

# 机器学习基础

## 目录

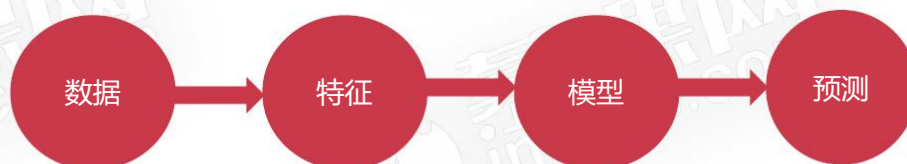
◆ 什么是机器学习

◆ 机器学习的发展历史

# 什么是机器学习

## 什么是学习

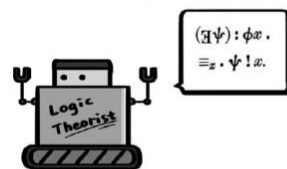
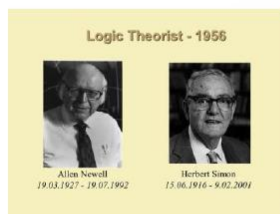
- 图灵奖与诺贝尔经济学奖得主Herbert Simon：学习是指数据系统通过执行某个过程改进性能



# 人工智能的两个重要阶段

## ◆ 逻辑推理，知识库与专家系统

1980年之前，人工建立规则、知识库



19世纪50年代诞生的第一个人工智能程序 "Logic Theorist"

## ◆ 机器学习

1980年之后，**计算机算法从数据中学习，并自动改进和优化算法，以便更好地预测和决策**



2016年，AlphaGo先后击败李世石与柯洁

# 人工智能与机器学习

## ◆ AI是计算机科学中获得图灵奖最多的方向之一，机器学习是主要实现方法

AI最近三次获得的图灵奖均属机器学习方向

2018 Hinton, LeCun, Bengio

2011 Judea Pearl (概率图模型)

2010 Leslie Valiant (PAC理论)



Judea Pearl



Leslie Valiant

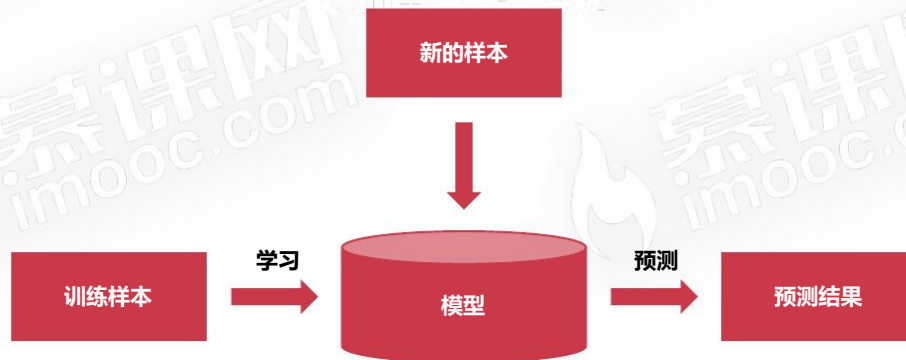


Geoff Hinton  
Yoshua Bengio  
Yann LeCun

# 什么是机器学习

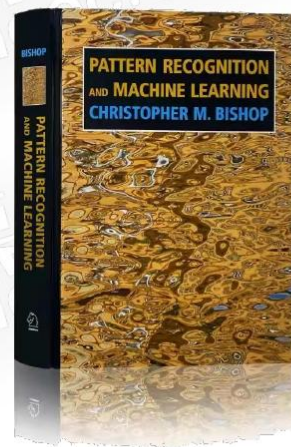
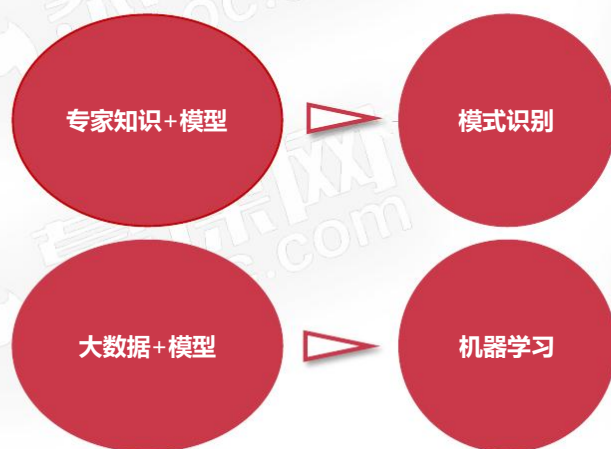
## ◆ Tom Mitchell对机器学习的定义

对于某类任务  $T$  (**机器学习问题**) 和性能度量  $P$  (**评估指标**)，如果一个计算机程序在  $T$  (**模型**) 上以  $P$  衡量的性能随着经验  $E$  (**学习数据**) 而自我完善，那么我们称这个计算机程序在从**经验  $E$  中学习**。



## 历史上的机器学习

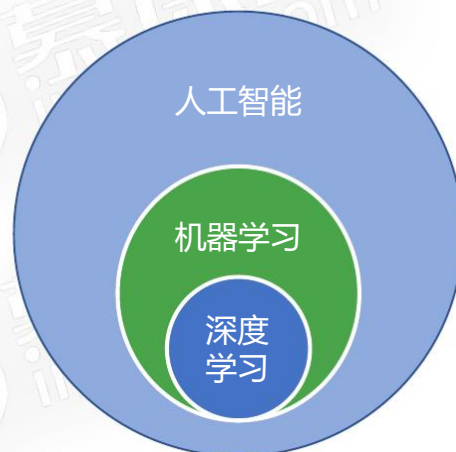
### ◆ 曾经的模式识别 (Pattern Recognition, 简称PR)，今天的机器学习 (Machine Learning, 简称ML)



# 机器学习与深度学习、人工智能的关系

## ◆ 人工智能 > 机器学习 > 深度学习

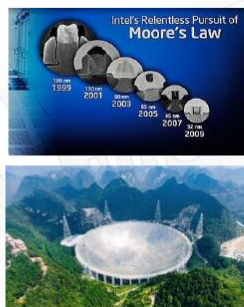
- **人工智能**是研究与开发用于模拟、延伸和扩展**人的智能**的综合性**技术科学**。
- **机器学习**属于人工智能的一个分支，是实现人工智能的方法。
- **深度学习**是一种机器学习方法，当前主要特指**深度神经网络**。



## 当前机器学习的背景

### ◆ 算法+数据+算力

- 算法成熟：神经网络、bayes方法、支撑向量机.....
- 大量可用数据：社交数据、金融数据、电商数据、科学研究数据.....
- 计算资源丰富：GPU算力增长迅速.....
- 竞争加剧：需要智能手段提供更方便的服务、发现商业规律.....



大数据 (训练样本)

计算能力 (GPU)

算法 (深度学习)



# 为什么需要机器学习

- ◆ 人工规则与专家系统的缺陷：可扩展性、通用性差，需要领域的专家知识



如何判断一张图像是不是猫？

对于一张512x512的黑白图像，每个像素取值0-255，总共的情况是

$256^{512 \times 512}$

$256^{262144}$

# 为什么需要机器学习

- ◆ 机器学习可以从样本数据中学习经验，将经验用于预测

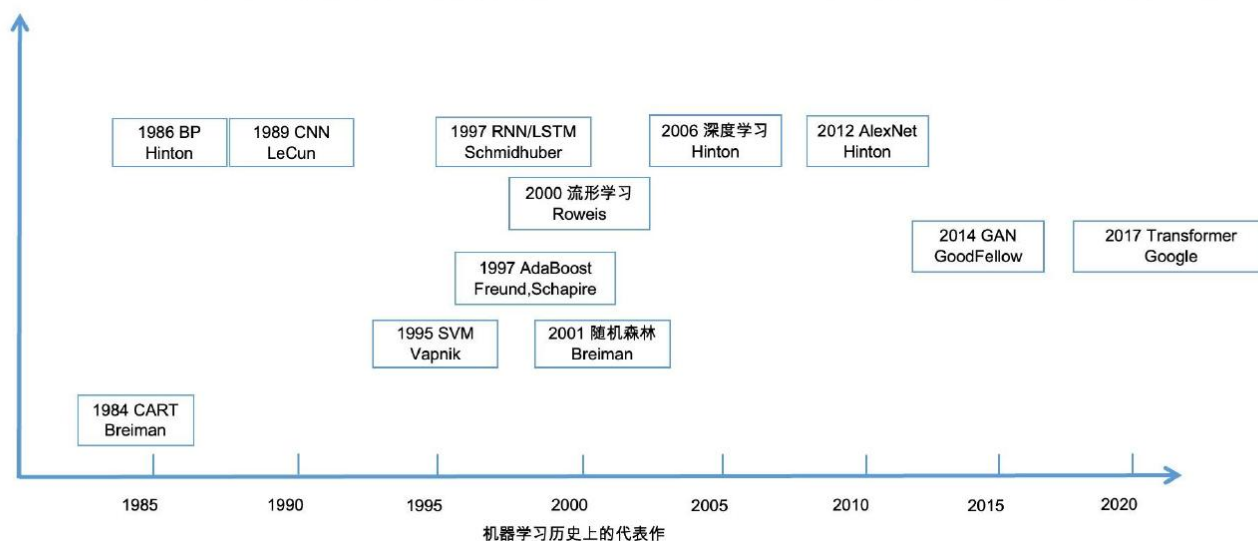


2012年，Google的一个由16000台电脑集群组成的人工神经网络通过YouTube上有关于猫的资料自行训练而能够识别出“猫”这一概念，与其把知识和经验总结好了告诉计算机，不如让计算机自己去学习知识和经验

# 机器学习的发展历史

## 机器学习的发展历史

◆ ANN, SVM, AdaBoost, CNN, Transformer等



# 机器学习的发展历史

## ◆ 机器学习的发展历程大致可以分为三个阶段

- 1980-1990年处于早期：成为独立的学科，诞生了**决策树**、**反向传播算法**，**卷积神经网络**等经典算法
- 1991-2011年处于蓬勃发展期：出现了大量方法，2次获得图灵奖。经典的算法包括支持**向量机**，**随机森林**，**Boosting**，**RNN/LSTM**，**流形学习**，**各种聚类算法**等
- 2012-至今处于深度学习时代：**深度卷积神经网络**，深度RNN，深度强化学习，生成对抗网络，图神经网络，**Transformer**，自动化机器学习/NAS等

# 机器学习相关学科

## ◆ 与其他学科的关系

学科	与机器学习关系
哲学	指导思想：众人拾柴火焰高-集成学习，中庸之道- tradeoff
信息论	很多信息论的方法也可以应用到机器学习场景；机器学习模型可解释性
心理学	研究人类认知学习的方法、模式与规律，为机器学习算法研究提供指导思想
神经生物学	深度学习的起源；类脑学习模仿人脑的工作机制
统计学	统计学习是机器学习的重要分支之一
.....	.....



**下次预告：机器学习特征**