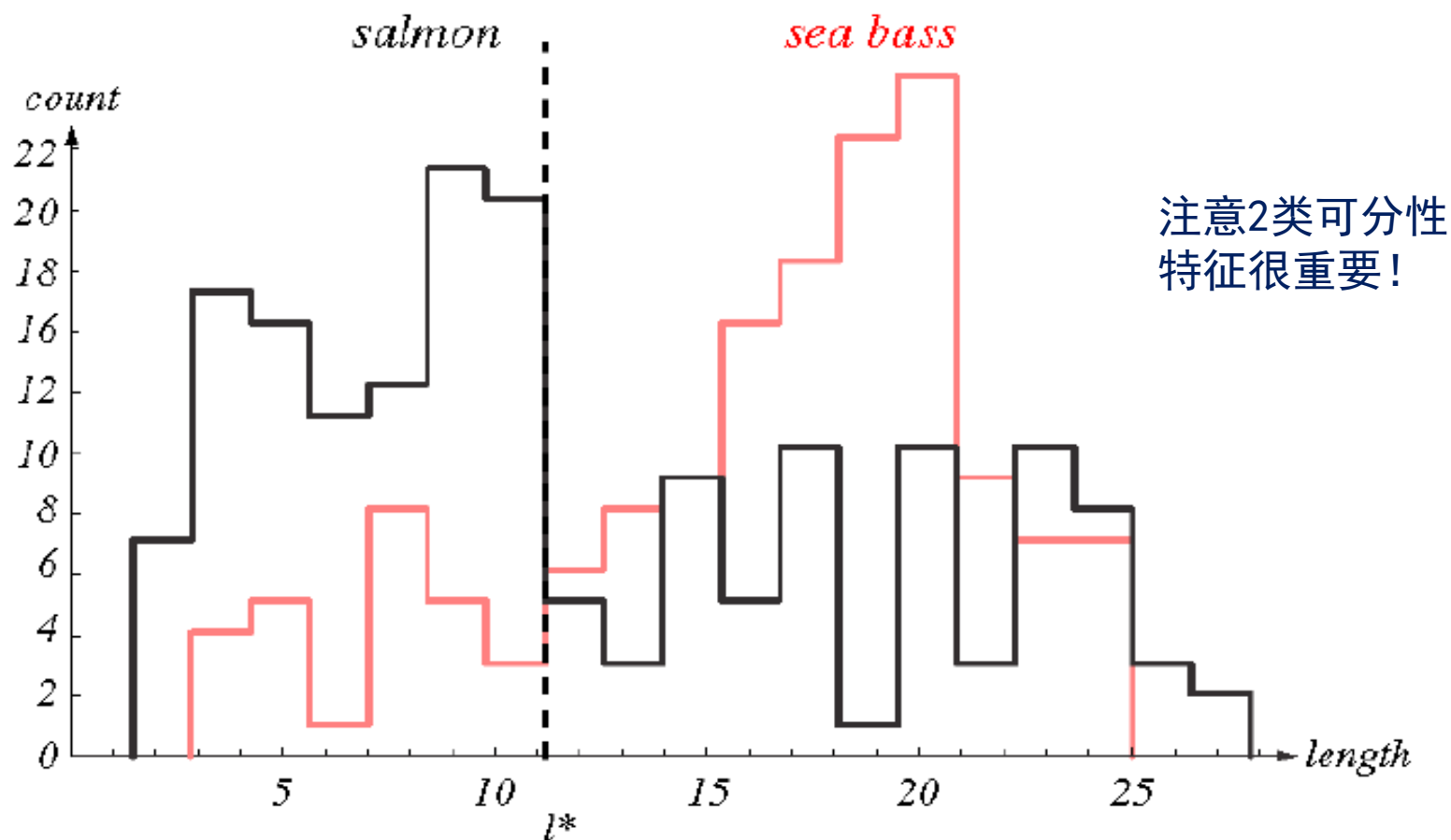


# 一个例子：鱼分拣

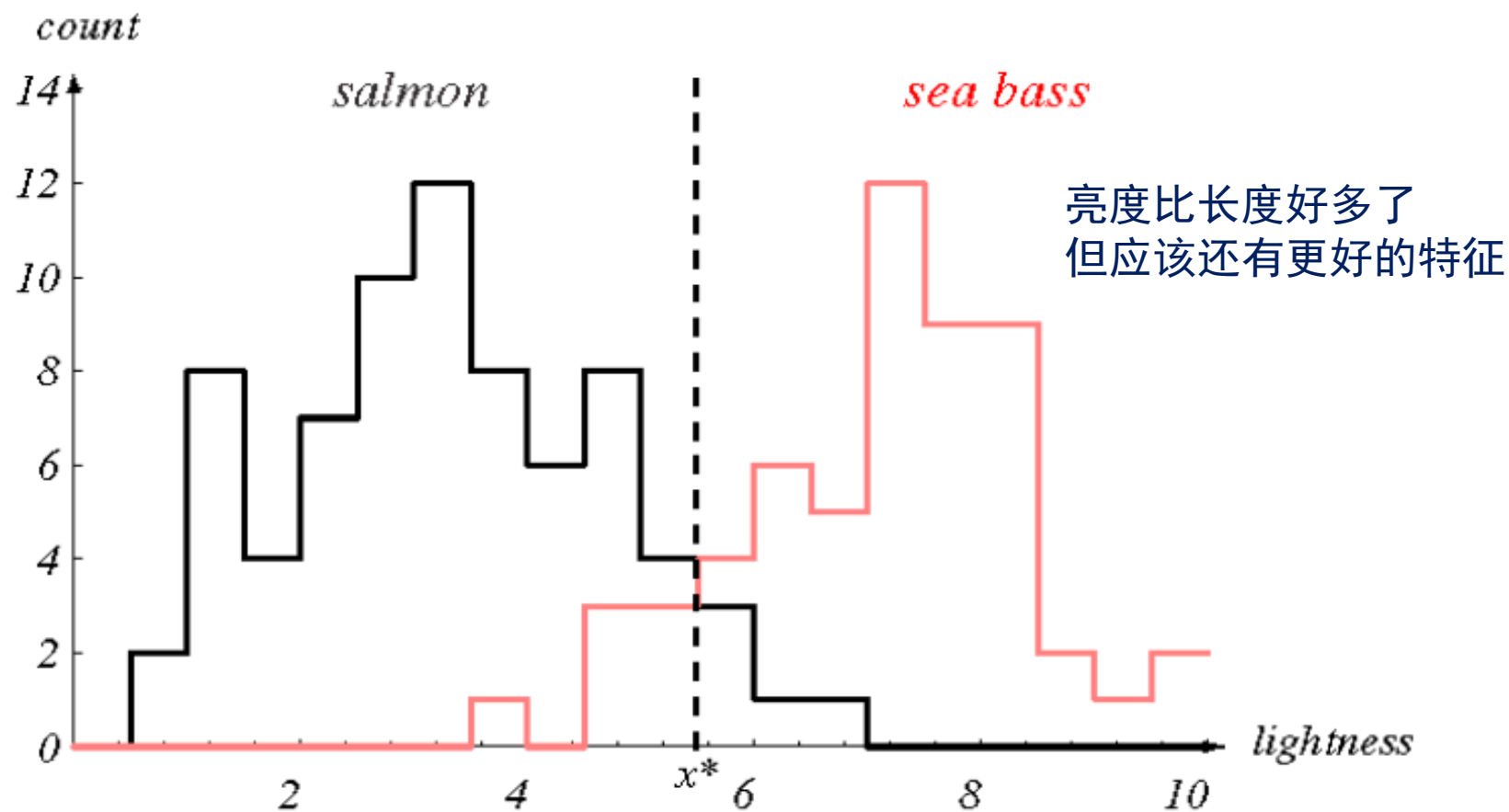
- 两类鱼
    - Sea bass（黑鲈）
    - Salmon（三文鱼）
- (*Pattern Classification*, 2001)



- 特征1：长度

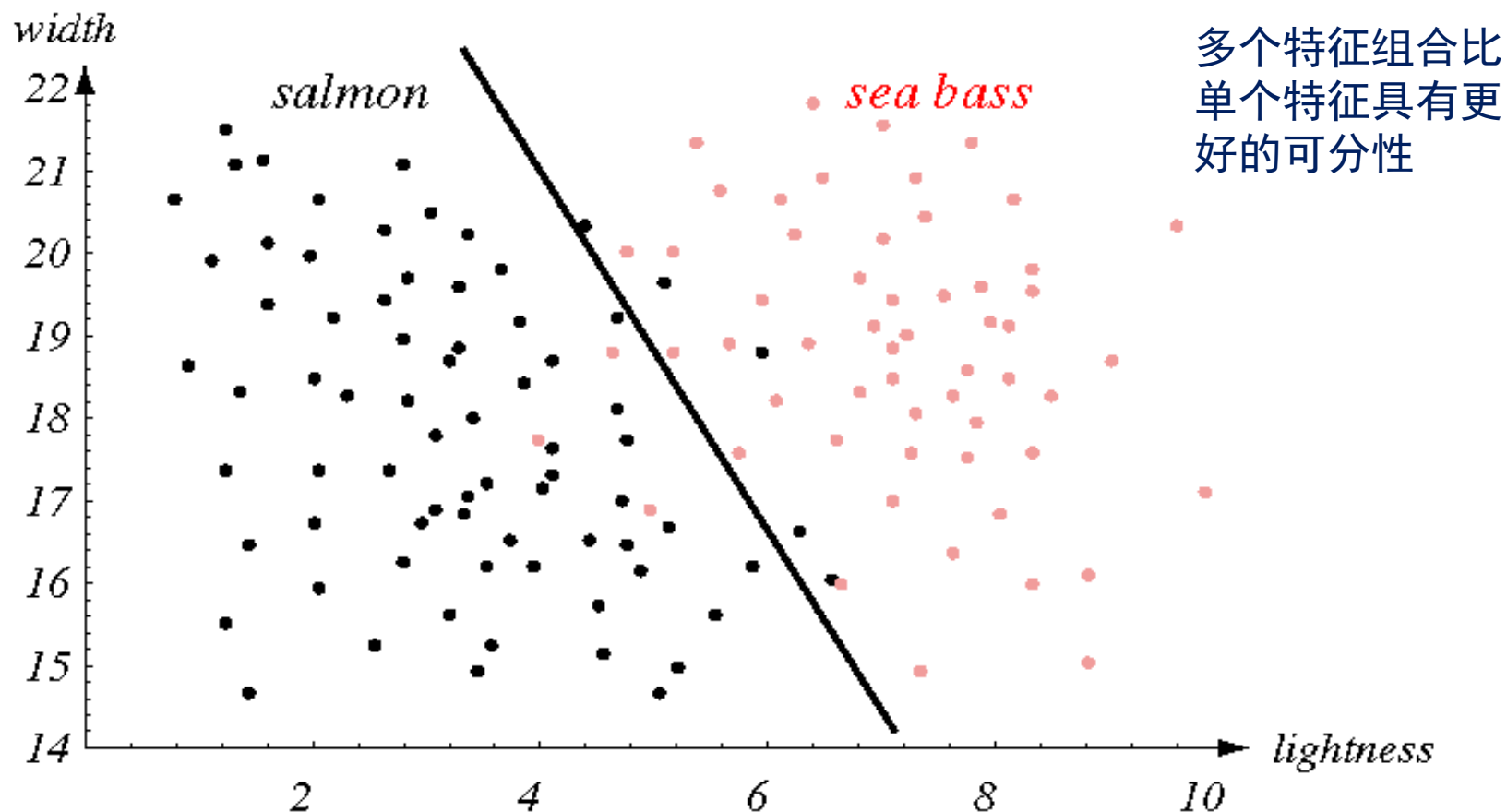


- 特征2：亮度

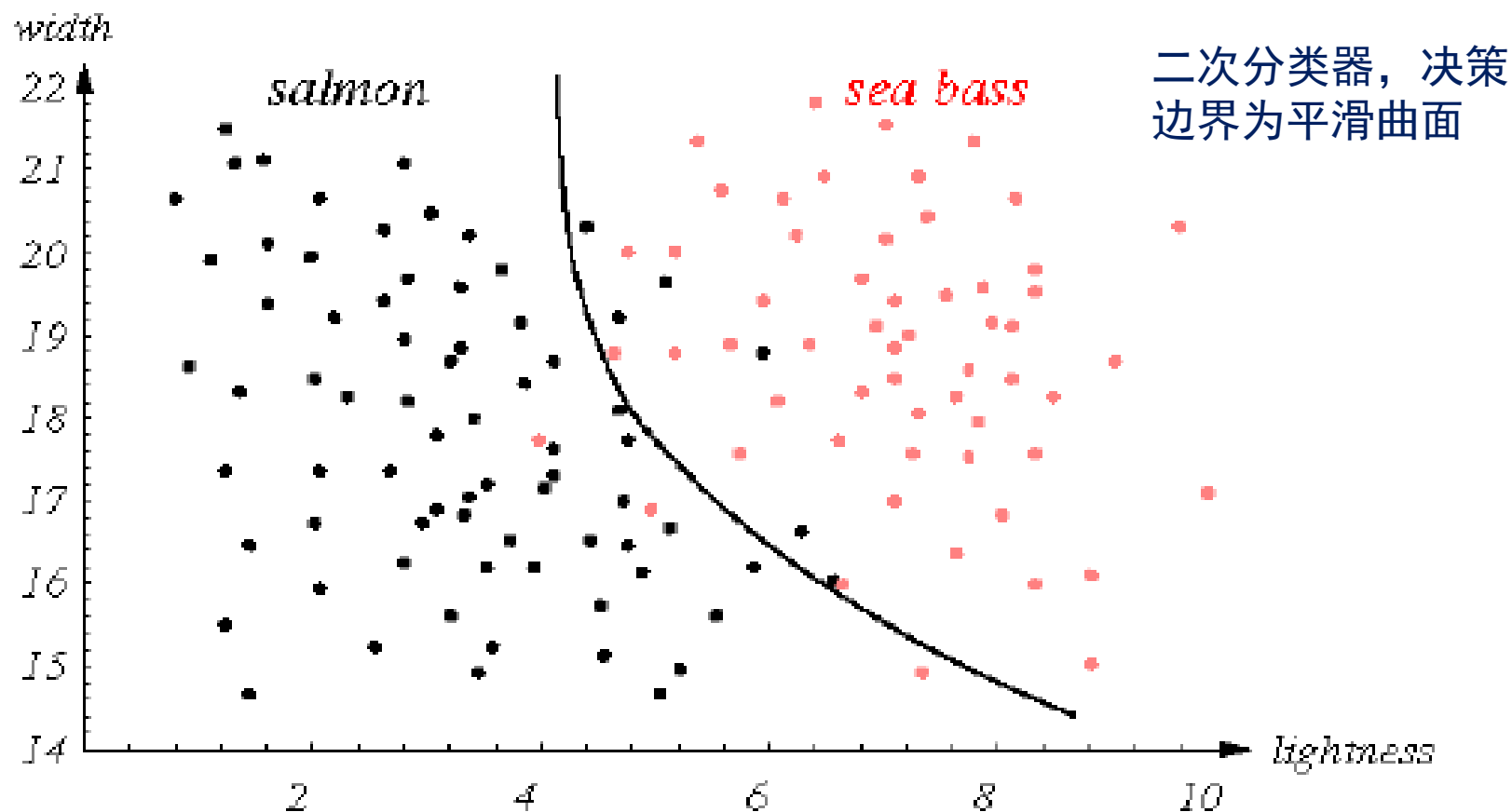




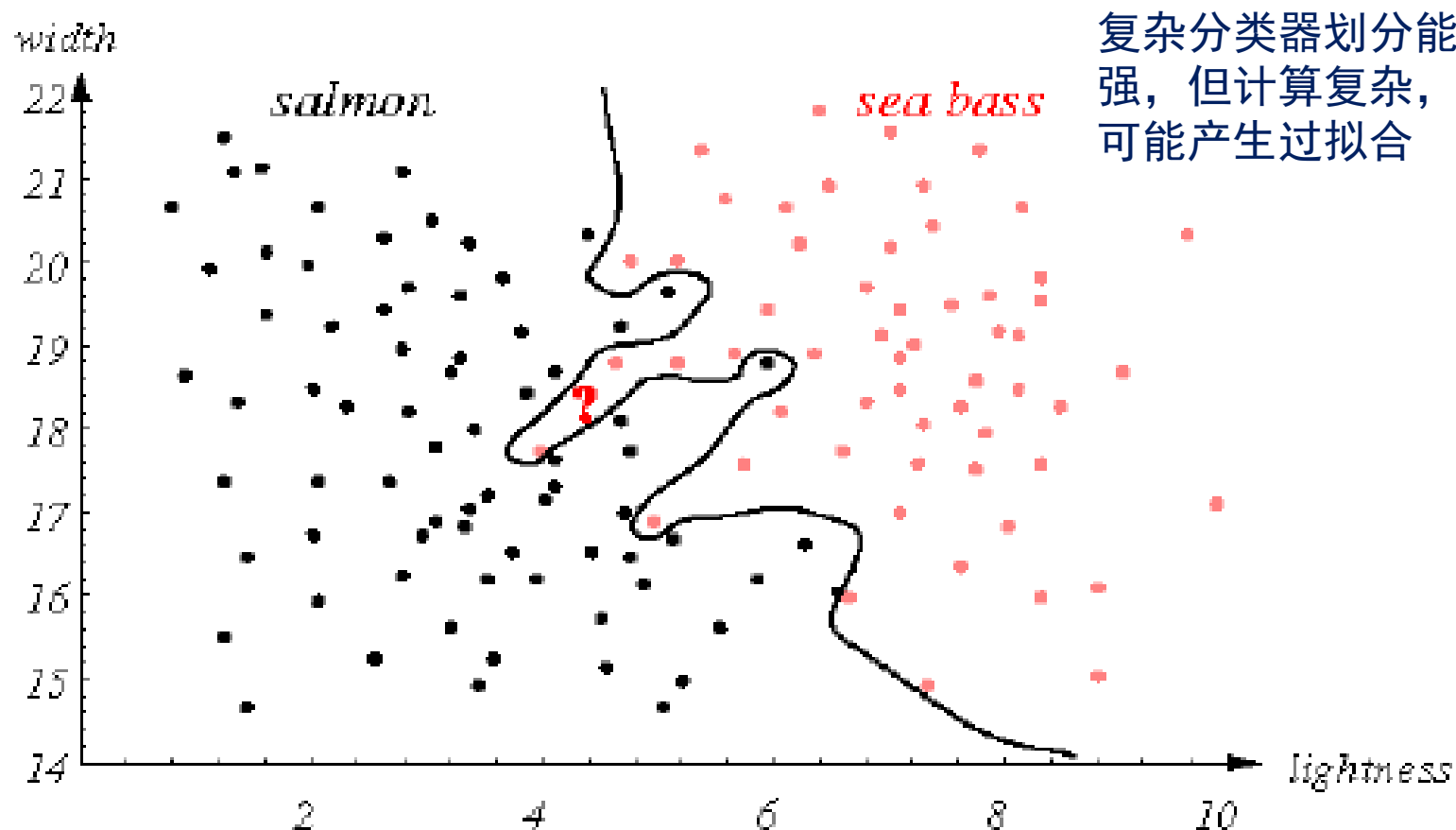
# 二维特征：线性分类器



# 二维特征：非线性分类器



# 二维特征：最近邻分类器

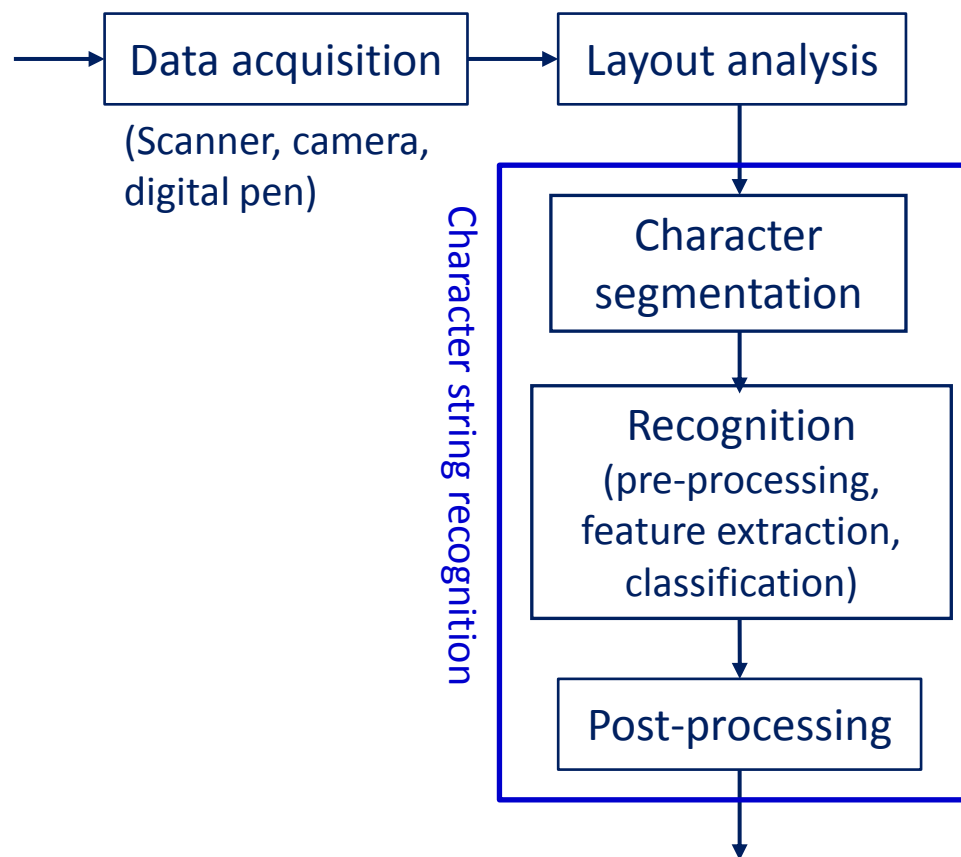


复杂分类器划分能力更强，但计算复杂，而且可能产生过拟合

# 模式识别系统流程

- 一个例子
- 完整识别流程
- 识别与训练的关系

# 一个例子：文档分析系统



## 中巴签署技术合作备忘录

新华社伊斯兰堡电 (记者李豪声) 中国科工委主任丁衡高中将和巴基斯坦负责国防的国务部长吴拉姆·萨尔瓦尔·奇马上校12月19日分别代表各自政府签署了一项有效期为10年的备忘录。

根据这项备忘录, 双方将在联合研究和开发、共同生产、技术转让方面进行合作。双方将向双方同意的第3国出口研究和开发的产品。

## 中缅签署经济技术合作协定

新华社仰光12月27日电  
由中国对外经贸部部长助理王文东率领的中国对外经贸代表团今天同缅甸政府签署了一项

### 中缅经济技术合作协定

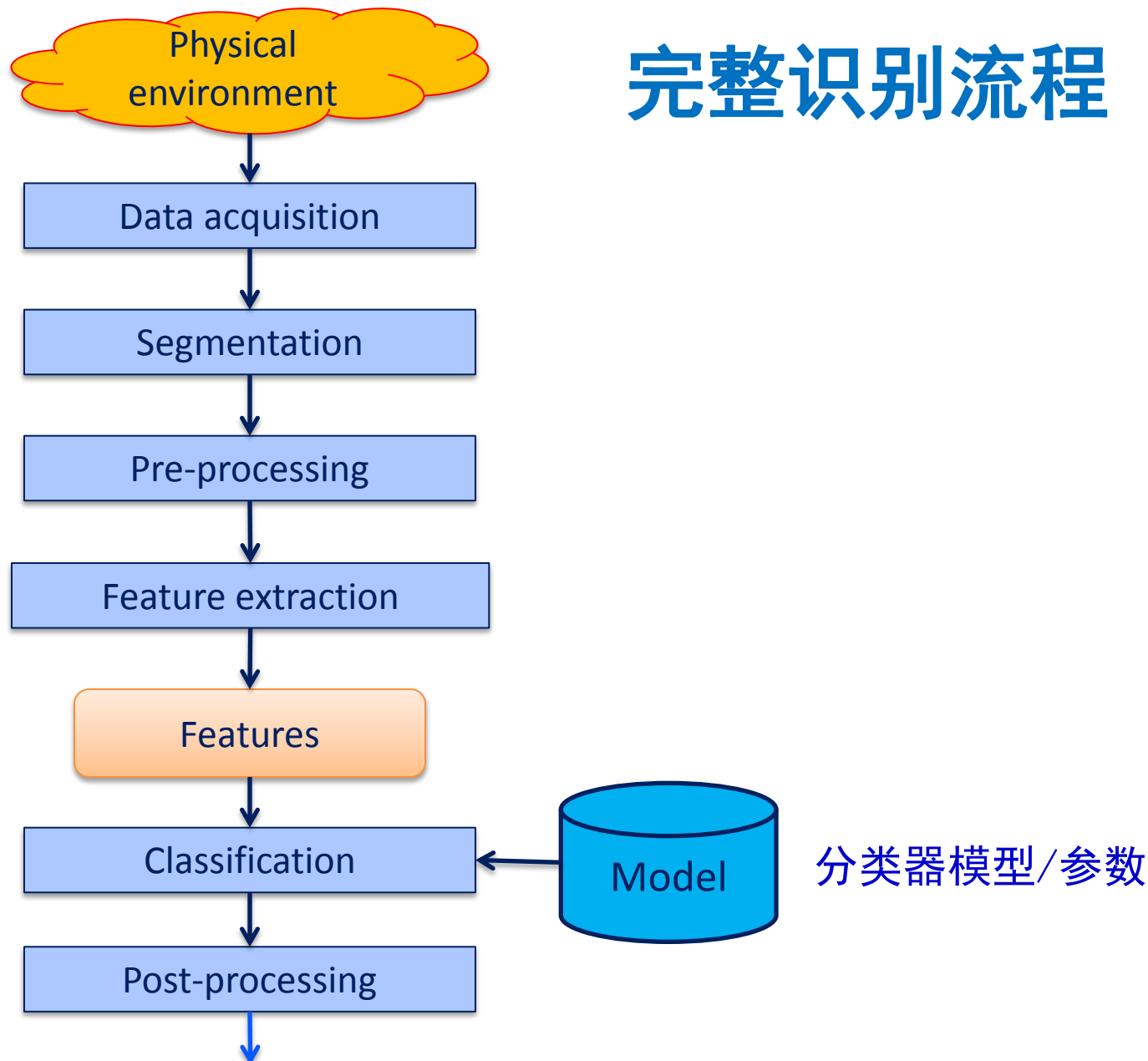
根据这项协定, 中国同意从1990年起5年内向缅甸提供一笔无息贷款。

中国对外经贸部

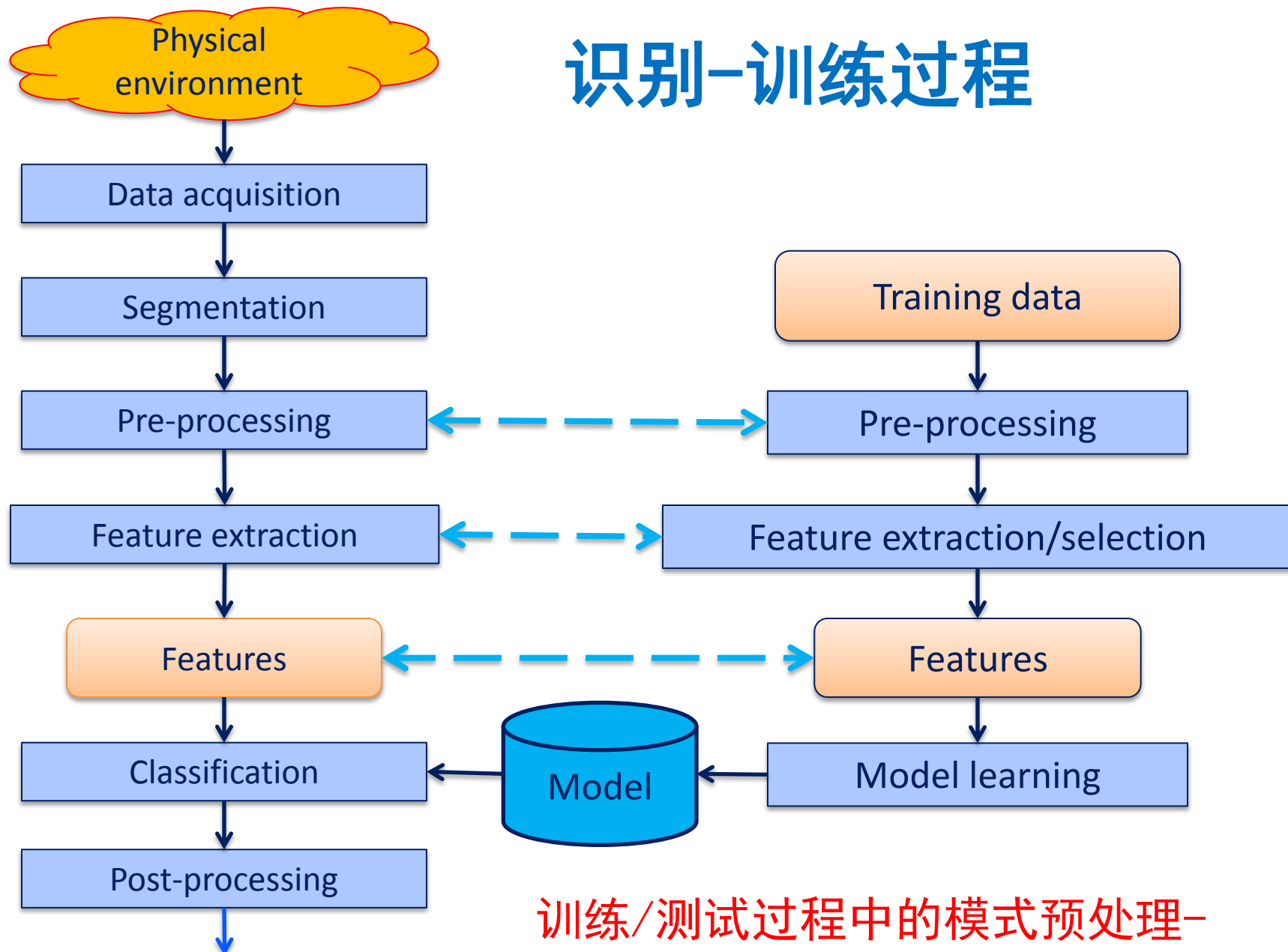
中国对外经**货**部

中国对外经贸部

# 完整识别流程



# 识别-训练过程



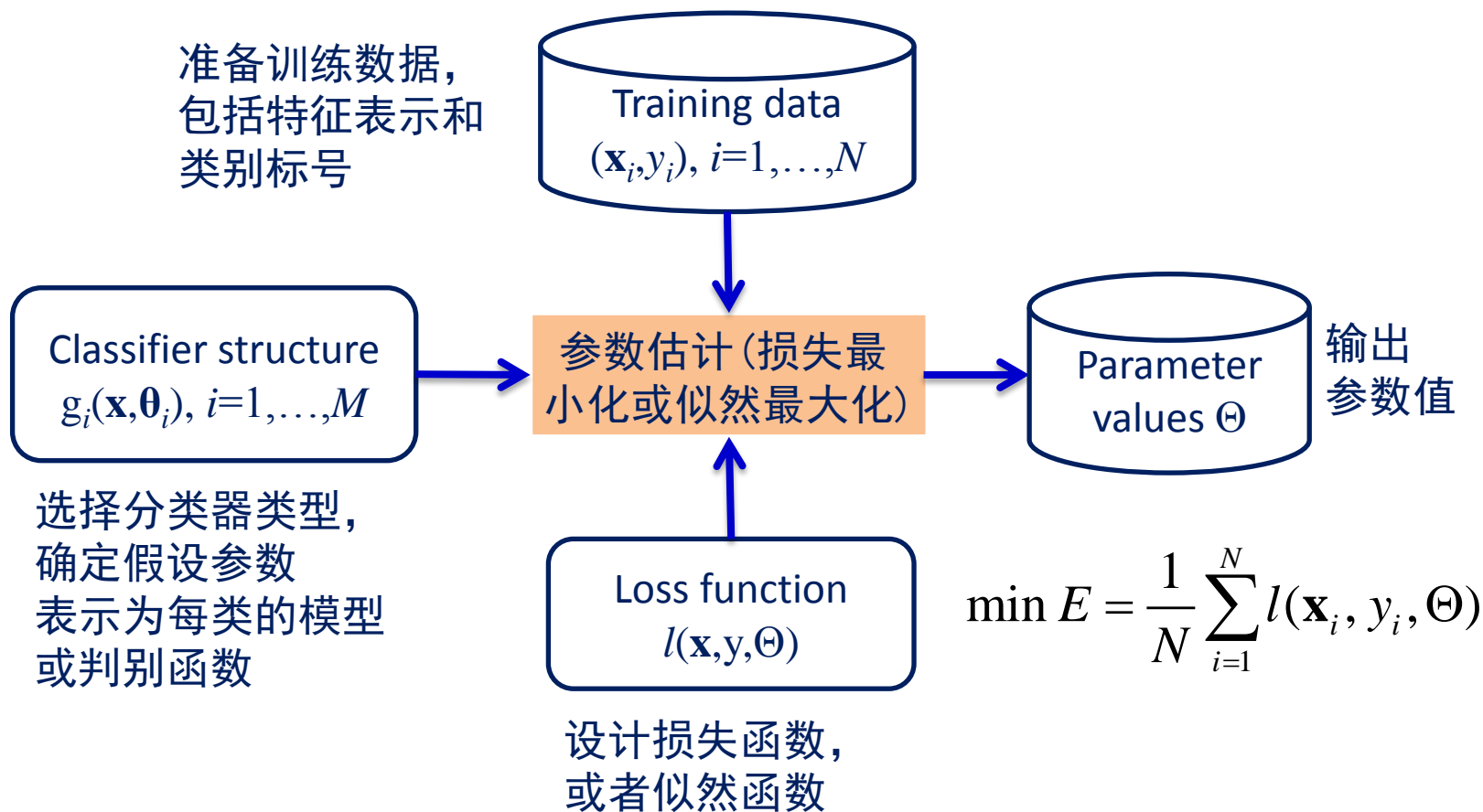
训练/测试过程中的模式预处理-特征提取必须完全一致



# 模式分类器设计

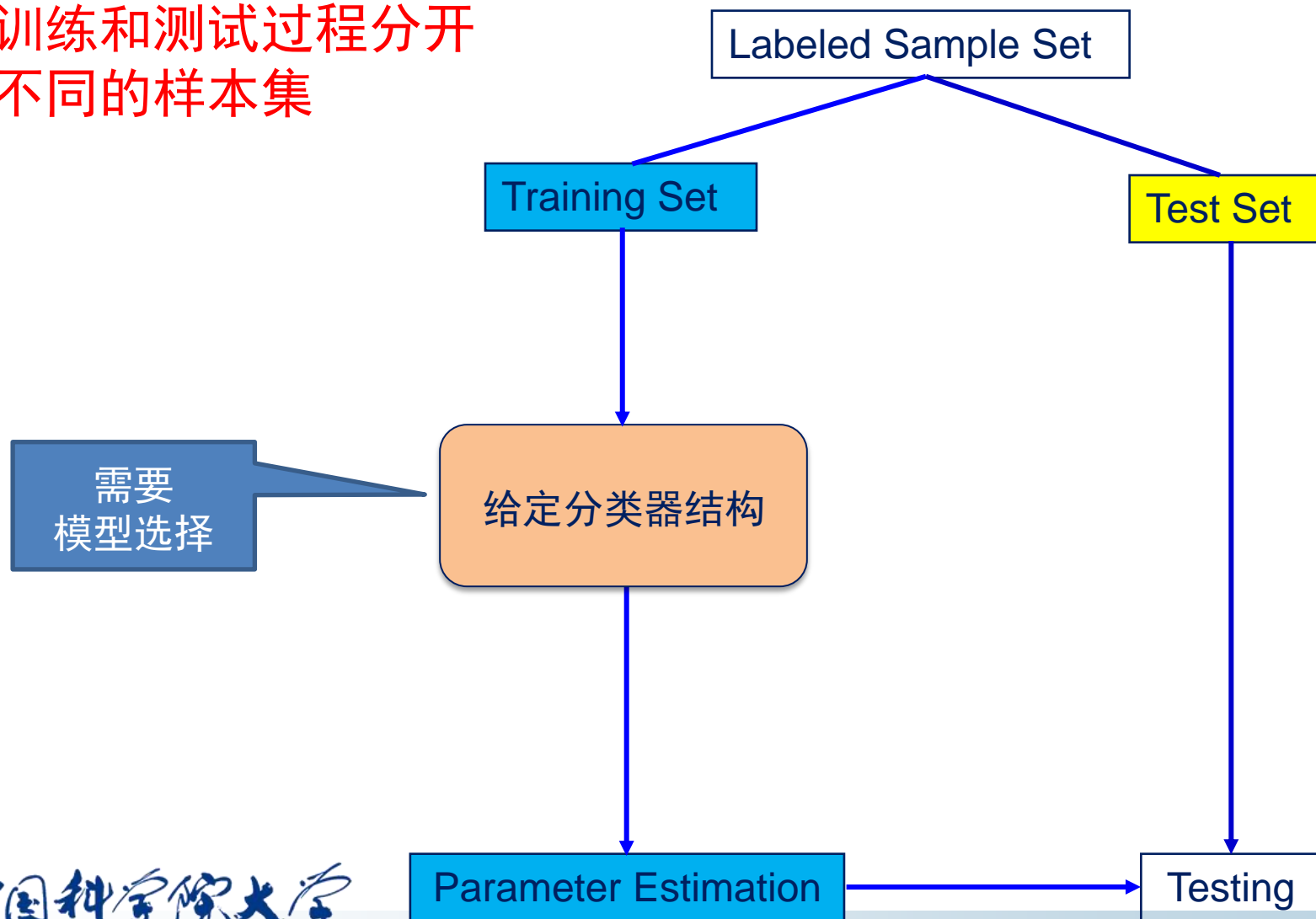
- 分类器训练过程
- 训练-评价流程
- 模型选择
- 数据划分方式
- 泛化性能

# 分类器训练过程



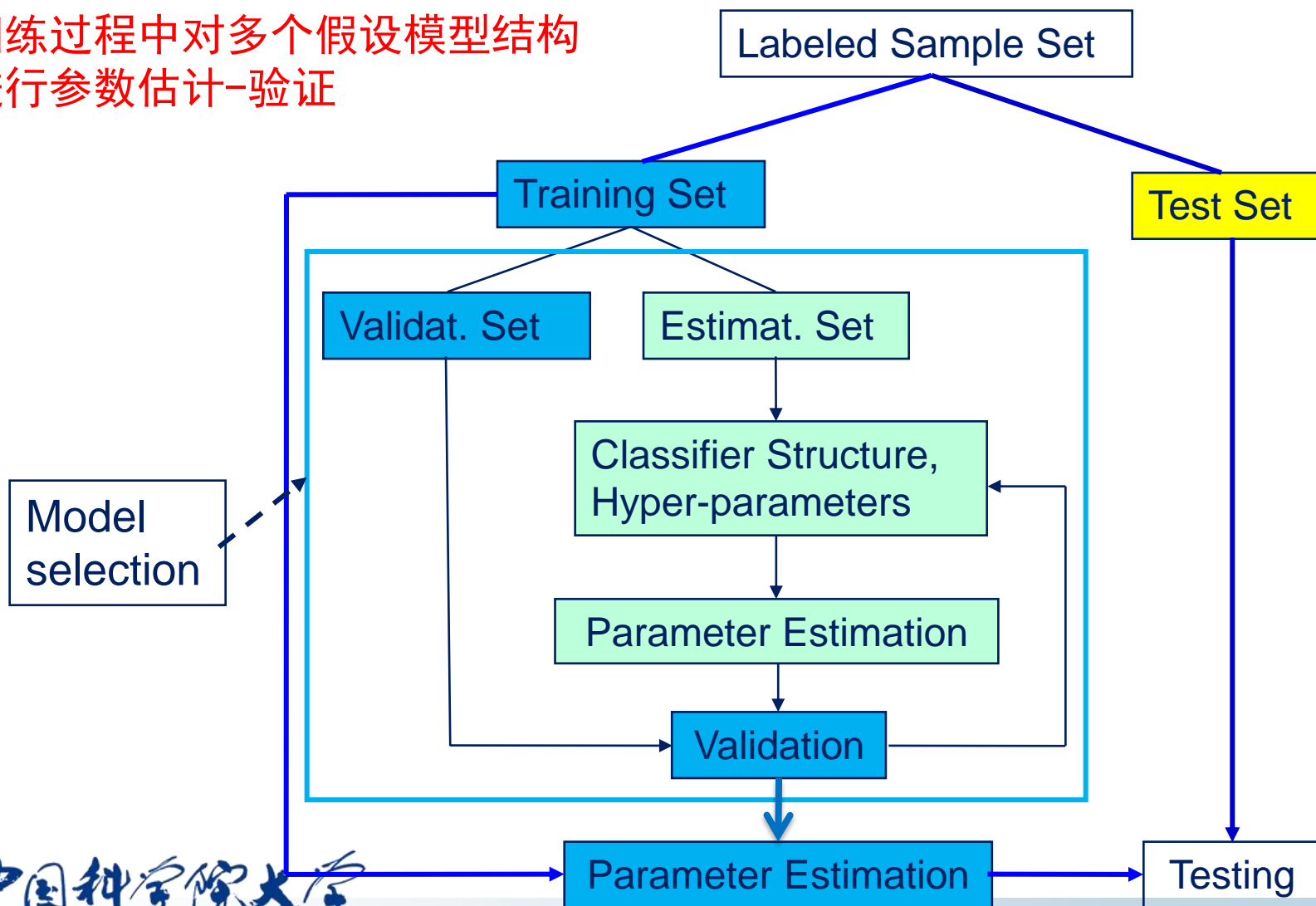
# 分类器训练-评价过程

训练和测试过程分开  
不同的样本集



# 模型选择和评价

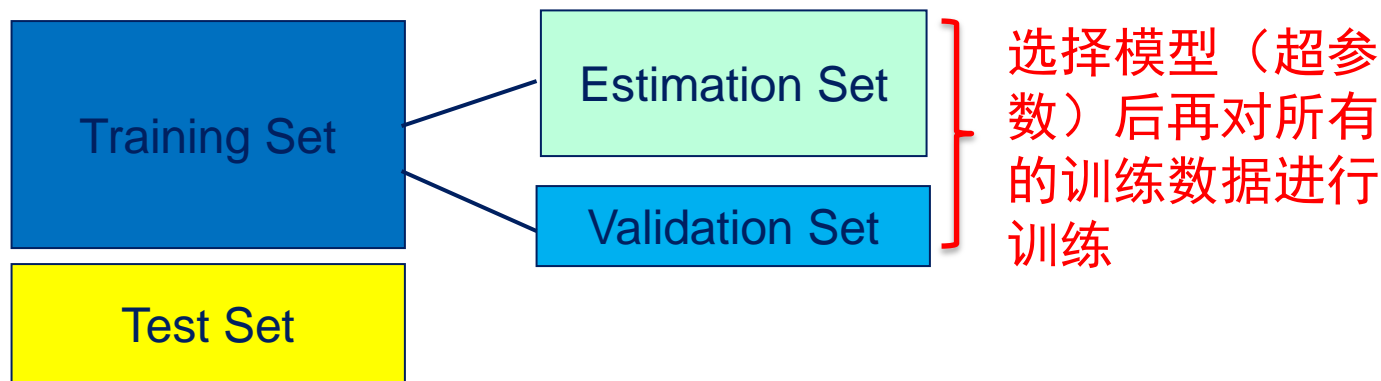
训练过程中对多个假设模型结构  
进行参数估计-验证



# 数据划分方式

- 两个层次的划分

- Performance evaluation: Training+Test
- Model selection: Estimation+Validation



- 划分方式

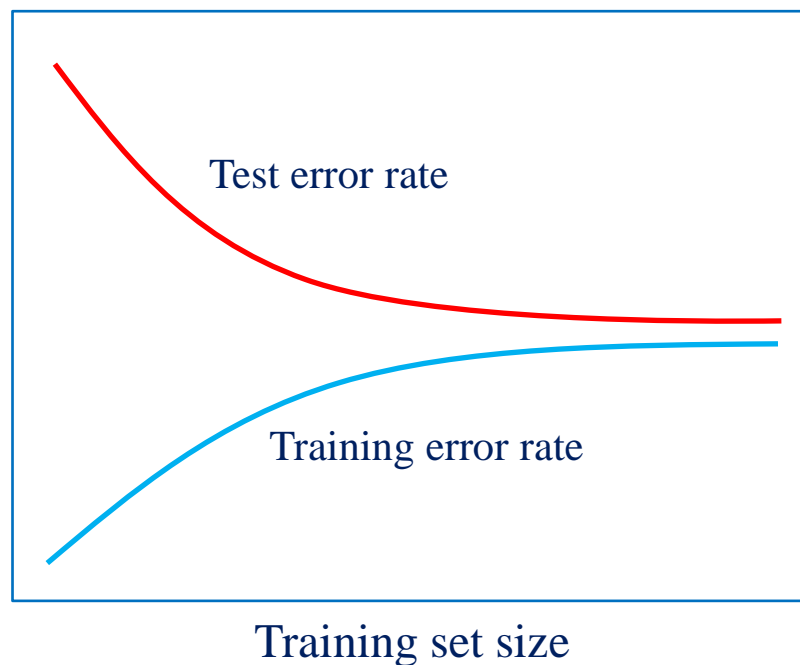
- Cross-validation (rotation)
  - N等份，每等份轮流做Test，其余部分用于训练
  - Leave-one-out (LOO)
- Holdout
- Random sampling



# 泛化性能

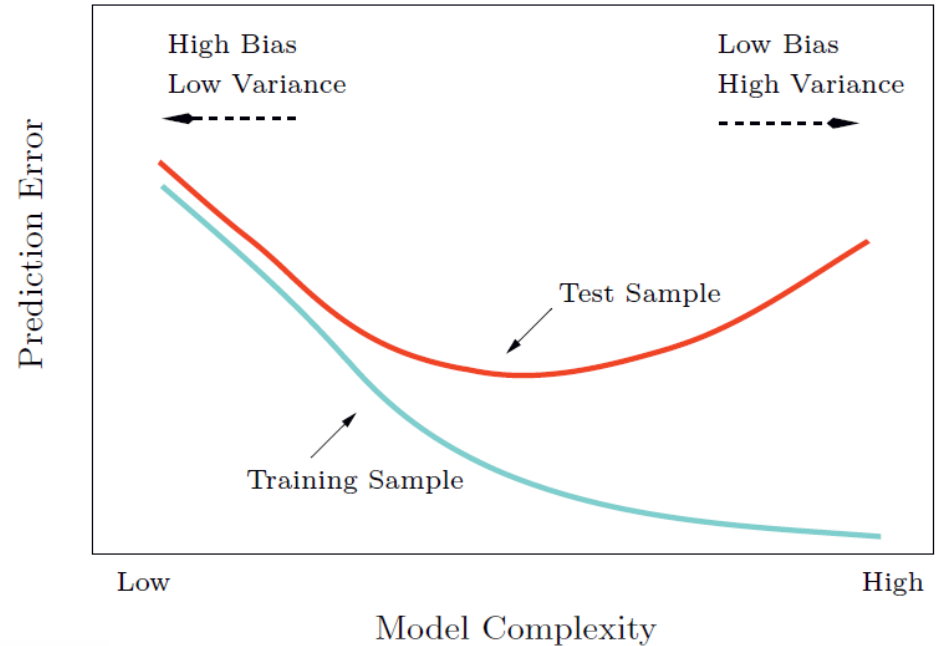
- 泛化性能(Generalization Performance): 测试数据上的分类性能
- 测试错误率跟训练错误率往往是有差异的
- 过拟合/过学习: 用复杂分类器能将训练数据分类错误率降到极低
- 训练数据越多、越有代表性, 则泛化性能越好

Error rate vs training set size

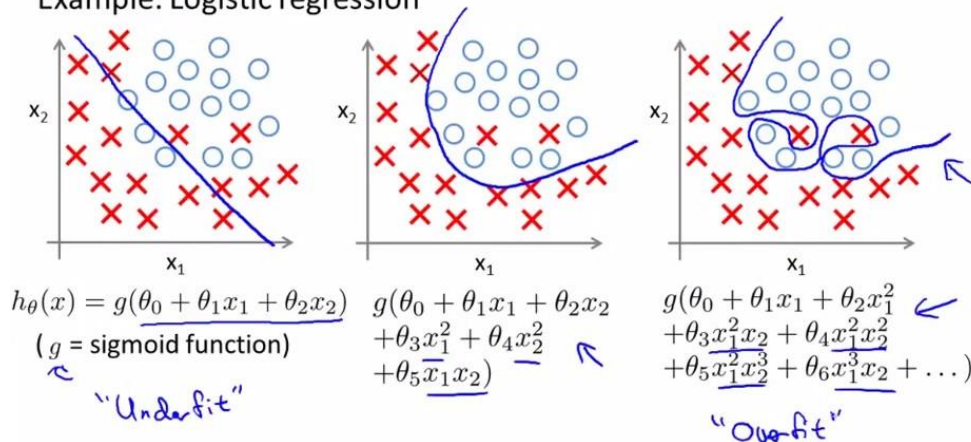


分类器（模型）复杂度  
对泛化性能的影响：

- 训练数据不变的情况下，分类器越复杂，对训练数据拟合程度越高
- 过拟合情况下，泛化性能会下降



Example: Logistic regression



T. Hastie, T. Tibshirani, J. Friedman (2001)  
*The Elements of Statistical Learning.*

过拟合的例子，来自  
[www.verydemo.com](http://www.verydemo.com)



# 模式识别方法分类

- 按特征表示分类
- 学习方法分类
- 生成/判别模型

# 模式识别方法分类

## 按模式/模型表示方式分类

- Statistical: 特征矢量
  - Parametric (Gaussian)
  - Non-parametric (Parzen window, k-NN)
  - Semi-parametric (GM)
  - Neural network
  - Logistic regression
  - Decision tree
  - Kernel (SVM)
  - Ensemble (Boosting)
- Structural: 句法、结构
  - Syntactic parsing
  - String matching, tree
  - Graph matching
  - Hidden Markov model (HMM)
  - Markov random field (MRF)
  - Structured prediction

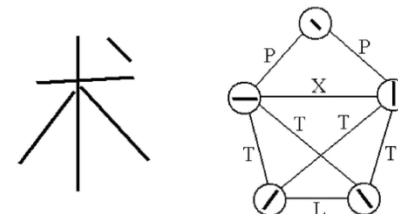
Hybrid Statistical-Structural: Statistical  
primitive/relationship

Attributed graphs, HMM and MRF/CRF are instances of hybrid models

# 为什么需要结构方法

- Problems Un-Solvable by Statistical Methods

- Need to **explore the structure** (e.g., strokes and radicals of a character)



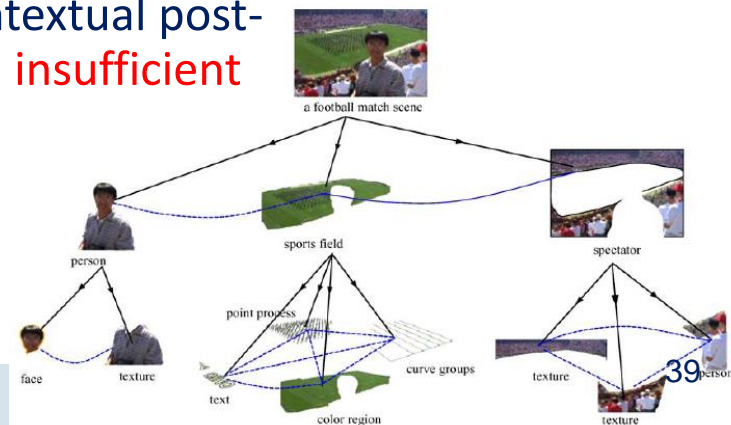
- Patterns of variable-size (e.g., character string)

- Holistic classification: **huge number of classes** (e.g.,  $10^6$  classes for 6-digit zipcodes)

**Please** *promote*

- Simultaneous classification of multiple related parts/objects

- Individual classification followed by contextual post-processing: **ambiguity of segmentation, insufficient utilization of context**



中国科学院大学

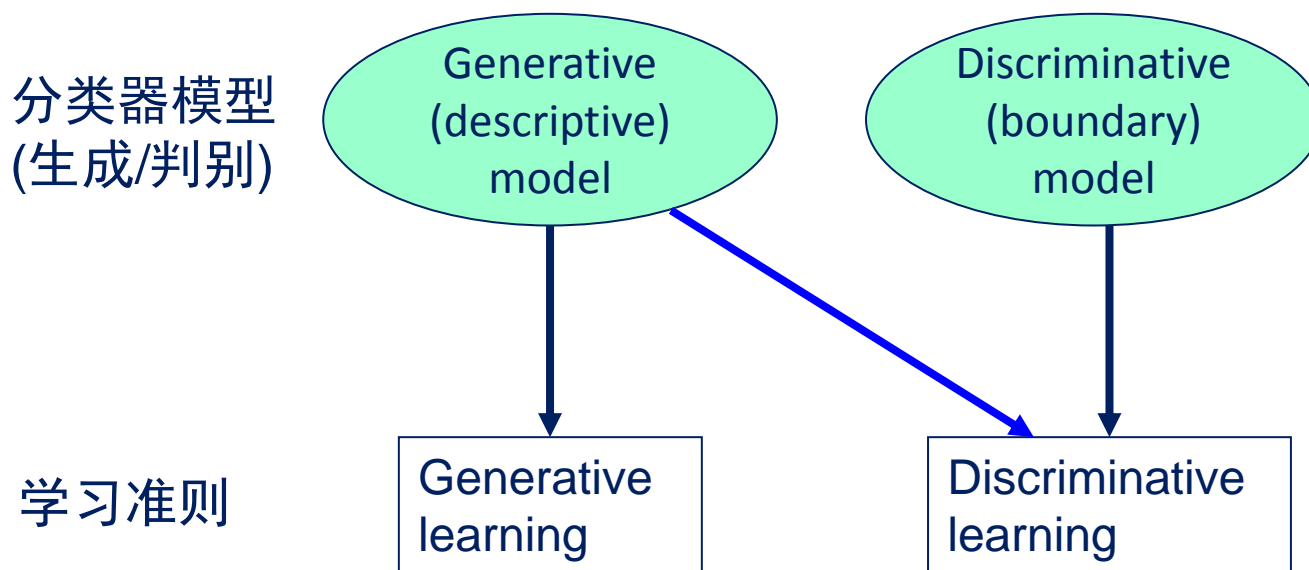
University of Chinese Academy of Sciences

# 学习方法分类

- 监督(Supervised)学习
  - 训练样本有类别标号
- 无监督(Unsupervised)学习
  - 训练样本无类别标号，得到数据结构表示或分布
- 半监督(Semi-supervised)学习
  - 训练样本一部分有类别标号，一部分没有
- Reinforcement learning
  - 学习过程中给出奖惩信号
  - 例如，Deep Mind（被Google收购）基于深度神经网络强化学习的玩视频游戏程序
- Domain Adaptation
  - 测试样本分布发生变化，分类器参数自适应

# 生成/判别模型

- 生成模型(Generative model): 表示各个类别内部结构或特征分布
- 判别模型(Discriminative model): 表示不同类别之间的区别, 一般为判别函数(Discriminant function)或边界函数



- 生成学习: 得到每个类别的结构描述或分布函数, 不同类别分别学习
- 判别学习: 得到判别函数或边界函数的参数, 所有类别样本同时学习

- Generative Models

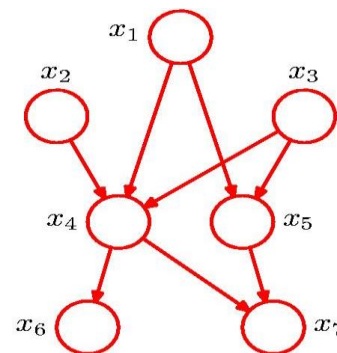
- Template (prototype)-based classifier
- Parametric probability density (Gaussian, GM)

$$p(\mathbf{x} | C) = f(\mathbf{x}, \theta)$$

- Bayesian network (directed graph)

概率密度函数树近似

$$p(\mathbf{x}) = \prod_{i=1}^n p(x_i | pa_i)$$

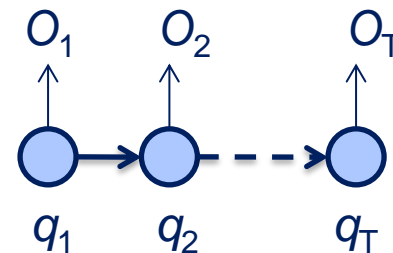


- Hidden Markov model (HMM)

特征矢量序列  
密度函数近似

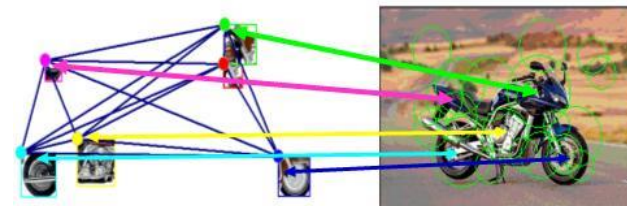
$$p(O | \lambda) = \sum_Q P(O | Q, \lambda) P(Q | \lambda)$$

$$= \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \dots a_{q_{T-1} q_T} b_{q_T}(O_T)$$



- Undirected graphs

- Attributed relational graph (ARG)
- Markov random field (MRF)



- Discriminative Models

- Artificial neural networks (ANN):  
discriminant function regardless of probability distribution

神经网络输出  
近似后验概率

$$y_i(\mathbf{x}) \doteq P(\omega_i | \mathbf{x})$$

- Support vector machine (SVM):  
hyperplane classifier (2-class)

- Decision boundary

$$\mathbf{w} \cdot \mathbf{x} + b >< 0$$

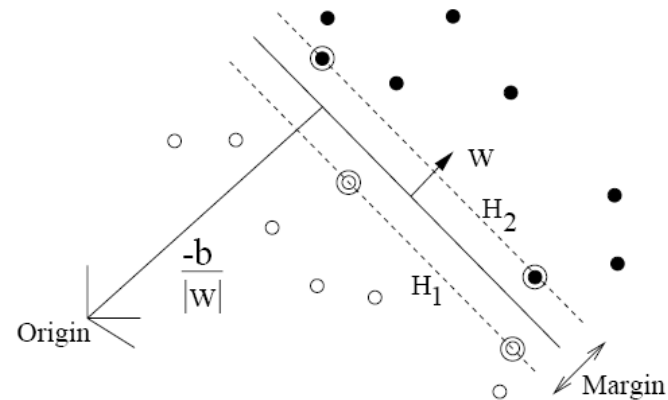
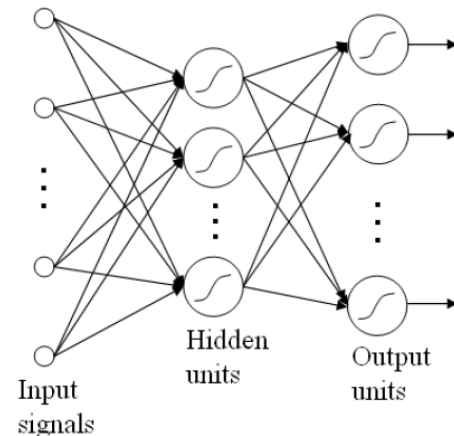
- Boosting: weighted combination of multi-discriminators (2-class)

Boosting判别函数是  
多个分类器加权和

$$F(\mathbf{x}) = \sum_{t=1}^T \alpha_t h_t(\mathbf{x})$$

- Conditional random field (CRF)

- Labeling by minimizing energy function, without assumption of conditional independence





# 本课程内容体系

# 本课程内容体系

12章，18次授课

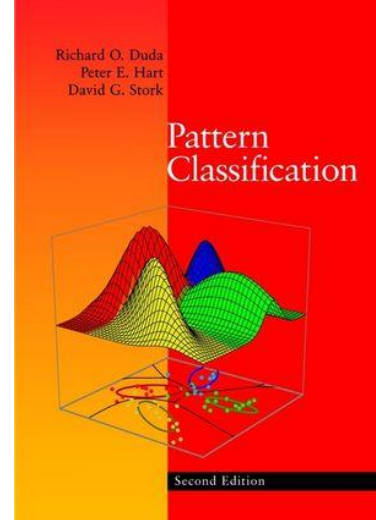
可看作  
非参数方法

一种结构方法

内容	课时	授课老师
绪论	5*3h	刘成林
贝叶斯决策理论		
概率密度参数估计		
非参数法		
线性分类器设计	7*3h	向世明
神经网络和深度学习		
模型选择(DHS第9章)		
聚类分析(非监督学习)		
支撑向量机与核方法	3*3h	王亮
决策树方法(DHS第8章)		
特征提取与选择	3*3h	刘智勇
概率图模型		



# 教科书



- 教材

- Richard O. Duda, Peter E. Hart, David G. Stork, *Pattern Classification*, 2nd Edition, John Wiley, 2001. (国内) 模式分类, 英文影印版

- 主要参考书和文献

- Andrew Webb, *Statistical Pattern Recognition*, 3rd Edition, John Wiley, 2011.
- Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- Yoshua Bengio, Ian J. Goodfellow, Aaron Courville, *Deep Learning*, <http://www.iro.umontreal.ca/~bengioy/dlbook>, 2015 (Book in preparation)
- K. Fodor, A Survey of Dimension Reduction Techniques, Center for Applied Scientific Computing, Lawrence Livermore National, Technical Report UCRL-ID-148494, 2002.
- Y. LeCun, L. Bottou, Y. Bengio, P. Haffner. Gradient-Based Learning Applied to Document Recognition, *Proceedings of IEEE*, Vol. 86, No. 11, pp. 2278-2324, 1998.
- A. Y. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm, In: *Advances in Neural Information Processing Systems*, 2001, pp. 849-856.
- 李航著, 统计学习方法, 清华大学出版社, 2012年3月出版 (第五、七章)
- Christopher J.C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery*, 2(2): 1-43, 1998.

# 下次课内容

- 第2章 贝叶斯决策理论（讲2次）
  - 导论：2类的例子
  - 最小风险决策
  - 判别函数和决策面
  - 高斯概率密度
  - 高斯密度下的判别函数
  - 错误率分析
  - 离散变量的贝叶斯决策
  - 复合模式分类