机器学习

01 绪论

李祎

liyi@dlut.edu.cn

大连理工大学 人工智能学院
School of Artificial Intelligence, Dalian University of Technology

01

PART ONE

引言

Introduction

课程介绍

■学习基础

➢矩阵与数值分析（线性代数），概率与统计，最优化理论

■交叉课程

➢图像处理，计算机视觉，数据挖掘，自然语言处理，多媒体技术

■参考书目

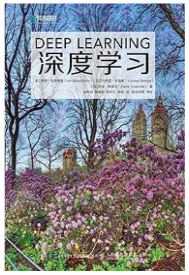
➢《机器学习》，周志华等，清华大学出版社

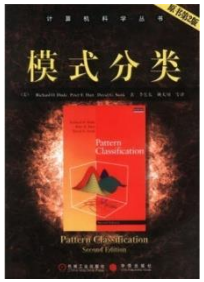
➢《统计学习方法》，李航等， 清华大学出版社

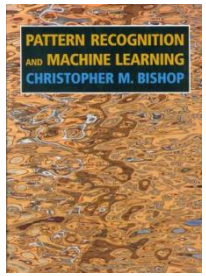
➢《深度学习》，Ian Goodfellow等，人民邮电出版社

➢《模式分类（第2版）》(Pattern Classification)，[美]迪达等著；

《Pattern Recognition and Machine Learning》, Christopher Bishop, Springer







■主要授课内容

- ➢机器学习基本概念
- ➢传统机器学习算法

  监督学习，无监督学习，半监督学习

  线性回归，逻辑斯蒂回归，Boosting，K-Means, Naive Bayes，

  支持向量机，...
- ➢稀疏表示与低秩矩阵，神经网络与深度学习，强化学习
- ➢机器学习前沿：迁移学习，对抗学习...

考核方法

■笔试（60%）

- 闭卷
- 考试时间：待定

■小作业（20%）

- 课程过程中会留4次课后作业
- 提交时间：每周一上课前
- 提交方式：电子版/扫描版

■大作业（20%）

- 线上汇报
- 书面报告

大作业说明

- 组队：上限4人一组（本周四前QQ私信给助教）
- 选题：机器学习应用问题（本周四前QQ私信给助教）
- 线上汇报：
  - 时长：5-8分钟
  - 内容：所选问题的研究背景、研究进展、应用情况等
  - 时间：第二周周一或周四
- 书面报告：
  - 围绕所选问题，实现一个算法，撰写报告
  - 题目和内容自拟（算法必须与课程内容相关）
  - 模板：使用本科生毕业论文模板
  - 篇幅：8-12页（不含封皮或说明性文字）；
  - 参考文献个数：不少于10个，注意在文中正确引用

（切勿出现只列参考文献，不在文中引用的情况！）

大作业说明

- 书面报告具体内容：
  - 摘要（只写中文摘要即可！！）
  - 绪论（研究内容和意义）

    注意在绪论第一段指出大作业与机器学习课程的关联！！
  - 相关工作（与研究内容直接相关的算法/参考文献）
  - 正文（所设计算法细节和流程，自己设计的算法指明创新点）
  - 实验（实验数据集，实验评测协议，实验结果）
  - 结论
  - 成员贡献（写清组内每位成员的具体工作和贡献，包括线上汇报）
  - 参考文献
- 提交时间：课程结束后两周内，提交方式待定。

■期刊

- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
- IEEE Transactions on Image Processing (TIP)
- Pattern Recognition (PR)
- IEEE Transactions on Neural Networks and Learning Systems (TNNLS)

■ 会议

- ICML (International Conference on Machine Learning)
- NeurIPS (Neural Information Processing Systems)
- CVPR (IEEE Conference on Computer Vision and Pattern Recognition)
- ICLR (International Conference on Learning Representations)
- ICCV (International Conference on Computer Vision)

….

- Arxiv: https://zh.wikipedia.org/wiki/ArXivarXiv

■会议或期刊

➢VALSE: http://valser.org

■ 机器学习

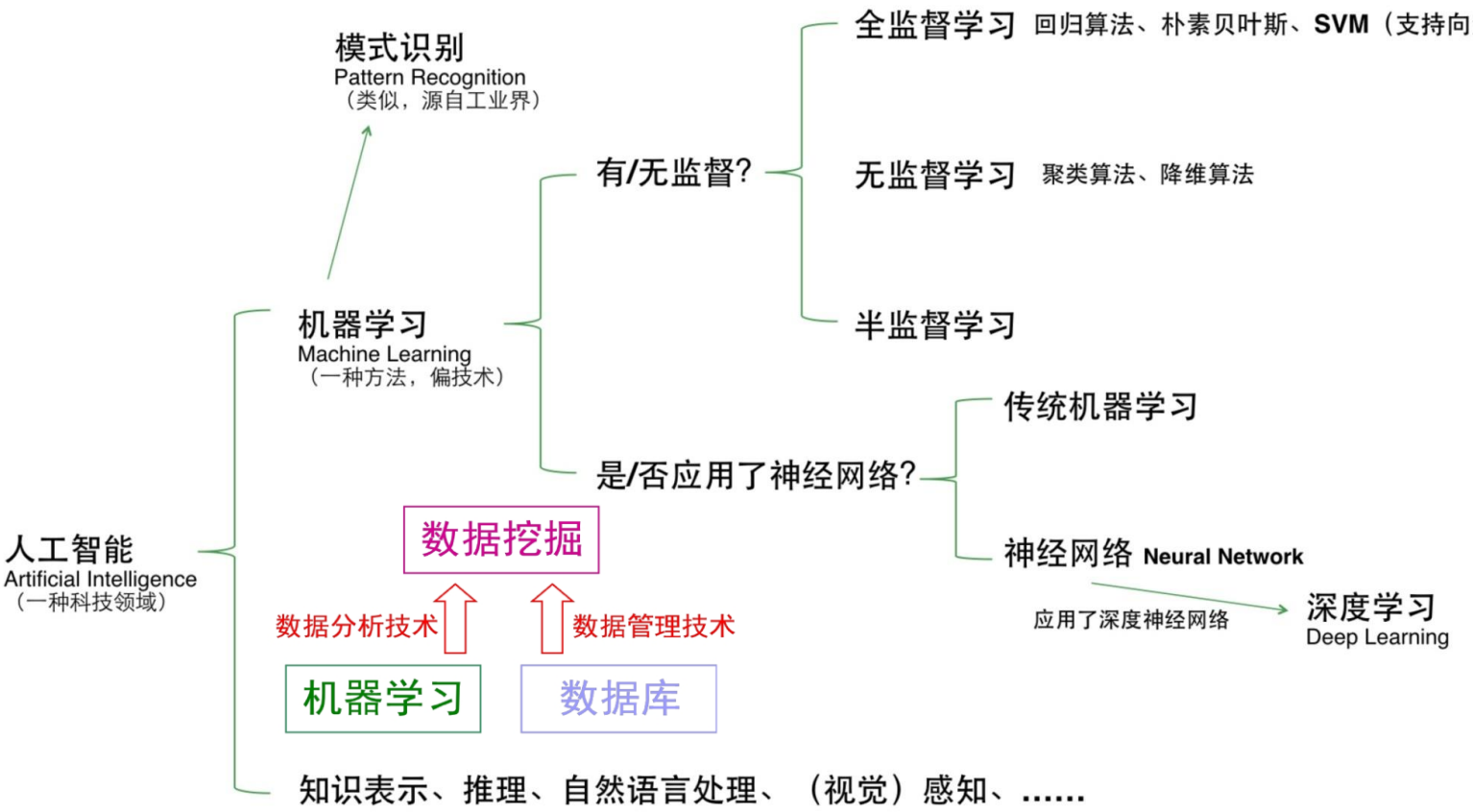
https://zh.wikipedia.org/wiki/%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0

机器学习是人工智能的一个分支。人工智能的研究历史有着一条从以“推理”为重点，到以“知识”为重点，再到以“学习”为重点的自然、清晰的脉络。显然，机器学习是实现人工智能的一个途径，即以机器学习为手段解决人工智能中的问题。机器学习在近30多年已发展为一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、计算复杂性理论等多门学科。机器学习理论主要是设计和分析一些让计算机可以自动“学习”的算法。

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。因为学习算法中涉及了大量的统计学理论，机器学习与推断统计学联系尤为密切，也被称为统计学习理论。算法设计方面，机器学习理论关注可以实现的，行之有效的学习算法。很多推论问题属于无程序可循难度，所以部分的机器学习研究是开发容易处理的近似算法。

机器学习已广泛应用于数据挖掘、计算机视觉、自然语言处理、生物特征识别、搜索引擎、医学诊断、检测信用卡欺诈、证券市场分析、DNA序列测序、语音和手写识别、战略游戏和机器人等领域。

什么是机器学习



➢**模式识别=机器学习**。两者的主要区别在于前者是从工业界发展起来的概念，后者则主要源自计算机学科。在著名的《Pattern Recognition And Machine Learning》这本书中，Christopher M. Bishop在开头是这样说的“模式识别源自工业界，而机器学习来自于计算机学科。不过，它们中的活动可以被视为同一个领域的两个方面，同时在过去的10年间，它们都有了长足的发展”。

什么是机器学习

机器学习是从人工智能中产生的一个重要学科分支，是实现智能化的关键

经典定义：利用经验改善系统自身的性能

经验 → 数据

随着该领域的发展，目前主要研究智能数据分析的理论和算法，并已成为智能数据分析技术的源泉之一

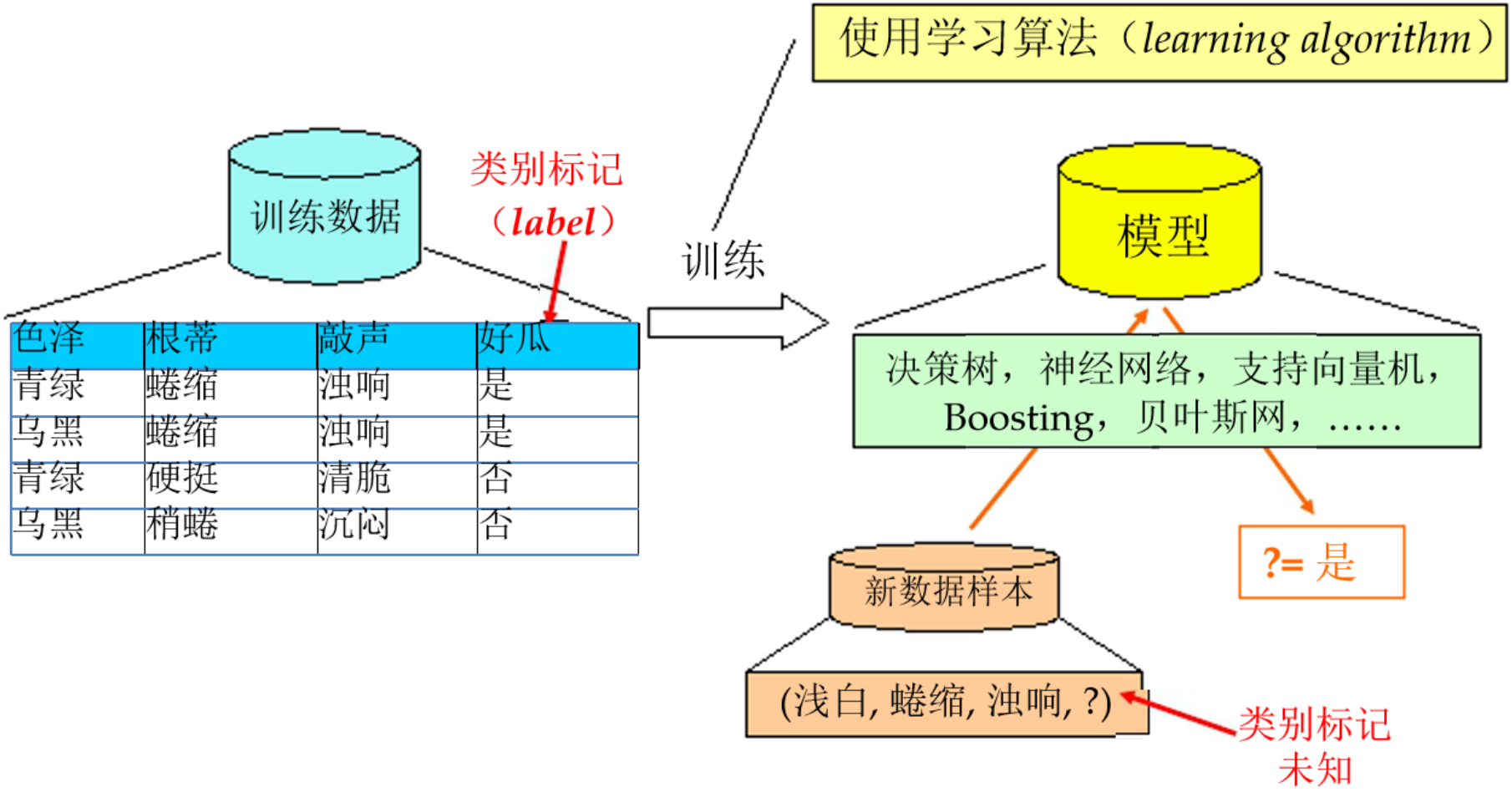
近期机器学习相关ACM图灵奖：

2011, Leslie Valiant, “计算学习理论”

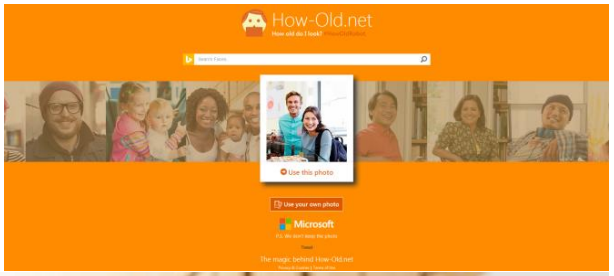
2012, Judea Pearl, “图模型学习方法”

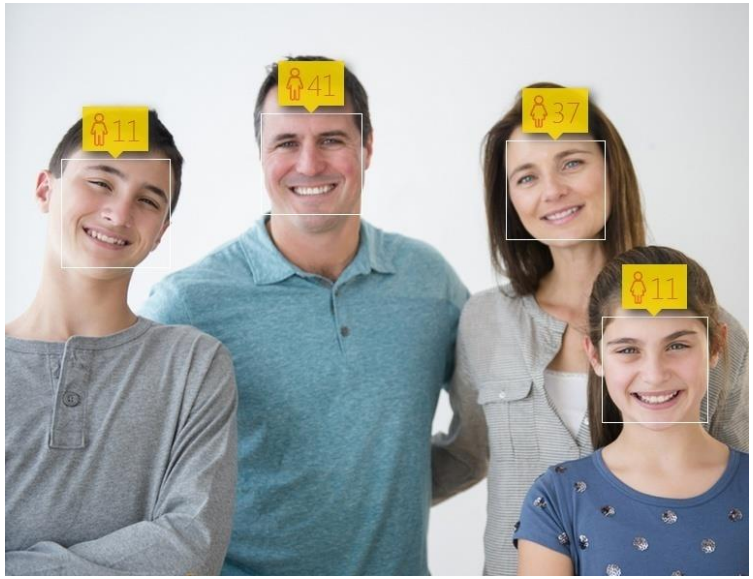
2018, **Geoffrey Hinton, Yoshua Bengio, Yann LeCun**, “神经网络与深度学习”

典型的机器学习过程


使用学习算法（learning algorithm）
训练数据
类别标记
（label）
训练
模型
色泽 根蒂 敲声 好瓜
青绿 蜷缩 浊响 是
乌黑 蜷缩 浊响 是
青绿 硬挺 清脆 否
乌黑 稍蜷 沉闷 否
决策树，神经网络，支持向量机，
Boosting，贝叶斯网，……
新数据样本
?= 是
(浅白, 蜷缩, 浊响, ?)
类别标记
未知

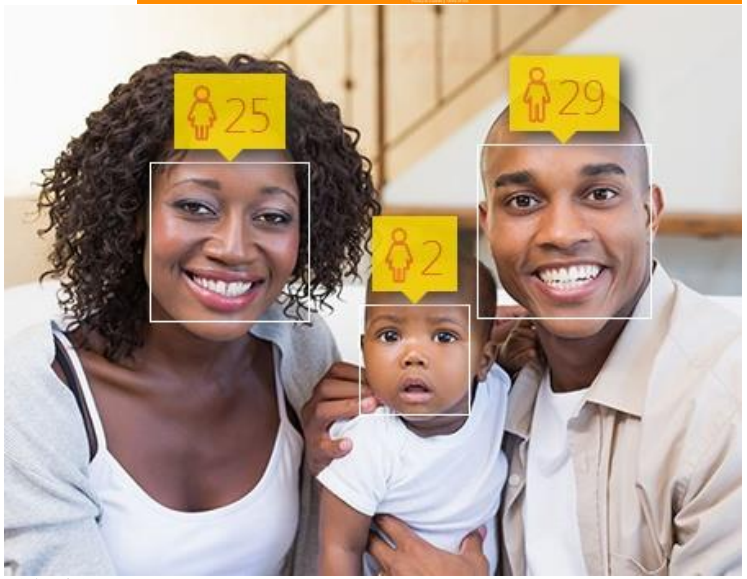
➢**how-old.net**





包含子问题：

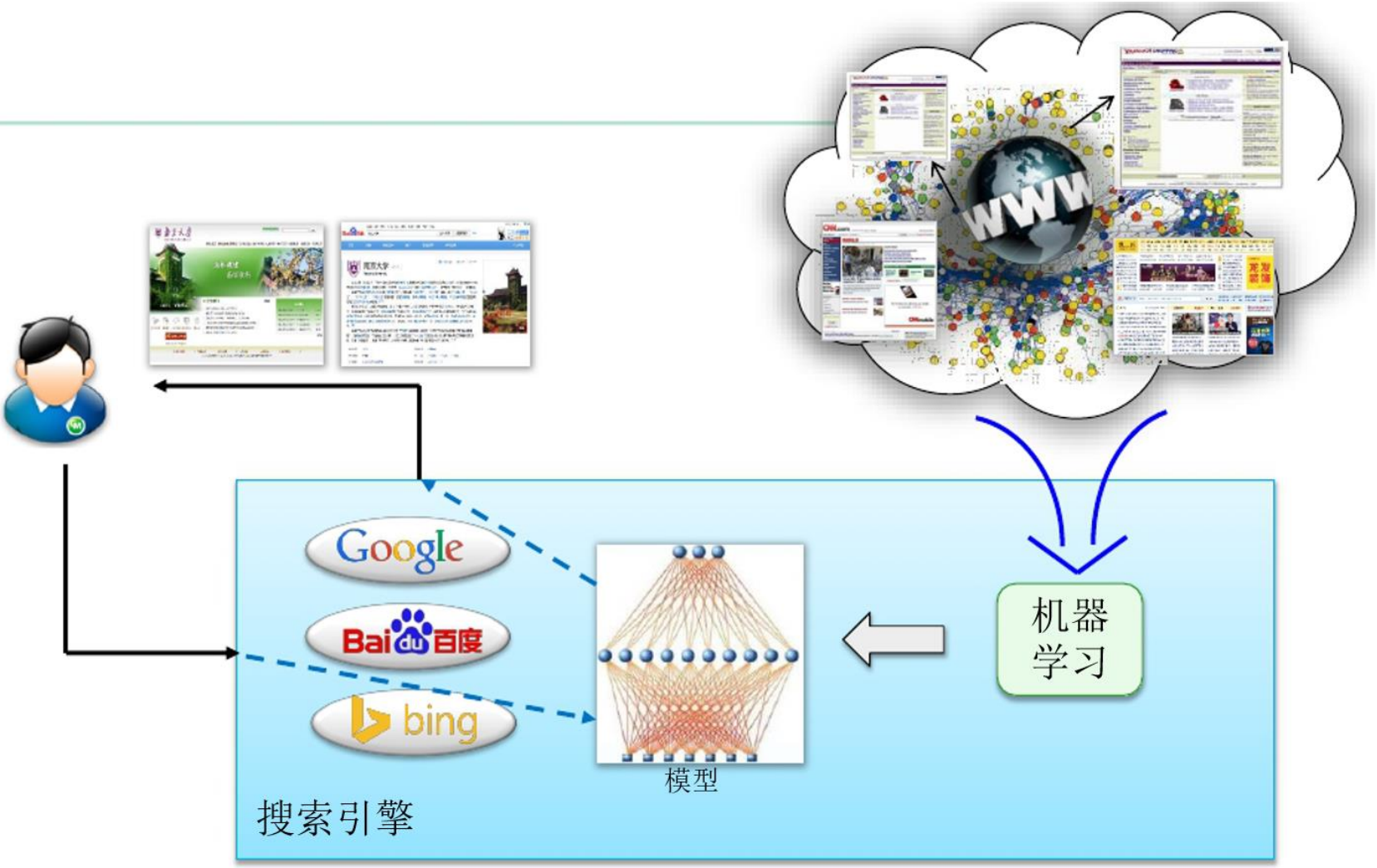
a) 人脸检测，Face Detection
b) 人脸对齐，Face Alignment
c) 年龄分类，Age Classification



其他子问题：

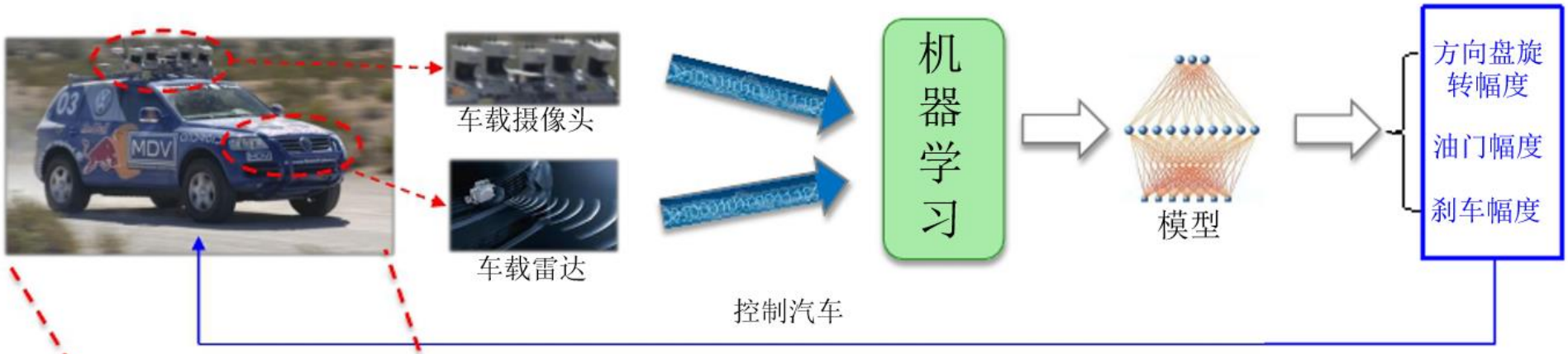
a) 人脸识别，Face Recognition
b) 性别识别，Gender Recognition
c) 表情识别，Expression Recognition
d) 种族识别，Race Recognition

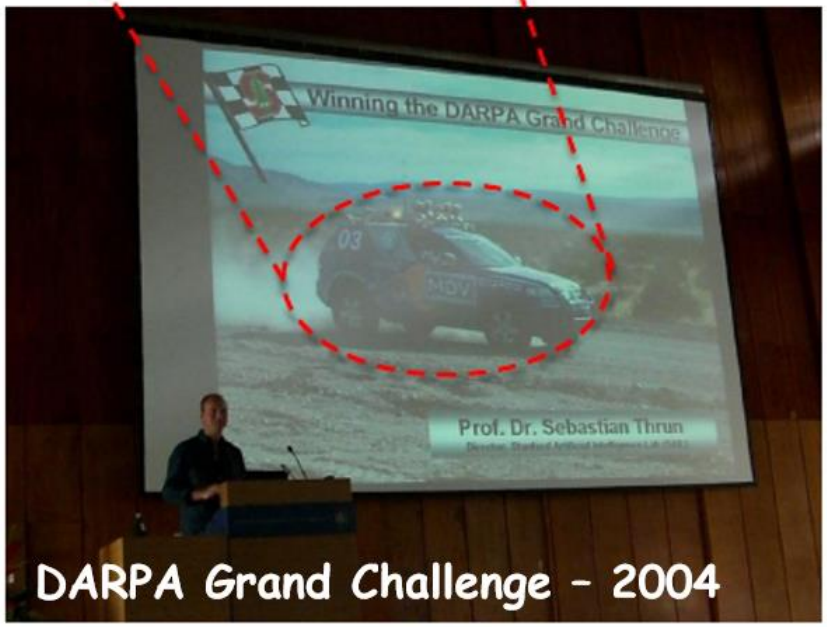
机器学习应用（网络搜索）

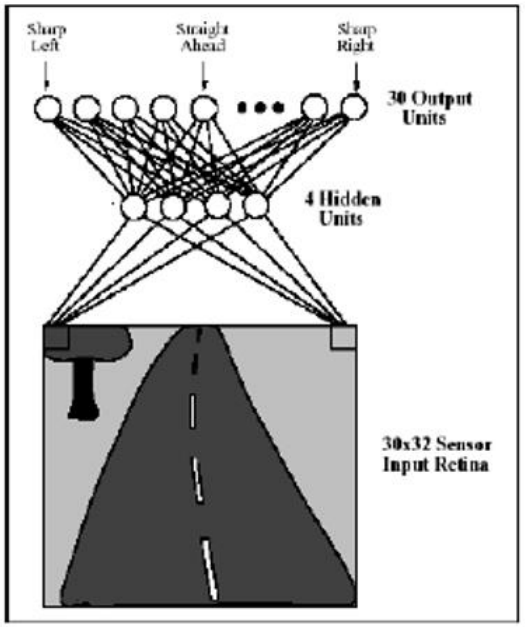


机器学习技术正在支撑着各种搜索引擎

机器学习应用（自动驾驶）







美国在20世纪80年代就开始研究基于机器学习的汽车自动驾驶技术

■ 机器学习源自“人工智能”

1956年夏　　美国达特茅斯学院

J. McCarthy, M. Minsky, N. Lochester, C. E. Shannon,
H.A. Simon, A. Newell, A. L. Samuel 等10余人

约翰 麦卡锡
(1927-2011)
“人工智能之父”
1971年图灵奖

达特茅斯会议标志着人工智能这一学科的诞生

John McCarthy (1927 - 2011):

1971年获图灵奖, 1985年获IJCAI终身成就奖。人工智能之父。他提出了“人工智能”的概念，设计出函数型程序设计语言Lisp，发展了递归的概念，提出常识推理和情境演算。出生于共产党家庭，从小阅读《10万个为什么》，中学时自修CalTech的数学课程，17岁进入CalTech时免修两年数学，22岁在Princeton获博士学位，37岁担任Stanford大学AI实验室主任。

■ 第一阶段：推理期

1956-1960s: Logic Reasoning

- 出发点: “数学家真聪明！”
- 主要成就: 自动定理证明系统 (例如，西蒙与纽厄尔的“Logic Theorist”系统)

渐渐地，研究者们意识到，仅有逻辑推理能力是不够的 ...

赫伯特 西蒙
(1916-2001)
1975年图灵奖

阿伦 纽厄尔
(1927-1992)
1975年图灵奖

■ 第二阶段：知识期

1970s -1980s: Knowledge Engineering

爱德华 费根鲍姆
(1936-)
1994年图灵奖

- ◆ 出发点: “知识就是力量！”
- ◆ 主要成就: 专家系统 (例如，费根鲍姆等人的“DENDRAL”系统)

渐渐地，研究者们发现，要总结出知识再“教”给系统，实在太难了 ...

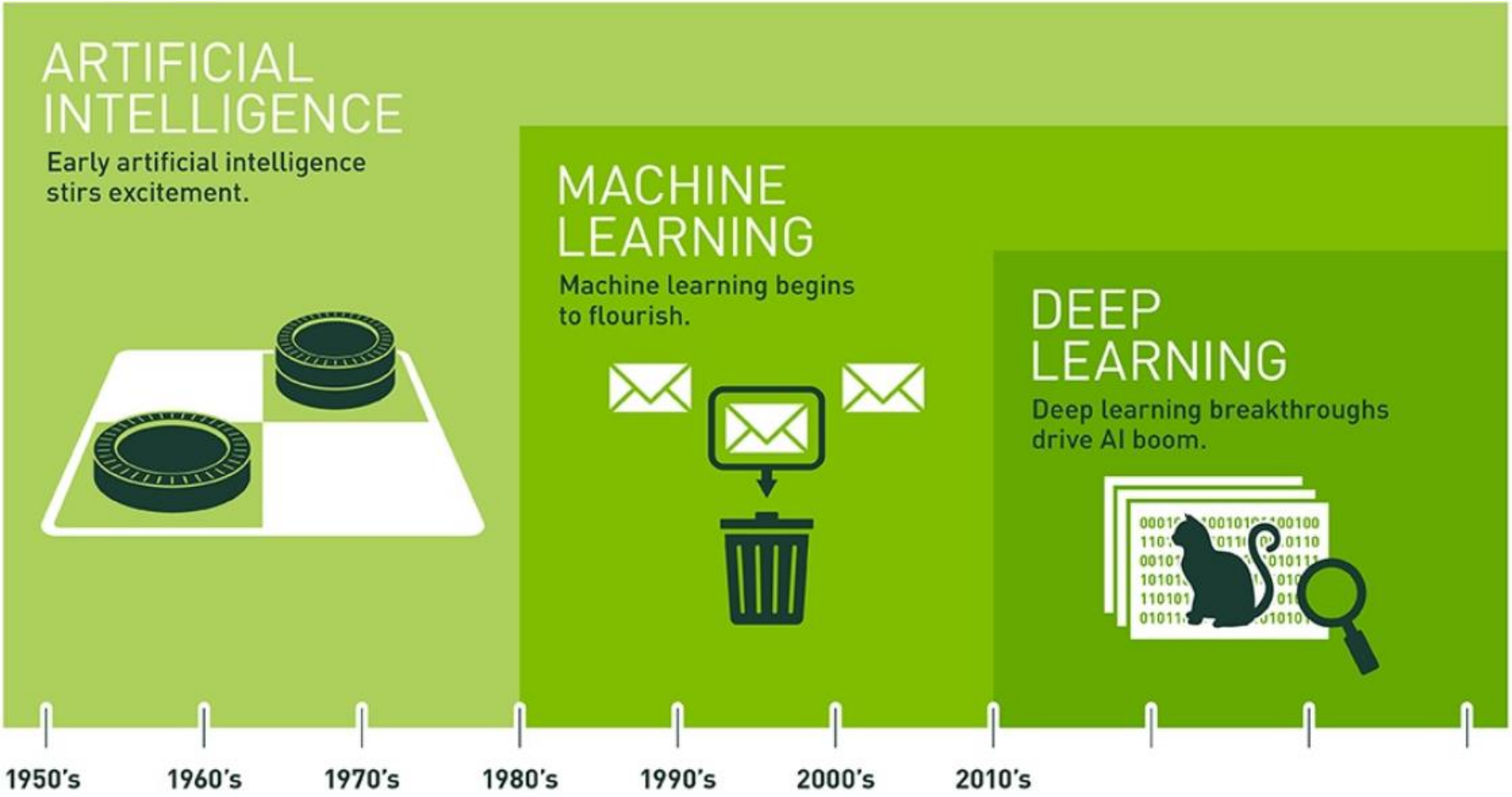
■ 第三阶段：学习期

1990s -now: Machine Learning

- ◆ 出发点: “让系统自己学！”
- ◆ 主要成就: .....

机器学习是作为“突破知识工程瓶颈”之利器而出现的

恰好在20世纪90年代中后期，人类发现自己淹没在数据的汪洋中，对自动数据分析技术——机器学习的需求日益迫切


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

奥巴马提出“大数据计划”后，美国NSF进一步加强资助UC Berkeley研究如何整合将”数据”转变为”信息”的三大关键技术——机器学习、云计算、众包(crowd sourcing)

National Science Foundation: In addition to funding the Big Data solicitation, and keeping with its focus on basic research, NSF is implementing a comprehensive, long-term strategy that includes new methods to derive knowledge from data; infrastructure to manage, curate, and serve data to communities; and new approaches to education and workforce development. Specifically, NSF is:

整合三大关键技术

- Encouraging research universities to develop interdisciplinary graduate programs to prepare the next-generation Expeditions in Computing project based at the University of California, Berkeley, that will integrate three powerful approaches for turning data into information - machine learning, cloud computing, and crowd sourcing;
- Funding, through the first round of grants to support "EarthCube" – a system that will allow geoscientists to access, analyze and share information about our planet;
- Providing the first round of grants to support "EarthCube" – a system that will allow geoscientists to access, analyze and share information about our planet;
- Issuing a $2 million award for a research training group to support training for undergraduates to use graphical and visualization techniques for complex data.
- Providing $1.4 million in support for a focused research group of statisticians and biologists to determine protein structures and biological pathways.
- Convening researchers across disciplines to determine how Big Data can transform teaching and learning.

大数据时代，机器学习必不可少！！

收集、传输、存储大数据的目的，

是为了“利用”大数据

没有机器学习技术分析大数据，

“利用”无从谈起

02

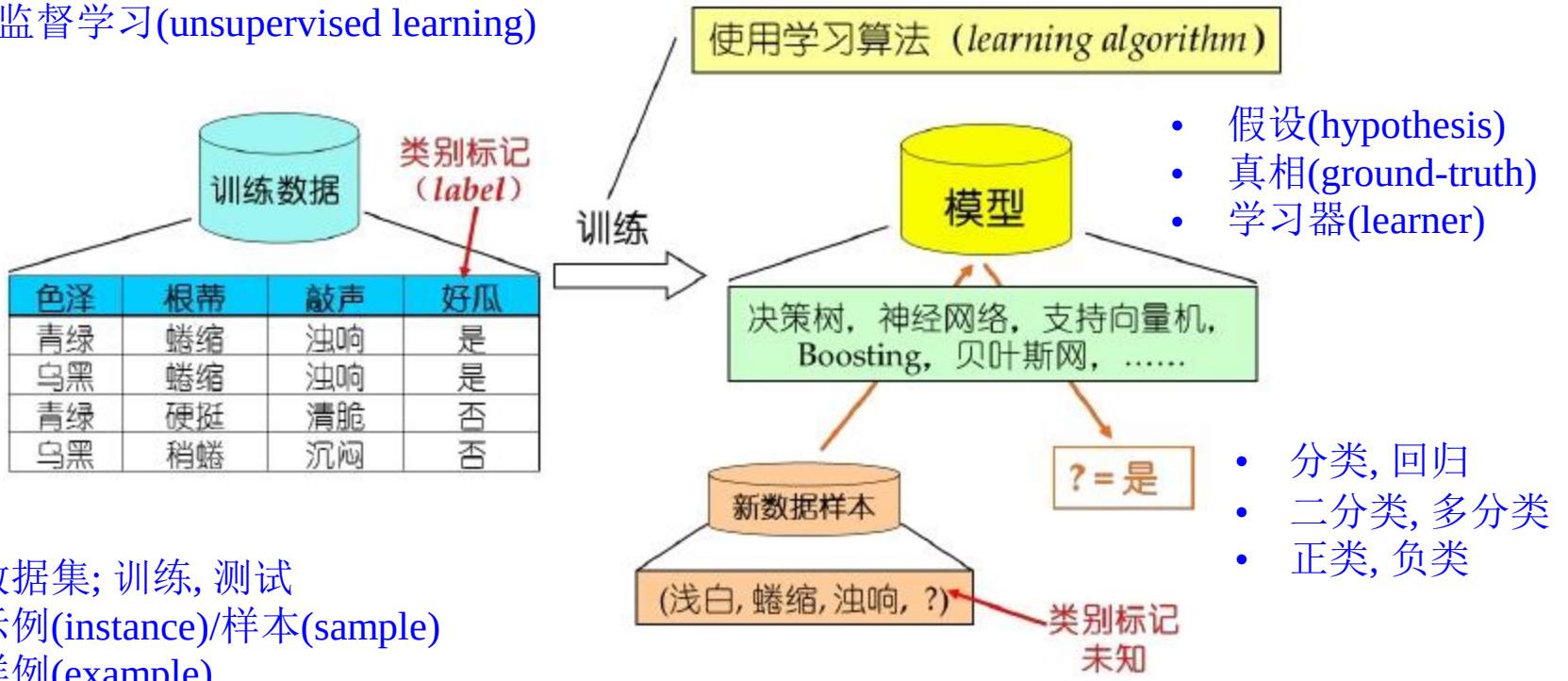
PART TWO

基本术语

Basic Notions

基本术语

- 监督学习(supervised learning)
- 无监督学习(unsupervised learning)



色泽	根蒂	敲声	好瓜
青绿	蜷缩	浊响	是
乌黑	蜷缩	浊响	是
青绿	硬挺	清脆	否
乌黑	稍蜷	沉闷	否

- 假设(hypothesis)
- 真相(ground-truth)
- 学习器(learner)

- 分类, 回归
- 二分类, 多分类
- 正类, 负类

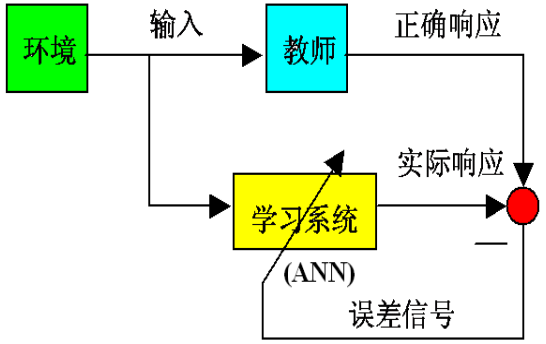
- 数据集; 训练, 测试
- 示例(instance)/样本(sample)
- 样例(example)
- 属性(attribute), 特征(feature); 属性值
- 属性空间, 样本空间, 输入空间
- 特征向量(feature vector)
- 标记空间, 输出空间

- 未见样本(unseen instance)
- 未知“分布”
- 独立同分布(i.i.d.)
- 泛化**(generalization)**

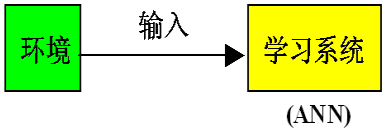
➢ **参考《机器学习》，周志华著，p.1-3**

监督学习/无监督学习

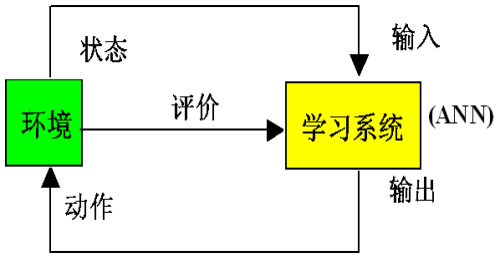
- **监督学习**(Supervised Learning)：监督学习是从标记的训练数据来推断一个功能的机器学习任务。如**分类、回归**。



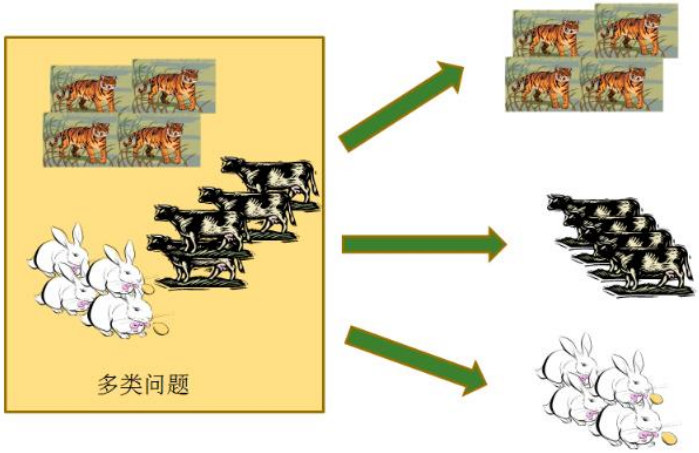
- **无监督学习**(Unsupervised Learning)：无监督学习的问题是，在未标记的数据中，试图找到隐藏的结构。如**聚类**、密度估计。



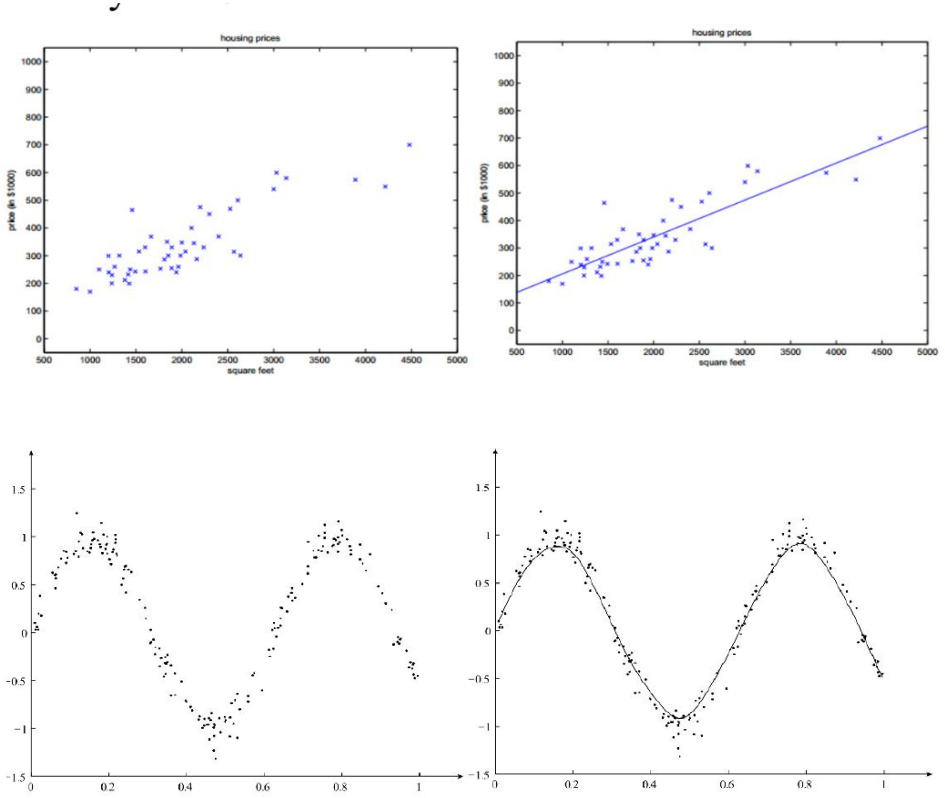
- **强化学习**(Reinforcement Learning)：强化学习是机器学习中的一个领域，强调如何基于环境而行动，以取得最大化的预期利益。



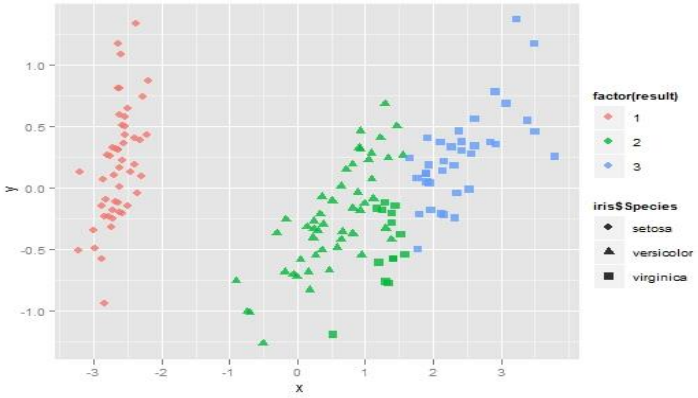
➢ 分类


多类问题

➢ 回归


housing prices
price (in $1000)
square feet
housing prices
price (in $1000)
square feet

➢ 聚类


factor(result)
1
2
3
iris$Species
setosa
versicolor
virginica
x
y

机器学习的基本过程

将数据对象进行特征(feature)化表示

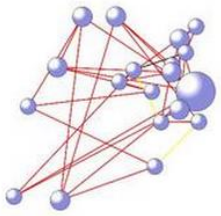
给定一个数据样本集，从中学习出规律（模型）

目标：该规律不仅适用于训练数据，也适用于未知数据(称为泛化能力)

对于一个新的数据样本，利用学到的模型进行预测

■ 向量表示法 $[x_1, \quad x_2, \ldots \quad x_n]$

■ 图表示法



➢ 天气预测：

样本：每一天

问题：如何把每天表示成一个向量？选取哪些特征？

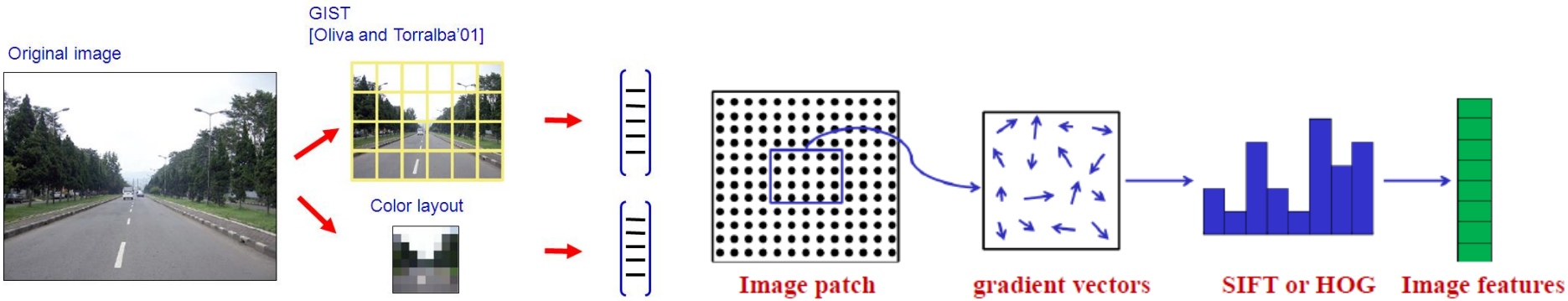
特征：温度，相对湿度，风向，风速，气压

➢ 判断好瓜坏瓜：

样本：每个瓜

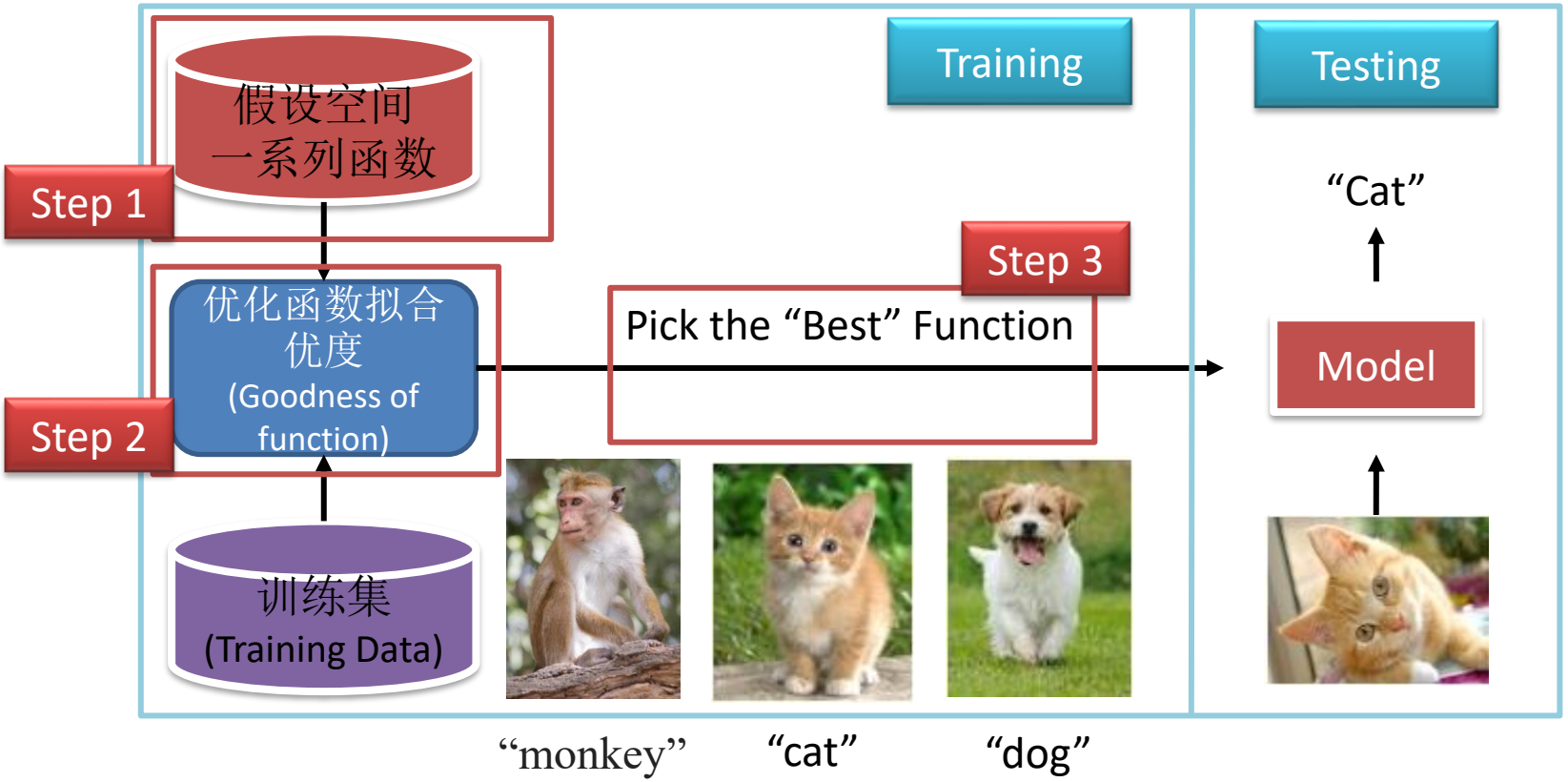
特征：色泽，根蒂，敲声

➢ 图像识别：

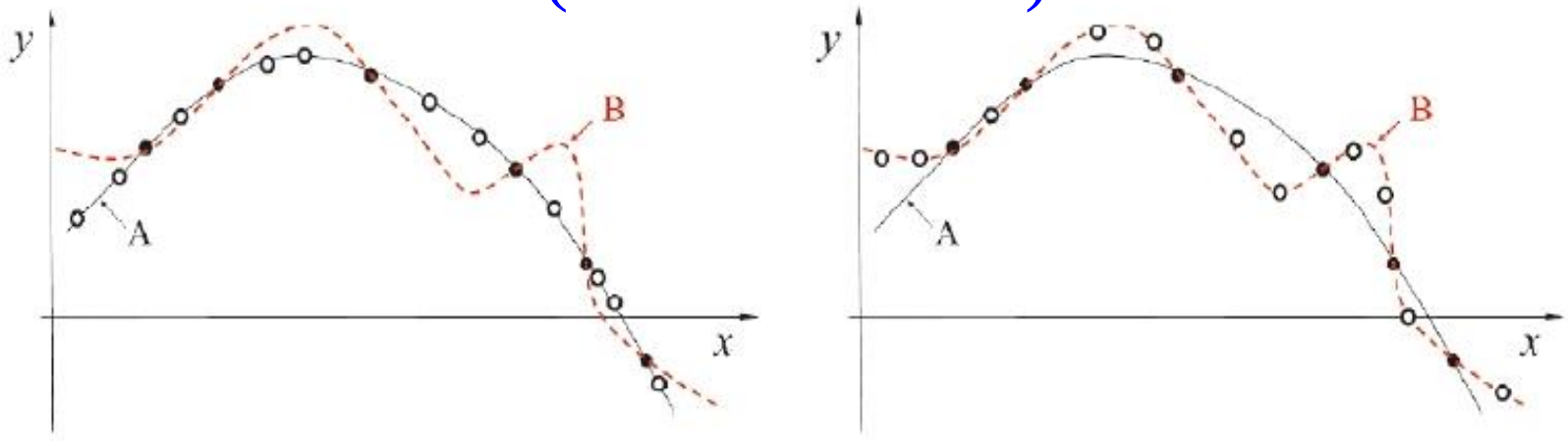


- 机器学习≈寻找函数“ f ”

➢ Image recognition



没有免费的午餐！
(No-free Lunch!)



(a) A 优于 B　　(b) B 优于 A

图 1.4 没有免费的午餐. (黑点: 训练样本; 白点: 测试样本)

NFL定理：一个算法 $\mathfrak{L}_a$ 若在某些问题上比另一个算法 $\mathfrak{L}_b$ 好，必存在另一些问题，$\mathfrak{L}_b$ 比 $\mathfrak{L}_a$ 好。

No free lunch定理寓意

NFL定理的重要前提：

所有“问题”出现的机会相同、或所有问题同等重要

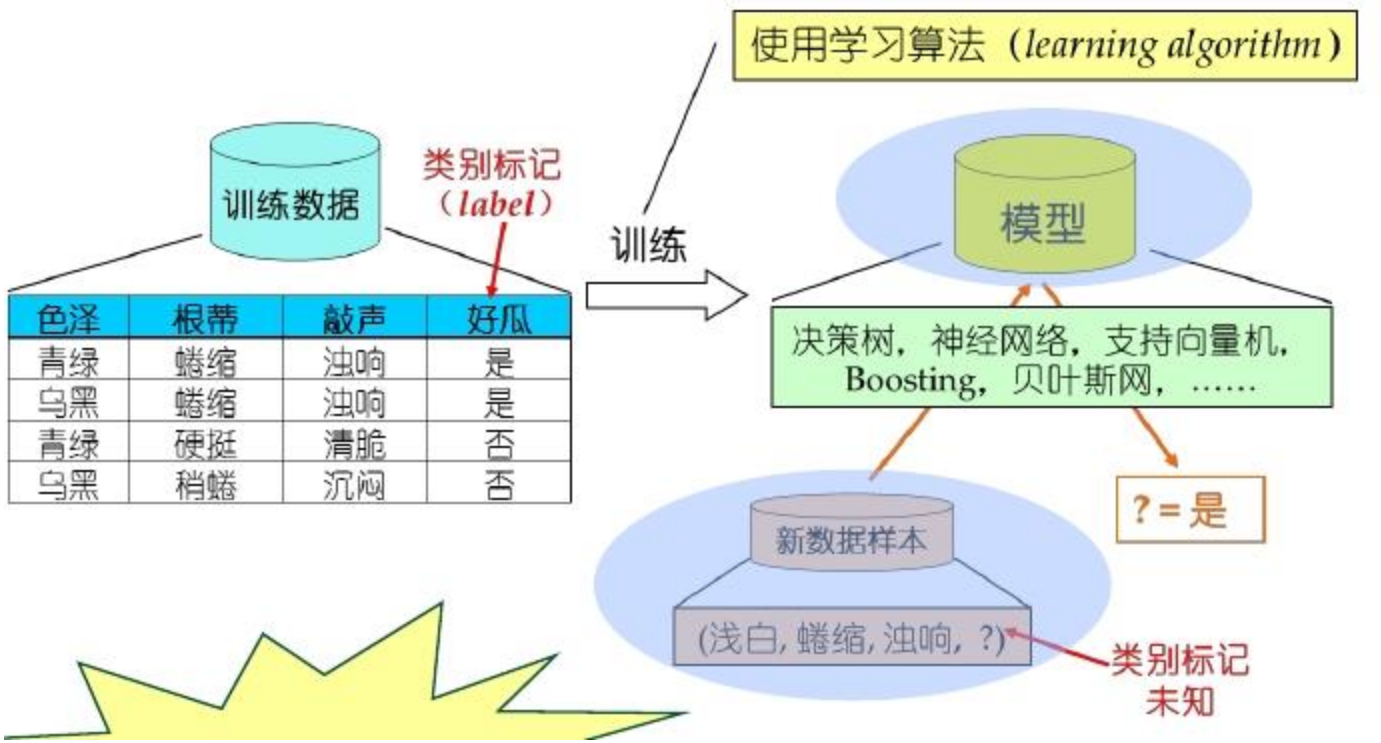
实际情形并非如此；我们通常只关注自己正在试图解决的问题

脱离具体问题，空泛地谈论“什么学习算法更好”

模型评估与选择

Model Evalution and Selection

典型的机器学习过程



色泽	根蒂	敲声	好瓜
青绿	蜷缩	浊响	是
乌黑	蜷缩	浊响	是
青绿	硬挺	清脆	否
乌黑	稍蜷	沉闷	否

泛化能力强！

能很好地适用于未知样本

例如，错误率低、精度高

然而，我们手上没有未知样本，……

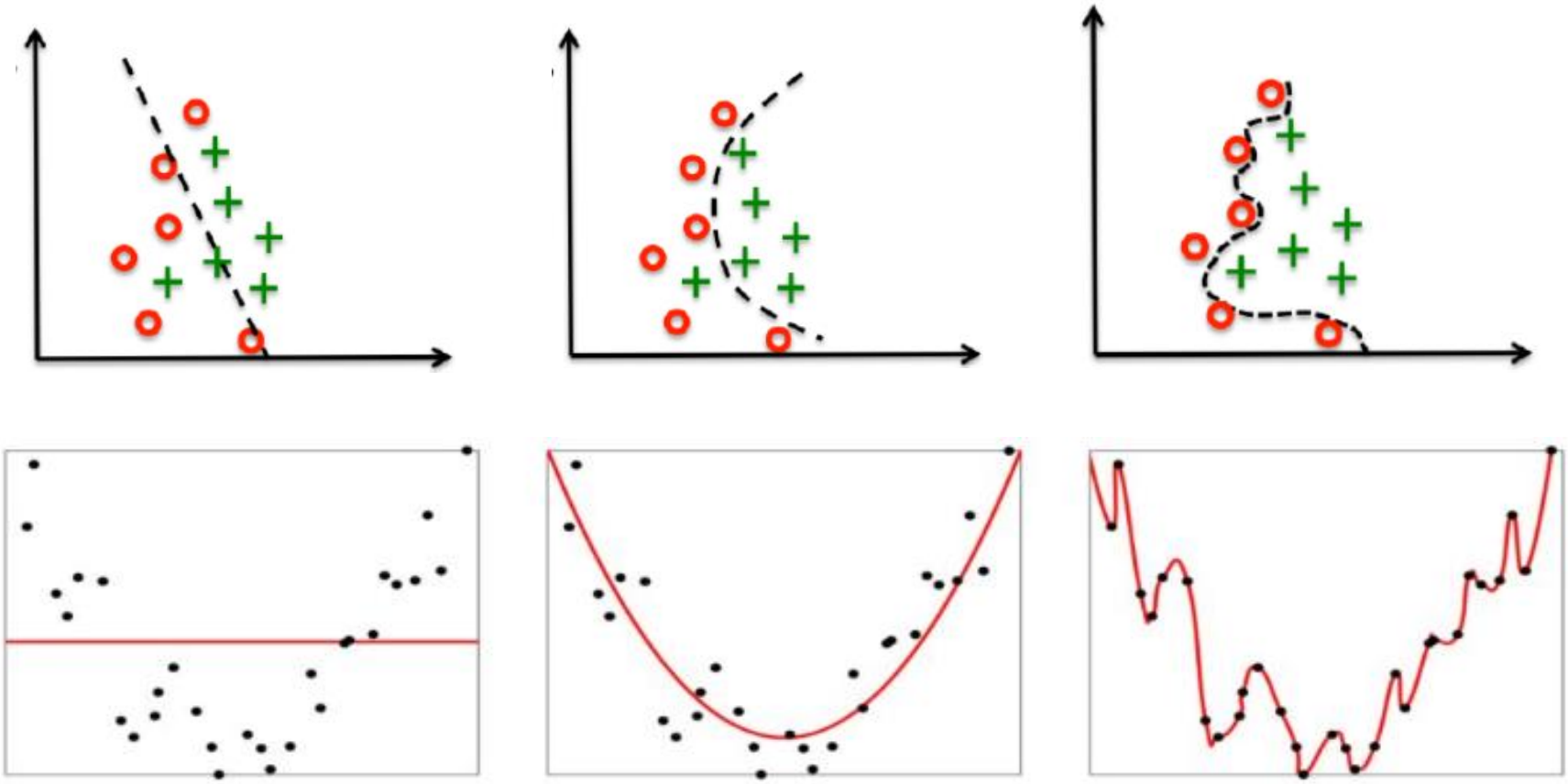
- 误差：样本真实输出与预测输出之间的差异
  - 训练(经验)误差：训练集上
  - 测试误差：测试集
  - 泛化误差：除训练集外所有样本

  - 泛化误差越小越好
  - 经验误差是否越小越好？

NO! 因为会出现“过拟合”(overfitting)

- 欠拟合：对训练样本的一般性质尚未学好
- 过拟合: 学习器把训练样本学习的“太好”，将训练样本本身的特点当做所有样本的一般性质，导致泛化性能下降

过拟合/欠拟合



欠拟合　　拟合　　过拟合

三个关键问题:

- 如何获得测试结果？ 评估方法
- 如何评估性能优劣？ 性能度量
- 如何判断实质差别？ 比较检验

关键：怎么获得“测试集”(test set)？

我们假设测试集是从样本真实分布中独立采样获得，将测试集上的“测试误差”作为泛化误差的近似，测试集应该与训练集“互斥”

常见方法：

- 留出法 (hold-out)
- **交叉验证法 (cross validation)**
- 自助法 (bootstrap)

通常将包含个m样本的数据集$D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, y_m)\}$
拆分成训练集 S 和测试集 T：

- 留出法：
  - 直接将数据集划分为两个互斥集合
  - 训练/测试集划分要尽可能保持数据分布的一致性
  - 一般若干次随机划分、重复实验取平均值
  - 训练/测试样本比例通常为2:1~4:1

k-折交叉验证法

- *k*-fold cross validation
- 将数据集分层采样划分为k个大小相似的互斥子集，每次用k-1个子集的并集作为训练集，余下的子集作为测试集，最终返回k个测试结果的均值，k最常用的取值是10.

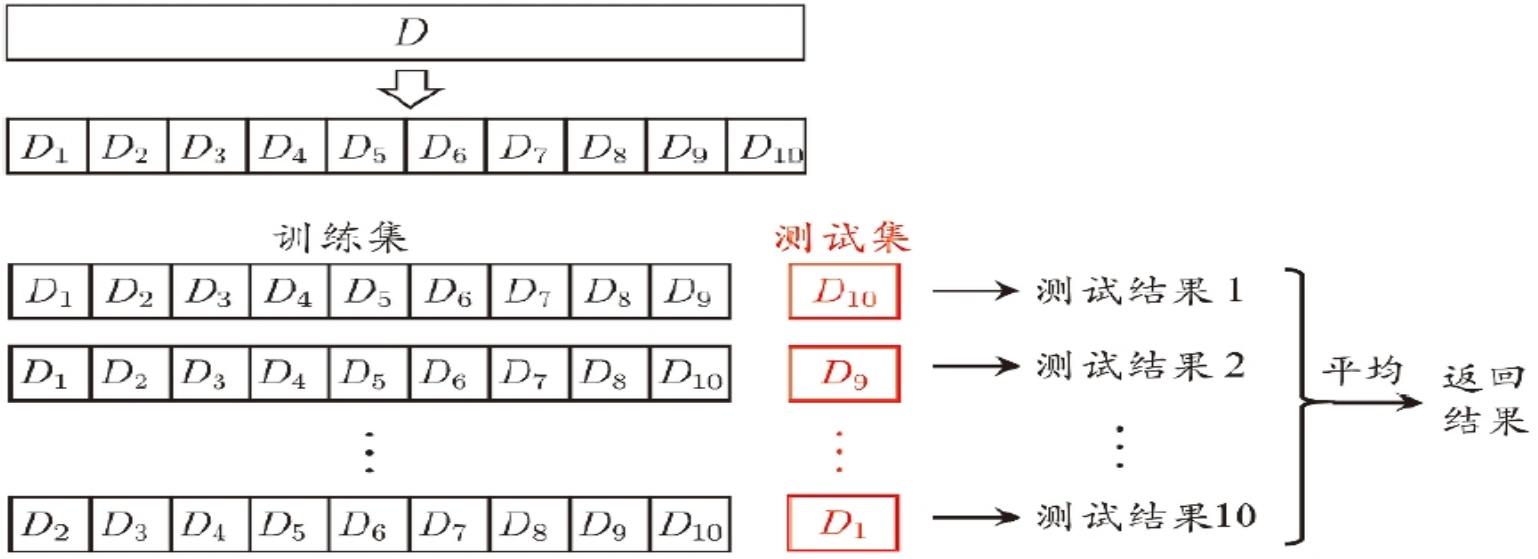


图 2.2 10 折交叉验证示意图

k-折交叉验证法

与留出法类似，将数据集D划分为k个子集同样存在多种划分方式，为了减小因样本划分不同而引入的差别，k折交叉验证通常随机使用不同的划分重复p次，最终的评估结果是这p次k折交叉验证结果的均值，例如常见的“10次10折交叉验证”

假设数据集D包含m个样本，若令 $k = m$ ，则得到留一法：

(leave-one-out, LOO)

- 不受随机样本划分方式的影响
- 结果往往比较准确
- 当数据集比较大时，计算开销难以忍受

自助法：

以自助采样法为基础，对数据集D有放回采样m次得到训练集D'，$D \setminus D'$用做测试集。

- 实际模型与预期模型都使用m个训练样本
- 约有1/3的样本没在训练集中出现
- 从初始数据集中产生多个不同的训练集，对集成学习有很大的好处
- 自助法在数据集较小、难以有效划分训练/测试集时很有用；由于改变了数据集分布可能引入估计偏差，在数据量足够时，留出法和交叉验证法更常用。

算法的参数：一般由人工设定，亦称“超参数”

模型的参数：一般由学习确定

调参过程相似：先产生若干模型，然后基于某种评估方法进行选择

参数调得好不好对性能往往对最终性能有关键影响

区别：训练集 vs. 测试集 vs. 验证集 (validation set)

算法参数选定后，要用“训练集+验证集”重新训练最终模型

三个关键问题:

- 如何获得测试结果？ 评估方法
- 如何评估性能优劣？ 性能度量
- 如何判断实质差别？ 比较检验

性能度量

- 在预测任务中，给定样例集 $D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, y_m)\}$
- 评估学习器的性能f 也即把预测结果$f(\boldsymbol{x})$和真实标记比较.

- 性能度量(performance measure)是衡量模型泛化能力的评价标准，反映了任务需求
- 使用不同的性能度量往往会导致不同的评判结果

对于回归(regression) 任务常用均方误差：

$$E(f; D) = \frac{1}{m} \sum_{i=1}^{m} (f(\boldsymbol{x}_i) - y_i)^2$$

什么样的模型是“好”的，不仅取决于算法和数据，还取决于任务需求

对于分类任务,错误率和精度是最常用的两种性能度量:

- 错误率：分错样本占样本总数的比例
- 精度：分对样本占样本总数的比率

- 错误率：

$$E(f;D)=\frac{1}{m}\sum_{i=1}^{m}\mathbb{I}\left(f\left(\boldsymbol{x}_i\right)\neq y_i\right)$$

- 精度：

$$\begin{aligned}\mathrm{acc}(f;D) &= \frac{1}{m}\sum_{i=1}^{m}\mathbb{I}\left(f\left(\boldsymbol{x}_i\right)=y_i\right)\\ &= 1-E(f;D)\ .\end{aligned}$$

什么样的模型是“好”的，不仅取决于算法和数据，还取决于任务需求

信息检索、Web搜索等场景中经常需要衡量正例被预测出来的比率或者预测出来的正例中正确的比率，此时查准率和查全率比错误率和精度更适合。

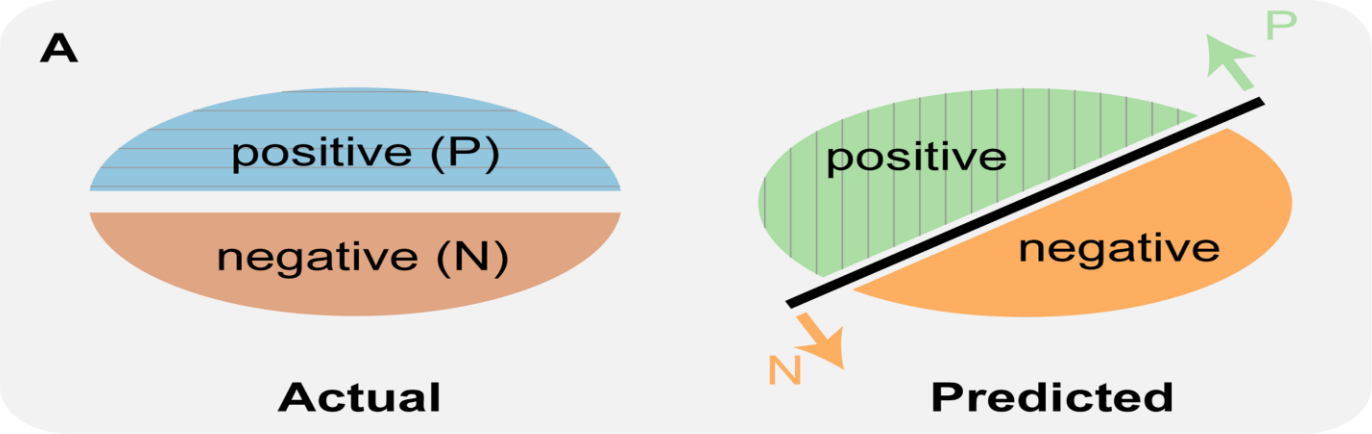
统计真实标记和预测结果的组合可以得到“混淆矩阵”

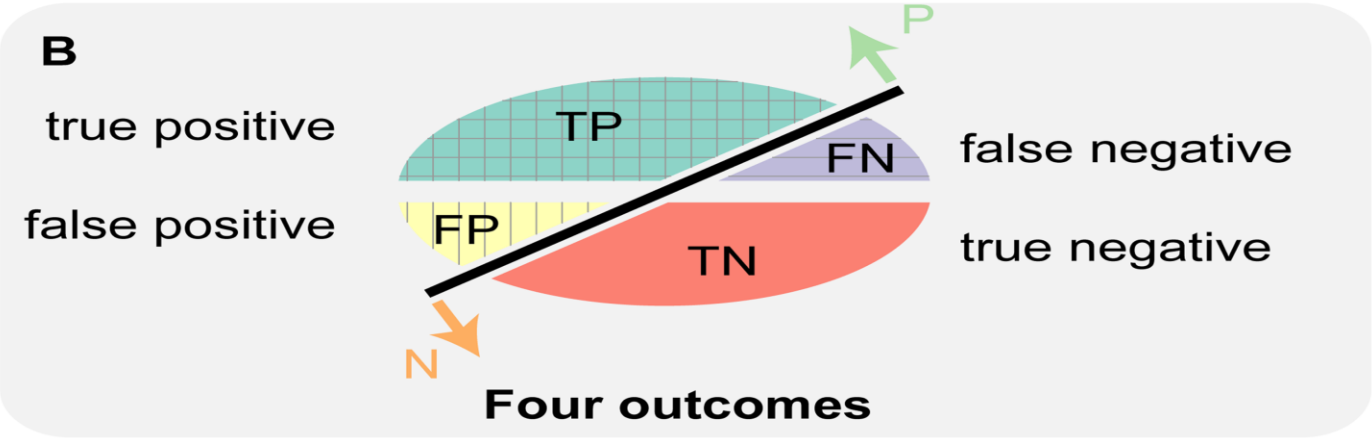
分类结果混淆矩阵

真实情况	预测结果	
	正例	反例
正例	TP (真正例)	FN (假反例)
反例	FP (假正例)	TN (真反例)

查准率 $P = \frac{TP}{TP + FP}$

查全率 $R = \frac{TP}{TP + FN}$





- 查准率（precision）：

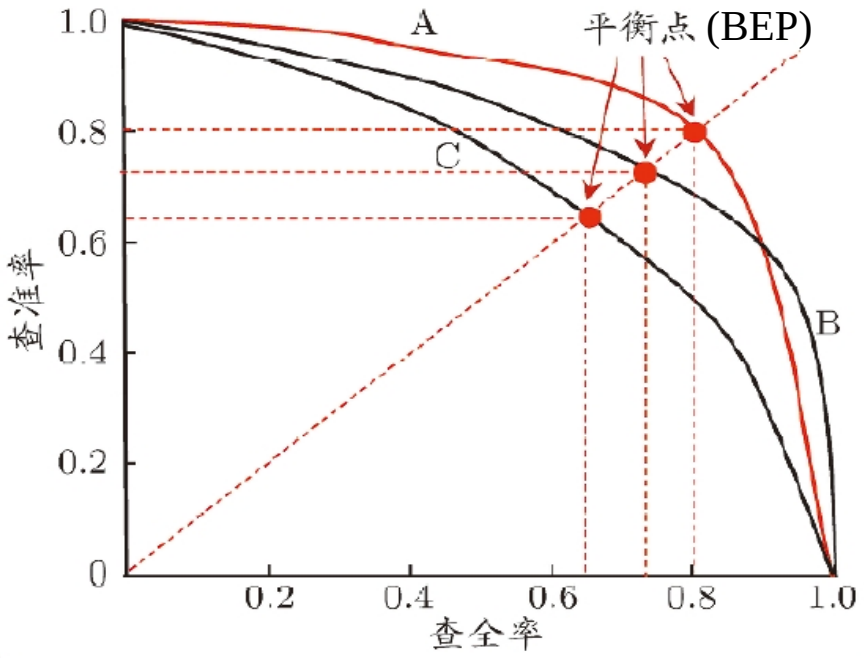
$$P = \frac{TP}{TP + FP}$$

- 查全率（Recall）：

$$R = \frac{TP}{TP + FN}$$

性能度量：P-R曲线、BEP

- 根据学习器的预测结果按正例可能性大小对样例进行排序，并逐个把样本作为正例进行预测，则可以得到查准率-查全率曲线，简称“P-R曲线”。
- 平衡点是曲线上“查准率=查全率”时的取值，可用来用于度量P-R曲线有交叉的分类器性能高低。



P-R曲线：

- 学习器 A 优于 学习器 C
- 学习器 B 优于 学习器 C
- 学习器 A ?? 学习器 B

BEP：

- 学习器 A 优于 学习器 B
- 学习器 A 优于 学习器 C
- 学习器 B 优于 学习器 C

性能度量：F-score

比 BEP 更常用的 F1 度量：

$$F1 = \frac{2 \times P \times R}{P + R} = \frac{2 \times TP}{\text{样例总数} + TP - TN}$$

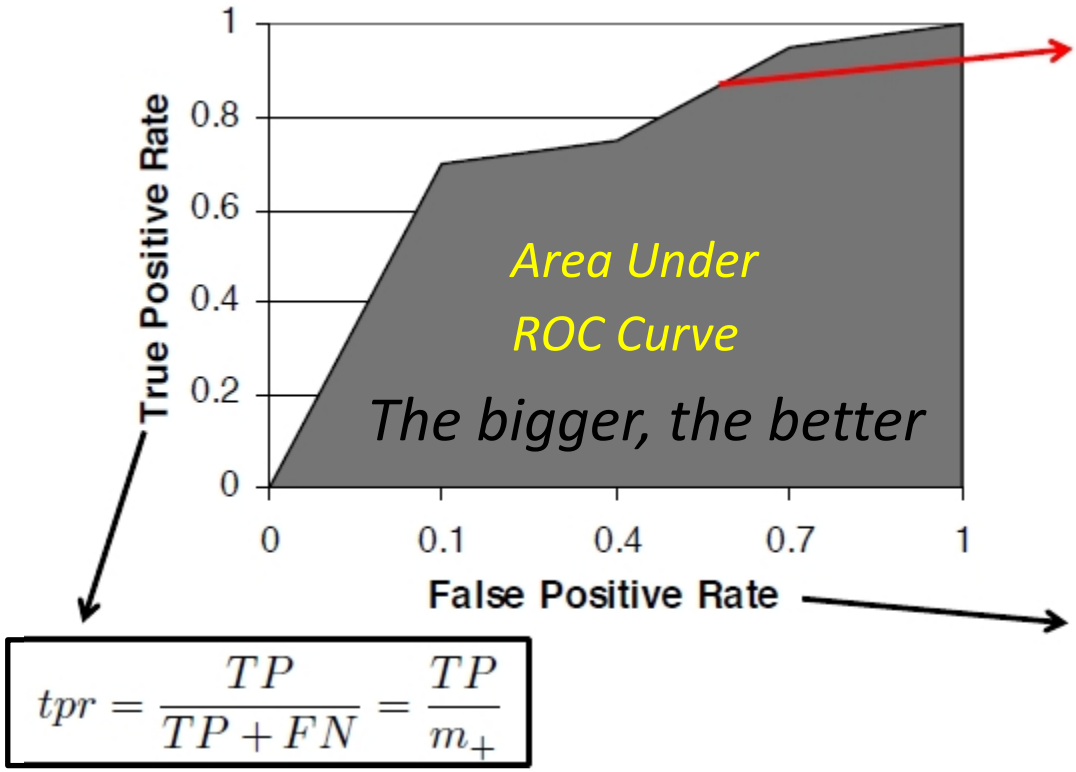
若对查准率/查全率有不同偏好：

比F1更一般的形式 F_β，

$$F_\beta = \frac{(1 + \beta^2) \times P \times R}{(\beta^2 \times P) + R}$$

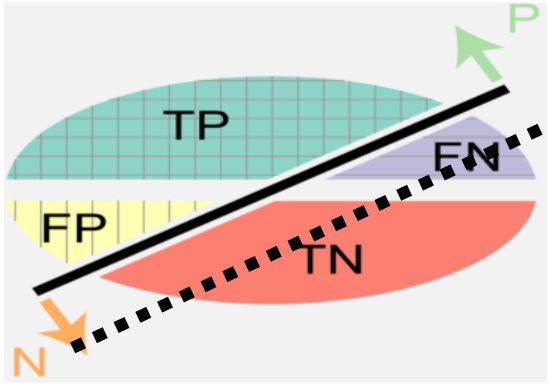
$\beta = 1$：标准F1
$\beta > 1$：偏重查全率(逃犯信息检索)
$\beta < 1$：偏重查准率(商品推荐系统)

AUC: **A**rea **U**nder the ROC **C**urve



$$tpr = \frac{TP}{TP+FN} = \frac{TP}{m_+}$$

ROC (Receiver Operating Characteristic) Curve [Green & Swets, Book 66; Spackman, IWML'89]



$$fpr = \frac{FP}{FP+TN} = \frac{FP}{m_-}$$

- 更多性能指标讨论：http://charleshm.github.io/2016/03/Model-Performance/

- 关于性能比较：
  - 测试性能并不等于泛化性能
  - 测试性能随着测试集的变化而变化
  - 很多机器学习算法本身有一定的随机性

直接选取相应评估方法在相应度量下比大小的方法不可取！

假设检验为学习器性能比较提供了重要依据，基于其结果我们可以推断出若在测试集上观察到学习器A比B好，则A的泛化性能是否在统计意义上优于B，以及这个结论的把握有多大。

参考《机器学习》周志华 第2.4节

“误差”包含了哪些因素？

换言之，从机器学习的角度看，

“误差”从何而来?

对回归任务，泛化误差可通过“偏差-方差分解”拆解为：

$$E(f;D) = bias^2(\boldsymbol{x}) + var(\boldsymbol{x}) + \varepsilon^2$$

期望输出与真实输出的差别

$$bias^2(\boldsymbol{x}) = \left(\bar{f}(\boldsymbol{x}) - y\right)^2$$

同样大小的训练集的变动，所导致的性能变化

$$var(\boldsymbol{x}) = \mathbb{E}_D\left[\left(f(\boldsymbol{x};D) - \bar{f}(\boldsymbol{x})\right)^2\right]$$

训练样本的标记与真实标记有区别

表达了当前任务上任何学习算法所能达到的期望泛化误差下界

$$\varepsilon^2 = \mathbb{E}_D\left[(y_D - y)^2\right]$$

泛化性能是由学习算法的能力、数据的充分性以及学习任务本身的难度共同决定

一般而言，偏差与方差存在冲突：

- 训练不足时，学习器拟合能力不强，偏差主导
- 随着训练程度加深，学习器拟合能力逐渐增强，方差逐渐主导
- 训练充足后，学习器的拟合能力很强，方差主导

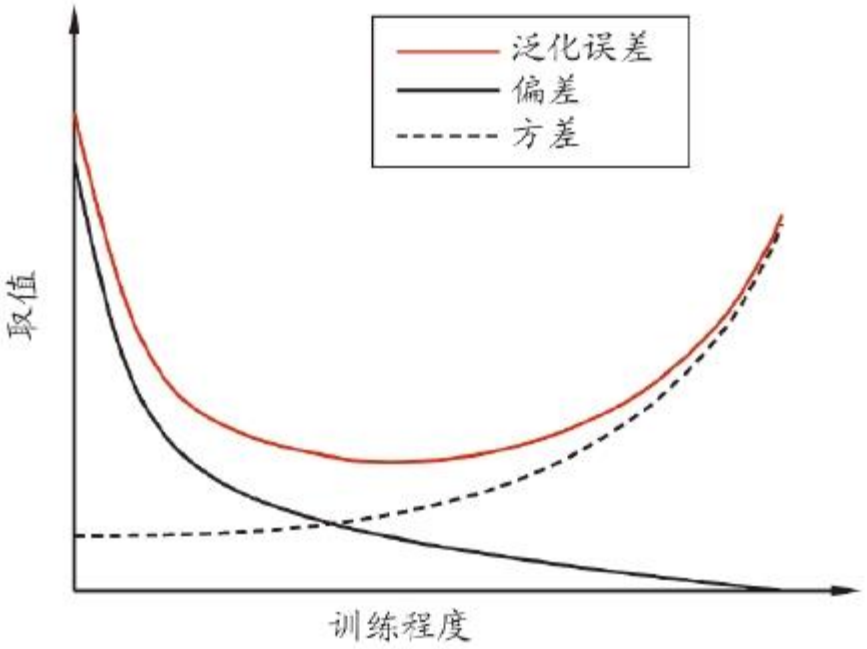


图 2.9 泛化误差与偏差、方差的关系示意图

➢ 更多偏差-方差的讨论：https://zhuanlan.zhihu.com/p/38853908

小结


机器学习
数据集
每条记录是关于一个事件或对象的描述，称为一个“示例”或“样本”
反映事件或对象在某方面的表现或性质的事项，称为“属性”或“特征”
属性张成的空间称为“属性空间”、“样本空间”或“输入空间”
训练数据集（训练样本）、测试数据集（测试样本）、标签
模型
从数据中学得的结果。从数据中学习的模型的过程称为“学习”或“训练”
学得模型后，使用其进行预测的过程称为“测试”
机器学习的目标是使学得的模型能很好地适用于“新样本”，而不是仅仅在训练样本上工作得很好。学得模型适用于新样本的能力，称为“泛化能力”
算法
监督学习：训练数据有标签
分类（K-NN, SVM, 决策树）
回归（线性回归）
非监督学习：训练数据没有标签
聚类
降维（主成分分析）

三个关键问题:

- 如何获得测试结果？

  评估方法：留出法、交叉验证法、自助法

- 如何评估性能优劣？

  性能度量：均方误差、错误率、精度、查准率、查全率、ROC曲线等

- 如何判断实质差别？

  比较检验

04

PART FOUR

参考资源

Resource

- 在线课程：

  吴恩达（Andrew Ng)：机器学习

  https://www.bilibili.com/video/av50747658?from=search&seid=9857444623866219885

  李宏毅（Hung-Yi Lee）：机器学习

  https://www.bilibili.com/video/av10590361?from=search&seid=10930215693466580647

- 公众号：
  - VALSE
  - 机器之心
  - 极市平台
  - 人工智能前沿讲习班
  - 深度学习大讲堂
  - SIGAI
  - …