

概念

数据挖掘

作为一种数据分析技术的统称，出现较早。那是有数据库处理能力局限等原因，强调从抽样数据出发分析数据全集

也就是data mining，是一个很宽泛的概念。字面意思就是从成吨的数据里面挖掘有用的信息。这个工作BI（商业智能）可以做，数据分析可以做，甚至市场运营也可以做。你用excel分析分析数据，发现了一些有用的信息，然后这些信息可以指导你的business，恭喜你，你已经会数据挖掘了。

机器学习

统计学习方法，是概率论与统计学的范畴；

计算机利用已有的数据(经验)，得出了某种模型(规律)，并利用此模型预测未来(结果)的一种方法。

机器学习严格来说与数据挖掘不是对等概念，仍属于数据挖掘范畴，只不过更多的基于大数据理念出发，直接在数据全集中进行分析，故而有"学习"一说

机器学习按照学习方式的维度划分，可以分为监督学习（输入数据有一个明确的标识或结果）、无监督学习、半监督学习、强化学习。当然，深度学习也包括有监督深度学习和无监督深度学习。

machine learning，是计算机科学和统计学的交叉学科，基本目标是学习一个 $x \rightarrow y$ 的函数（映射），来做分类或者回归的工作。之所以经常和数据挖掘合在一起讲是因为现在好多数据挖掘的工作是通过机器学习提供的算法工具实现的，例如广告的ctr预估，PB级别的点击日志在通过典型的机器学习流程可以得到一个预估模型，从而提高互联网广告的点击率和回报率；个性化推荐，还是通过机器学习的一些算法分析平台上的各种购买，浏览和收藏日志，得到一个推荐模型，来预测你喜欢的商品。

范围

模式识别=机器学习 数据挖掘=机器学习+数据库 统计学习近似等于机器学习 计算机视觉=图像处理+机器学习
语音识别=语音处理+机器学习 自然语言处理=文本处理+机器学习

深度学习

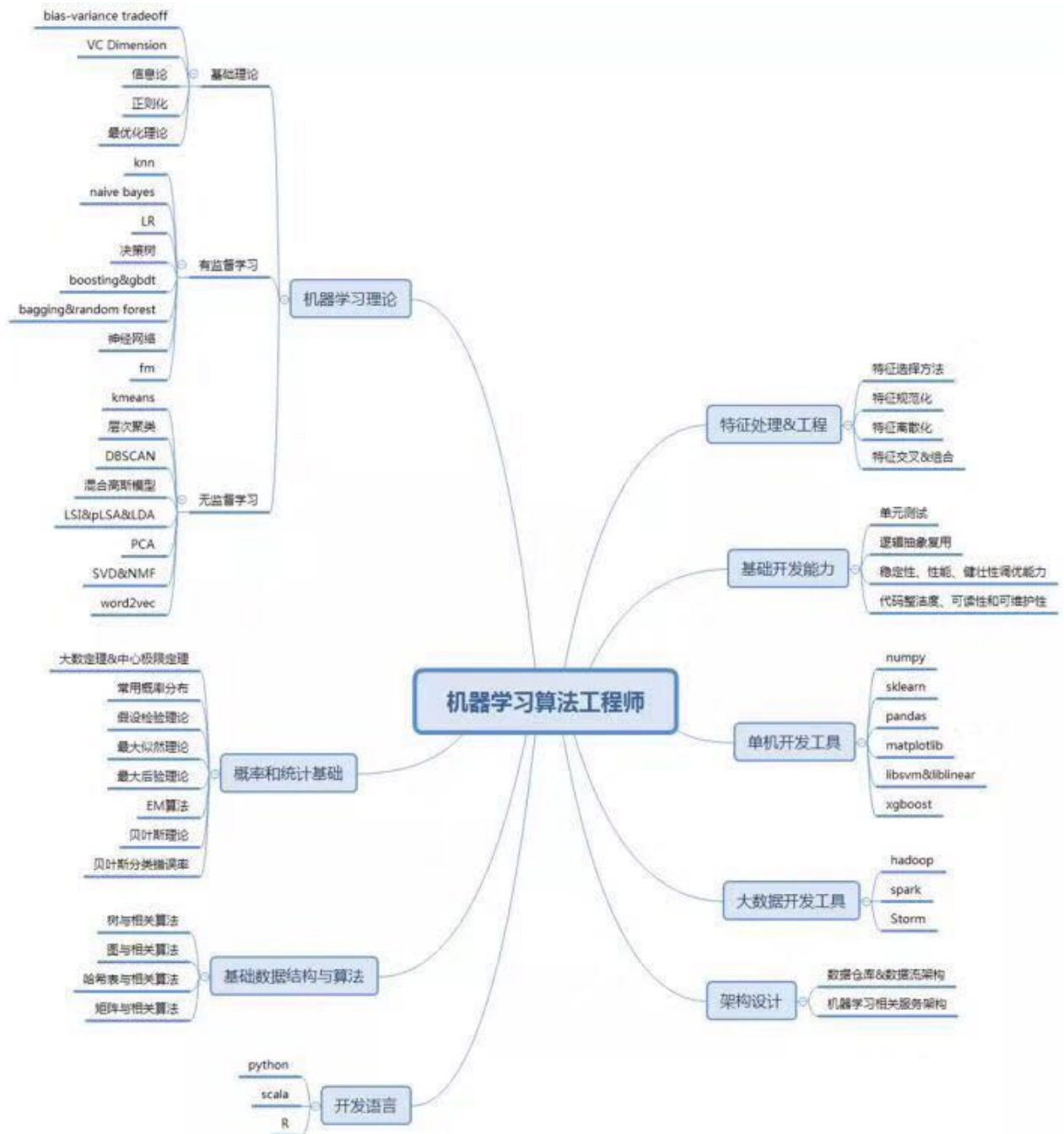
基于神经网络的，是基于神经认知科学发展的产物。

是 deep learning 的中译，原文是个形容词，用来修饰 "神经网络算法"

如果把数据挖掘认为是基于局部数据猜测全局状态的技术，而强调机器学习是直接在全局数据中习得关联规律的话，深度学习在两种场景下都有应用，但在大数据时代更加出彩

deep learning，机器学习里面现在比较火的一个topic（大坑），本身是神经网络算法的衍生，在图像，语音等富媒体的分类和识别上取得了非常好的效果，所以各大研究机构和公司都投入了大量的人力做相关的研究和开发。

导图



-
- 一个程序员能不能用2年挣到50万+,看他是否在做这件事就知道了
- 我是学Java的, 想尝试大数据和数据挖掘, 该怎么规划学习?
- 机器学习入门课程
- 机器学习该怎么入门?
- 机器学习应该准备哪些数学预备知识?
- 怎样进行大数据的入门级学习?
- 机器学习中用到哪些矩阵知识, 如果要补这些知识, 求推荐合适的书籍资料?
- 机器学习需要哪些理论知识, 哪些比较成体系的书籍可以看?
- 从入门到高阶, 读懂机器学习需要哪些数学知识 (附网盘) 机器之心
- 机器学习从入门到出家
- 如何系统入门机器学习?
- 机器学习入门导引
- 我的机器学习入门学习清单及路线
- 机器学习入门教程分享
- 机器学习入门书单

- [机器学习入门2：逻辑回归](#)
- [机器学习的入门“秘籍”](#)
- [8个入门级趣味机器学习项目（附教程和数据源）](#)
- [机器学习入门一定要避开的几大陷阱](#)
- [机器学习进阶笔记之一 | TensorFlow安装与入门](#)
- [入门机器学习到底需要多少数学知识](#)

连接

- [数据挖掘、机器学习、和人工智能\(AI\)的区别是什么](#)
- [如何向外行解释机器学习和数据挖掘](#)
- [机器学习常见算法分类汇总](#)
- [Poll的笔记](#)
- [机器学习常见算法分类汇总](#)
- [从决策树学习到贝叶斯分类算法、EM、HMM](#)
- [数据挖掘领域十大经典算法初探](#)
- [从贝叶斯方法谈到贝叶斯网络](#)
- [数据挖掘-决策树ID3分类算法的C++实现](#)
- [数据挖掘-基于贝叶斯算法及KNN算法的新sgroup18828文本分类器的JAVA实现（下）](#)
- [推荐引擎算法学习导论](#)
- [HMM学习最佳范例](#)

博客

- [结构之法 算法之道](#)
- [CodingLabs](#)
- [Coding for Dreams](#)
- [Porter算法](#)
- [漫谈Clustering之k-means](#)

书籍

- [统计自然语言处理，宗成庆编著](#)
- [机器学习，Tom M.Mitchell著](#)
- [数据挖掘导论，\[美\] Pang-Ning Tan / Michael Steinbach / Vipin Kumar 著](#)
- [支持向量机导论，\[美\] Nello Cristianini / John Shawe-Taylor](#)

其他

- [1 机器学习的应用----大数据 机器学习的父类----人工智能 机器学习的子类----深度学习](#)
- [2](#)

数据挖掘概念火爆的时候，数据仓库正当家 机器学习是随着大数据概念为人熟知 深度学习是神经网络算法的新近发展，随着机器学习概念被广知

- [3](#)

“用机器学习的方法来进行数据挖掘。”机器学习是一种方法；数据挖掘是一件事情；还有一个相似的概念就是模式识别，这也是一件事情。

- 4

数据挖掘（Data Mining）是从海量数据中“挖掘”隐藏信息；数据挖掘采用的一个重要方法，是机器学习（Machine Learning），即通过程序积累经验，但机器学习是另一门学科，并不从属于数据挖掘，二者相辅相成；而深度学习（Deep Learning）是机器学习的一个子集，就是用复杂、庞大的神经网络进行机器学习。

数据挖掘，顾名思义就是从海量数据中“挖掘”隐藏信息，按照教科书的说法，这里的数据是“大量的、不完全的、有噪声的、模糊的、随机的实际应用数据”，信息指的是“隐含的、规律性的、人们事先未知的、但又是潜在有用的并且最终可理解的信息和知识”。在商业环境中，企业希望让存放在数据库中的数据能“说话”，支持决策。所以，数据挖掘更偏向应用。为了做好数据挖掘，企业又要建立数据仓库。

机器学习，就是利用计算机、概率论、统计学等知识，通过给计算机程序输入数据，让计算机学会新知识，是实现人工智能的途径，但这种学习不会让机器产生意识。机器学习的过程，就是通过训练数据寻找目标函数。数据质量会影响机器学习精度，所以数据预处理非常重要。

深度学习是机器学习的一种，现在深度学习比较火爆。在传统机器学习中，手工设计特征对学习效果很重要，但是特征工程非常繁琐。而深度学习能够从大数据中自动学习特征，这也是深度学习在大数据时代受欢迎的一大原因。

传统数据挖掘主要针对相对少量、高质量的样本数据，机器学习更多的是针对海量、混杂的大数据。但机器学习并不一定要全局数据，只是在大数据时代，堆数据、堆机器的方法在工业界成本低而见效快，被广泛采用。学术界当前有很多研究在结合小数据学习与大数据学习，比如香港科技大学杨强教授的迁移学习。