人工智能与机器学习

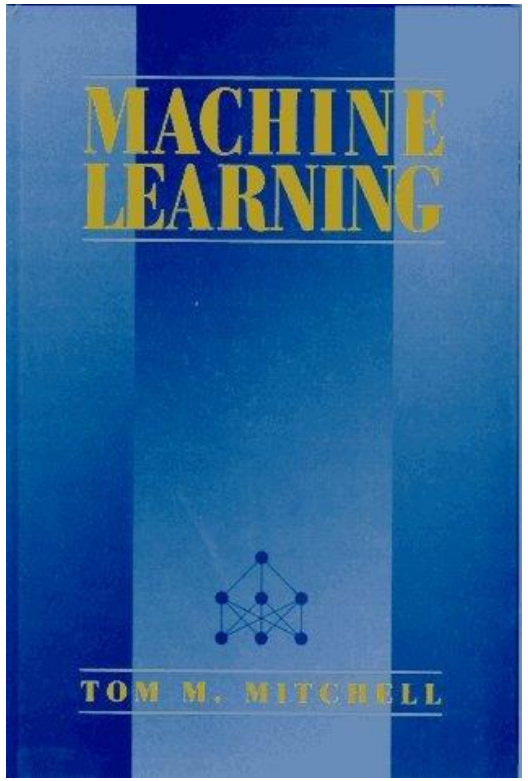
叶琦

工业控制研究所

杭州 • 浙江大学 • 2022

参考书

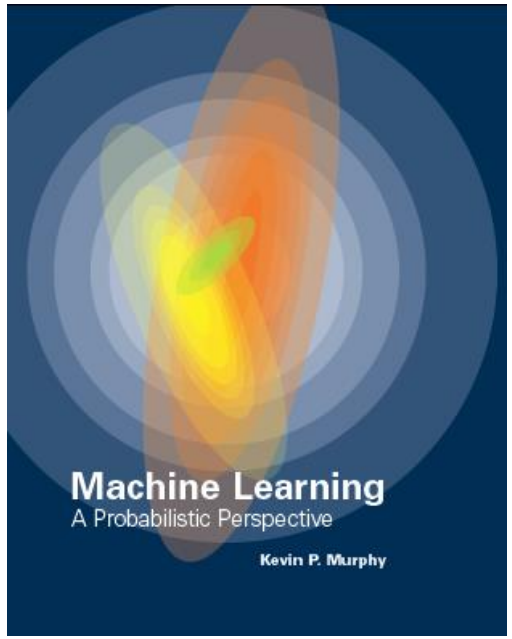
- 《机器学习》，TomM.Mitchell（汤姆·米切尔）著，曾华军，张银华等译，机械工业出版社，2003年 。



其它参考书

- 《机器学习及其应用》，周志华，王钰主编，清华大学出版社，2009。
- 《神经网络与机器学习》，Simon Haykin著，机械工业出版社，2010。
- 《机器学习导论》，Ethem Alpaydin著，机械工业出版社，2009。
- 《Machine Learning—— A Probabilistic Perspective》

  Kevin P. Murphy, 2012



第1章　机器学习概述

内容

- 什么是机器学习?
- 机器学习的组成部分
- 机器学习的应用和影响
- 机器学习简要发展历史回顾

内容

- 什么是机器学习?
- 机器学习的组成部分
- 机器学习的应用和影响
- 机器学习简要发展历史回顾

什么是机器学习?

- 计算机程序如何随着经验积累自动提高性能，系统自我改进的过程; 或，计算机利用经验改善系统自身性能的行为。——米切尔

- 案例:
  - 医疗记录学习治疗新疾病的方法
  - 用户用电模式记录学习降低能源消耗的策略
  - 用户浏览记录学习广告精准投放

什么是机器学习

Machine learning: what and why?

We are drowning in information and starving for knowledge. — John Naisbitt.

We are entering the era of **big data**. For example, there are about 1 trillion web pages[1]; one hour of video is uploaded to YouTube every second, amounting to 10 years of content every day[2]; the genomes of 1000s of people, each of which has a length of 3.8×10^9 base pairs, have been sequenced by various labs; Walmart handles more than 1M transactions per hour and has databases containing more than 2.5 petabytes (2.5×10^{15}) of information (Cukier 2010); and so on.

This deluge of data calls for automated methods of data analysis, which is what **machine learning** provides. In particular, we define machine learning as a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty (such as planning how to collect more data!).

Machine Learning: A Probabilistic Perspective. Kevin P. Murphy

什么是机器学习

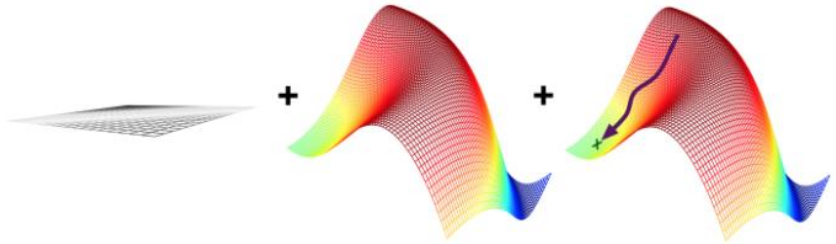
- 中科院王珏研究员给出的定义：
  - 令W是给定世界的有限或无限所有观测对象的集合，由于我们的观测能力有限，我们只能获得这个世界的一个子集 $Q \subset W$，称为样本集。机器学习就是根据这个样本集，推算这个世界W的模型，使它对这个世界（尽可能地）为真。
- 三个重要的理论问题：
  - 一致：W与Q有相同的性质。eg. i.i.d
  - 划分：设样本定义于d维空间，要寻找在这个空间上的决策分界面
  - 泛化（推广能力）：对未知样本的判断能力

什么是机器学习

- Pedro Domingos对学习理解

2. LEARNING = REPRESENTATION + EVALUATION + OPTIMIZATION

Suppose you have an application that you think machine learning might be good for. The first problem facing you is the bewildering variety of learning algorithms available. Which one to use? There are literally thousands available, and hundreds more are published each year. The key to not getting lost in this huge space is to realize that it consists of combinations of just three components. The components are:



Representation: The model landscape

Evaluation: Preferences over the landscape

Optimization: Strategy for fulfilling preferences

Representation. A classifier must be represented in some formal language that the computer can handle. Conversely, choosing a representation for a learner is tantamount to choosing the set of classifiers that it can possibly learn. This set is called the *hypothesis space* of the learner. If a classifier is not in the hypothesis space, it cannot be learned. A related question, which we will address in a later section, is how to represent the input, i.e., what features to use.

Evaluation. An evaluation function (also called *objective function* or *scoring function)* is needed to distinguish good classifiers from bad ones. The evaluation function used internally by the algorithm may differ from the external one that we want the classifier to optimize, for ease of optimization (see below) and due to the issues discussed in the next section.

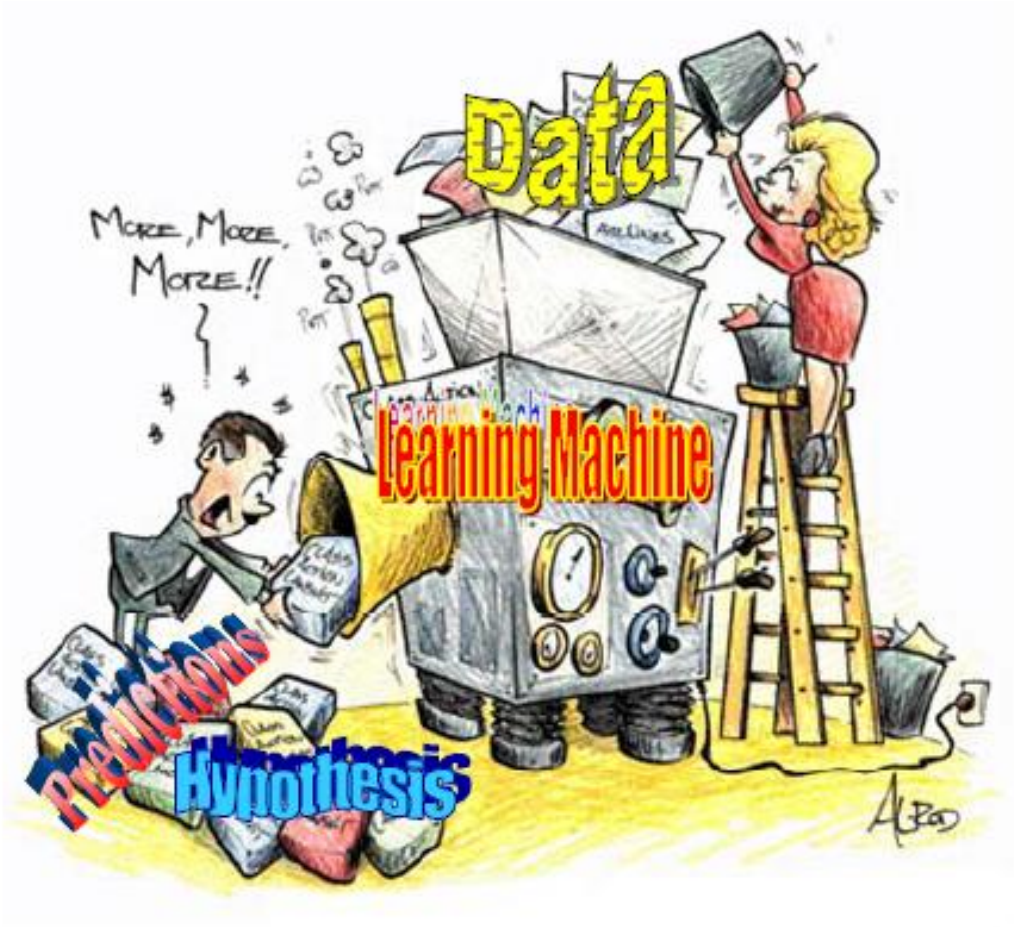
Optimization. Finally, we need a method to search among the classifiers in the language for the highest-scoring one. The choice of optimization technique is key to the efficiency of the learner, and also helps determine the classifier produced if the evaluation function has more than one optimum. It is common for new learners to start out using off-the-shelf optimizers, which are later replaced by custom-designed ones.

[A Few Useful Things to Know About Machine Learning](#)
[Machine Learning = Representation + Evaluation + Optimization](#)

Table 1. The three components of learning algorithms.

Representation	Evaluation	Optimization
Instances	Accuracy/Error rate	Combinatorial optimization
K-nearest neighbor	Precision and recall	Greedy search
Support vector machines	Squared error	Beam search
Hyperplanes	Likelihood	Branch-and-bound
Naive Bayes	Posterior probability	Continuous optimization
Logistic regression	Information gain	Unconstrained
Decision trees	K-L divergence	Gradient descent
Sets of rules	Cost/Utility	Conjugate gradient
Propositional rules	Margin	Quasi-Newton methods
Logic programs		Constrained
Neural networks		Linear programming
Graphical models		Quadratic programming
Bayesian networks		
Conditional random fields		

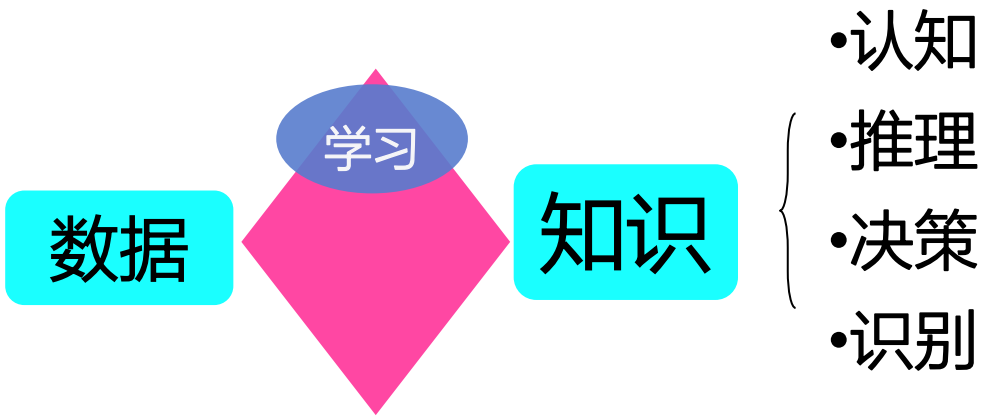
什么是机器学习



引用自CMU Dr. Eric Xing的Lecture Notes

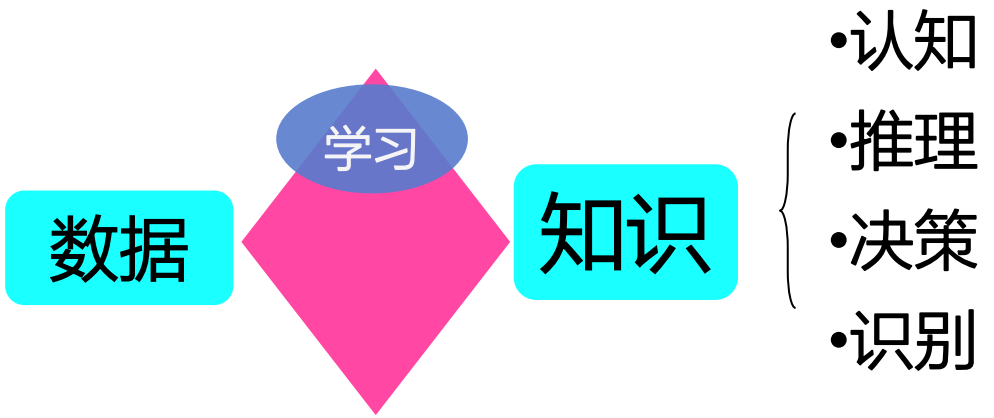
学习与智能

- 学习现象
  - 语言、文字的认知识别
  - 图像、场景、自然物体的认知识别
  - 规则
    - 下雨天要带雨伞
  - 复杂的推理、判断能力（智能）
    - 好人与坏人？
    - 好猫与坏猫？



什么是机器学习?

- 使得计算机具备和人类一样的学习能力
  - 决策
  - 推理
  - 认知
  - 识别
  - …… 等智能
- 给定数据（样本、实例）和一定的学习规则，从数据中获取知识的能力



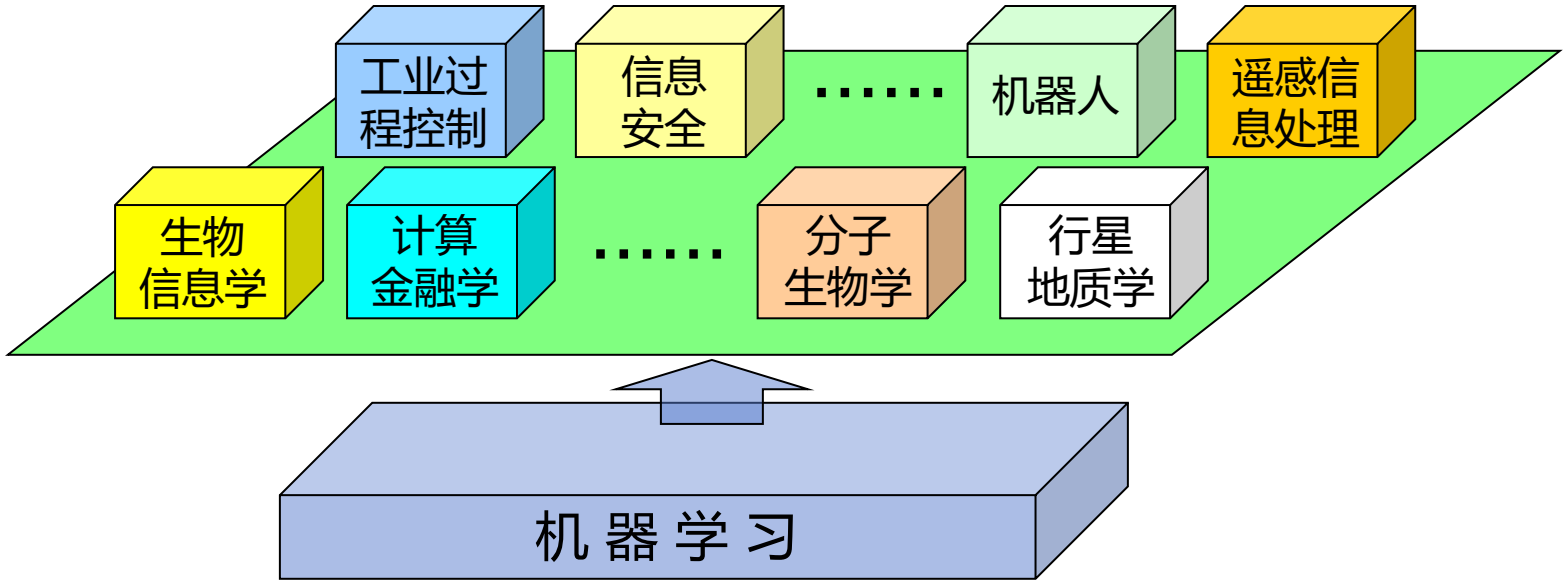
机器学习与人工智能

- 自然智慧的伟大与奥妙
  - 举例：婴儿的认知能力（声音、人脸、汽车...）
  - 重要的二个特点： 容错性，推广能力（举一反三）
- 机器智能：希望用机器实现部分智能
- 基于数据的机器学习问题（引自清华张学工教授）
  - 根据已知样本估计数据之间的依赖关系，从而对未知或无法测量的数据进行预测和判断
  - 关键：推广能力

内容

- 什么是机器学习?
- 机器学习的应用
- 机器学习的组成部分
- 机器学习简要发展历史回顾

机器学习的应用



摘自南京大学周志华教授

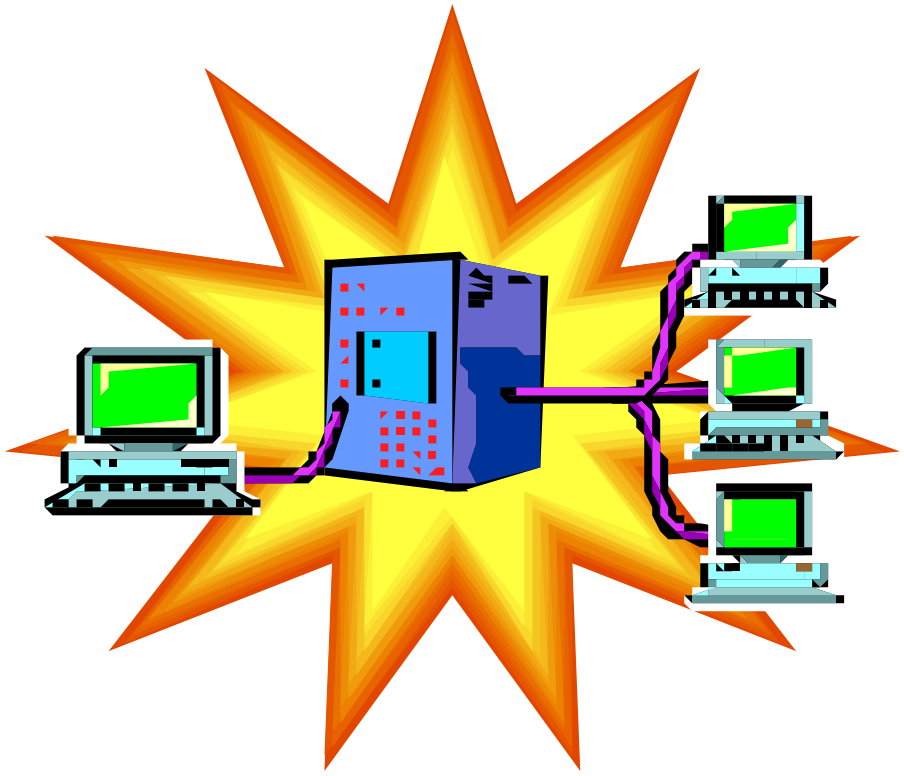
多学科交叉

- 机器学习也是一个多学科交叉的产物，它吸取了人工智能、概率统计、神经生物学、认知科学、信息论、控制论、计算复杂性理论、哲学等学科的成果。
- 机器学习在很多应用领域发挥了重要的实用价值，特别是在数据挖掘、语音识别、图像处理、机器人、车辆自动驾驶、生物信息学、信息安全、遥感信息处理、计算金融学、工业过程控制。

机器学习应用举例-网络安全

入侵检测：

是否是入侵？是何种入侵？



如何检测？

•历史数据：以往的正常访问模式及其表现、以往的入侵模式及其表现……

•对当前访问模式分类

这是一个典型的预测型机器学习问题

常用技术：

神经网络	决策树
支持向量机	k近邻
序列分析	聚类
……	……

摘自南京大学周志华教授

机器学习应用举例-搜索引擎

Google的成功，使得Internet搜索引擎成为一个新兴的产业

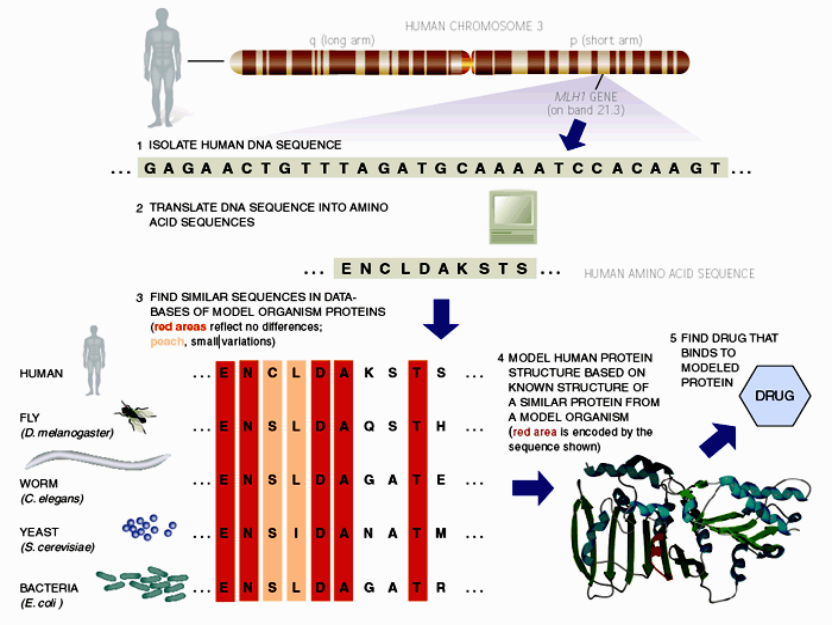
不仅有众多专营搜索引擎的公司出现（例如专门针对中文搜索的就有慧聪、百度等），而且Microsoft等巨头也开始投入巨资进行研发

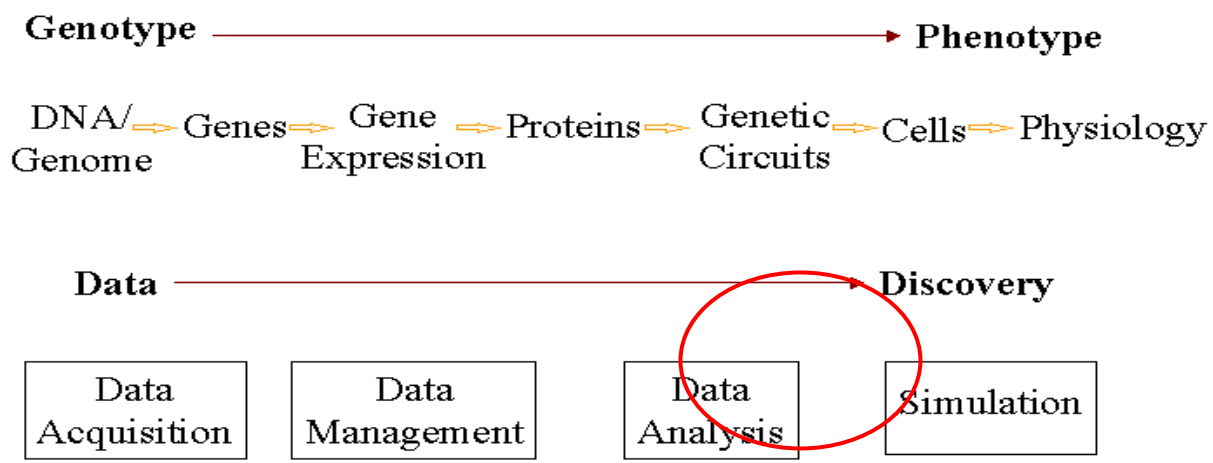
Google掘到的第一桶金，来源于其创始人Larry Page和Sergey Brin提出的PageRank算法

机器学习技术正在支撑着各类搜索引擎（尤其是贝叶斯学习技术）

摘自南京大学周志华教授

机器学习应用举例-生物信息学





常用技术：

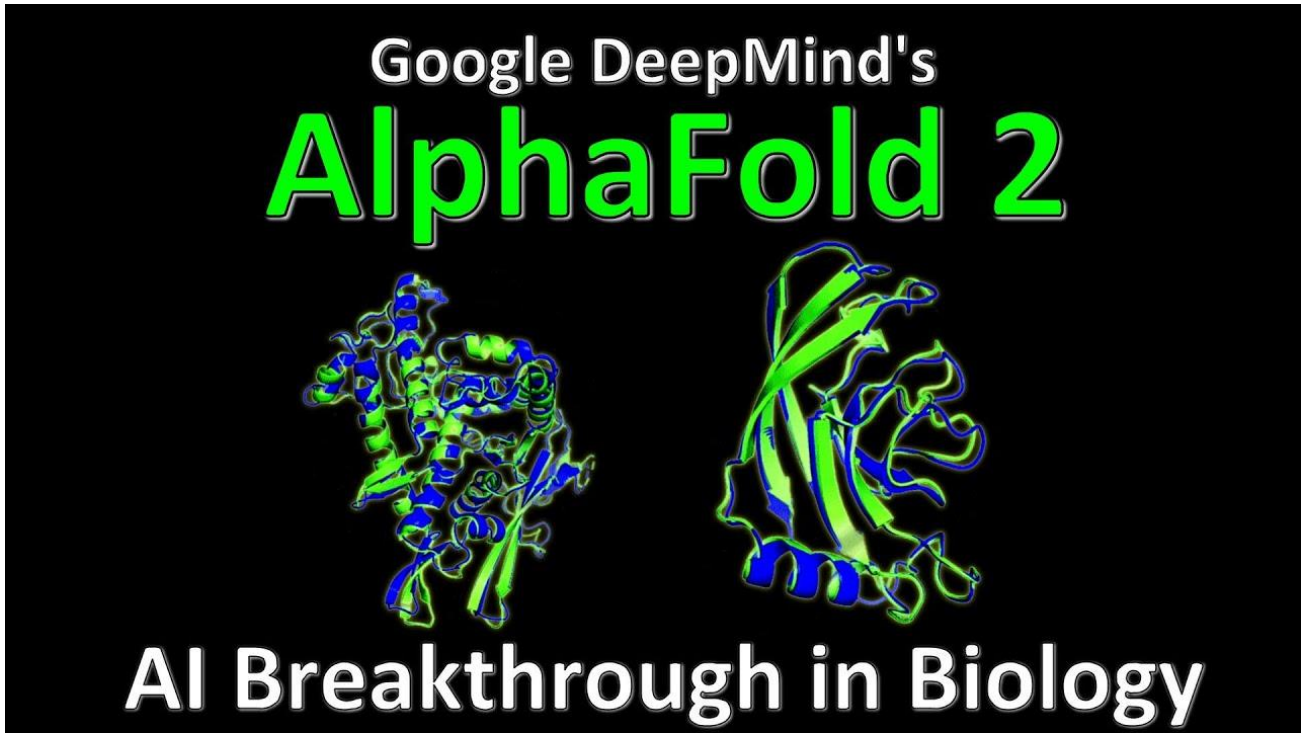
神经网络　支持向量机

隐马尔可夫模型

k近邻　　决策树

序列分析　聚类

……　……



机器学习应用举例-数据驱动控制

2.3.2 基于懒惰学习的控制方法

懒惰学习 (Lazy learnig, LL) 是一种有监督的机器学习算法 (Supervised machine learning). 1994 年 Schaal 和 Atkeson 首先将懒惰学习应用于控制[56]. 与其他有监督的机器学习算法一样, 懒惰学习的目的是: 从一个由输入输出数据对组成的训练数据集中, 找到输入与输出的映射关系.

与其他监督学习算法不同, 懒惰学习具有以下三个特点[57]: 1) 推迟对训练集中数据的处理, 直到算法被要求估计某个输入值的输出值, 该输入值通常被称作查询点 (Query piont), 数据通常只被简单地存放, 以备将来使用; 2) 在估计查询点处的输出值时, 只需要训练集中的数据提供的信息, 而不需要其他额外的信息; 3) 在得到查询点处输出值的估计值后, 只保留该估计值, 其他的中间结果都被丢弃掉. 在一些文献里也将懒惰学习称为基于实例的学习方法, 是机器学习领域的一个分支, 受到很多学者的关注, 如 1997 年期刊 *Artificial Intelligence Review* 出版了一期懒惰学习专刊[57], 对该方法进行了一次重要的总结.

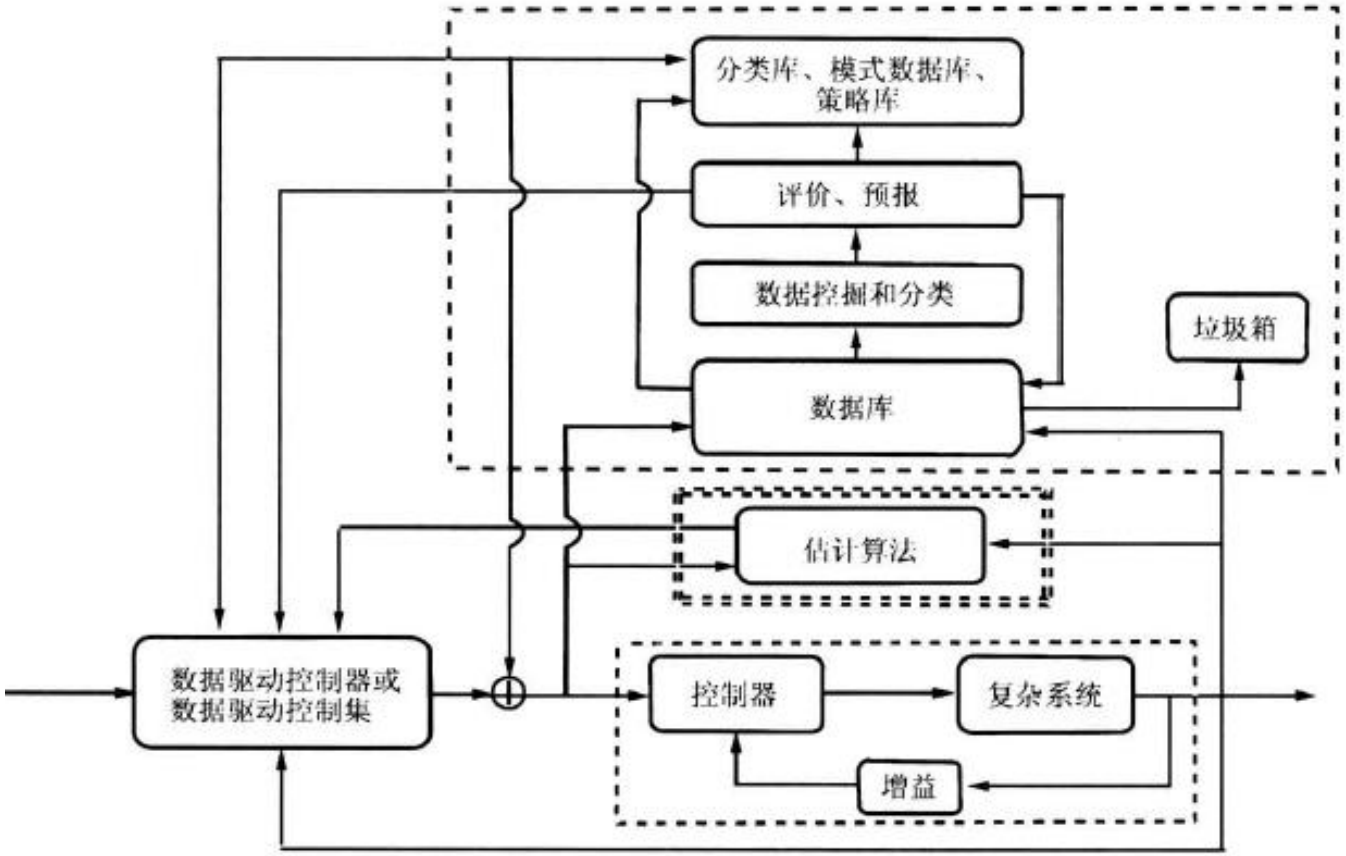


图 7 一类数据驱动控制系统结构

基于机器学习的模型预测控制 Learning-based MPC

- 1）利用ML算法（例如非线性回归）从数据中学习预测模型，以提高预测模型的精度，进而实现更优的控制；
- 2）利用ML算法（例如强化学习）从数据中直接学习并生成MPC控制策略，以及MPC控制参数标定；
- 3）利用ML算法（例如深度神经网络）从数据中学习和重构不可直接测量，而需要估计的MPC控制所需的状态量

相关学科对机器学习的影响

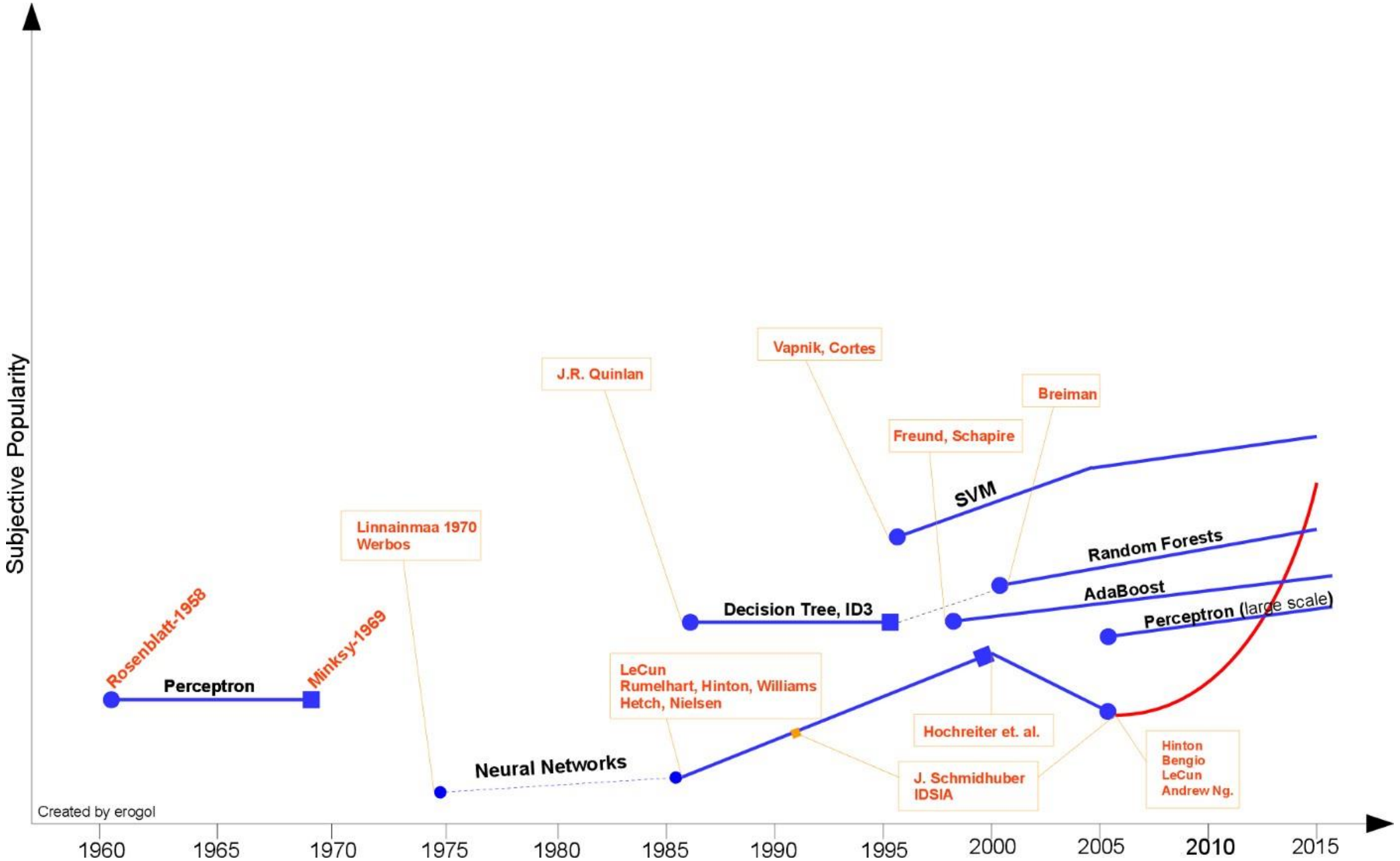
- 人工智能： 学习的概念符号表示
- Bayes 方法
- 统计学：统计学习理论 (SLT)
- 计算复杂性理论
- 控制论
- 信息论：最小描述长度
- 哲学： “Occam’s Razor原则”， “没有免费午餐”
- 心理学和神经生物学： Neural Networks（神经网络）

机器学习目前主要的一些研究领域

- 符号机器学习： Eg. 决策树，ID3，…
- 计算学习理论（统计学习理论）：PAC，SVM
- 监督学习，非监督学习，半监督学习
- 集群机器学习：Ensemble Learning， Boosting/Bagging
- 流行（Manifold）学习
- 强化学习
- 排序（Ranking）学习
- 聚类学习…

内容

- 什么是机器学习?
- 机器学习的应用和影响
- 机器学习简要发展历史回顾
- 机器学习的组成部分

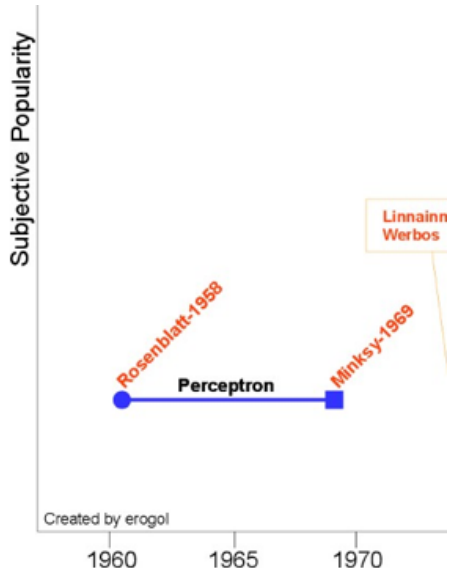


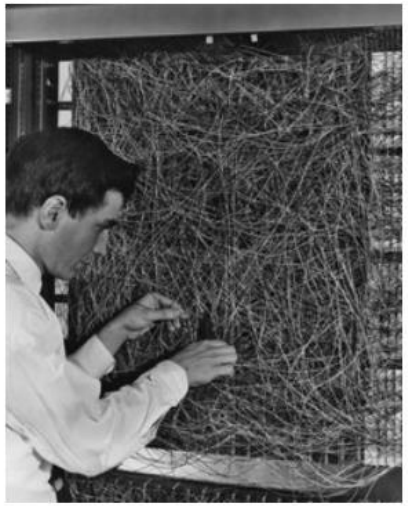
Eren Golge, Brief History of Machine Learning

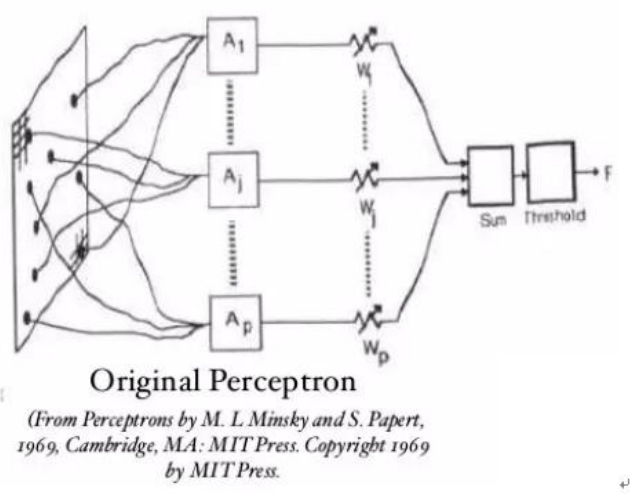
机器学习发展历史

- 1950s：神经科学的理论基础
  - James关于神经元是相互连接的发现
  - McCulloch & Pitts的神经元模型
  - Hebb 学习律（相互连接强弱度的变换规则）

- 1960s：感知器（Perceptron）时代
  - 1957年Rosenblatt首次提出
  - 1969年：《Perceptron》出版，提出著名的XOR问题







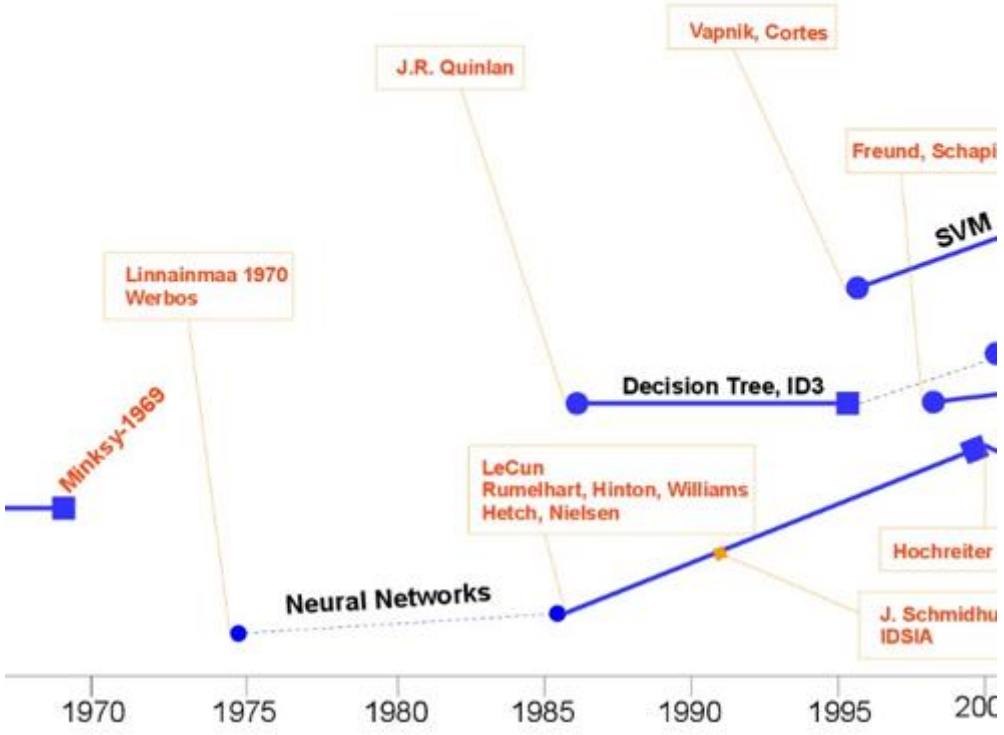
Original Perceptron

(From Perceptrons by M. L Minsky and S. Papert, 1969, Cambridge, MA: MIT Press. Copyright 1969 by MIT Press.

Rosenblatt感知器

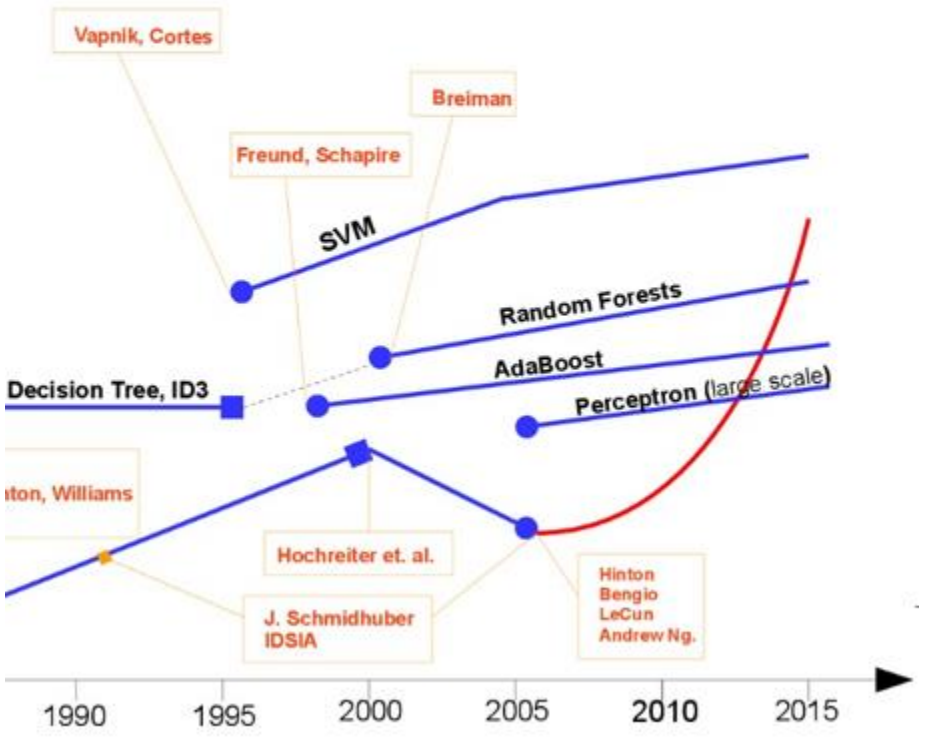
机器学习发展历史

- 1970s：符号主义，逻辑推理
- 1980s：MLP+BP算法成功解决XOR问题，从此进入神经网络时代（连接主义）
- 1960s-1970s: 统计学习理论创立
  - VC维的基本概念
  - 结构风险最小化原则
  - 概率空间的大数定律



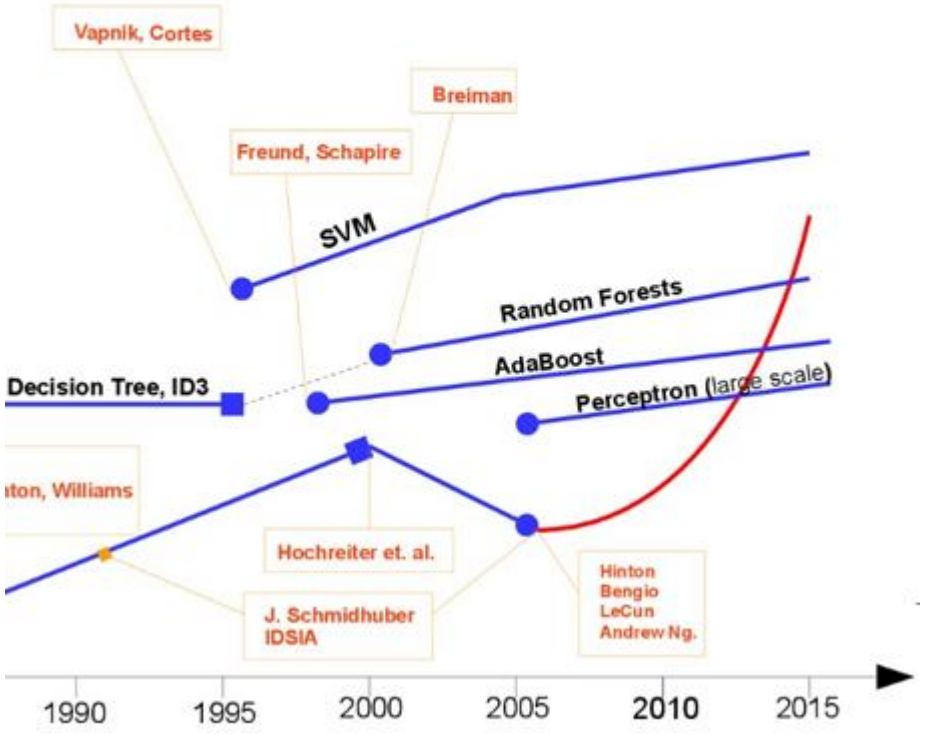
机器学习发展历史

- 1990s：统计学习理论的发展及完善
  - 典型代表：SVM （Vapnik，Bell实验室）
  - 结构风险最小化
  - 最小描述长度原则
  - 小样本问题
  - 核函数、核空间变化
  - PAC理论下的弱可学习理论的建立
  - 支持向量机
  - …



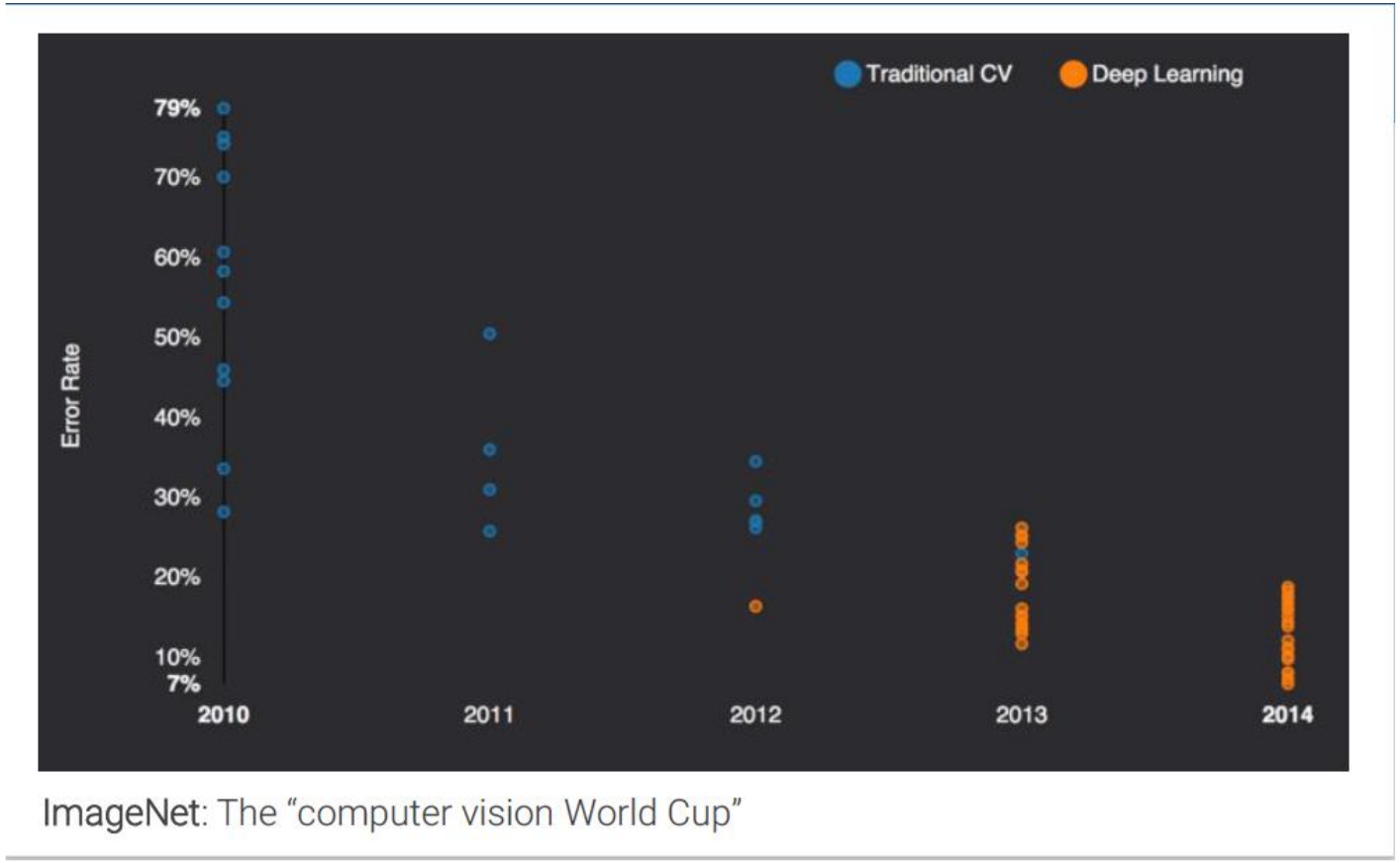
机器学习发展历史

- 2000s：各种机器学习理论及算法得以充分发展
  - 符号机器学习
  - 计算机器学习（统计学习理论，典型例子：SVM）
  - 集群机器学习（典型代表：Boosting）
  - 强化机器学习
  - 流行机器学习
  - 监督学习，非监督学习
  - 半监督学习、
  - ....



机器学习发展历史

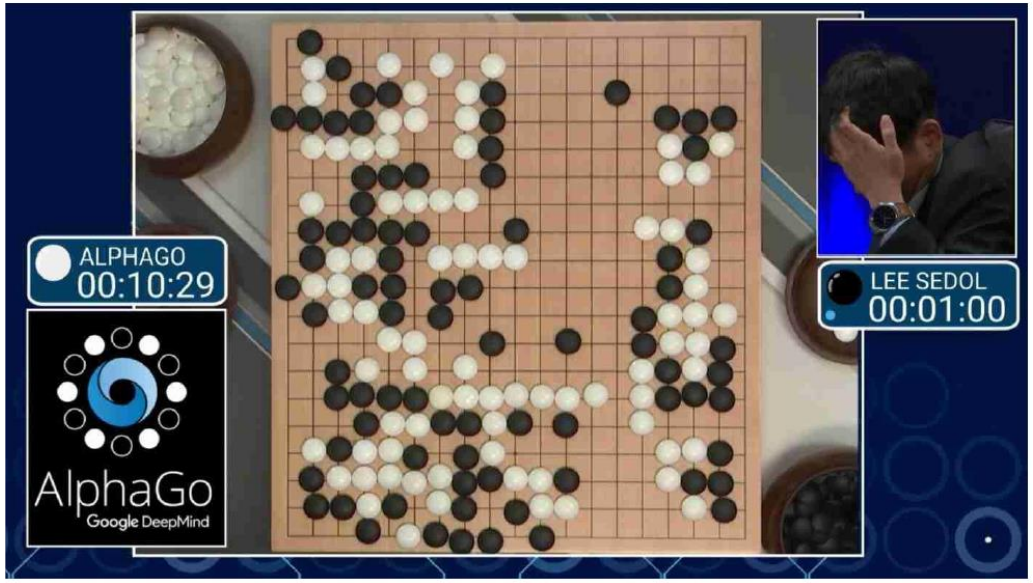
- 2012：深度学习的兴起



2012年, ImageNet中深度学习方法取得了最佳效果，显著提升了性能

机器学习发展历史

- 深度学习的标志性成就



History is made: Google's AlphaGo wins the match against Go champion Lee Sedol

2016年, 谷歌的AlphaGo战胜了李世石

未来发展趋势

- 机器实际上是一个应用驱动的学科，其根本的驱动力是：
  - “更多、更好地解决实际问题”
- 由于近20年的飞速发展，机器学习已经具备了一定的解决实际问题的能力，似乎逐渐开始成为一种基础性、透明化的“支持技术、服务技术”
  - 基础性：在众多的学科领域都得以应用（“无所不在”）
  - 透明化：用户看不见机器学习，看见的是防火墙、生物信息、搜索引擎；（“无所不在”）
  - “机器更好用了”

内容

- 什么是机器学习?
- 机器学习的应用和影响
- 机器学习简要发展历史回顾
- 机器学习的基本结构和过程

学习问题的标准描述

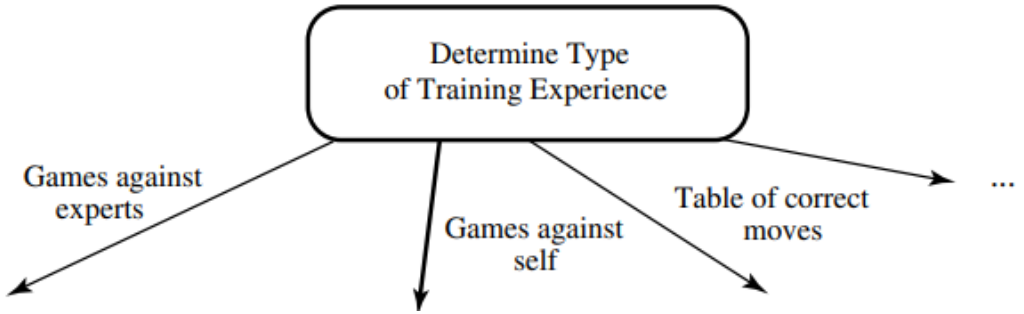
- 定义：对某类任务T和性能度量P，如果一个计算机程序在T上以P衡量的性能随着经验E而我完善，那么我们称这个计算机程序在从经验E中学习。
- 定义学习问题的三个特征：任务种类T，衡量任务提高的标准P，经验来源E
- 西洋跳棋学习问题的解释
  - E，和自己下棋
  - T，参与比赛
  - P，比赛成绩（或赢棋能力，击败对手的百分比）
- 手写识别学习问题
- 机器人驾驶学习问题

机器学习的基本设计过程

- 选择训练经验类型
- 选择目标函数
- 选择目标函数的表示
- 选择函数逼近算法

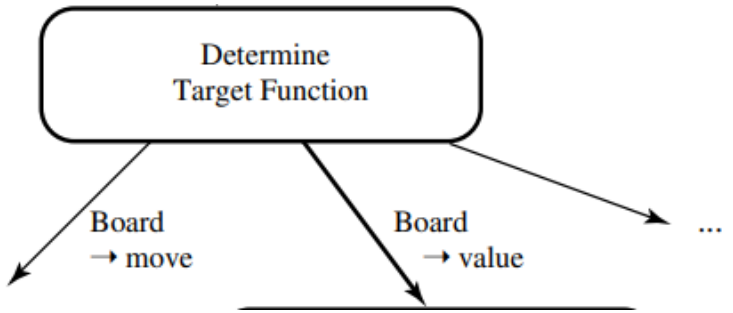
选择训练经验类型

- 直接信息还是间接信息?
- 有施教者还是自我学习?
- 训练样本的分布能多好地表示实例分布?



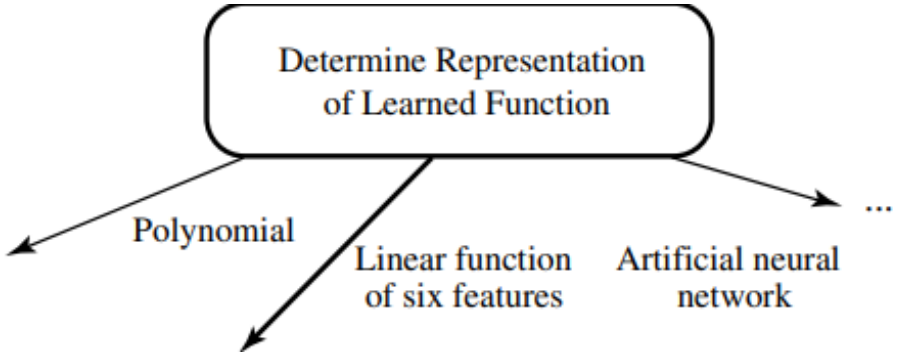
选择目标函数

- 给定棋局状态，从合法的走子集合中产生走子
- 给定棋局，赋予一个评分



选择目标函数的表示

- 规则的集合
- 神经网络
- 多项式
- 决策树
- ...



$$w_0+w_1\cdot bp(b)+w_2\cdot rp(b)+w_3\cdot bk(b)+w_4\cdot rk(b)+w_5\cdot bt(b)+w_6\cdot rt(b$$

- $bp(b)$: number of black pieces on board b
- $rp(b)$: number of red pieces on b
- $bk(b)$: number of black kings on b
- $rk(b)$: number of red kings on b
- $bt(b)$: number of red pieces threatened by black (i.e., which can be taken on black's next turn)
- $rt(b)$: number of black pieces threatened by red

多项式函数举例

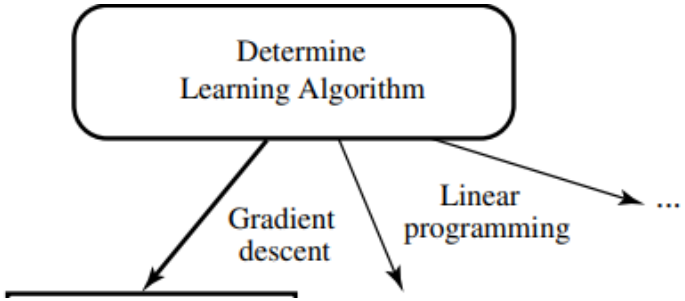
选择函数逼近算法

- 训练数据

  - $V(b)$: the true target function
  - $\hat{V}(b)$: the learned function
  - $V_{train}(b)$: the training value

  One rule for estimating training values:

  - $V_{train}(b) \leftarrow \hat{V}(Successor(b))$



- 最小均方误差（Least Mean Squares）

  LMS Weight update rule:

  Do repeatedly:

  - Select a training example b at random

    1. Compute $error(b)$:

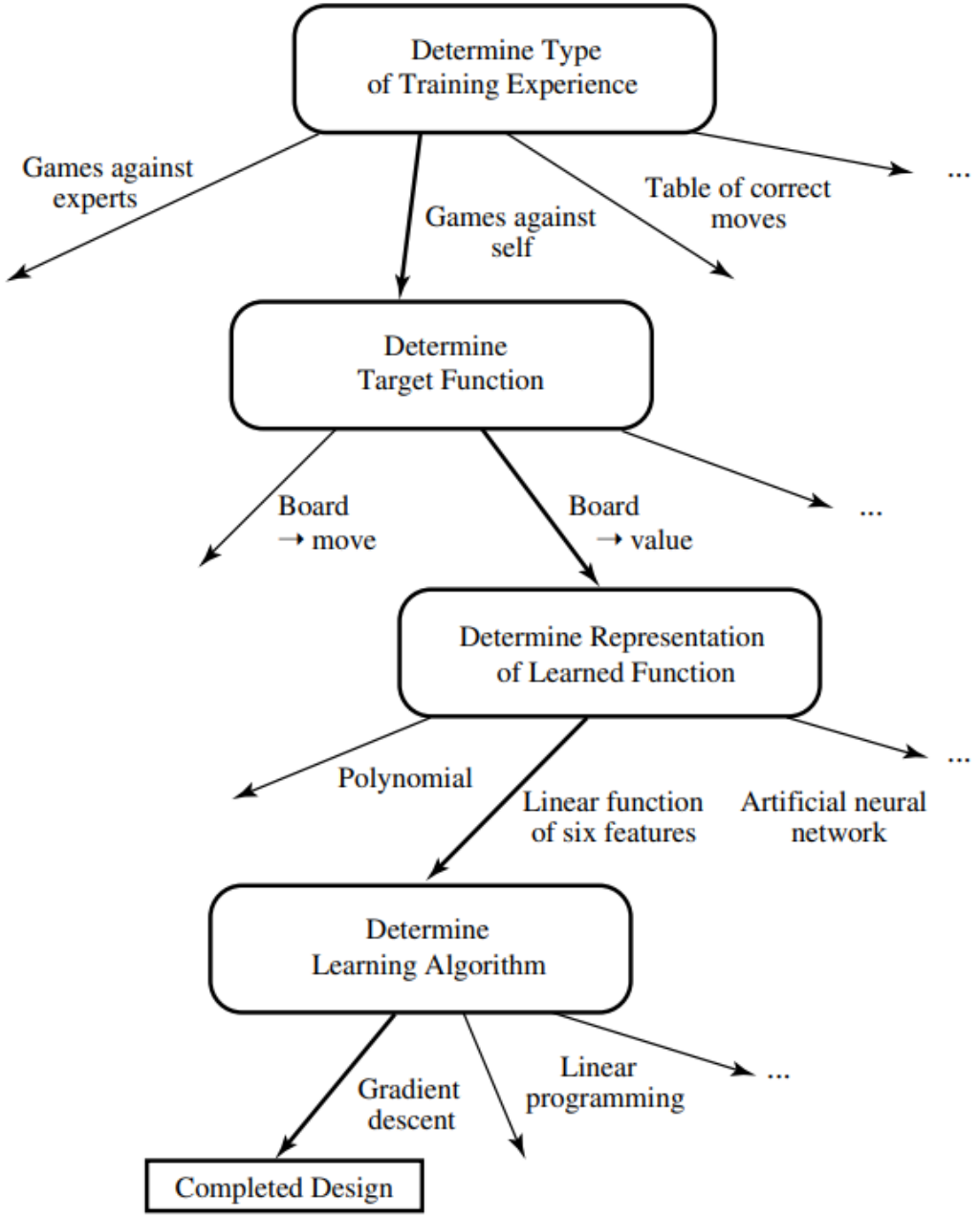
    $$error(b) = V_{train}(b) - \hat{V}(b)$$

    2. For each board feature f_i, update weight w_i:

    $$w_i \leftarrow w_i + c \cdot f_i \cdot error(b)$$

  c is some small constant, say 0.1, to moderate the rate of learning

机器学习的基本设计过程

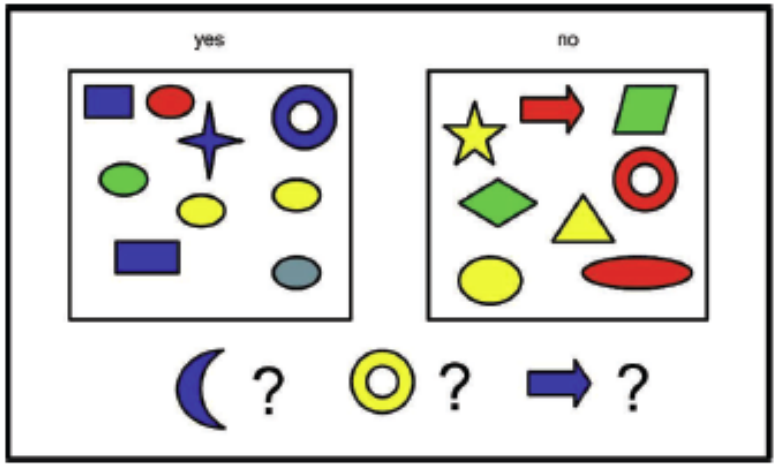


学习任务分类

- 根据训练数据是否有标记信息，分为两类
  - 监督学习（Supervised Learning）
  - 无监督学习（Unsupervised Learning）

有监督学习

- 有监督的学习方法
- 在样本标签已知的情况下，可以统计出各类训练样本不同的描述量，如其概率分布，或在特征空间分布的区域等，利用这些参数进行分类器设计，称为有监督的学习方法。



(a)

D features (attributes)

N cases

Color	Shape	Size (cm)	Label
Blue	Square	10	1
Red	Ellipse	2.4	1
Red	Ellipse	20.7	0

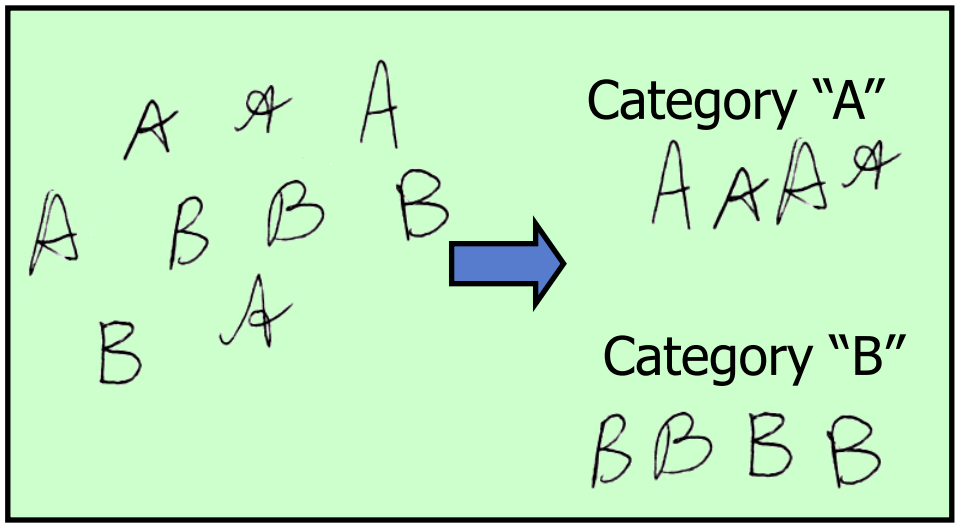
(b)

无监督学习

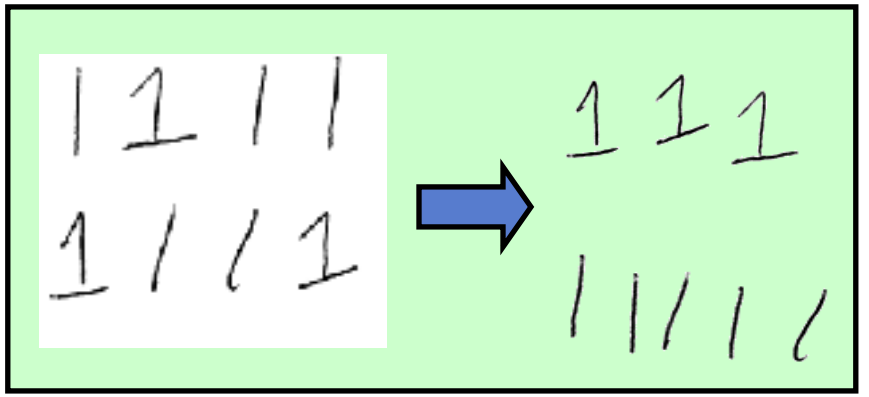
- 无监督学习
  - 然而在实际应用中，不少情况下无法预先知道样本的标签，也就是说没有训练样本
  - 因而只能从原先没有样本标签的样本集开始进行分类器设计，这就是通常说的无监督学习方法。
- 对一个具体问题来说有监督与无监督的作法是不相同的

有监督VS无监督

- 将一个模式标示为一个我们已知或者熟悉的分类
  - 分类（Classification，已知类别），监督学习
  - 聚类（Clustering，学习类别），无监督学习



Classification



识别：猫、狗

有监督学习

训练样本已经标注（labeled）

无监督学习

训练样本无标注（unlabeled）

机器学习的问题

- 存在什么样的算法能从特定的训练数据学习一般的目标函数呢？如果提供了充足的训练数据，什么样的条件下，会使特定的算法收敛到期望的函数？哪个算法对哪些问题和表示的性能最好？
- 多少训练数据是充足的？怎样找到学习到假设的置信度与训练数据的数量及提供给学习器的假设空间特性之间的一般关系？
- 学习器拥有的先验知识是怎样引导从样例进行泛化的过程的？当先验知识仅仅是近似正确时，它们会有帮助吗？
- 关于选择有效的后验训练经验，什么样的策略最好？这个策略的选择会如何影响学习问题的复杂性。
- 怎样把学习任务简化为一个或多个函数逼近问题？换一种方式，系统该试图学习哪些函数？这个过程本身能自动完成吗？
- 学习器怎样自动地改变表示法来提高表示和学习目标函数的能力？

参考期刊与会议

- 相关杂志
  - Machine Learning
  - Neural Computation
  - Journal of the American Statistical Association
  - IEEE transactions on Pattern Analysis & Machine Intelligence
- 国际会议
  - 国际机器学习会议ICML
  - 神经信息处理系统会议NeuIPS
  - 计算学习理论会议CCLT
  - 国际遗传算法会议ICGA

参考学术期刊及国际会议

一些比较公认的相关权威刊物和会议（仅供参考）

相关权威刊物（字母序）：

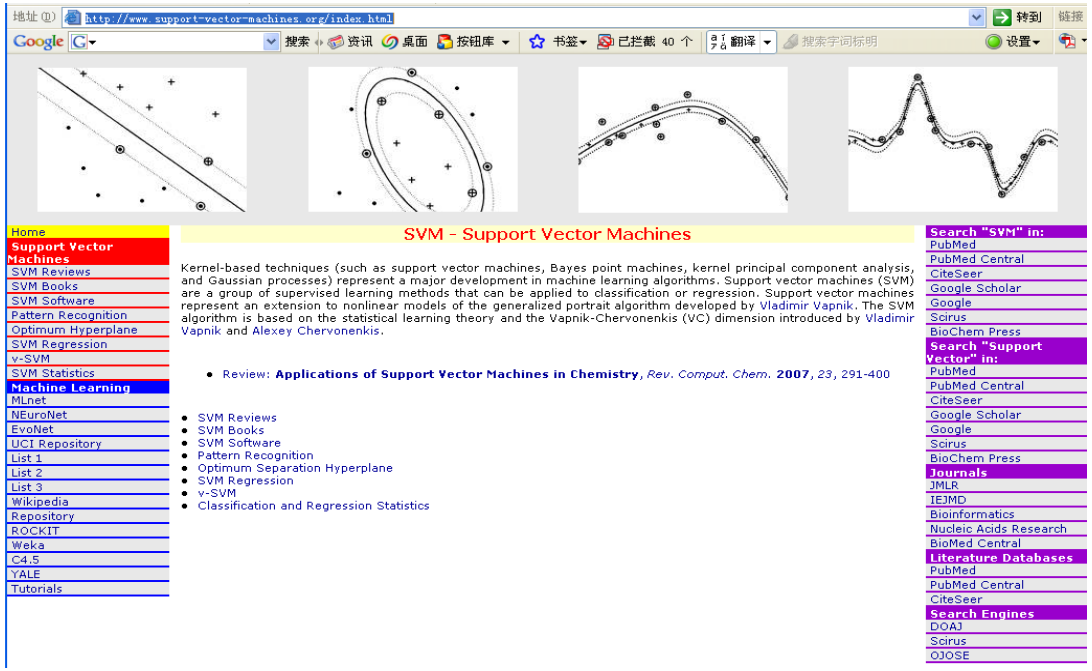
- **AIJ** (Artificial Intelligence)
- **JAIR** (Journal of Artificial Intelligence in Medicine)
- **JMLR** (Journal of Machine Learning Research)
- **MLJ** (Machine Learning)
- **NCJ** (Neural Computation)
- **TKDE** (IEEE Trans Knowledge and Data Engineering)
- **TOIS** (ACM Trans Information Systems)
- **TPAMI** (IEEE Trans Pattern Analysis and Machine Intelligence)

相关权威会议（字母序）：

- **AAAI** (National Conf. AI)
- **COLT** (ACM Ann. Conf. Learning Theory)
- **ECML** (Euro. Conf. Machine Learning)
- **ICML** (Intl. Conf. Machine Learning)
- **IJCAI** (Intl. J. Conf. AI)
- **NIPS** (Ann. Conf. Neural Information Processing Systems)
- **SIGKDD** (ACM Intl. Conf. Knowledge Discovery and Data Mining)
- **UAI** (Intl. Conf. Uncertainty in AI)

一些网络资源 (1)

- http://machine-learning.martinsewell.com
- AAAI Machine Learning Topics： www.aaai.org/AITopics/html/machine.html -
- Support Vector Machines：http://www.support-vector-machines.org/index.html



一些网络资源(2)

- http://www.cs.cmu.edu/~tom/10701_sp11/lectures.shtml
  - Machine Learning (Spring 2011) @ CMU
  - Tom Mitchell
  - Video Lecture & Slides

- Machine Learning Resources : http://home.earthlink.net/~dwaha/research/machine-learning.html

一些网络资源(3)

- Weka：Data Mining (ML) software in Java: http://www.cs.waikato.ac.nz/ml/weka/
- LibSVM -- A Library for Support Vector Machines：www.csie.ntu.edu.tw/~cjlin/libsvm
- MLC++: http://www.sgi.com/tech/mlc/ :A library of C++ classes for supervised machine learning
- UCI - Machine Learning information, software and databases: http://archive.ics.uci.edu/ml/

一些网络资源(4)

- Kernel Machines: http://www.kernel-machines.org/
- http://mloss.org/software/: Machine Learning Open Source Software
- http://www3.ntu.edu.sg/home/aswduch/ai-ml.html
- **数据挖掘研究院：** http://www.chinakdd.com/