机器学习及优化

中国矿业大学 计算机科学与技术学院

知行合一

Talk is cheap. Show me the code.

——Linus Torvalds

联系方式

13813451993

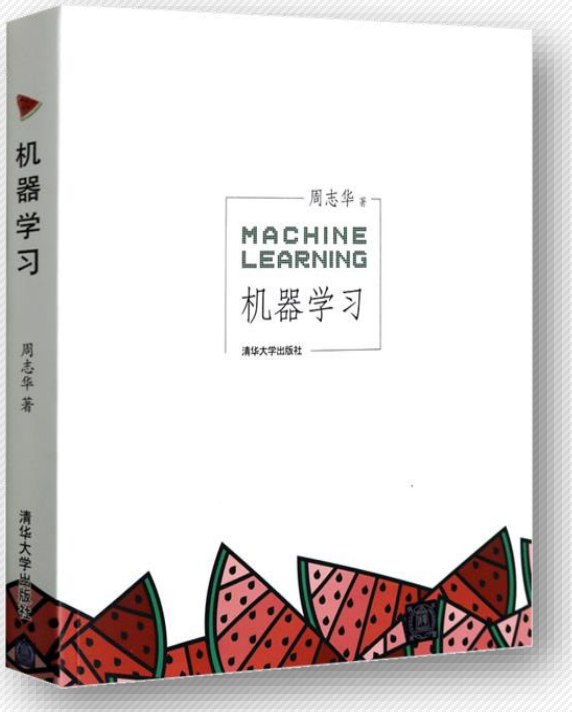
8366828

8366828@qq.com

群名称：机器学习2023

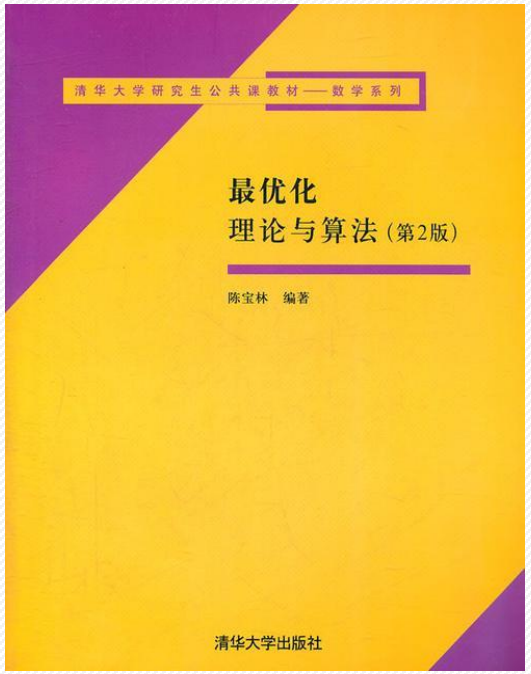
群　号：319772047

参考书目



机器学习，周志华，清华大学出版社

机器学习方法，李航，清华大学出版社

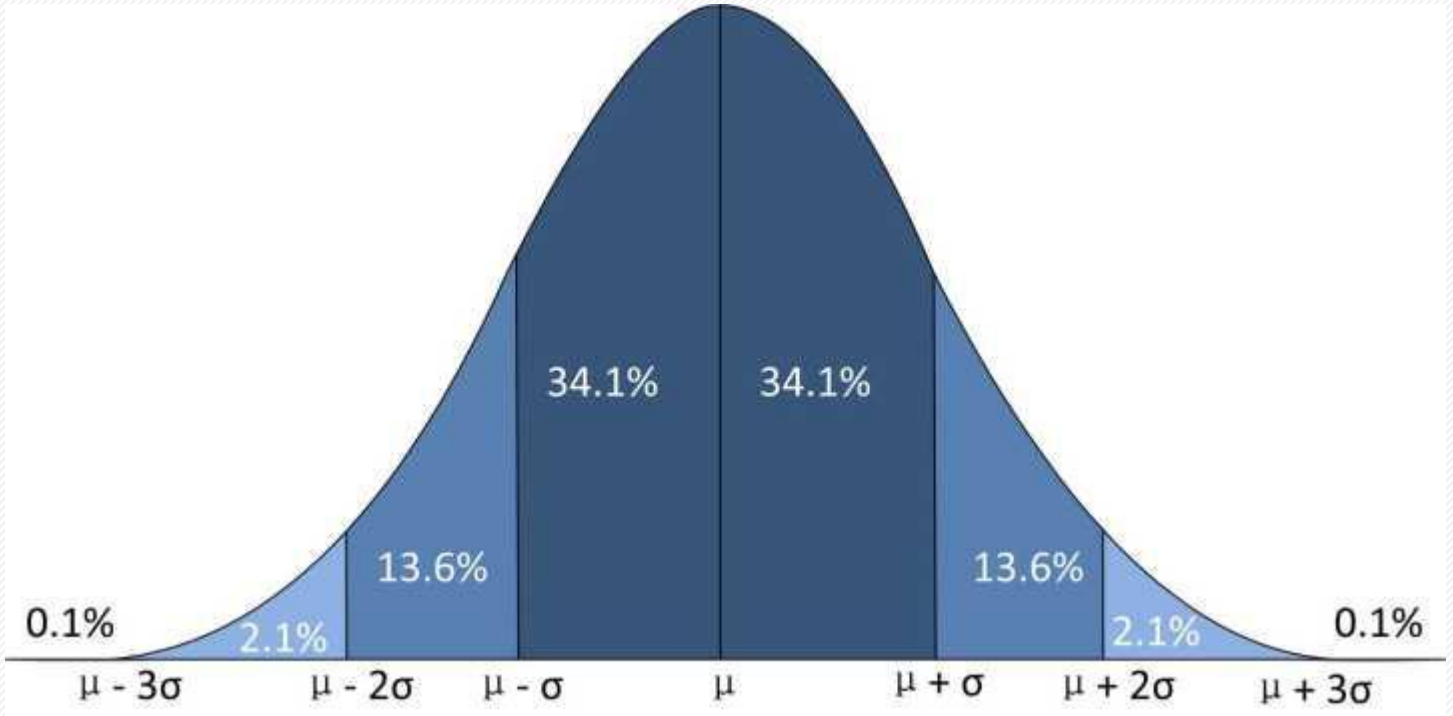


最优化理论与算法（第2版），陈宝林，清华大学出版社

■ ■ ■ ■ ■ ■

课程考核

- **平时（20%）：** 出勤、表现
- **实验（30%）：** 4次实验
- **期末考试（50%）：** 闭卷



教学内容

1. 绪论
2. 模型评估与选择
3. 线性模型
4. ~~决策树~~
5. 神经网络
6. 支持向量机
7. ~~贝叶斯分类器~~
8. 最优化理论及方法
9. 集成学习
10. 图神经网络（GNN）

实验内容（不局限于下述内容）

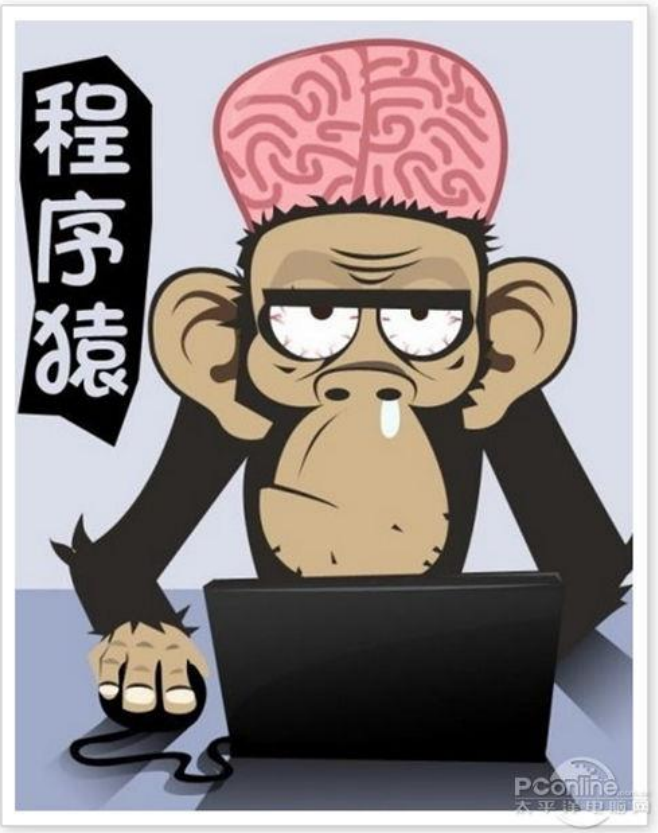
序号	实验名称	内容及要求	学时
1	梯度下降法与线性回归	编程实现用梯度下降法对波士顿房价进行线性拟合，开展模型优化，绘制拟合曲线。	2
2	使用SVM开展乳腺癌检测	通过载入数据、模型库导入，基于sklearn库使用SVM分类器开展乳腺癌检测，比较不同核函数对分类结果的影响。	2
3	中文情绪分类器	构建多层前馈神经网络模型实现中文情绪分类。	2
4	自选实验		2

提交服务器IP地址：219.219.61.252

前导知识

- **自然语言基础**
  - （英语）听说读写
- **数学基础**
  - 高数、线代与概率
  - 统计分析
  - 最优化

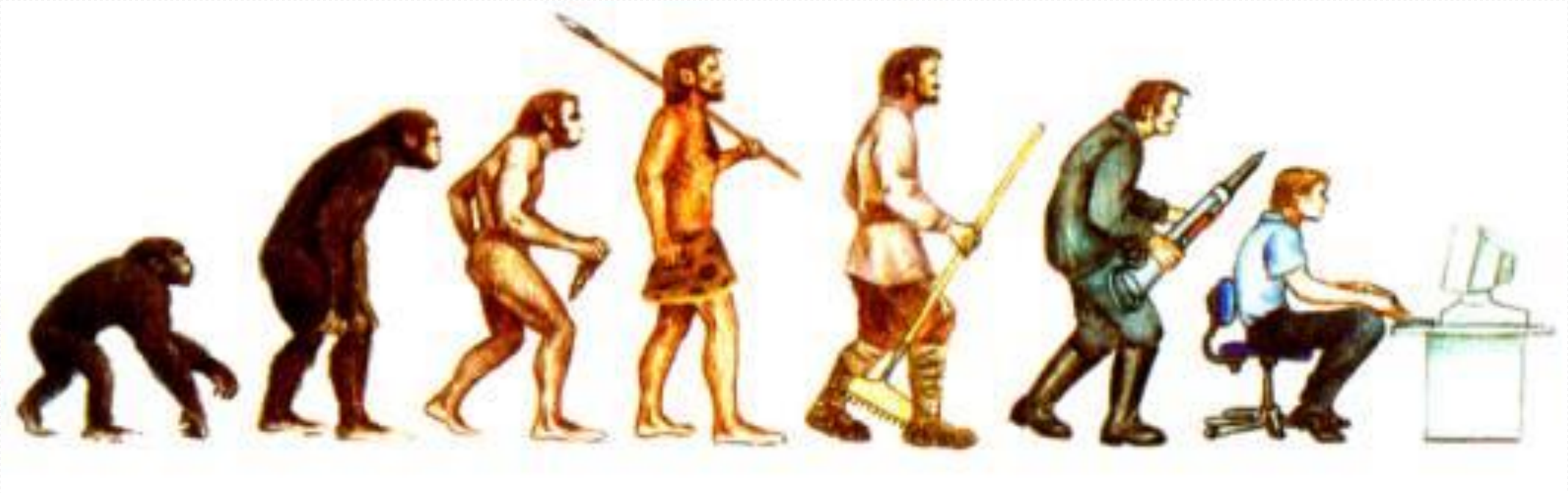
- **实现**
  - Python
  - Matlab
  - Java
  - Perl
  - R
  - C/C++
  - Windows
  - Linux



1 绪论

1 Introduction

学习的能力，才是智能的本质

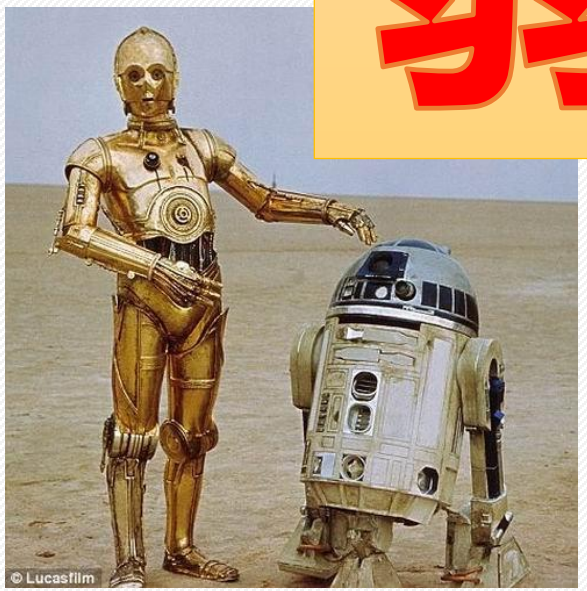


机器学习——赋予机器“学习”的灵魂

- 400多年前，发明了望远镜，拓展了“视觉”的能力；
- 150多年前，发明了电话机，提升了“听觉”的能力；
- 100多年前，从热气球到莱特兄弟的飞机，具备了“飞行”的能力；
- 如何让机器像智能生物一样，获得学习的能力？
- 早些时候的科学家试图让机械拥有真正意义上的智能，进而产生智能行为——学习，但最终这个方向并没有走通。
- 今天的科学家走出了思维的桎梏，放弃纯粹的生物模仿，而是利用科学理论完成仿生——用数学模拟智能生物学习的过程。

人们头脑中的机器人

弱人工智能



1.1引言

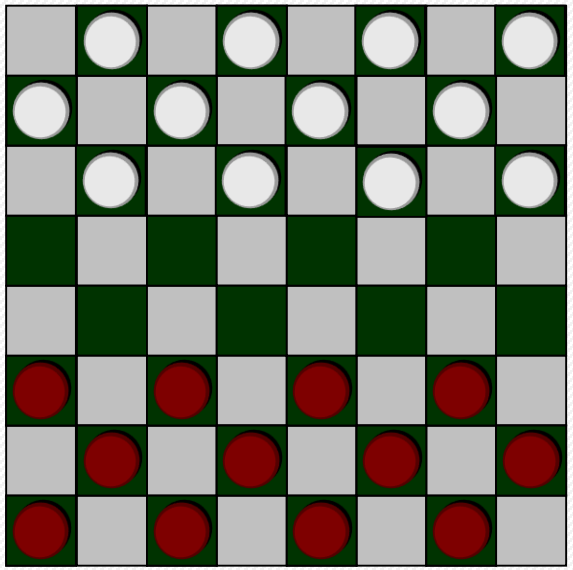
- 傍晚小街路面上沁出微雨后的湿润，和煦的细风吹来，抬头看看天边的晚霞，嗯，明天又是一个好天气。走到水果摊旁，挑了个根蒂蜷缩、敲起来声音浊响的青绿西瓜，一边满心期待着皮薄肉厚瓤甜的爽落感，一边愉快地想着，这学期狠下了工夫，基础概念弄得清清楚楚，算法作业也是信手拈来，这门课成绩一定差不了！
- 为什么看到微湿路面、感到和风、看到晚霞，就认为明天是好天呢？——**经验**
- 为什么色泽青绿、根蒂蜷缩、敲声浊响，就能判断出是正熟的好瓜？——**经验**
- 为什么下足了工夫、弄清了概念、做好了作业，自然会取得好成绩？——**经验**
- 上面对经验的利用是靠我们人类自身完成的。计算机能帮忙吗？

什么是机器学习？

■**Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.**

■**机器学习**：不需要确定性编程就可以赋予机器某项技能的研究领域。

Arthur Samuel (1959)



什么是机器学习？

■"... the design and development of **algorithm** that allow computers to **evolve** behaviors based on **empirical data**..."

机器学习的定义

■**A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.**

Tom M. Mitchell (1997)

■设定：

- **P**：计算机程序在某任务类T上的性能。
- **T**：计算机程序希望实现的任务类。
- **E**：表示经验，即历史的数据集。

■若该计算机程序通过利用经验E在任务T上获得了性能P的改善，则称该程序对E进行了学习。

机器学习 VS 数据挖掘

- Machine learning focuses on prediction, based on known properties learned from the training data, data mining focuses on the discovery of (previously) unknown properties in the data (this is the analysis step of knowledge discovery in databases).
- Data mining uses many machine learning methods, but with different goals; on the other hand, machine learning also employs data mining methods as "unsupervised learning" or as a preprocessing step to improve learner accuracy.
- They often employ the same methods and overlap significantly.

机器学习的定义

■**机器学习**：又称为统计机器学习，是关于计算机基于数据构建概率统计模型并运用模型对数据进行预测与分析的一门学科。

■**机器学习的特点**：

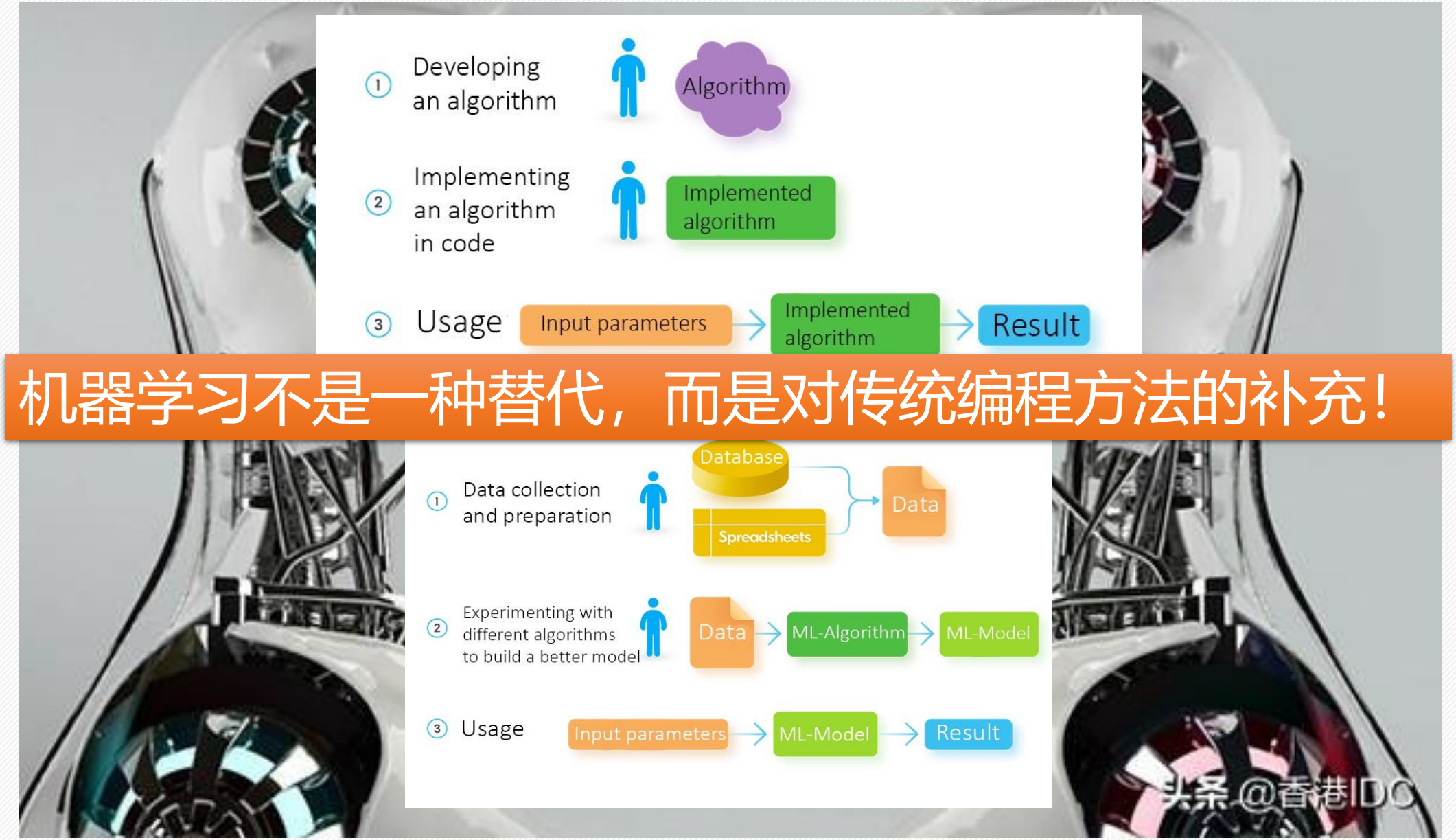
1）以**数据**为研究对象，是数据驱动的科学；

2）目的是对数据进行**预测与分析**，特别是对未知新数据；

3）以**方法**为中心；

4）是概率论、统计学、信息论、计算理论、最优化理论及计算机科学等多个领域的**交叉学科**，并在发展中逐步形成独自的理论体系和方法论。

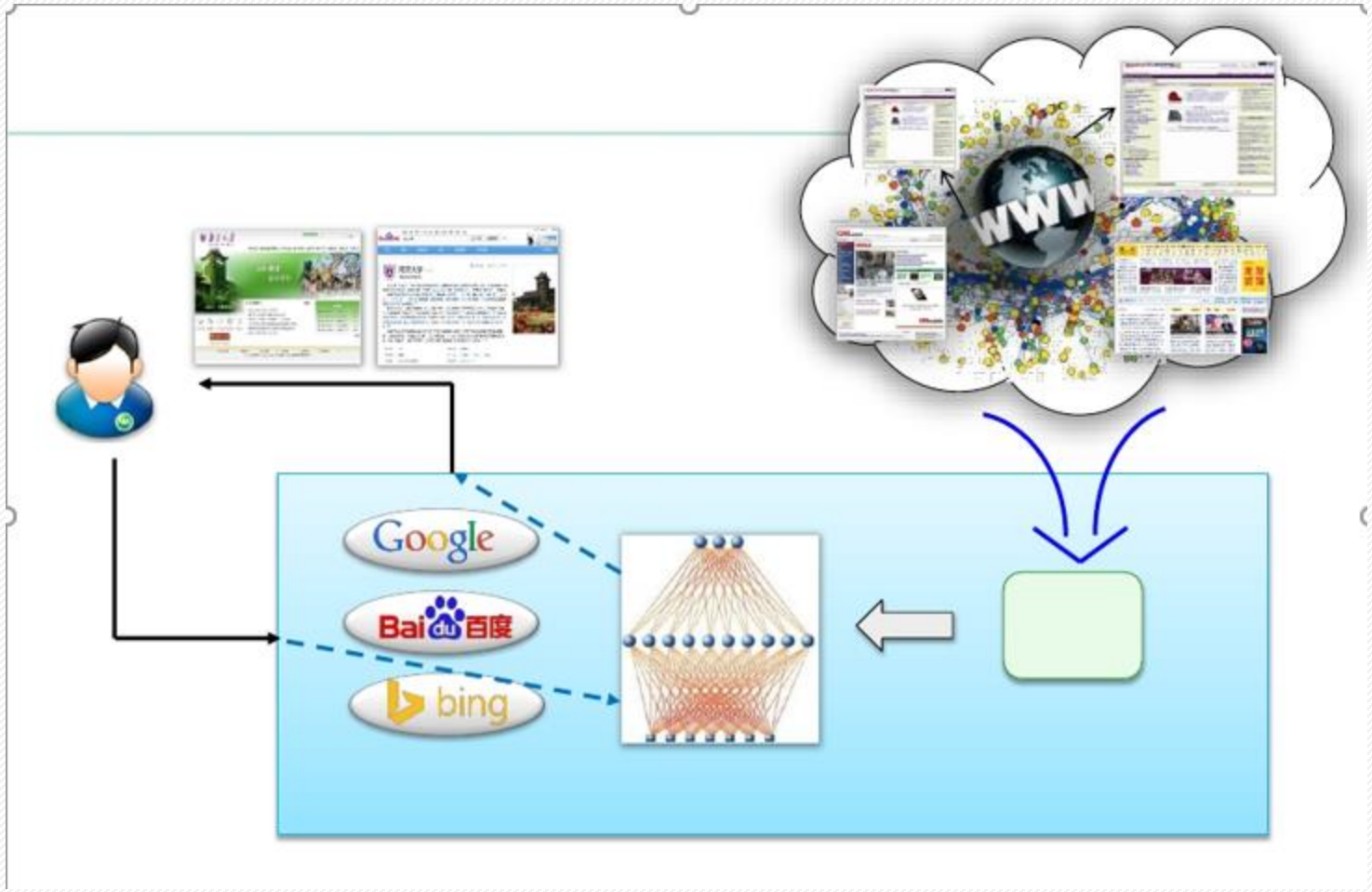
传统编程模式 VS 机器学习



机器学习能做什么？

机器学习已经渗透到社会生活的方方面面！

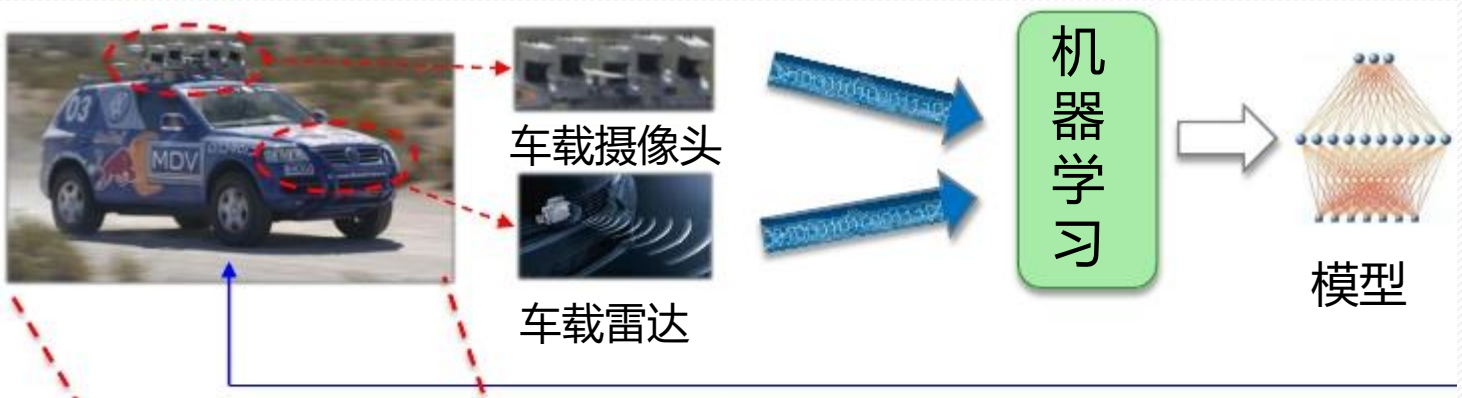
例，互联网搜索

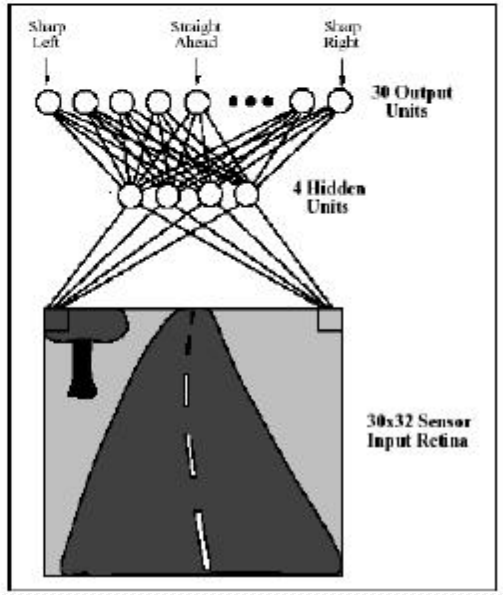


例，垃圾邮件拦截


垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件
垃圾邮件

例：自动驾驶汽车





美国早在20世80年代已开始研究基于机器学习的汽车自动驾驶技术。

距离自动驾驶还有多远？

咨询公司麦肯锡（McKinsey）的一份报告显示，拥有世界上最大汽车行业的中国有一天可能成为全球最大的自动驾驶汽车市场。它预测到2040年，中国可以从自动驾驶服务中产生多达1.1万亿美元的收入。

自动驾驶分级

自动驾驶分级		名称	定义	驾驶操作	周边监控	接管	应用场景
NHTSA	SAE						
L0	L0	人工驾驶	由人类驾驶者全权驾驶汽车。	人类驾驶员	人类驾驶员	人类驾驶员	无
L1	L1	辅助驾驶	车辆对方向盘和加减速中的一项操作提供驾驶，人类驾驶员负责其余的驾驶动作。	人类驾驶员和车辆	人类驾驶员	人类驾驶员	限定场景
L2	L2	部分自动驾驶	车辆对方向盘和加减速中的多项操作提供驾驶，人类驾驶员负责其余的驾驶动作。	车辆	人类驾驶员	人类驾驶员	
L3	L3	条件自动驾驶	由车辆完成绝大部分驾驶操作，人类驾驶员需保持注意力集中以备不时之需。	车辆	车辆	人类驾驶员	
L4	L4	高度自动驾驶	由车辆完成所有驾驶操作，人类驾驶员无需保持注意力，但限定道路和环境条件。	车辆	车辆	车辆	
	L5	完全自动驾驶	由车辆完成所有驾驶操作，人类驾驶员无需保持注意力。	车辆	车辆	车辆	所有场景

百度Apollo GO

2013年起步，2017年带着一张北京五环的罚单闯入大众视线。

在路试落地北京之前，无人驾驶出租车已在长沙、沧州进行试运营，接送超过10万名乘客

在Navigant Research为自动驾驶企业制定的竞争力2019排名榜单中，百度也首次进入到“Leader”评级，与Waymo与Cruise处于同级

宣称已经发展到接近L4级别的自动驾驶

运营时段受限在周一到周日的10：00-16：00，避开了早晚高峰以及夜晚的驾驶环境；其次，无法行驶在预设外的路线，只能在事前计划好的15个站点上下车；车速限制在60km/h以下；每台车只能在后座乘坐1-2名的乘客，乘坐者年龄符合18-60周岁之间；最重要的是，每辆车都配备有安全员，随时接管车辆行驶。

狗咬人 VS 人咬狗

2018年3月18日晚10时左右，Uber无人驾驶汽车在自动驾驶模式中撞到了49岁的伊莱恩·赫茨伯格（Elaine Herzberg），致其身亡。这是全球首例无人驾驶汽车导致行人死亡的事故

例，辅助总统胜选

How Obama's data crunchers helped him win

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

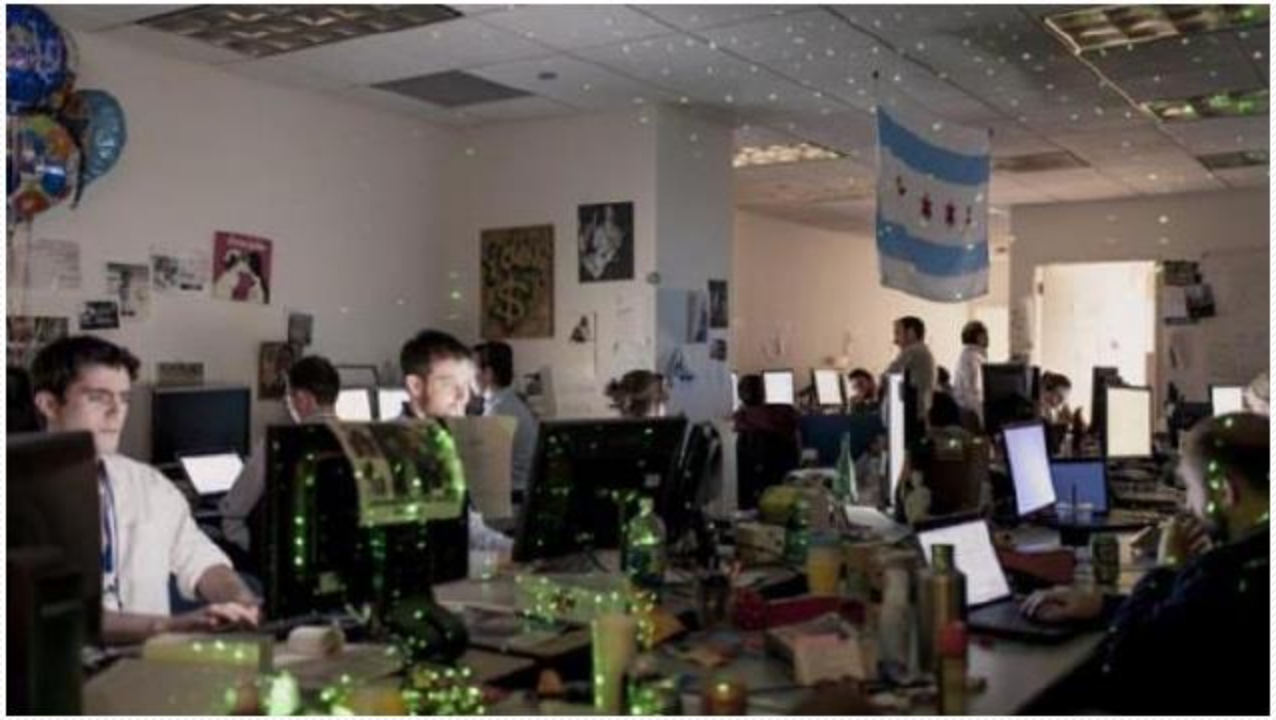
◆在总统候选人第一次辩论后，分析出哪些**选民**将倒戈，为每位选民找出一个最能说服他的理由

◆精准定位不同选民群体，建议购买冷门广告**时段**，广告资金效率比2008年提高14%

◆向奥巴马推荐，竞选后期应当在什么**地方**展开活动 —— 那里有很多争取对象

◆借助模型帮助奥巴马筹集到创纪录的10亿美元

例：帮助奥巴马胜选（政治）

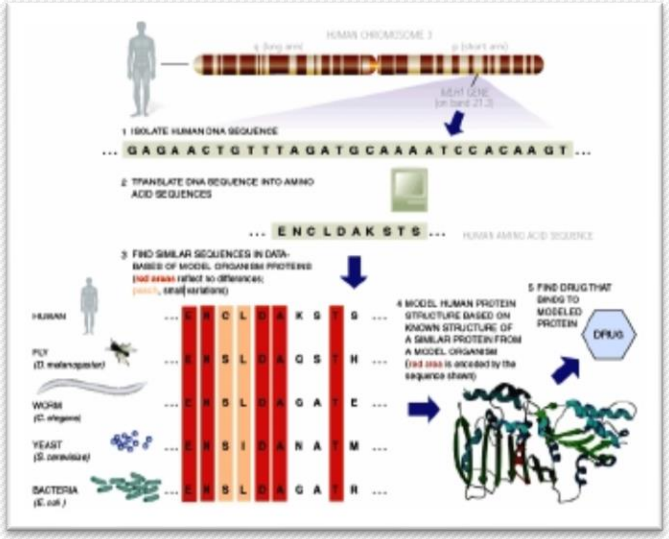


队长：Rayid Ghani

卡内基梅隆大学机器学习系
首任系主任Tom Mitchell
教授的博士生

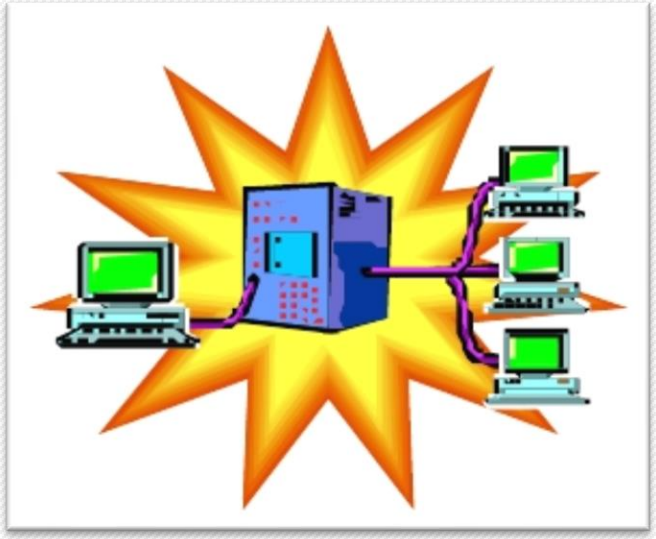
这个团队行动保密，定期向奥巴马报送结果；
被奥巴马公开称为总统竞选的
“核武器按钮”(They are our nuclear codes)

机器学习已经“无处不在”

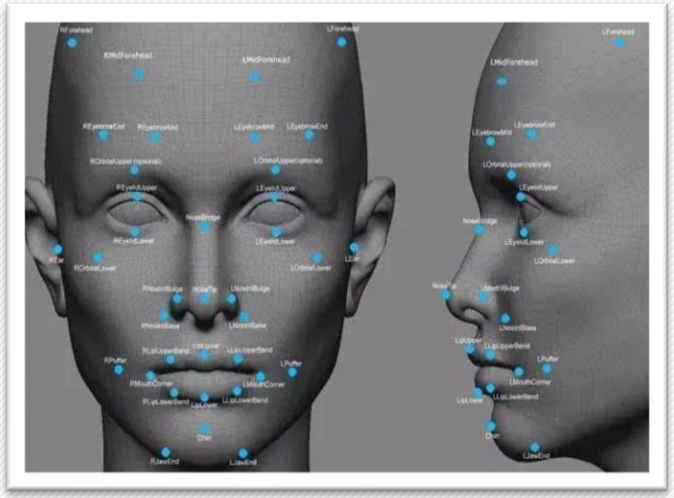


生物信息学

网络购物



入侵检测



人脸识别

火星机器人

决策助手

机器学习源自“人工智能”

机器学习源自“人工智能”

Artificial Intelligence (AI), 1956 -

1956年夏　美国达特茅斯学院夏季研讨会

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

达特茅斯会议标志着人工智能这一学科的诞生

约翰　麦卡锡
(1927-2011)
“人工智能之父”
1971年图灵奖

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。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。中学时自修CalTech的数学课程，17岁进入CalTech时免修两年数学，23岁在Princeton获博士学位，37岁担任Stanford大学AI实验室主任。

第三阶段：学科形成（20世纪80年代）

- 20世纪80年代，机器学习成为一个独立学科领域并快速发展，各种机器学习技术百花齐放。
- 1980年美国卡内基梅隆大学举行第一届机器学习研讨会。
- 1990年《机器学习：范型与方法》出版。

第四阶段：学习期

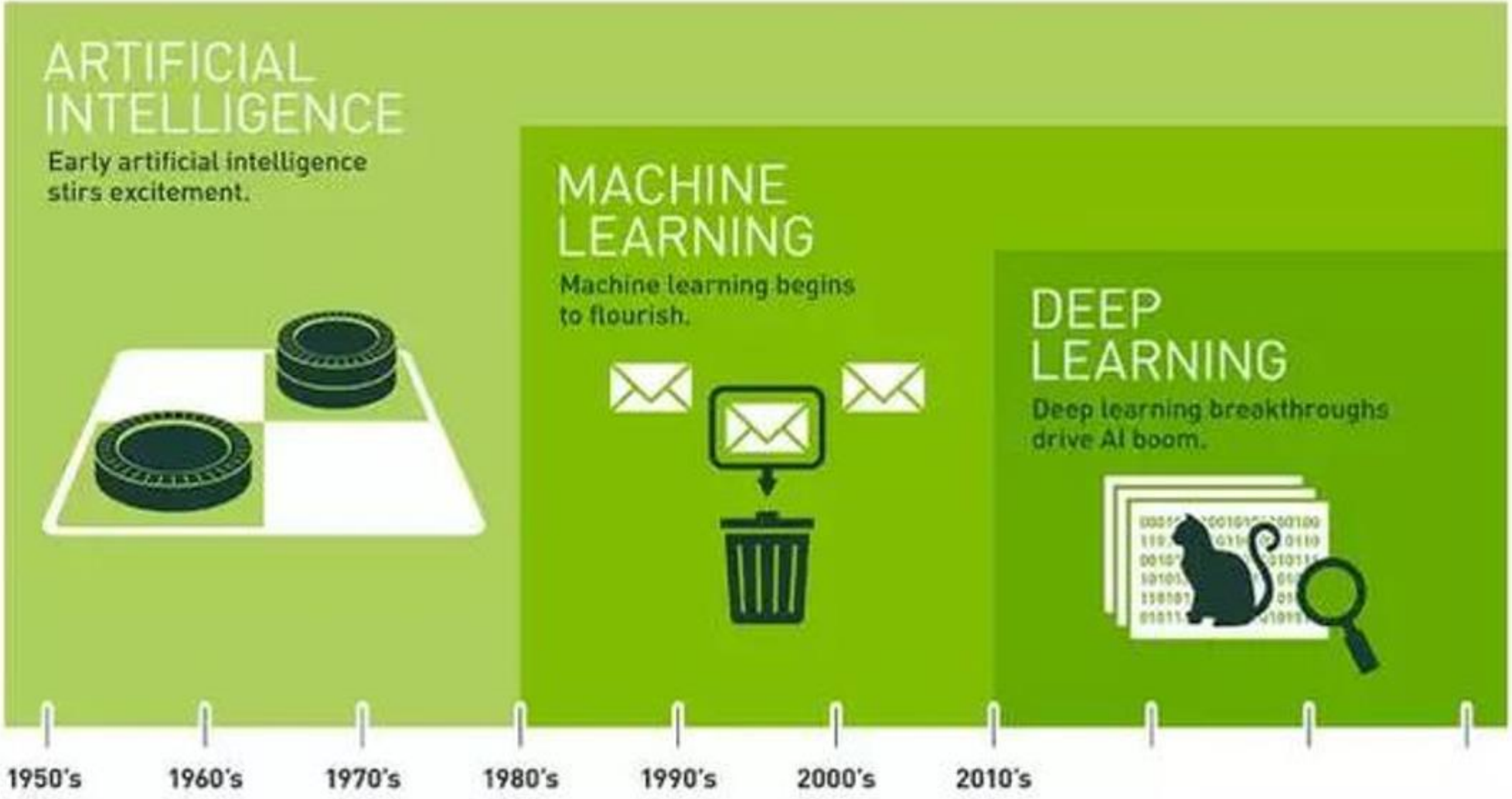
■1990s ~: Machine Learning

- 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。

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

人工智能、机器学习、深度学习的关系



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

当下的机器学习

今天的“机器学习”已经是一个广袤的学科领域

第32届国际机器学习大会的“主题领域”

2006年，美国CMU(卡内基梅隆大学)成立“机器学习系”

- Active Learning
- Approximate Inference
- Bayesian Nonparametric Methods
- Bioinformatics
- Causal Inference
- Clustering
- Computational Social Sciences
- Cost-Sensitive Learning
- Digital Humanities
- Ensemble Methods
- Feature Selection and Dimensionality Reduction
- Finance
- Gaussian Processes
- Graphical Models
- Inductive Logic Programming and Relational Learning
- Information Retrieval
- Kernel Methods
- Large-Scale Machine Learning
- Latent Variable Models
- Learning for Games
- Learning Theory
- Manifold Learning
- Network and Graph Analysis
- Neural Networks and Deep Learning
- Planning and Control
- Privacy, Anonymity, and Security
- Ranking and Preference Learning
- Recommender Systems
- Reinforcement Learning
- Robotics
- Rule and Decision Tree Learning
- Semi-Supervised Learning
- Sparsity and Compressed Sensing
- Spectral Methods
- Speech Recognition
- Statistical Relational Learning
- Structured Output Prediction
- Supervised Learning
- Sustainability, Climate, and Environment
- Time-Series Analysis

经常被谈到的“深度学习”(Deep Learning)仅是机器学习中的一个小分支

大数据时代，机器学习必不可少

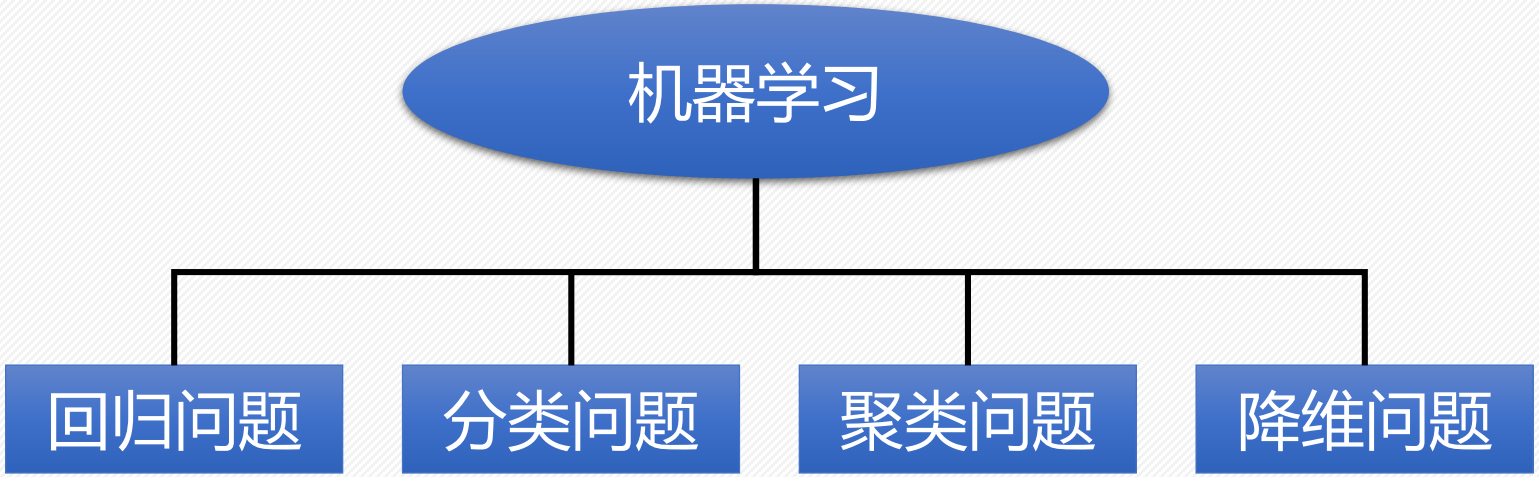
收集、传输、存储大数据的目的，

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

按任务类型分类

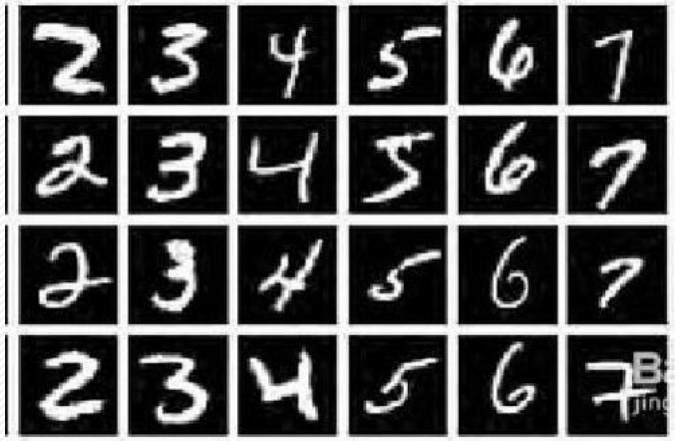


按任务类型分类

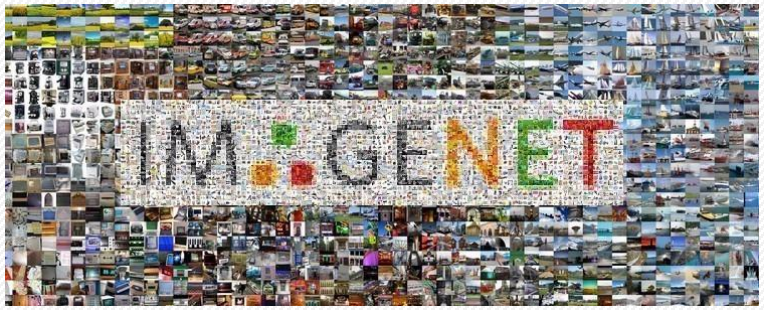
■**回归问题**：对已有的数据样本点进行拟合，再根据拟合出来的函数，对未来进行预测。

按任务类型分类

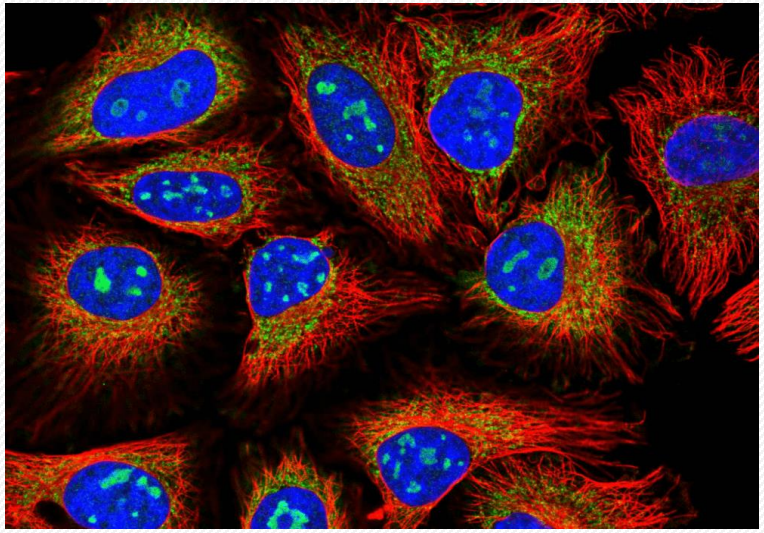
■**分类问题**：比如图片识别、情感分析等领域会经常用到分类任务。



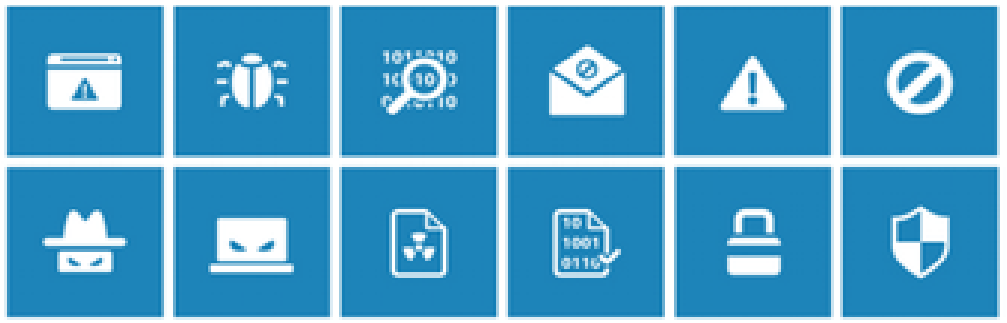
几种典型的分类任务



ImageNet：对自然对象进行分类（1000 类）



将人类蛋白质显微镜图像分为28类



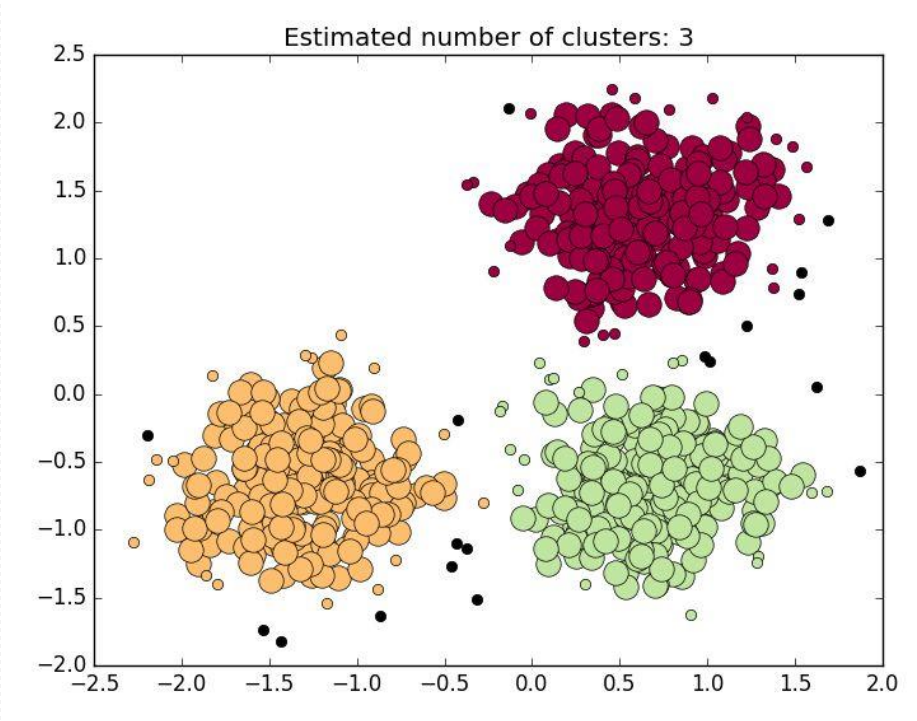
将恶意软件分为9类

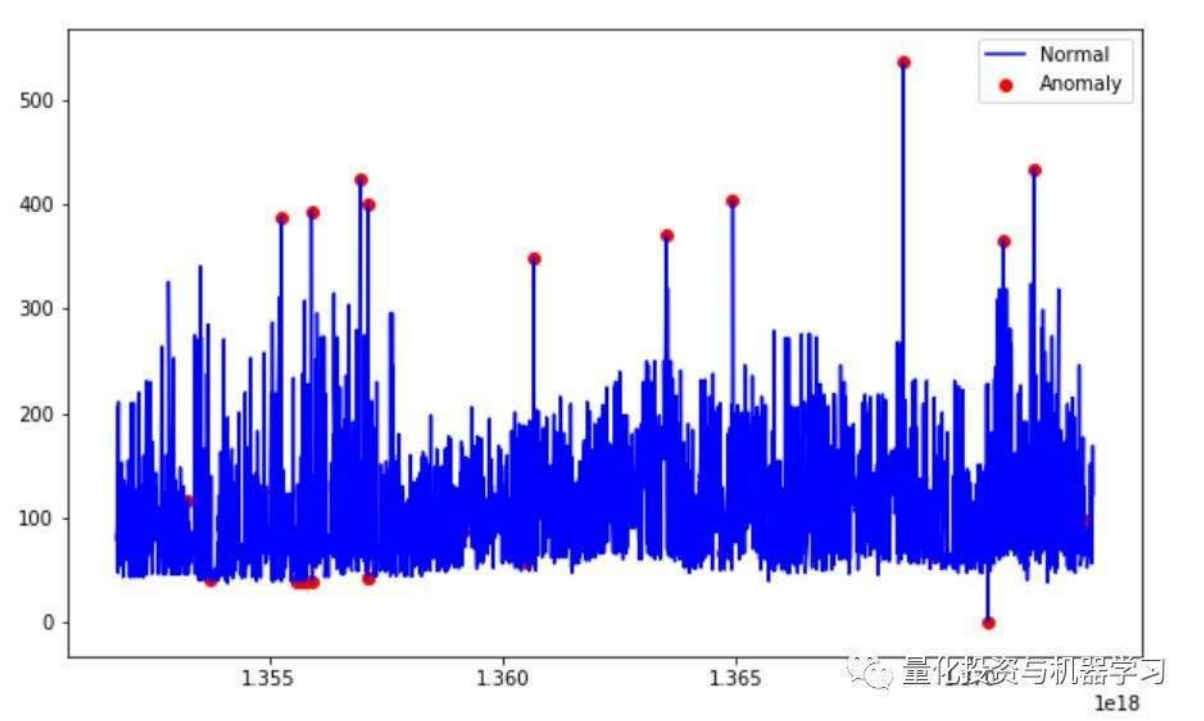
comment_text	toxic	severe_toxic	obsc
Explanation\nWhy the edits made under my usern...	0	0	0
D'aww! He matches this background colour I'm s...	0	0	0
Hey man, I'm really not trying to edit war. It...	0	0	0
"\nMore\nI can't make any real suggestions on ...	0	0	0
You, sir, are my hero. Any chance you remember...	0	0	0

将维基百科上的恶语评论分为7类

按任务类型分类

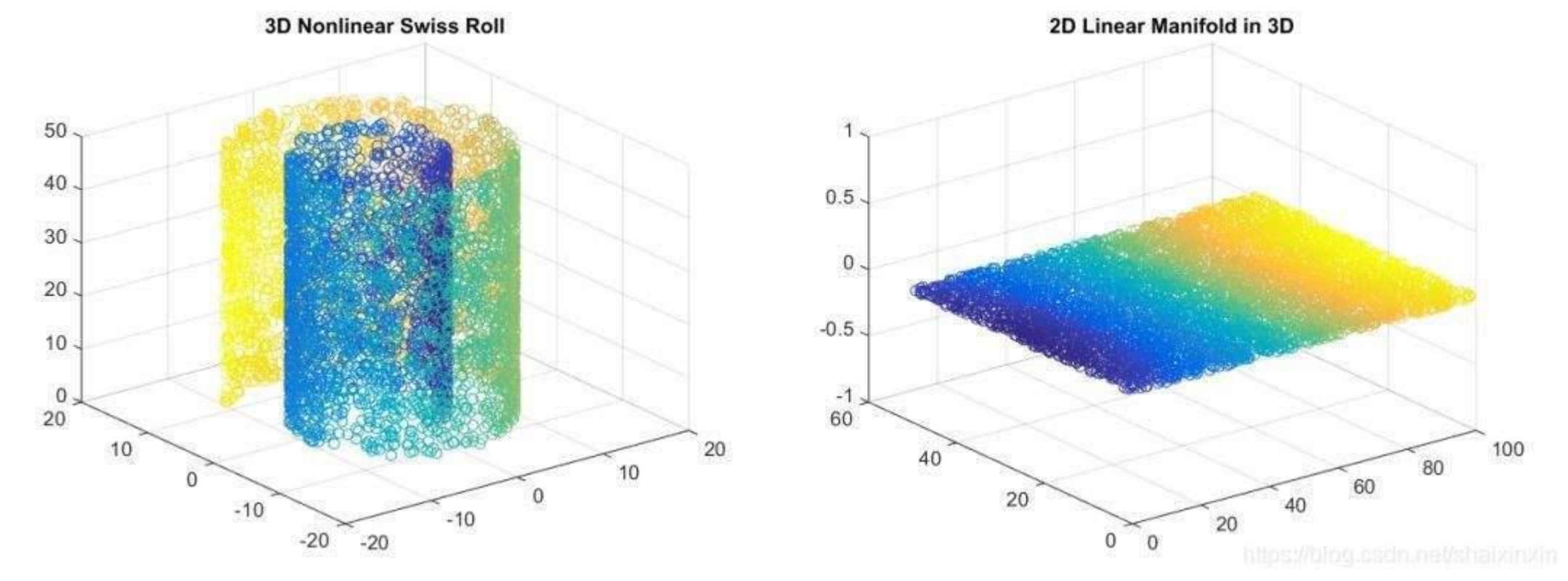
■**聚类问题**：目标是将样本划分为子集或簇。简单的说，希望利用模型将样本数据集聚合成若干个类。



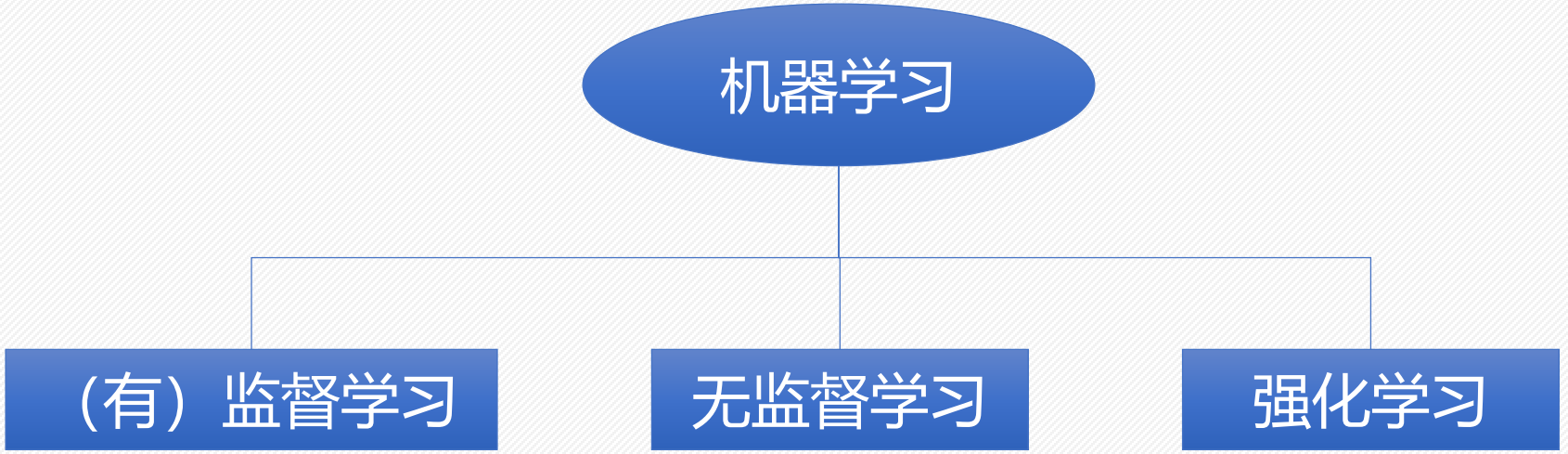


按任务类型分类

■**降维问题**：减少数据的维度，对数据进行降噪、去冗余。



按学习方式分类



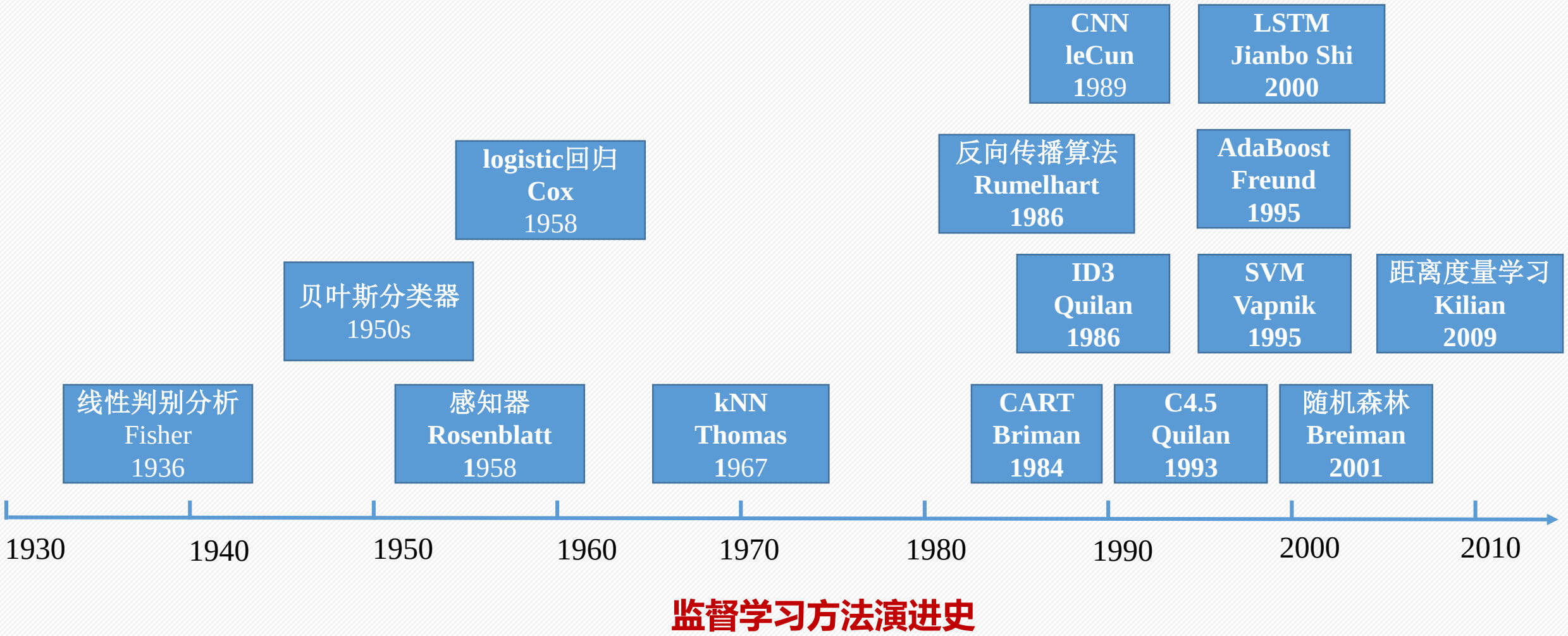
按学习方式分类

■1）**有监督学习**（supervised learning）：简称**监督学习**，是指从标注数据中学习预测模型的机器学习问题。常见的监督学习任务有分类（classification）和回归（regression）。

■2）**无监督学习**（unsupervised learning）：在无监督学习中，训练样本的结果信息没有被标注，即训练集的结果标签是未知的。学习的目标是通过这些无标记训练样本的学习来揭示数据的内在规律，发现隐藏在数据之下的内在模式。比较常见的无监督学习任务：聚类（clustering）和降维（dimension reduction）。

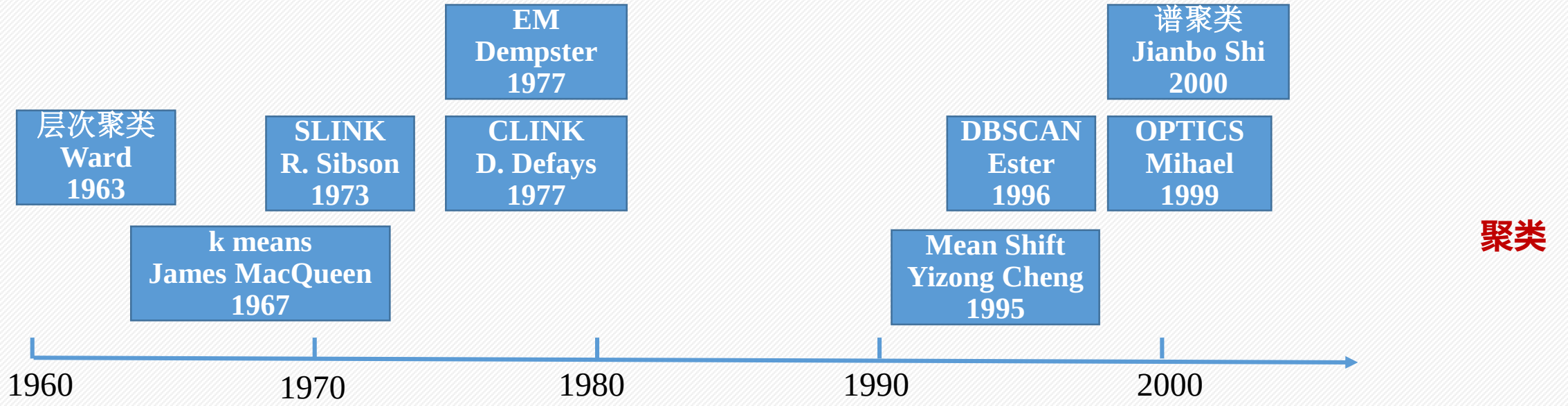
■3）**强化学习**（Reinforcement learning）：让模型在一定的环境中学习，每次行动会有对应的奖励，目标是使奖励最大化。

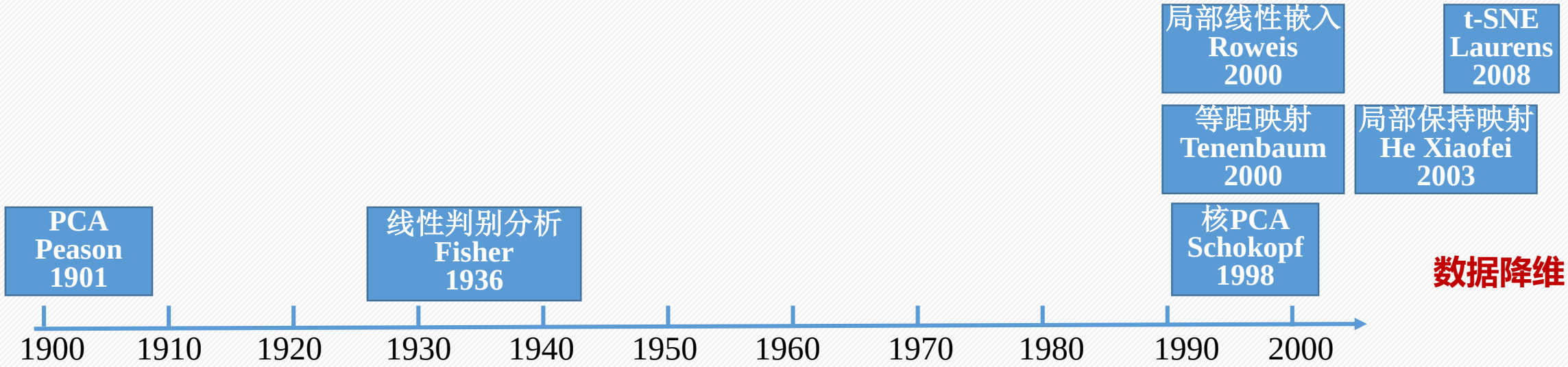
机器学习算法——监督学习



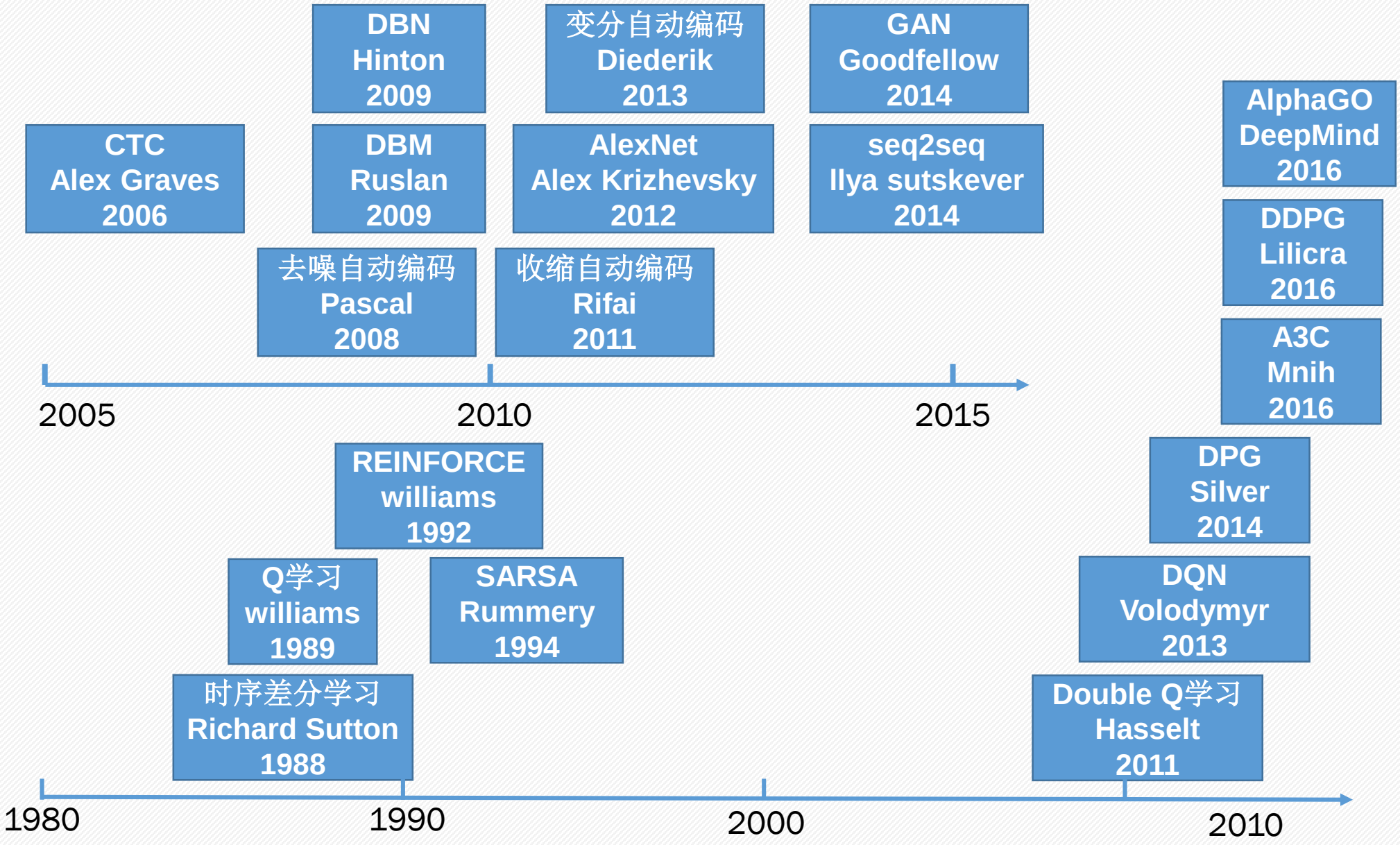
监督学习方法演进史

机器学习算法——无监督学习





机器学习——深度学习


DBN
Hinton
2009
变分自动编码
Diederik
2013
GAN
Goodfellow
2014
AlphaGO
DeepMind
2016
CTC
Alex Graves
2006
DBM
Ruslan
2009
AlexNet
Alex Krizhevsky
2012
seq2seq
Ilya sutskever
2014
DDPG
Lilicra
2016
去噪自动编码
Pascal
2008
收缩自动编码
Rifai
2011
A3C
Mnih
2016
2005
2010
2015
REINFORCE
williams
1992
DPG
Silver
2014
Q学习
williams
1989
SARSA
Rummery
1994
DQN
Volodymyr
2013
时序差分学习
Richard Sutton
1988
Double Q学习
Hasselt
2011
1980
1990
2000
2010

1.2 基本术语

序号	色泽	根蒂	敲声
1	青绿	蜷缩	浊响
2	乌黑	稍蜷	沉闷
3	浅白	硬挺	清脆
4	乌黑	稍蜷	沉闷

- 每行称为一个**记录**，所有记录的集合称为**数据集**（data set）。
- 每一条记录称为一个**实例**（**示例**）（instance）或**样本**（sample）。
- 反映事件或对象在某方面的表现或性质，称为**特征**（feature）或**属性**（attribute）。例如：色泽、根蒂、敲声。
- **属性值**（attribute value）：属性的取值。例如，青绿、清脆。
- **属性空间**（**样本空间**、**输入空间**）：属性张成的空间。
- **特征向量**（feature vector）：即每个示例。
- **维度**（dimensionality）：样本的特征数。

1.2 基本术语

- 令$\boldsymbol{D}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_N\}$表示$N$个示例的数据集，每个示例有$d$个属性组成。每个示例$\boldsymbol{x}_i=(x_{i1};x_{i2};\ldots;x_{id})$是$d$维空间上的一个向量。$\boldsymbol{x}_i\in\mathcal{X}$，其中$x_{ij}$是$\boldsymbol{x}_i$在第$j$个属性上的取值。$\mathcal{X}$称为**属性空间**。
- 从数据中学习模型的过程称为“**学习**”（learning）或“**训练**”（training）。
- 训练过程中使用的样本称为“**训练样本**”（training sample）。
- **训练集**（training set）：训练样本的集合。
- **假设**（hypothesis）：学得模型对应的关于数据的某种潜在规律。
- **真相或真实**（ground truth）：这种潜在规律本身。
- 学习的过程就是为了找出或逼近真相。

1.2 基本术语

- **标记/标签**（label）：示例的结果信息。
- **样例**（example）：拥有了标记的示例。(x_i, y_i)表示第i个样例，其中$y_i \in \mathcal{Y}$，$\mathcal{Y}$是所有标记的集合，称“**标记空间**”或“**输出空间**”。
- **分类**（classification）:预测值为离散值的问题。例如，好瓜、坏瓜。
- **回归**（regression）：预测值为连续值的问题。例如，西瓜成熟度，0.95、0.37等。
- **二分类**（binary classification）：只涉及两个类别，通常将其中一个称为**正类**（positive class），另一个称为**反类**（negative class）
- **多分类**（multi-class classification）：涉及多个类别。

1.2 基本术语

■通过对训练集$\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_N, y_N)$进行学习，建立一个从输入空间$\boldsymbol{x}$到输出空间$\boldsymbol{y}$的映射$f: \boldsymbol{x} \mapsto \boldsymbol{y}$。

■对于二分类任务，通常令$\boldsymbol{y} = \{-1, +1\}$或$\{0, 1\}$；对于多分类任务，$|\boldsymbol{y}| > 2$；对于回归任务，$\boldsymbol{y} = \mathbb{R}$，$\mathbb{R}$为实数集。

■**测试**（**testing**）：使用模型进行预测的过程。

■**测试样本**（**testing sample**）：被测试的样本。

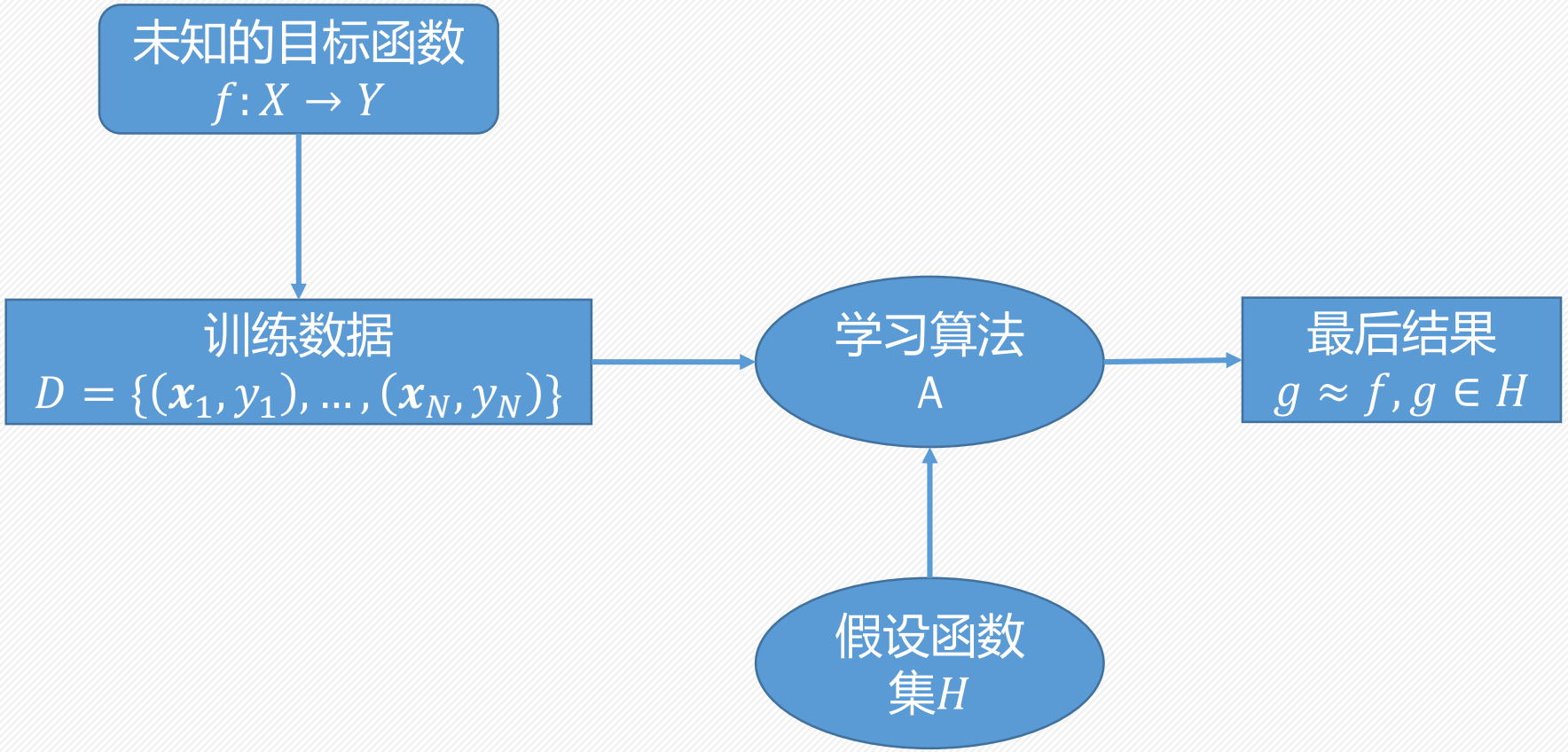
■**泛化能力**（**generalization**）：学得的模型适用于新样本的能力。

■**独立同分布**（**independent and identically distributed，IID**）：通常假设样本空间中全体样本服从一个未知分布（**distribution**），获得的每个样本都独立地从这个分布上采样获得，即“独立同分布”。

机器学习再认识

- 假设任意一个现象背后都存在规律。该规律可以看作一个复杂函数f。从机器学习和数学角度去看，f是目标函数。
- 从数学角度看，所观察的现象就是目标函数f产生的样本集$\boldsymbol{D}$。我们通过不断地观察现象、进行总结，会得到规律函数g，因为现实中所观察到的现象往往包含误差或干扰，并且样本数不可能无限多，所以g只能趋近f。g越趋近f，说明我们的总结归纳越好、理论越完备。
- 机器学习就是让机器代替人类去观察样本、求解函数g的过程。机器学习算法A负责从数据样本集$\boldsymbol{D}$中找出统计规律（在假设函数集H中找出规律函数g），找到的规律函数g与目标函数f越相似，找到的规律就越可靠。

机器学习再认识



机器学习流程示意图

1.3 假设空间

- **归纳**（induction）：从特殊到一般的泛化（generalization）过程，也就是从具体的事实归纳出一般性规律。
- **演绎**（deduction）：从一般到特殊的特化（specialization）过程，也就是从基础原理推演出具体状况。
- **归纳学习**（inductive learning）：从样例中学习就是归纳的过程。
- **广义的归纳学习**：从样例中学习。
- **狭义的归纳学习**：从训练数据中学习概念，因此称为**概念生成**/**概念（concept）学习**。

1.3 假设空间

表1.1 西瓜数据集

序号	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

好瓜↔ (色泽=?)∧(根蒂=?)∧(敲声=?)

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”的假设

假设空间的大小：$4 \times 4 \times 4 + 1 = 65$

版本空间

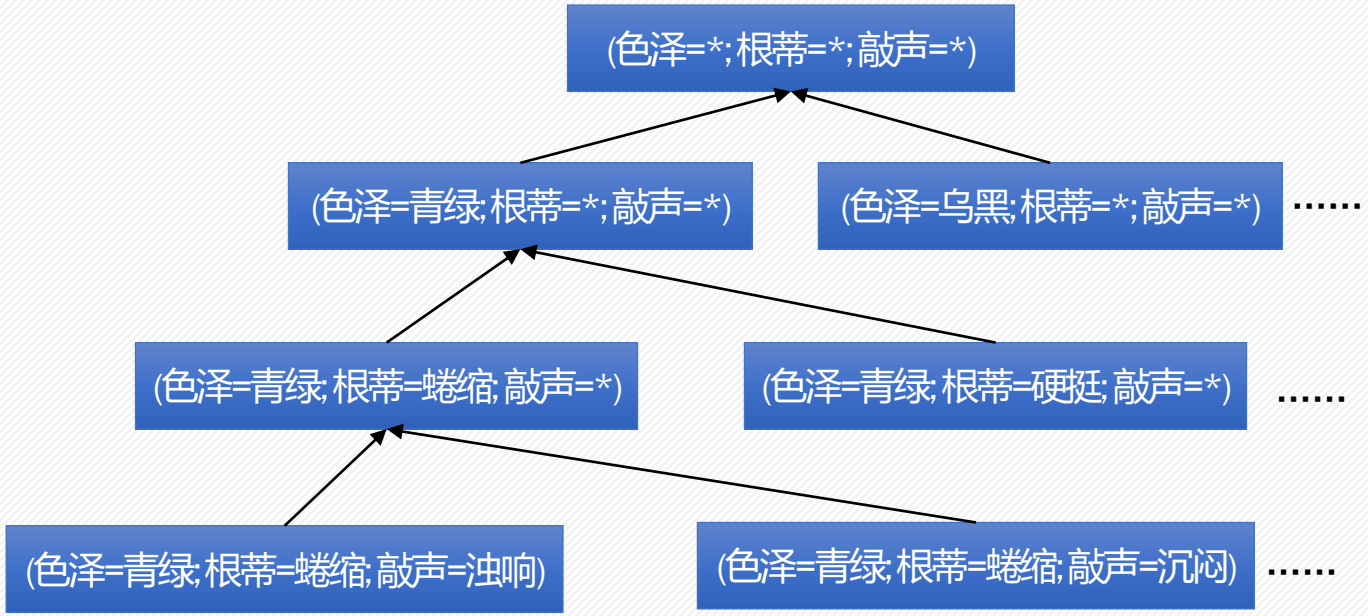


图1.1 西瓜问题的假设空间

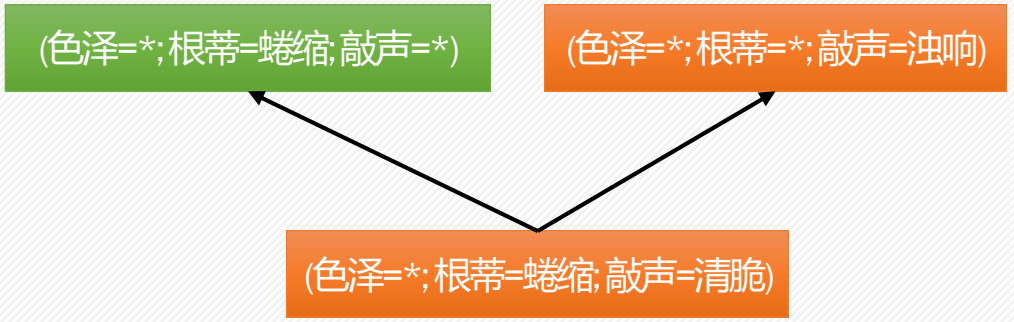


图1.2 西瓜问题的版本空间

新瓜： (色泽=青绿； 根蒂=蜷缩； 敲声=沉闷)

应该采用哪一个模型(假设)?

版本空间(version space): 与训练集一致的假设集合。

例: (青绿; 蜷缩; 沉闷)

在面临新样本时，会产生不同的输出。

1.4 归纳偏好(inductive bias)

归纳偏好：机器学习算法在学习过程中对某种类型假设的偏好。
任何一个有效的机器学习算法必有其偏好。

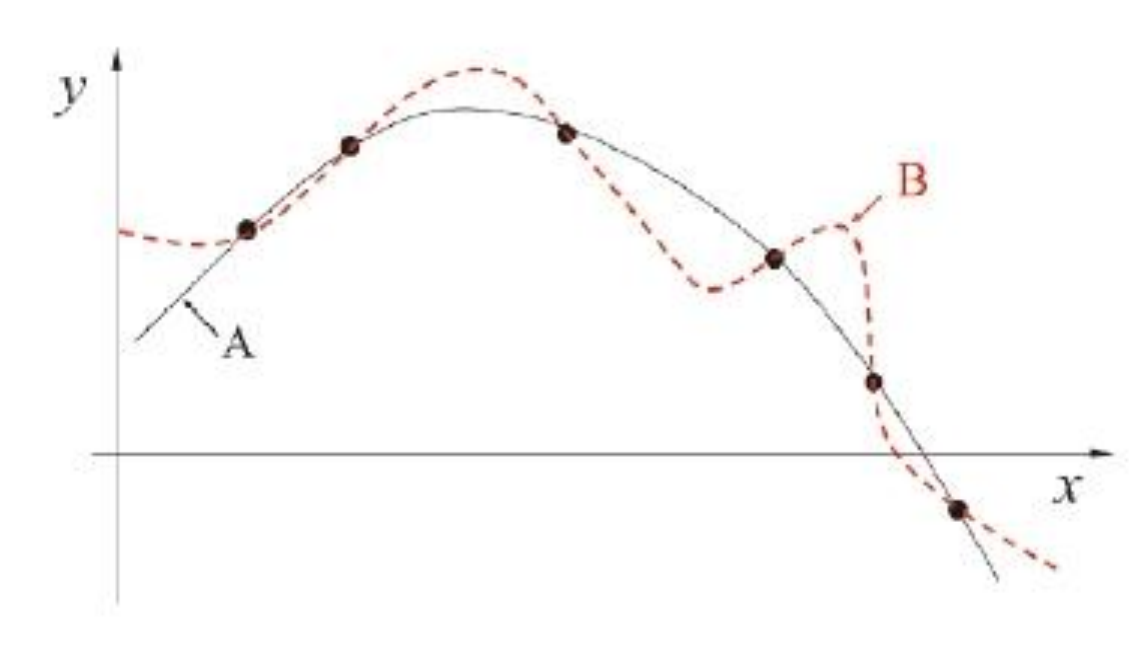
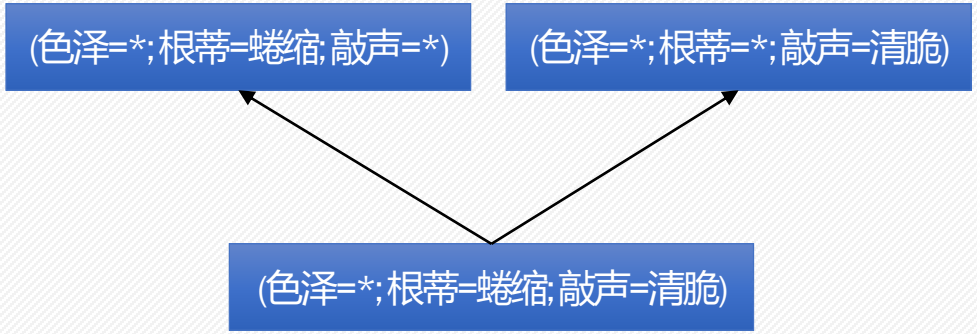


图1.3 存在多条曲线与有限样本训练集一致

奥卡姆剃刀（Occam razor）原理

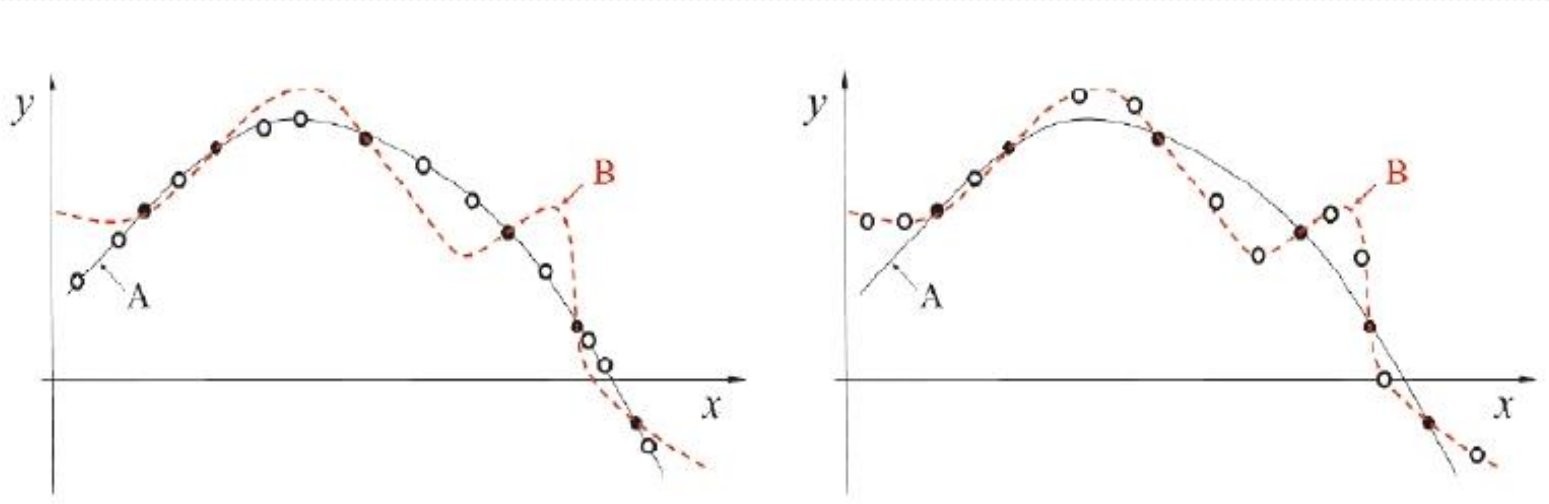


- 由14世纪逻辑学家、奥卡姆的威廉（William of Occam，约1285-1349）提出。这个原理称为“如无必要，勿增实体”，即“简单有效原理”。
- 在所有可能选择的模型中，能够很好地解释已知数据并且最简单才是最好的模型，也就是应该选择的模型。
- 在具体现实问题中，算法的归纳偏好是否与问题本身匹配，大多数时候直接决定了算法能否取得好的性能。

哪个算法更好？

■没有免费的午餐！（No Free Lunch Theorem，NFL）

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$\mathfrak{L}_a$算法好。



(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

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

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

考虑二分类问题，真实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

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

NFL定理

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

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

f为均匀分布

$$=\sum_{\boldsymbol{x}\in\mathcal{X}-\boldsymbol{X}}P(\boldsymbol{x})\sum_h P(h|\boldsymbol{X},\mathfrak{L}_a)\frac{1}{2}2^{|\boldsymbol{x}|}$$

概率之和为1

$$=\frac{1}{2}2^{|\boldsymbol{x}|}\sum_{\boldsymbol{x}\in\mathcal{X}-\boldsymbol{X}}P(\boldsymbol{x})\sum_h P(h|\boldsymbol{X},\mathfrak{L}_a)$$

$$=2^{|\boldsymbol{x}|-1}\sum_{\boldsymbol{x}\in\mathcal{X}-\boldsymbol{X}}P(\boldsymbol{x})\cdot 1$$

总误差与学习算法无关！

所有算法一样好！

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；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！

现实的免费午餐：美国政府为低收入家庭的中小学学生提供免费午餐（School meal programs in the United States），以确保不同收入家庭的儿童皆可得到充足营养。

我们中国也有！

我矿也有！

在路上……

- ML开始从概念走向应用，未来ML将无处不在……
- 挑战依然存在，不可期望ML可以解决所有问题。
- 目前ML技术还在路上，远非处于AI的神奇大爆炸时代，还远未成为一个理论完备的学科，任重道远。
- “预测未来真的很难，尤其是提前预测更是难上加难”[MIT AI Lab&CSAIL Rodney Brooks]。
- “在一个合理分工的社会里，提高生产力对每个人都有好处，问题不在于技术，而是利益如何分配”[深度学习之父，G. Hinton，2018图灵奖获得者]。

The end