

数据库系统概论新技术篇

大数据与机器学习

Part I：机器学习概述

卢志武

中国人民大学信息学院

2017年4月

目录

- ❖ 机器学习的基本概念
- ❖ 机器学习的发展历程
- ❖ 机器学习的前沿方向



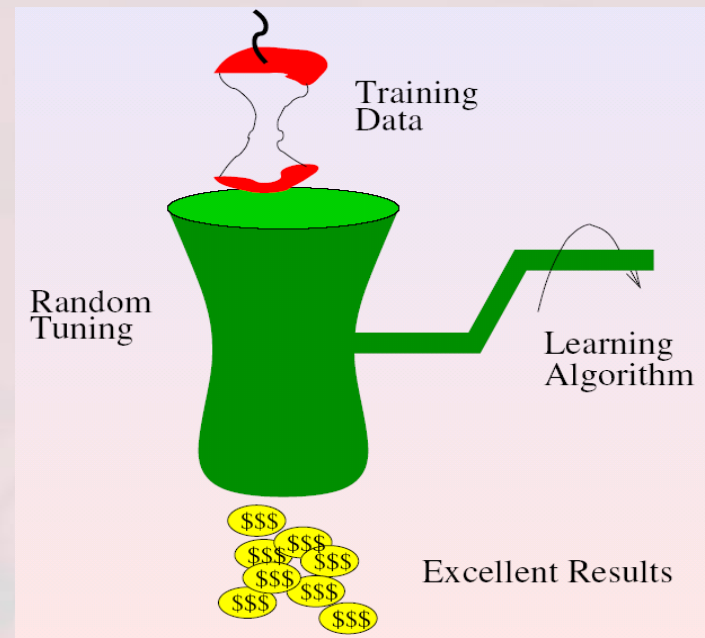
什么是机器学习？

❖ 不同的定义：

- “机器学习是一门人工智能的科学，主要研究对象是人工智能，特别是如何在**经验**学习中**改善**具体算法的性能。”
- “机器学习是对能通过**经验**自动**改进**的计算机算法的研究。”
- “机器学习是使用**数据**或以往的**经验**，以此**优化**计算机程序的性能标准。”

❖ 总结如下：

- 人工智能的重要组成部分。
- 利用经验数据改善系统自身性能。



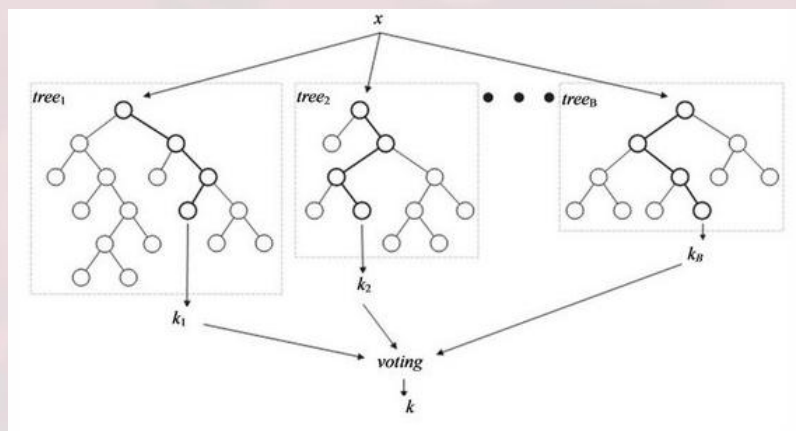
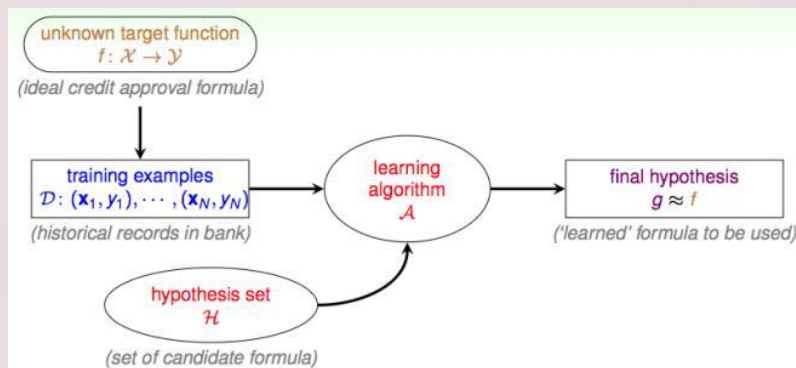
机器学习的问题与模型

❖ 三大类机器学习问题

- 有监督：分类、回归
- 无监督：概率密度估计、聚类
- 半监督：EM、Co-training

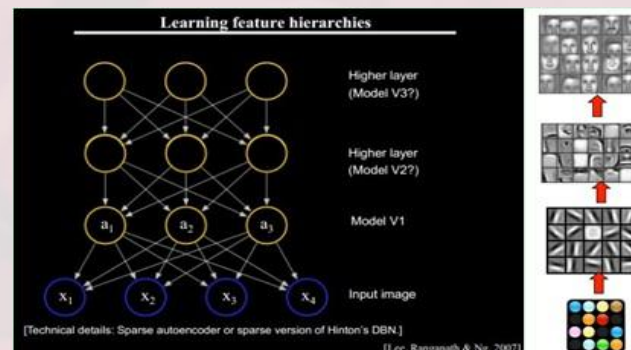
❖ 两大类机器学习模型

- 单模型：SVM、k-means
- 混合模型：bagging、boosting



深度学习与机器学习

- ❖ 深度学习是机器学习领域中一系列试图使用多重非线性变换对数据进行多层抽象的算法，**本质上是机器学习中特征学习方法的一类**。
- ❖ 至今已有多多种深度学习框架：**深度神经网络，卷积神经网络，深度信念网络**。
- ❖ 已经被应用于多个领域：**计算机视觉，语音识别，自然语言处理**。



大数据与机器学习

- ❖ 机器学习无处不在，即便我们没有专程调用它，也经常出现在大数据应用中。机器学习对大数据应用的贡献主要在于：
 - 促进数据科学家们的多产性
 - 发现一些被忽视的方案
- ❖ 上述价值来自于机器学习的核心功能：**让分析算法无需人类干预和显式程序即可对最新数据进行学习。**



目录

- ❖ 机器学习的基本概念
- ❖ 机器学习的发展历程
- ❖ 机器学习的前沿方向



机器学习的派别

❖ 机器学习的两大派别：

- 第一派别，是把机器学习看作人工智能分支的一个群体，这群人的主体是计算机科学家。
- 第二派别，是把机器学习看作“应用统计学”的一个群体，这群人的主体是统计学家。

❖ 综合来看，机器学习 = 神经科学与认知科学 + 数学 + 计算

Symbolists



Tom Mitchell



Steve Muggleton



Ross Quinlan

Bayesians



David Heckerman

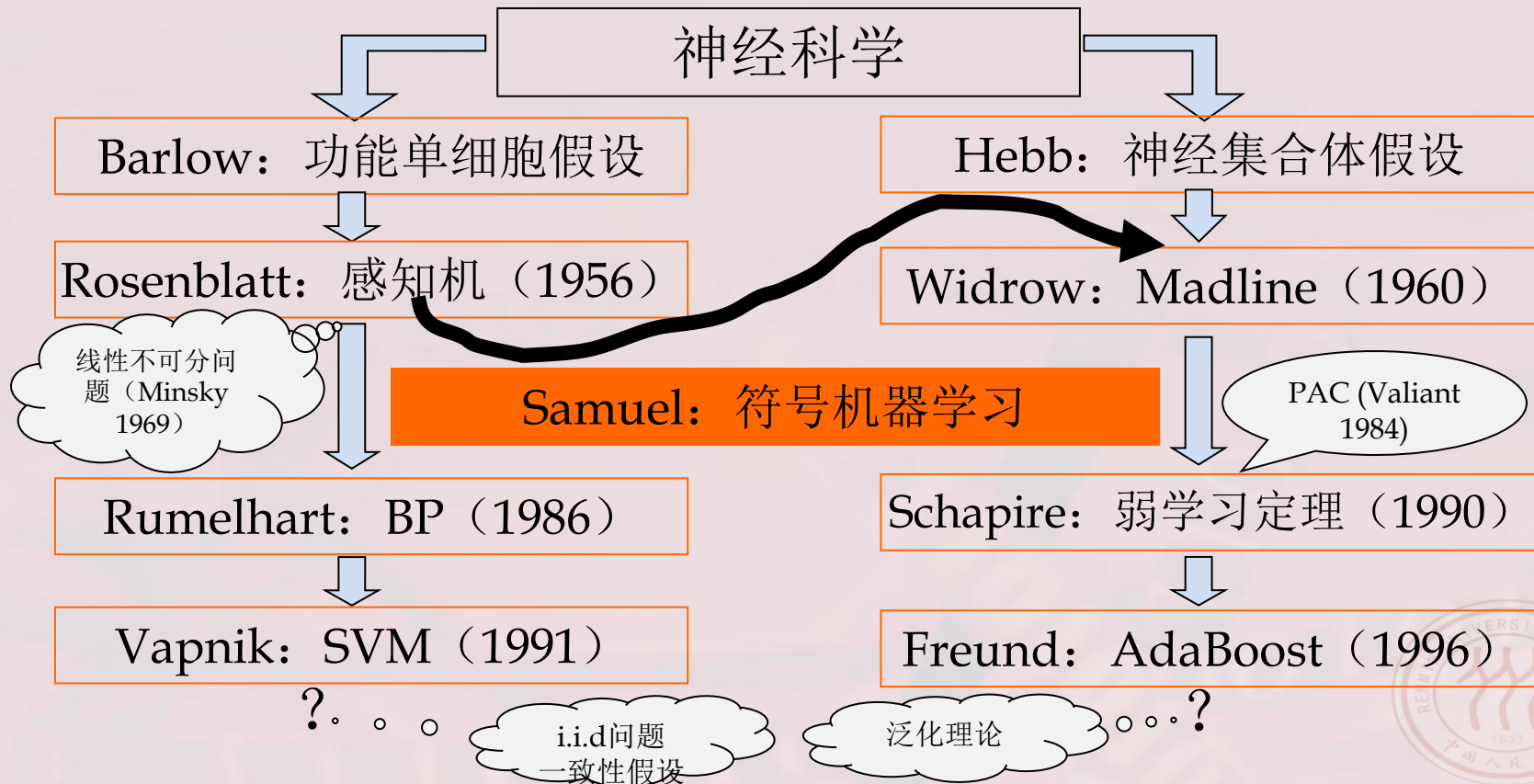


Judea Pearl



Michael Jordan

机器学习研究历程



目录

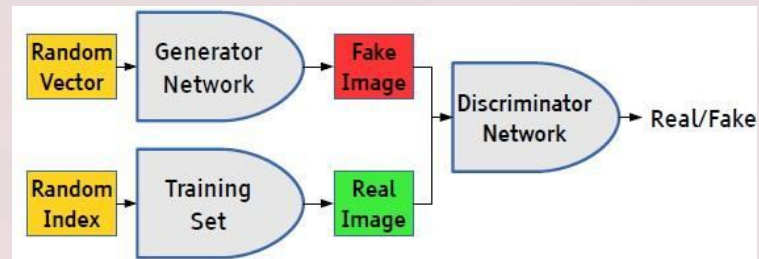
- ❖ 机器学习的基本概念
- ❖ 机器学习的发展历程
- ❖ 机器学习的前沿方向



机器学习两大热点

❖ 无监督深度学习

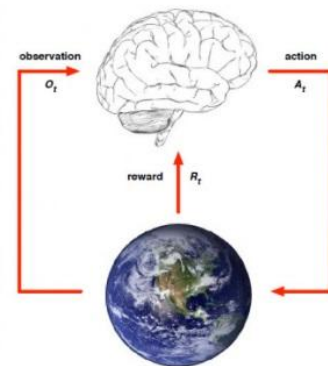
- 深度学习的原意是**特征学习**，应该是无监督的。但是，目前表现最好的深度学习算法都是监督的，在实际使用中有很多限制。
- 无监督深度学习是一种类似人类的学习方式，目前已经取得初步的研究进展，特别是**生成对抗网络 (GAN)** 在很多应用中表现出色。



❖ 深度强化学习

- 强化学习就是智能系统从环境到行为映射的学习，以使奖励信号(强化信号)函数值最大。不同于监督学习中的教师信号，环境提供的强化信号是**对产生动作的好坏作一种评价**，并没有告诉如何去产生正确的动作。
- **深度学习与强化学习的结合**，给强化学习带来了新的活力，目前深度强化学习最成功的应用例子是**AlphaGo**，已经在围棋比赛中横扫人类高手。

Agent and Environment



- At each step t the agent:
 - Executes action A_t
 - Receives observation O_t
 - Receives scalar reward R_t
- The environment:
 - Receives action A_t
 - Emits observation O_{t+1}
 - Emits scalar reward R_{t+1}
- t increments at env. step

面向大数据的机器学习

❖ 传统机器学习有如下局限性:

- 数据量小，可以一次性全部处理。
- 过分注重算法的精度，而对于算法效率关注不够。
- 假设数据独立同分布，没有考虑数据分布发生变化的情况。

❖ 针对大数据的特点，需要从**算法效率、模型假设**等方面进行机器学习的研究。这正是**Part II**的授课内容。



❖ 主讲人：卢志武

❖ Email: luzhiwu@ruc.edu.cn

❖ 电话：010-62515670

