

一、绪论

主讲教师：周志华

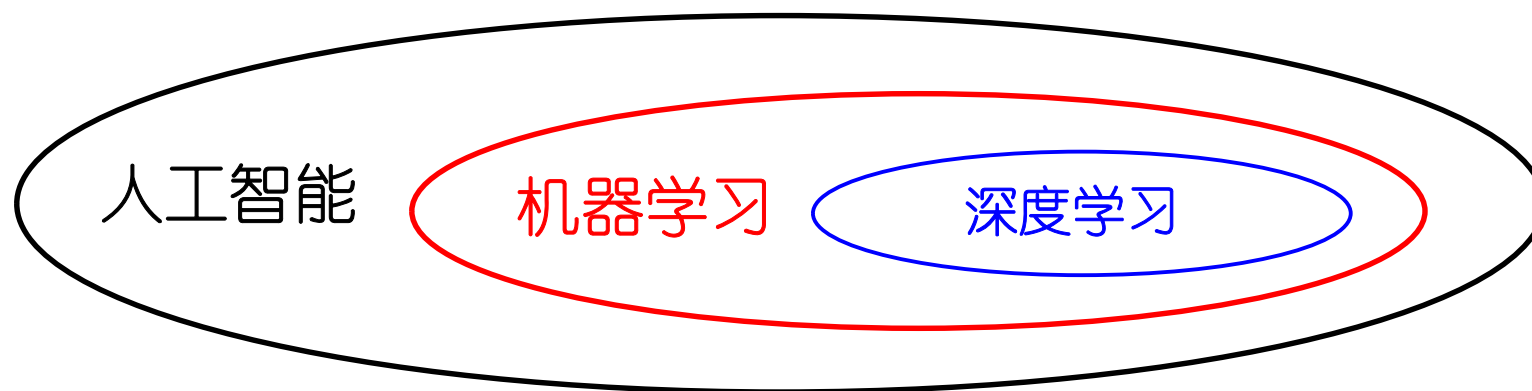
“机器学习”与“人工智能”

人工智能从1956年正式成为一个学科

机器学习是人工智能的核心研究领域（之一）

今天的“人工智能热潮”

正是由于机器学习、尤其深度学习技术取得了巨大进展
基于大数据、大算力发挥出巨大威力



两种不同的“人工智能”

□ 强人工智能（“科幻人工智能”）

研制出和人一样聪明，甚至比人更聪明的机器



重要特征：

- 具有自主意识
- 全面达到、甚至超过人类智能水平
-

两种不同的“人工智能”

□ 弱人工智能（“科学人工智能”）

让机器做事时聪明一点

“人工智能就是让机器来完成那些如果由人来做则需要智能的事情的科学”



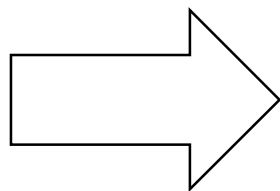
马文·闵斯基
(1927-2016)
人工智能奠基者之一
1969年图灵奖

解读：

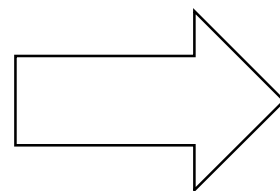
- 如果某件事情需要智能，通过机器来做，就是人工智能
- 不要求“全面”达到人类智能水平
- “做事”就行，不必具备“自主意识”……
- ……

“智能”与“人工智能”的关系

一个类比



(人的)智能行为



人工智能

人工智能重要，是因为能造出“智能工具”（类比：飞机）

- 造飞机的人不会关心飞机有没有“自主意识”
- 更不会关心飞机是否“全面达到”鸟的能力（例如：下蛋）

注意

“强人工智能” 与 “弱人工智能”

区别：

并非在于 “能力有多强”，
而在于 “是否拥有自主意识”

简单地说：

“强人工智能” 目的是 “造人”

“弱人工智能” 目的是 “造工具”



我们讨论的“人工智能”

从“教育”角度出发，所关注的“人工智能”内容应该是：

- 有助于为人类社会谋福祉的
 - 有助于解决产业创新需求的
 - 有助于学生未来职业发展的
- } 弱人工智能

人工智能 \neq 人造智能

人工智能 \approx

Intelligence-inspired computing



人工智能的诞生

Artificial Intelligence (AI), 1956 -



1956年夏 美国达特茅斯学院



J. McCarthy
“人工智能之父”
图灵奖(1971)



M. Minsky
图灵奖(1969)



C. Shannon
“信息论之父”



H. A. Simon
图灵奖(1975)
诺贝尔经济学奖(1978)



A. Newell
图灵奖(1975)

.....
.....

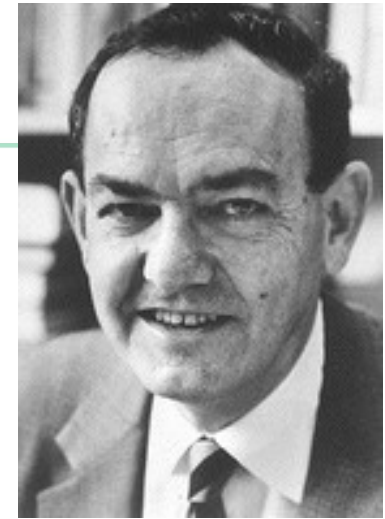
达特茅斯会议标志着人工智能这一学科的诞生

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

- ◆ 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 ...



爱德华·费根鲍姆
(1936-)
1994年图灵奖



瑞吉·芮迪
(1937-)
1994年图灵奖

第三阶段：学习期

1990s -now: Machine Learning

- ◆ 出发点：“让系统自己学！”
- ◆ 主要成就：……

科学界极为关注



美国两院院士
M. I. Jordan、
工程院院士**T. Mitchell**共同
指出：“机器学习是当前发展
最迅速的科学技术领域之一”



冯·诺依曼奖章得主**A.-L. Barabási**
称：“大数据分析建模推动了自然与计算科学的蓬勃发展，而机器学习是未来研究的关键领域”

各国政府高度重视



国务院**2015年8月**印发的《促进大数据发展行动纲要》明确指出，机器学习是提升大数据分析处理能力的**关键**



美国政府**2016年5月**公布的《联邦大数据研发战略计划》中，将机器学习作为支撑大数据研发战略的**核心技术**

工业界大力投入



Google、微软、IBM、亚马逊等投入巨资研发机器学习平台，以满足公司对机器学习技术的迫切需求



美国军工重镇**洛克希德·马丁公司**将机器学习作为新一代电子战致胜的**关键技术**进行研究与**应用**

图灵奖在十年中三次授予在该领域取得突出成就的学者



Leslie Valiant

“计算学习理论”奠基人

2010
年度



Judea Pearl

“图模型学习方法”先驱

2011
年度



Geoff Hinton



Yann LeCun



Yoshua Bengio

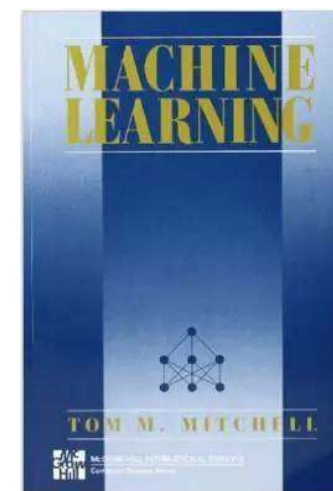
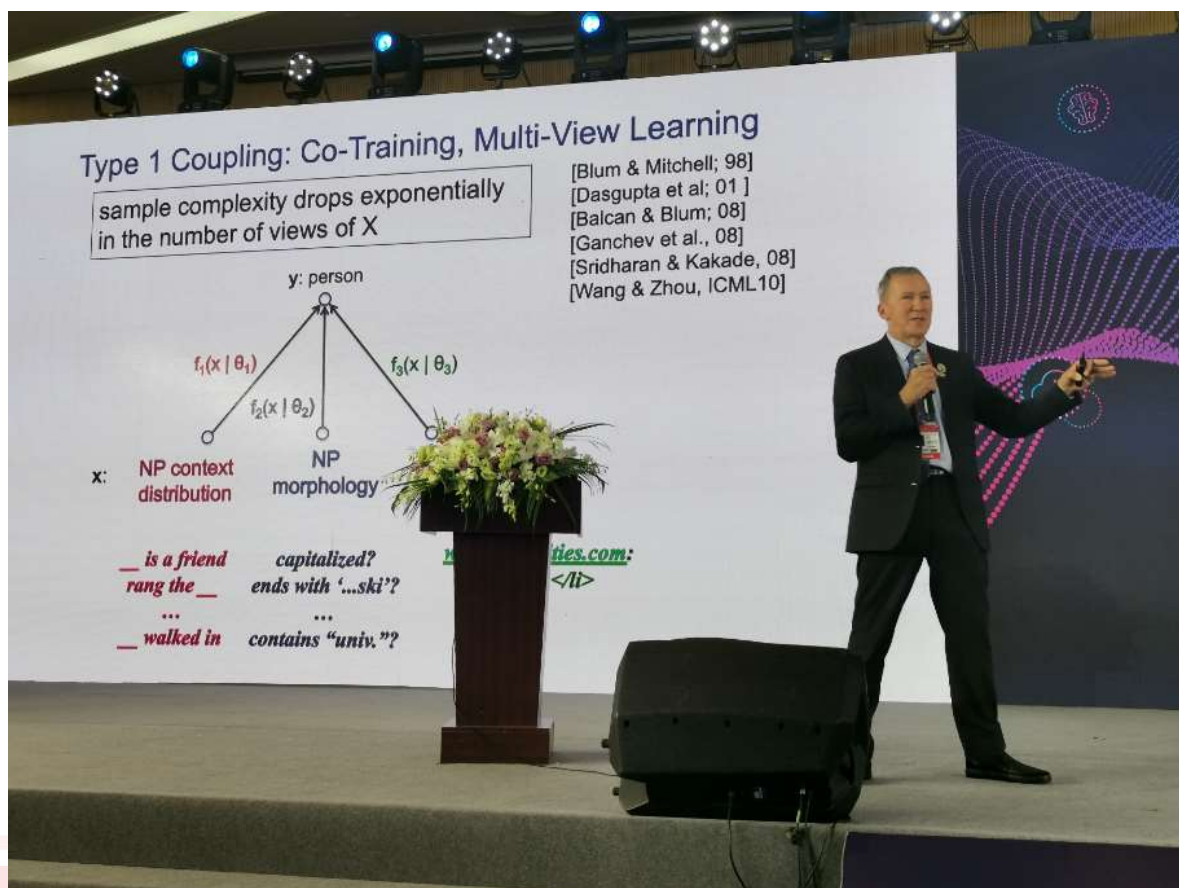
“深度学习”三架马车

2018
年度

机器学习

经典定义：利用经验改善系统自身的性能

[T. Mitchell 教科书, 1997]



机器学习领域奠基人

CMU机器学习系首任系主任

世界第一本《机器学习》教科书作者

T. Mitchell 教授

机器学习

经典定义：利用经验改善系统自身的性能

[T. Mitchell 教科书, 1997]

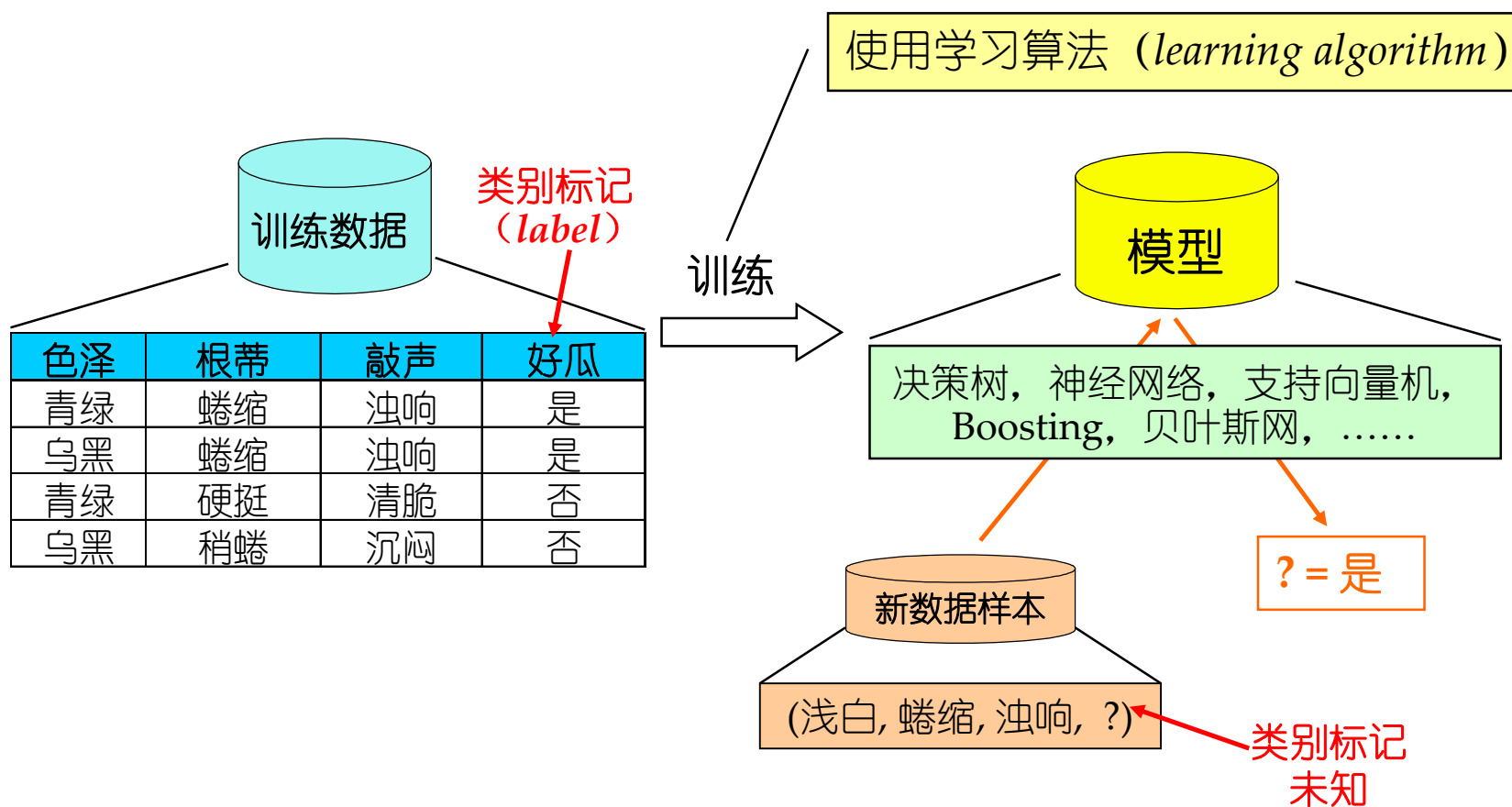


经验 → 数据



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

典型的机器学习过程



机器学习有什么用？

大数据时代



大数据 \neq 大价值

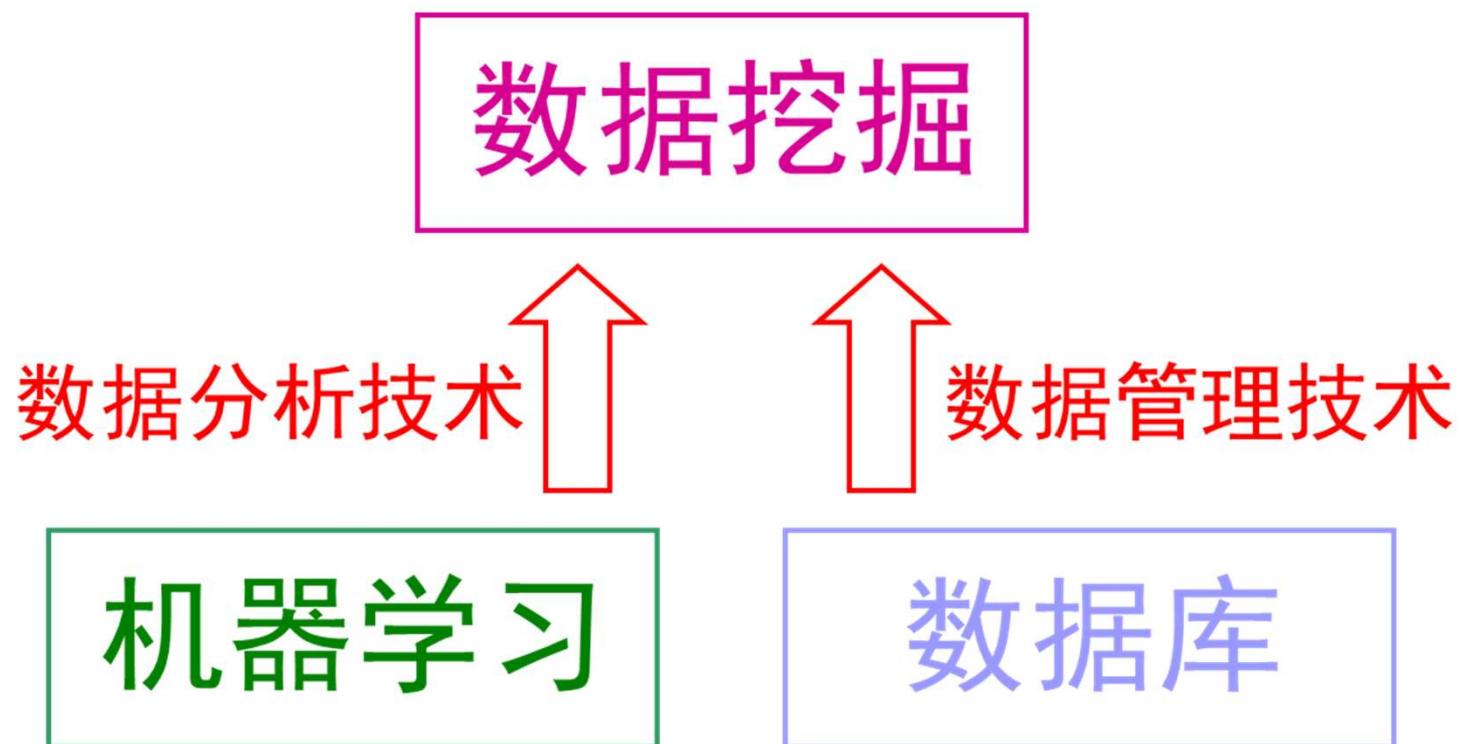


机器学习

有效的数据分析



机器学习与数据挖掘



机器学习已经“无处不在”



互联网搜索



火星机器人



生物特征识别

How Obama's data crunchers helped him win

By Michael Scherer
November 8, 2012 -- Updated 1645 GMT (0045 HKT) | Filed under: Web



美国总统选举



汽车自动驾驶



军事决策助手 (DARPA)

机器学习“无所不能”吗？NO!

并非“一切皆可学”，例如：

- ◆ 特征信息不充分

- 例如，重要特征信息没有获得

- ◆ 样本信息不充分

- 例如，仅有很少的数据样本

机器学习有坚实的理论基础

计算学习理论

Computational learning theory



Leslie Valiant
(莱斯利·维利昂特)
(1949-)
2010年图灵奖



2018年，计算机楼

机器学习有坚实的理论基础

计算学习理论

Computational learning theory

最重要的理论模型：

PAC (Probably Approximately Correct,
概率近似正确) learning model [Valiant, 1984]

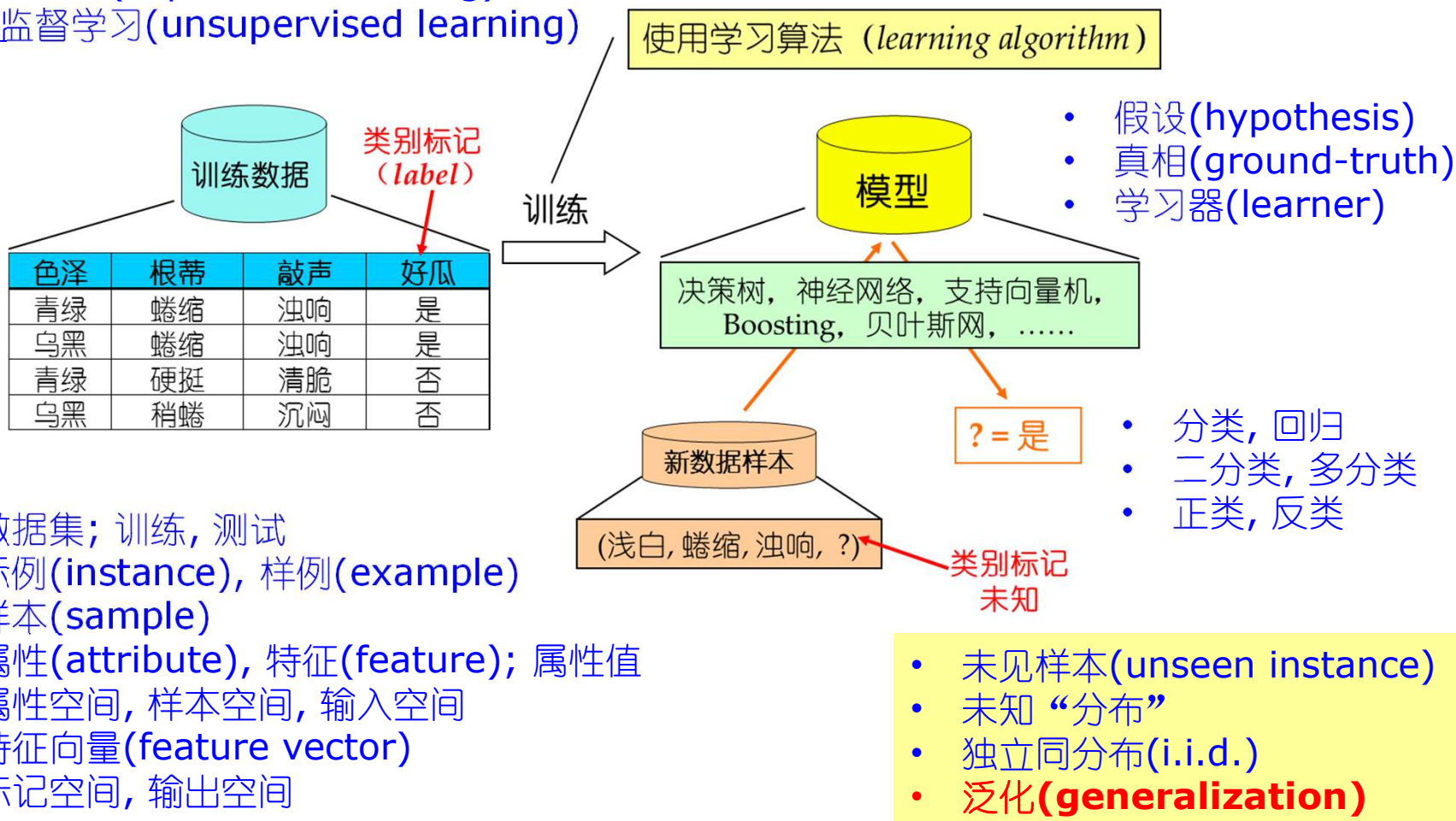
$$P(|f(\mathbf{x}) - y| \leq \epsilon) \geq 1 - \delta$$



Leslie Valiant
(莱斯利·维利昂特)
(1949-)
2010年图灵奖

基本术语

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



假设空间

表 1.1 西瓜数据集

| 编号 | 色泽 | 根蒂 | 敲声 | 好瓜 |
|----|----|----|----|----|
| 1 | 青绿 | 蜷缩 | 浊响 | 是 |
| 2 | 乌黑 | 蜷缩 | 浊响 | 是 |
| 3 | 青绿 | 硬挺 | 清脆 | 否 |
| 4 | 乌黑 | 稍蜷 | 沉闷 | 否 |

$(\text{色泽}=?)\wedge(\text{根蒂}=?)\wedge(\text{敲声}=?)\leftrightarrow\text{好瓜}$

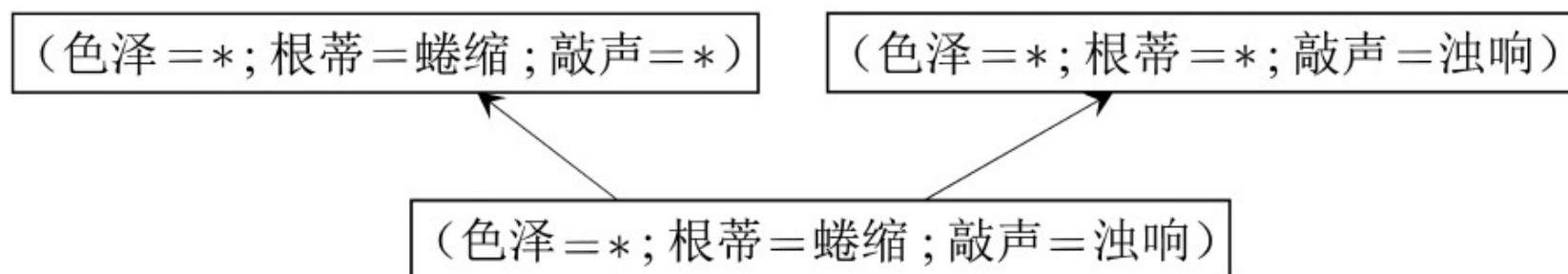
学习过程 → 在所有假设(hypothesis)组成的空间中进行搜索的过程

目标：找到与训练集“匹配”(fit)的假设

假设空间的大小： $(n_1+1) \times (n_2+1) \times (n_3+1) + 1$

版本空间

版本空间(version space): 与训练集一致的假设集合



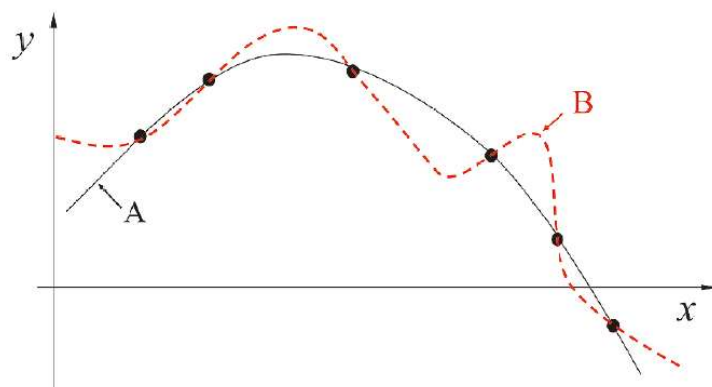
在面临新样本时, 会产生不同的输出

例如: (青绿; 蜷缩; 沉闷)

应该采用哪一个
模型(假设)?

归纳偏好 (inductive bias)

机器学习算法在学习过程中对某种类型假设的偏好



A更好 ?
B更好 ?

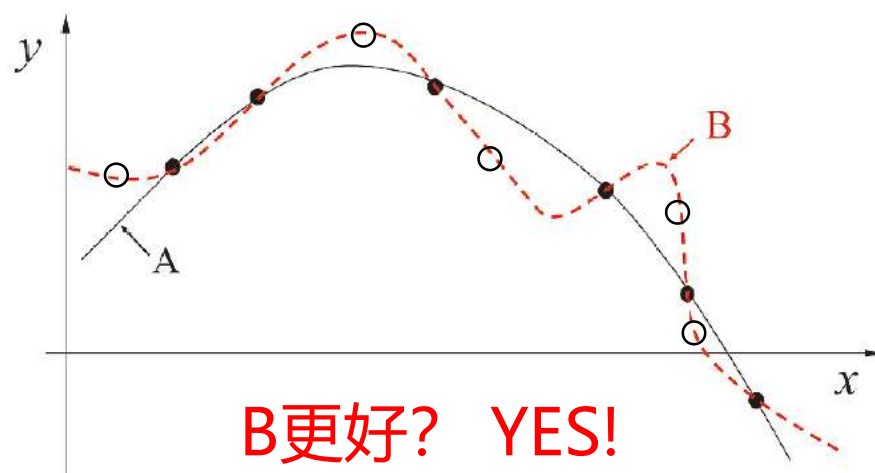
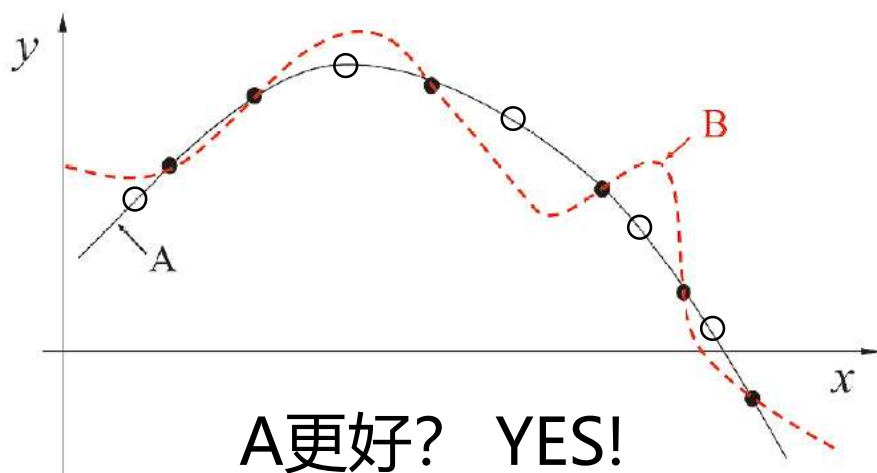
一般原则：
奥卡姆剃刀
(Ocam's razor)

任何一个有效的机器学习算法必有其偏好

**学习算法的归纳偏好是否与问题本身匹配，
大多数时候直接决定了算法能否取得好的性能！**

哪个算法更好？

黑点：训练样本；白点：测试样本



没有免费的午餐!

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

NFL定理

简单起见，假设样本空间 \mathcal{X} 和假设空间 \mathcal{H} 离散，令 $P(h|X, \mathfrak{L}_a)$ 代表算法 \mathfrak{L}_a 基于训练数据 \mathbf{X} 产生假设 h 的概率， f 代表要学的目标函数， \mathfrak{L}_a 在训练集之外所有样本上的总误差为

$$E_{ote}(\mathfrak{L}_a | X, f) = \sum_h \sum_{\mathbf{x} \in \mathcal{X} - X} P(\mathbf{x}) \mathbb{I}(h(\mathbf{x}) \neq f(\mathbf{x})) P(h | X, \mathfrak{L}_a)$$

考虑二分类问题，目标函数可以为任何函数 $\mathcal{X} \mapsto \{0, 1\}$ ，函数空间为 $\{0, 1\}^{|\mathcal{X}|}$ ，对所有可能的 f 按均匀分布对误差求和，有

$$\sum_f E_{ote}(\mathfrak{L}_a | X, f) = \sum_f \sum_h \sum_{\mathbf{x} \in \mathcal{X} - X} P(\mathbf{x}) \mathbb{I}(h(\mathbf{x}) \neq f(\mathbf{x})) P(h | X, \mathfrak{L}_a)$$

NFL定理

考虑二分类问题，目标函数可以为任何函数 $\mathcal{X} \mapsto \{0,1\}$ ，函数空间为 $\{0,1\}^{|\mathcal{X}|}$ ，对所有可能的 f 按均匀分布对误差求和，有

$$\begin{aligned}\sum_f E_{ote}(\mathcal{L}_a | X, f) &= \sum_f \sum_h \sum_{\mathbf{x} \in \mathcal{X} - X} P(\mathbf{x}) \mathbb{I}(h(\mathbf{x}) \neq f(\mathbf{x})) P(h | X, \mathcal{L}_a) \\&= \sum_{\mathbf{x} \in \mathcal{X} - X} P(\mathbf{x}) \sum_h P(h | X, \mathcal{L}_a) \sum_f \mathbb{I}(h(\mathbf{x}) \neq f(\mathbf{x})) \\&= \sum_{\mathbf{x} \in \mathcal{X} - X} P(\mathbf{x}) \sum_h P(h | X, \mathcal{L}_a) \frac{1}{2} 2^{|\mathcal{X}|} \\&= \frac{1}{2} 2^{|\mathcal{X}|} \sum_{\mathbf{x} \in \mathcal{X} - X} P(\mathbf{x}) \sum_h P(h | X, \mathcal{L}_a) \\&= 2^{|\mathcal{X}|-1} \sum_{\mathbf{x} \in \mathcal{X} - X} P(\mathbf{x}) \cdot 1\end{aligned}$$

总误差与学习算法无关! \Rightarrow 所有算法同样好!

NFL定理的寓意

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”
毫无意义！

具体问题，具体分析！



现实机器学习应用中

把机器学习的“十大算法”“二十大算法”都弄熟，
逐个试一遍，是否就“止于至善”了？

NO !

机器学习并非“十大套路”“二十大招数”的简单堆积
现实任务千变万化，
以有限的“套路”应对无限的“问题”，焉有不败？

最优方案往往来自：**按需设计、度身定制**

Three horizontal bars at the bottom of the slide: a thin light pink bar, a medium light red bar, and a thin light green bar.

前往第二站.....

