

## 关于本课

# 主讲教师

---



Han-Jia Ye, 叶翰嘉

Contact me:

<http://www.lamda.nju.edu.cn/yehj/>  
[yehj@nju.edu.cn](mailto:yehj@nju.edu.cn)

**LAMDA**

Learning And Mining from Data

<http://www.lamda.nju.edu.cn>

# 授课教材



ISBN: 978-7-302-206853-6

2016年1月第1次印刷

2020年11月第35次印刷

周志华 著. 机器学习,  
北京: 清华大学出版社,  
2016年1月.

425页, 62.6万字

16 章, 3 附录

附录请自行阅读

本学期讲授前 9-10? 章

# 建议使用方式

## 1. 初学机器学习的第一本书：

通读、速读；细节不懂处略过

了解机器学习的疆域和基本思想，理解基本概念  
“观其大略”

## 2. 阅读其他关于机器学习具体分支的读物（三月、半年？）

## 3. 再读、对“关键点”的理解

理解技术细冗后的本质，升华认识  
“提纲挈领”

## 4. 对机器学习多个分支有所了解（1-3年？）

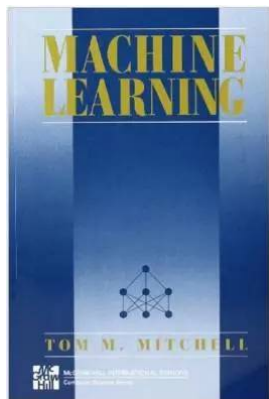
## 5. 再读、细思：

不同内容的联系，不同的描述方式、出现位置蕴涵的意义、……  
个别字句的启发，可能自行摸索数年不易得  
“疏通经络”



# 几本参考读物

## 初入门阶段



### 1997 Book

第一本机器学习教科书  
帮助读者建立领域整体知识框架；无学派偏见  
(最接近本书意图)

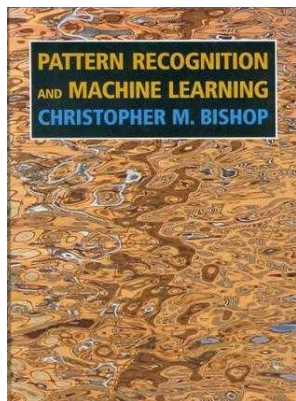


### “蓝皮书”

具体算法着眼  
适合希望快速了解一些著名算法的读者

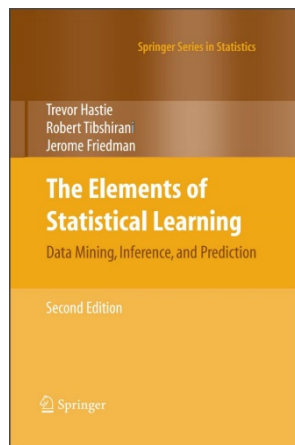
...

## 提高阶段



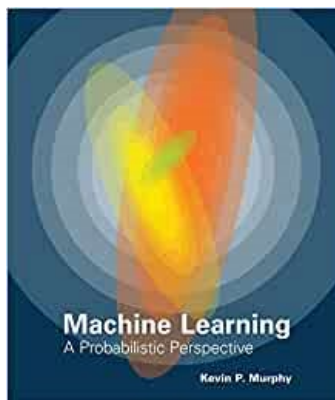
### PRML

贝叶斯学派视角



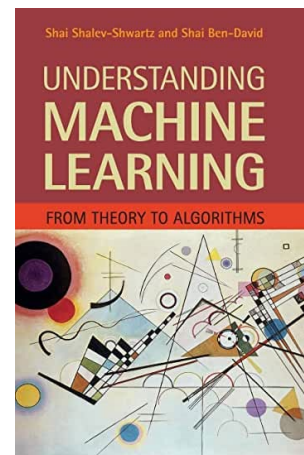
### ESL

统计学派(频率主义)视角



### MLAPP

概率学派视角



### UML

适合具有理论偏好的读者

...



周志华 著. 机器学习, 北京: 清华大

(ISBN 978-7-302-206853-6) 425页, 62.6

[\[前言&目录\]](#) [\[样章\]](#)

[{出版社网址}](#) [{网购网址1}](#) [{网购网址2}](#) [{网](#)

- 本课程使用的slides 与公开的不同

- 这是本班的红利

- 仅限本班使用, **请勿外传!**

2016年6月起将为本使用本书授课的教师提供PPT (免费; 不提供后续支持)

需要的老师请填写[申请表](#) [\[仅供教师使用\]](#) [\[后续提供部分习题参考答案\]](#)

[\[如何使用本书\]](#) [\[勘误修订\]](#)

对初学者, 建议使用方式:

- 先通读, 了解机器学习概貌 (不懂的细节地方跳过去)
- 通过其他书籍材料对感兴趣的若干方面进一步学习
- 再返回阅读本书, 会有新收获

如何使用本书 (写在第十次印刷之际): [\[PDF\]](#)

本书 2016 年 1 月底出版, 首印 5000 册一周内竟告  
榜首. 出乎预料的销量和受欢迎程度, 意味着本书读者  
使用本书需注意的一些事项. 因此, 在第 10 次印刷之

勘误修订 (Latex格式)

[本书因颇受欢迎, 出版社提出重印, 于是作者借机要求在每次重印时加入新的修订, 省却让读者等待第二版的麻烦. 为方便读者, 所有修订内容都列举在此. 其中部分修订是为了更便于读者理解, 并非原文有误]

(第一版第35次印刷, 2020年11月):

- p. 59, 倒数第二行: "其第 $t+1$ 轮 $\rightarrow$ "从当前 $\beta$ 生成下
- p. 59, 式 (3.29):  $\beta^{t+1} \rightarrow \beta^{\prime}$ ,  $\beta$
- p. 327, 倒数第10至倒数第4行: 两处" $P$ "  $\rightarrow$  " $p$ ", 4处" $V$ "
- p. 337, 14.6节第3段: 5个" $N$ "  $\rightarrow$  " $d$ "

- 经常有更新, 请自行查阅

- 欢迎各位同学发现问题后邮件告知**

- 对一般读者, 非勘误的学习问题恕难回复

# 课程主页

<https://www.lamda.nju.edu.cn/yehj/ml2022/ml2022.html>

[\[Home\]](#)

---

## 机器学习导论

---

课程代码:	30000150
授课对象:	人工智能学院、匡亚明学院等
学生人数:	120
上课时间:	2022年春季学期, 每周一, 14:00 - 15:50
上课地点:	南京大学仙林校区 仙II-304
教学用书:	周志华 著. 机器学习, 北京: 清华大学出版社, 2016年1月. <a href="#">{教材勘误}</a>
讲义作业:	<a href="#">{内部网站}</a> (本班同学校内访问)
主讲老师:	<a href="#">叶翰嘉</a> 博士

---

[\[Home\]](#)

# 课程作业

**6 次作业，每2-3周一次**

**Deadline:** 每次作业布置后 一般两周截止，请看作业网站的规定

[\[Home\]](#)

---

## 机器学习导论

---

课程代码:	30000150
授课对象:	人工智能学院、匡亚明学院等
学生人数:	120
上课时间:	2022年春季学期，每周一，14:00 - 15:50
上课地点:	南京大学仙林校区 仙II-304
教学用书:	周志华 著. 机器学习，北京：清华大学出版社，2016年1月. <a href="#">{教材勘误}</a>
讲义作业:	<a href="#">{内部网站}</a> (本班同学校内访问)
主讲老师:	<a href="#">叶翰嘉</a> 博士

---

[\[Home\]](#)



# 课程成绩

---

□能力测试：**20%**

**6**次作业中，各人自选**1**次

□平时成绩：**40%**

其他**5**次作业中，各人自选**4**次之和

□期末考试：**40%**

**Deadline**之后提交的作业，以此次**0**分计算

博士生、硕士生助教：  
胡文超、高恩豪、施意、张逸凯  
负责作业、答疑、考试、评分

前往第一站.....



## 一、绪论

# “机器学习”与“人工智能”

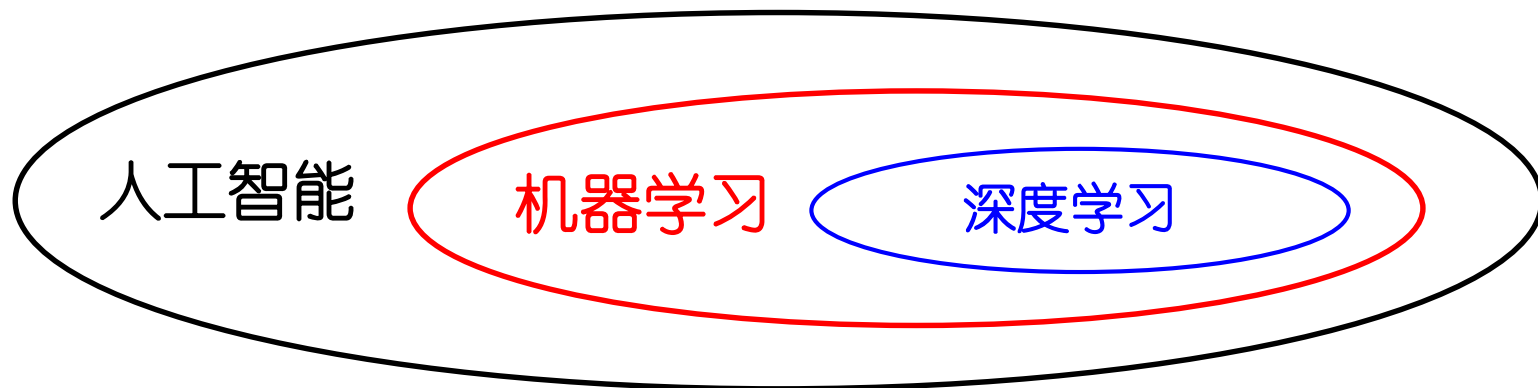
---

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



# 人工智能

## 科幻电影中的“人工智能”

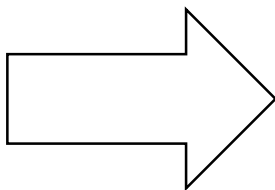


**常有人问：“比人类聪明的AI何时出现？”**

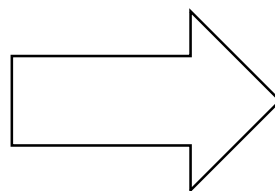
**这是把“人工智能”理解为“人造智能”**

# “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“意识”、会不会“疼”
- 更不会关心飞机是否“全面达到”鸟的能力（例如：下蛋）

# 我们讨论的“人工智能”

---

人工智能  $\neq$  人造智能

(Artificial Intelligence  $\neq$  Man-made intelligence)

人工智能 = **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



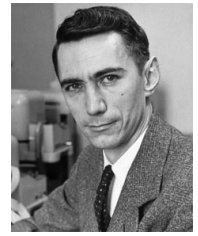
J. McCarthy

“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)

.....  
.....

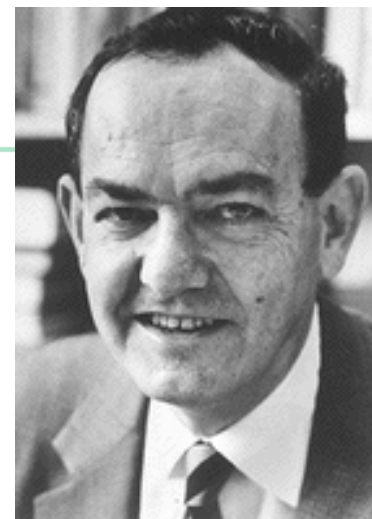
达特茅斯会议标志着人工智能这一学科的诞生

# 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

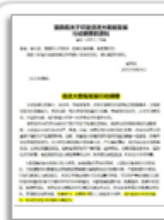


美国两院院士  
**M. I. Jordan**、  
工程院院士**T. Mitchell**共同  
指出：“机器学习是当前发展  
最迅速的科学技术领域之一”



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

## 图灵奖在近十年中三次授予在该领域取得突出成就的学者



2010  
年度

**Leslie Valiant**

“计算学习理论”奠基人



2011  
年度

**Judea Pearl**

“图模型学习方法”先驱



**Geoff Hinton**



**Yann LeCun**



**Yoshua Bengio**

2018  
年度

“深度学习”三架马车



# 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

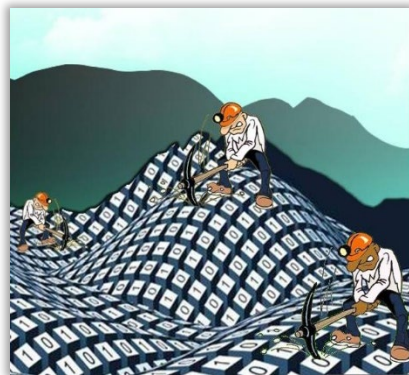


随着该领域的发展，目前主要研究智能数据分析的理论和方法，并已成为智能数据分析技术的源泉之一

大数据时代



大数据  $\neq$  大价值



智能  
数据分析

机器  
学习



# 机器学习 (Machine Learning)

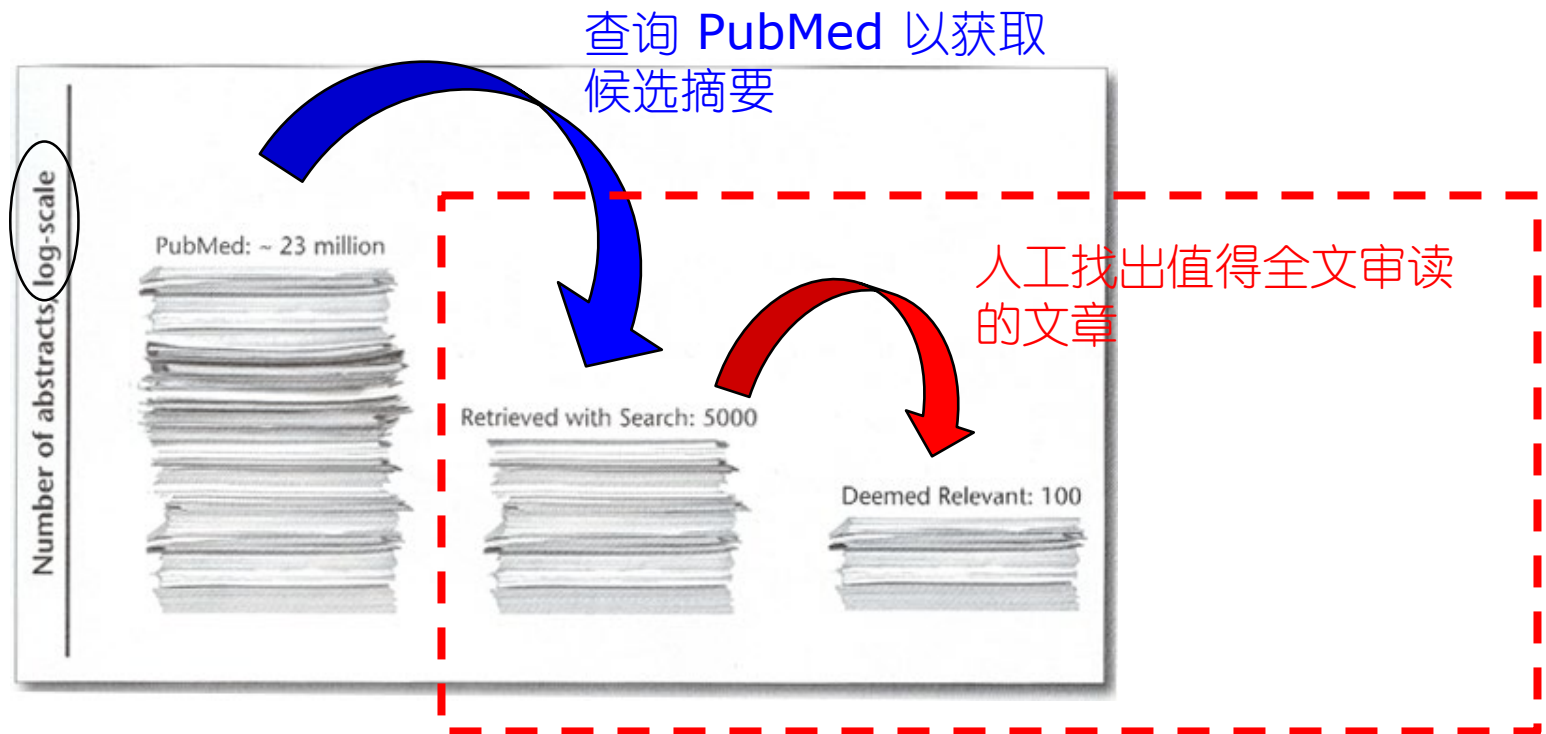
究竟是什么东东？



看两个例子 ⇨

# 医学文献筛选

在“循证医学”（evidence-based medicine）中，针对特定的临床问题，先要对相关研究报告进行详尽评估



# 医学文献筛选

在一项关于婴儿和儿童残疾的研究中，美国Tufts医学中心筛选了约 **33,000** 篇摘要

尽管Tufts医学中心的专家效率很高，对每篇摘要只需 **30** 秒钟，但该工作仍花费了 **250** 小时



a portion of the 33,000 abstracts

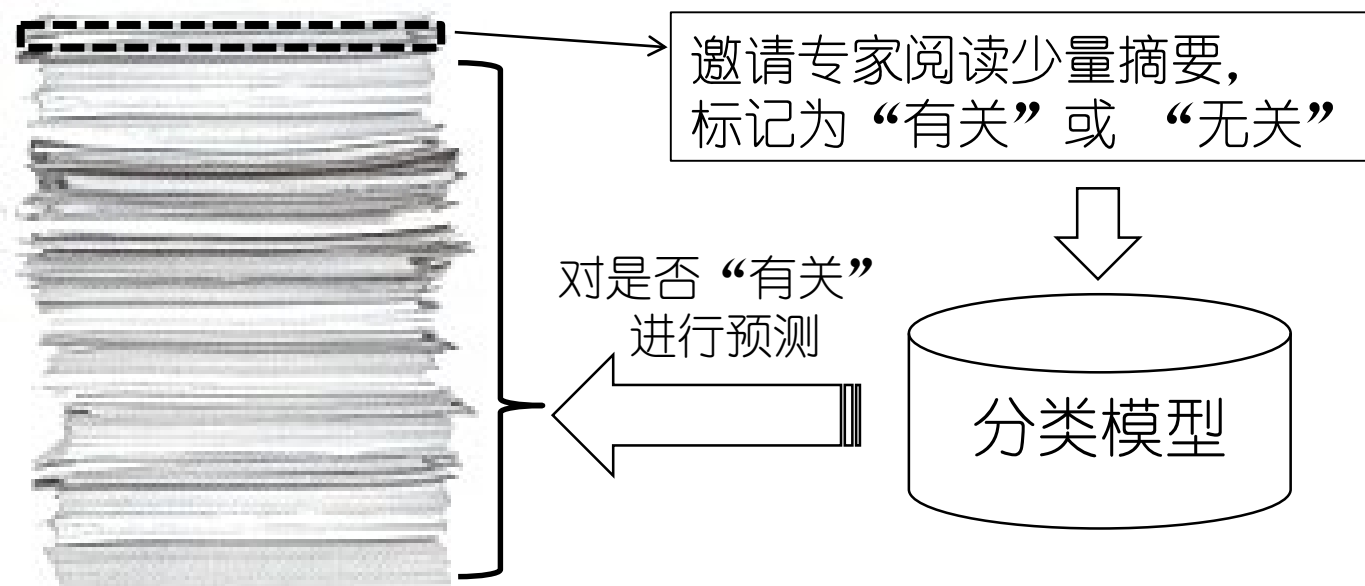
**每项新的研究都要重复  
这个麻烦的过程！**

**需筛选的文章数在不断显著增长！**



# 医学文献筛选

为了降低昂贵的成本, Tufts医学中心引入了机器学习技术



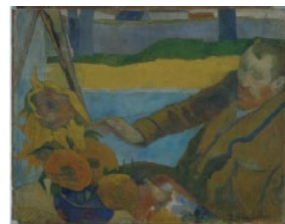
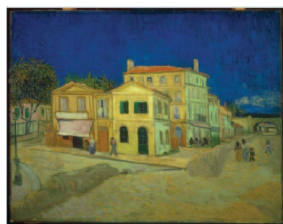
人类专家只需阅读 **50** 篇摘要, 系统的自动筛选精度就达到 **93%**  
人类专家阅读 **1,000** 篇摘要, 则系统的自动筛选敏感度达到 **95%**  
(人类专家以前需阅读 **33,000** 篇摘要才能获得此效果)

# 画作鉴别

画作鉴别(painting authentication): 确定作品的真伪



勃鲁盖尔 (1525-1569) 的作品？



梵高 (1853-1890) 的作品？

该工作对专业知识要求极高

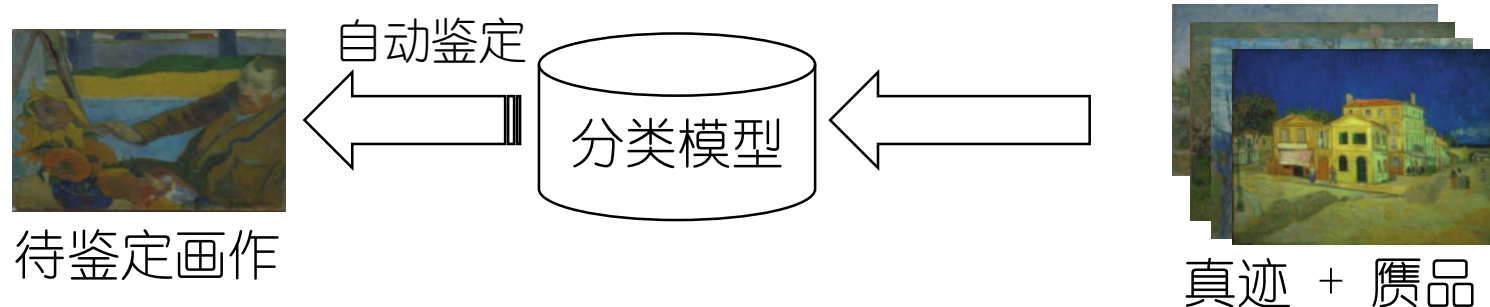
- 具有较高的绘画艺术修养
- 掌握画家的特定绘画习惯

只有少数专家花费很大精力  
才能完成分析工作！

很难同时掌握不同时期、不同流派多位画家的绘画风格！

# 画作鉴别

为了降低分析成本，**机器学习**技术被引入

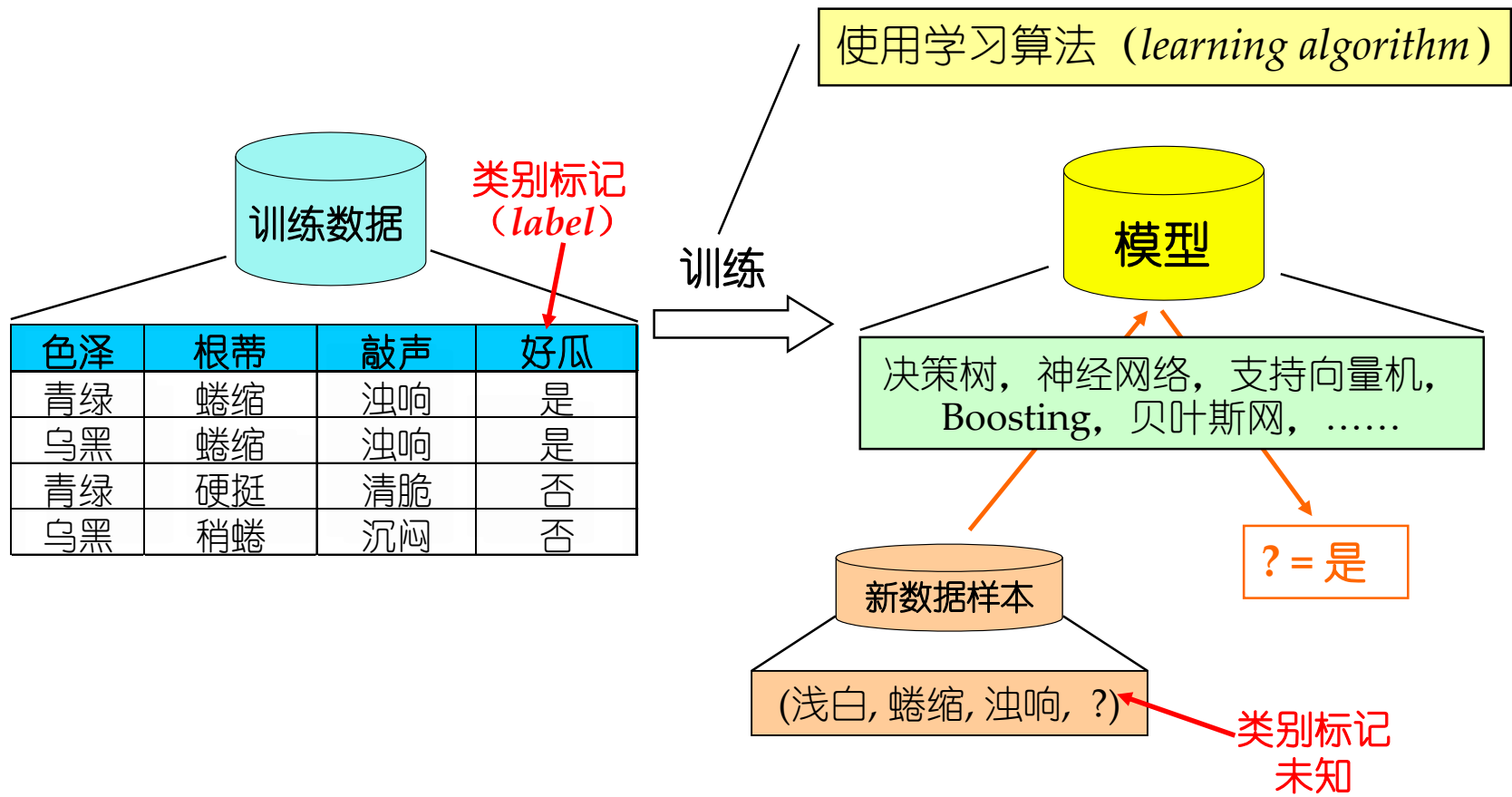


Kröller Müller美术馆与Cornell等大学的学者对82幅梵高真迹和6幅赝品进行分析，自动鉴别精度达 **95%** [C. Johnson et al., 2008]

Dartmouth学院、巴黎高师的学者对8幅勃鲁盖尔真迹和5幅赝品进行分析，自动鉴别精度达 **100%** [J. Hughes et al., 2009][J. Mairal et al., 2012]

(对用户要求低、准确高效、适用范围广)

# 典型的机器学习过程



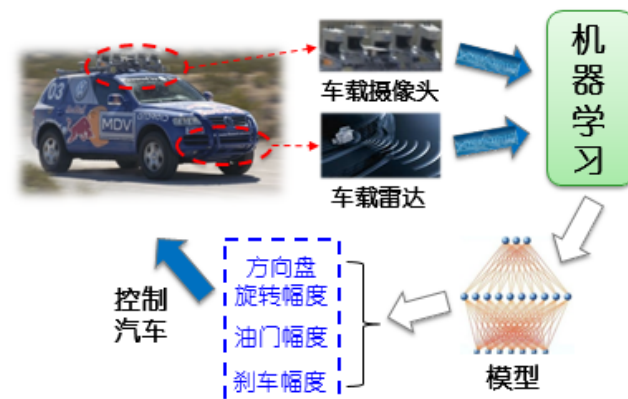
# 机器学习已经“无处不在”



互联网搜索



生物特征识别



汽车自动驾驶



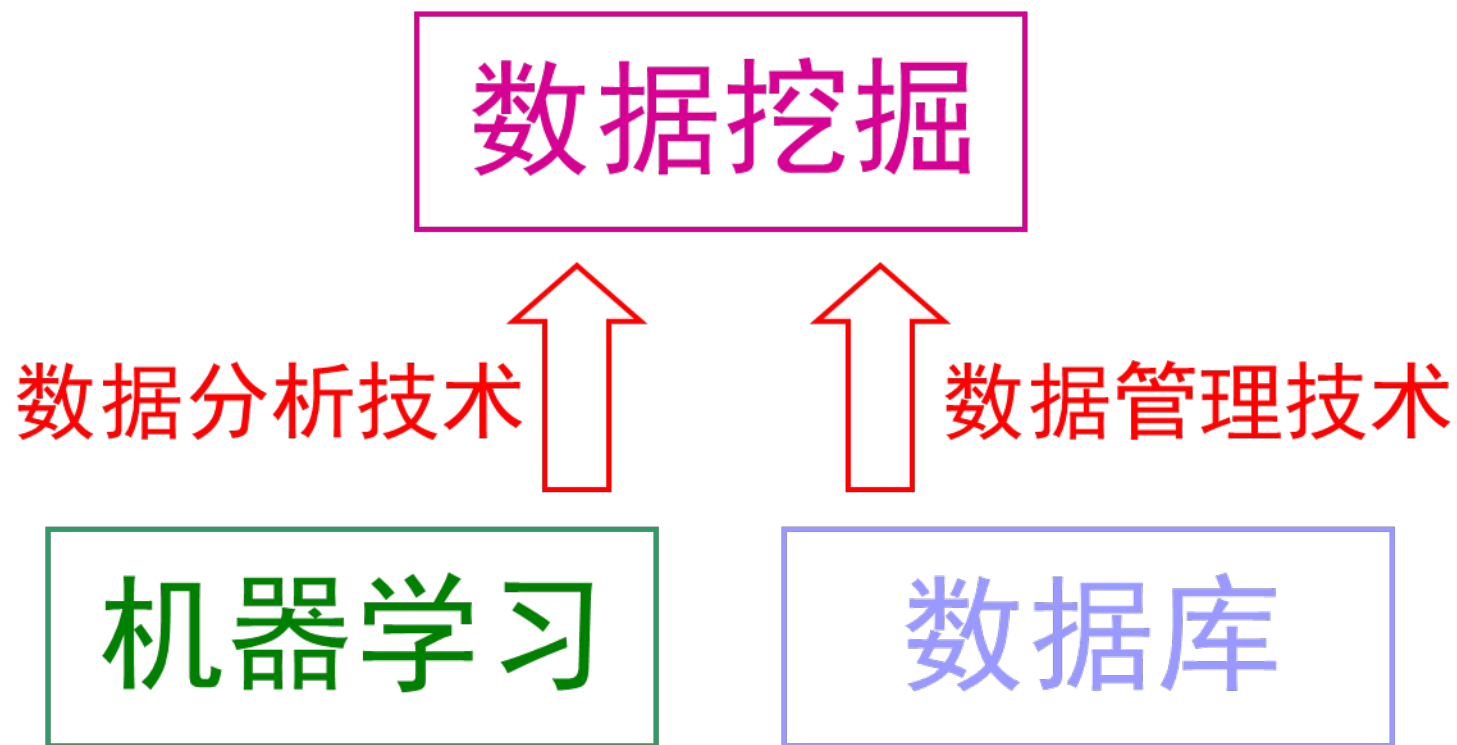
火星机器人



美国总统选举



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

最重要的理论模型：

**PAC** (Probably Approximately Correct,

概率近似正确) learning model [Valiant, 1984]

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

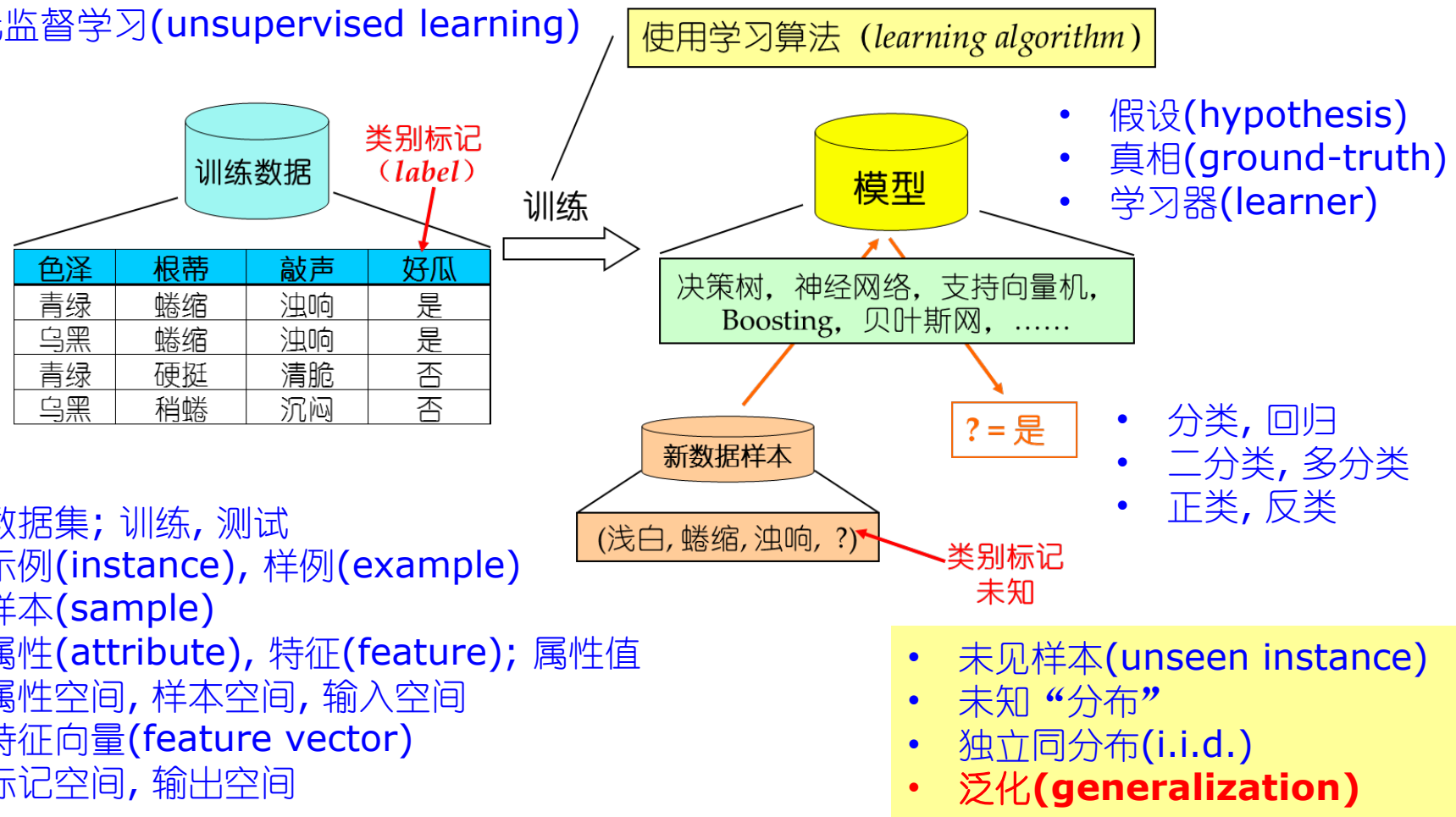


Leslie Valiant  
(莱斯利·维利昂特)  
(1949- )  
2010年图灵奖



# 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

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

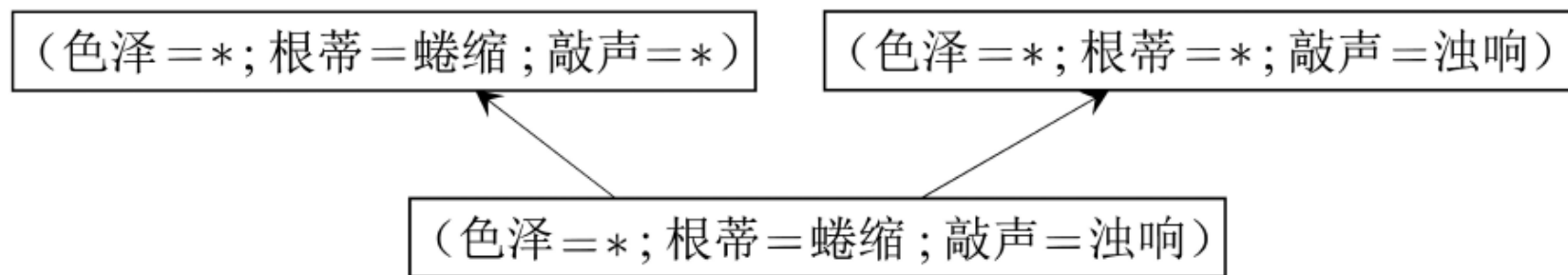
学习过程 → 在所有假设(hypothesis)组成的空间中进行搜索的过程

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)

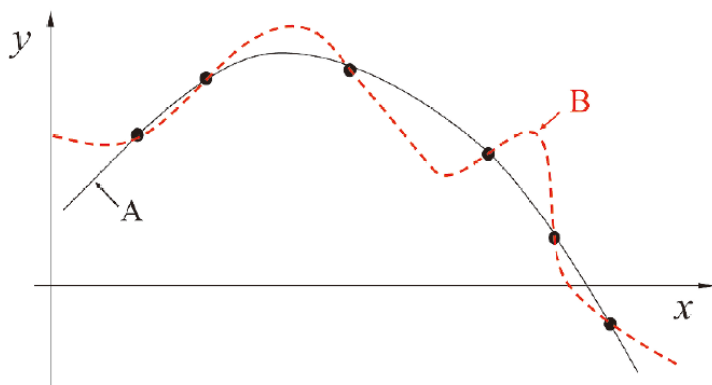
应该采用哪一个  
模型(假设)?

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

# 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？

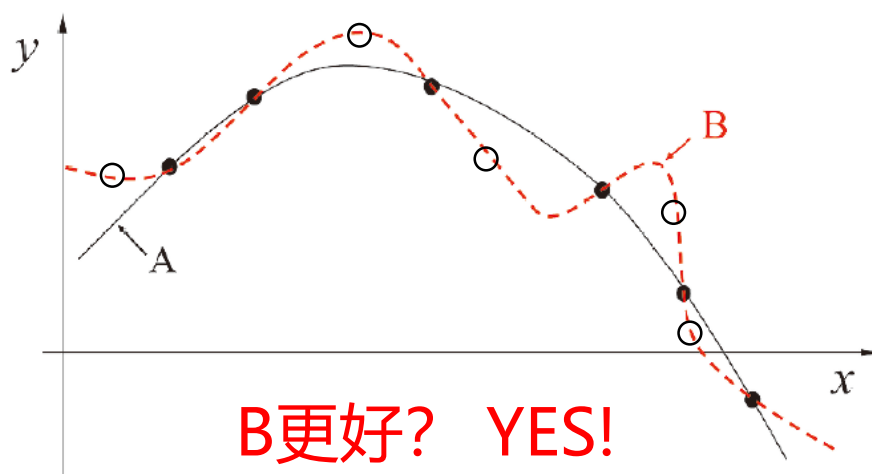
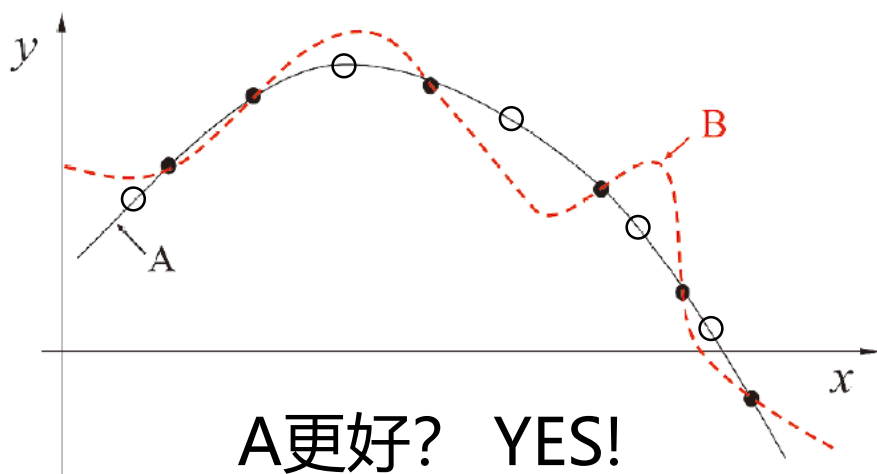
一般原则：  
奥卡姆剃刀  
(Occam's razor)

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  $P(h|X, \mathcal{L}_a)$  代表算法  $\mathcal{L}_a$  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， $\mathcal{L}_a$  在训练集之外所有样本上的总误差为

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

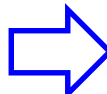
考虑二分类问题，目标函数可以为任何函数  $\mathcal{X} \mapsto \{0, 1\}$ ，函数空间为  $\{0, 1\}^{|\mathcal{X}|}$ ，对所有可能的  $f$  按均匀分布对误差求和，有

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

# 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!  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**

前往第二站.....

