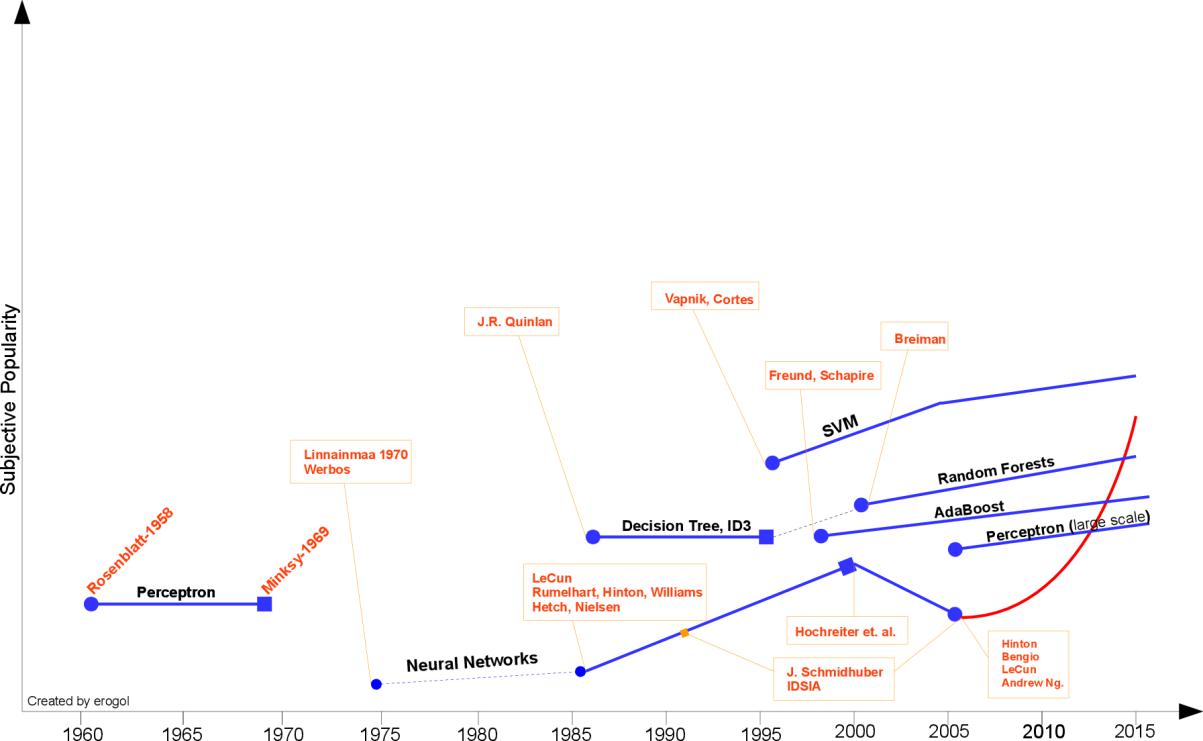
1. 深度学习发展历史



1.人工智能的萌芽

作为人工智能研究的核心内容和方向，机器学习人工智能研究阶段发展到一定程度时的必然产物。在1950年图灵（被称为“人工智能之父”）提出图灵测试的文章中，就已经提到了机器学习的可能性。从 20 世纪50 年代到 70 年代初，人工智能研究处于“推理期”，人们认为只要给机器赋予逻辑推理能力，机器就能具有智能。这一时期的代表有“逻辑推理家”程序，由 Allen Newell, Herbert A. Simon和Cliff Shaw所编写于1955年。这是第一个“人工智能程序”，在怀特海德和罗素的《数学原理》中证明前52个定理中的38个，其他的一些也相继被证明，其中一些定理甚至比罗素和怀特海证明得更巧妙。

2.机器学习的出现

1949 年，Donald Hebbian 提出的赫布理论——解释了学习过程中大脑神经元所发生的变化，标志着机器学习领域迈出的第一步。概括来讲，赫布理论研究的是循环神经网络（RNN）中各节点之间的关联性。而里面提到的 RNN 具有把相似神经网络连接在一起的特征，并起到类似于记忆的作用。Hebbian是如此描述赫布理论的：

我们可以假定，反射活动的持续与重复（或称“痕迹”）会引起神经元稳定性的持久性提升……当神经元 A 的轴突与神经元 B 之间距离非常近，且 A 对 B 进行重复持续的刺激时，这两个神经元或者它们的其中之一便会发生某些生长过程或代谢的变化，从而使得 A 对刺激 B 的效率得到提高。[[1]](#endnote-0)

机器学习第一次进入大众的主流视野是在1952年，Arthur Samuel（被誉为“机器学习之父”）设计了一款可以学习的西洋跳棋程序。这是世界上第一个可以自主学习的程序。它能通过观察棋子的走位来构建新的模型，并用其提高自己的下棋技巧。

Arthur Samuel对“机器学习”的定义是：不需要确定性编程就可以赋予机器某项技能的研究领域。“机器学习”由此产生。

3.基于神经网络的“连接主义”的出现

20世纪50年代开始的机器学习相关的研究工作，主要集中在基于神经网络的“连接主义”学习方面，代表性工作主要有1957 年F. Rosenblatt 提出的感知机模型，这是第二个有着神经系统科学背景的机器学习模型。Rosenblatt对于感知器算法的解释是：

感知器算法的作用是，在不用深入理解只对一些特定生物有机体有效的未知条件的前提下，说明了通用智能系统一些基础特点。[[2]](#endnote-1)

1969年，异或（XOR）问题的提出，浇灭了感知机的热度。机器学习的发展在此之后一度停滞。尽管 1970 年时，反向模式自动微积分算法就已经首次完整的出现（反向传播算法 BP 的雏形），但在当时并没有引起重视。直到 Werbos于 1981 年提出将反向传播算法应用于神经网络以建立多层感知器后，神经网络的发展才得以提速。接着在 1985 到 1986 两年间，多位神经网络学者也相继提出了使用反向传播算法来训练多层感知器的相关想法。

在1986 年, Rumelhart 等人的BP 算法解决了XOR 问题, 沉寂近二十年的感知机研究方向重新获得认可,人们自此重新开始关注这个研究方向。

1986年，J.R.Quinlan提出了另一个同样著名的ML算法——决策树算法（ID3），决策树作为一个预测模型，代表的是对象属性与对象值之间的一种映射关系，而且紧随其后涌现出了很多类似或者改进算法，如ID4，回归树，CART等。

ID3算法是一种贪心算法，用来构造决策树。ID3算法起源于概念学习系统（CLS），以信息熵的下降速度为选取测试属性的标准，即在每个节点选取还尚未被用来划分的具有最高信息增益的属性作为划分标准，然后继续这个过程，直到生成的决策树能完美分类训练样例。

4.基于逻辑表示的“符号主义”的蓬勃发展

在 20 世纪 6、70 年代，多种学习技术得到了初步发展，例如以决策理论为基础的统计学习技术以及强化学习技术等，代表性工作主要有 A.L. Samuel 的跳棋程序以及 N.J. Nilson 的“学习机器”等。20 多年后红极一时的统计学习理论的一些重要结果也是在这个时期取得的。在这一时期，基于逻辑或图结构表示的符号学习技术也开始出现，代表性工作有 P. Winston的“结构学习系统”、R.S. Michalski等人的“基于逻辑的归纳学习系统”、E.B. Hunt 等人的“概念学习系统”等。

5.统计学习理论的奠基

尽管以Open 问题驱动的BP 算法研究大大推动了感知机研究方向的发展, 然而, 近十年计算机科学与技术的快速发展,使得人们获得数据的能力大大提高, BP 这类算法已不能完全适应这种需求, 同时,Minsky 的算法设计原则愈显重要。Vapnik的“概率近似正确”学习的考虑在机器学习的发展中扮演了一个重要的角色。1984 年,Vapnik 提出了机器学习的一个重要考虑, 他建议评价机器学习算法应该以“概率近似正确(PAC)”为基础,而不是以传统模式识别理论中以概率为1 成立为基础,由此, 他引入了类似在数学分析中的ε-δ语言来描述PAC, 这个考虑对近代机器学习研究产生了重要的影响。首先, 统计机器学习理论中泛化不等式的推导均以这个假设为基础;其次, 基于这个考虑的“弱可学习理论”,为研究基于Hebb 路线的学习算法设计奠定了理论基础, 并产生被广泛应用的集群机器学习理念。

Vapnik 的研究主要涉及机器学习中两个相互关联的问题, 泛化问题与表示问题。前者包含两个方面的内容: 其一, 有限样本集合的统计理论; 其二, 概率近似正确的泛化描述。而后者则主要集中在核函数, 由此, 将算法设计建立在线性优化理论之上。

自1992年开始,Vapnik 将有限样本统计理论介绍给全世界, 并出版了统计机器学习理论的著作尽管这部著作更多地是从科学、哲学上讨论了机器学习的诸多问题, 但是, 其暗示的算法设计思想对以后机器学习算法研究产生了重要的影响。

1995 年 Vapnik 和 Cortes[10]提出的支持向量机实现了机器学习的重大突破。这种算法不但有坚实的理论基础，还有出色的实验结果。从那之后，机器学习这一领域便分成了两大流派，神经网络及支持向量机。2000年左右，随着核方法的提出，SVM大占上风，在很多领域上都超过了NN模型。除此之外，SVM还发展了一系列的针对NN模型的基础理论，包括凸优化、泛化间隔理论和核方法。可以说，在这个时段，SVM的发展无论是理论还是实践都占尽天时地利，因而发展速度极快。

6.深度学习

2006年，Hinton和他的学生在《Nature》上发表了一片文章，同一年另外两个团队也实现了深度学习，从此开启了深度学习（Deep Learning）阶段，掀起了深度神经网络即深度学习的浪潮。

Hinton在2006年提出的深度置信网络（Deep Belief Network，DBN）开启了深度学习新纪元，他提出了用神经网络给数据降维，就是所谓的稀疏化编码或者自动编码。他提出了一个新的神经网络模型叫做深度置信网络（Deep Belief Network，DBN），该网络可以看作是限制玻尔兹曼机像垒积木一样一层一层垒起来的。

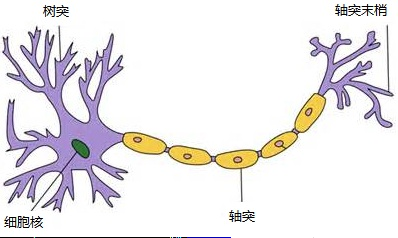
1. 人工智能，机器学习，深度学习有什么区别和联系

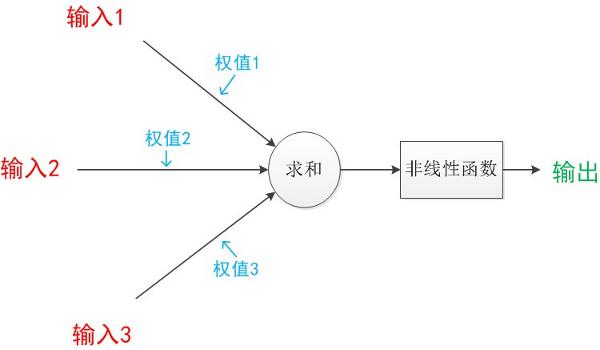


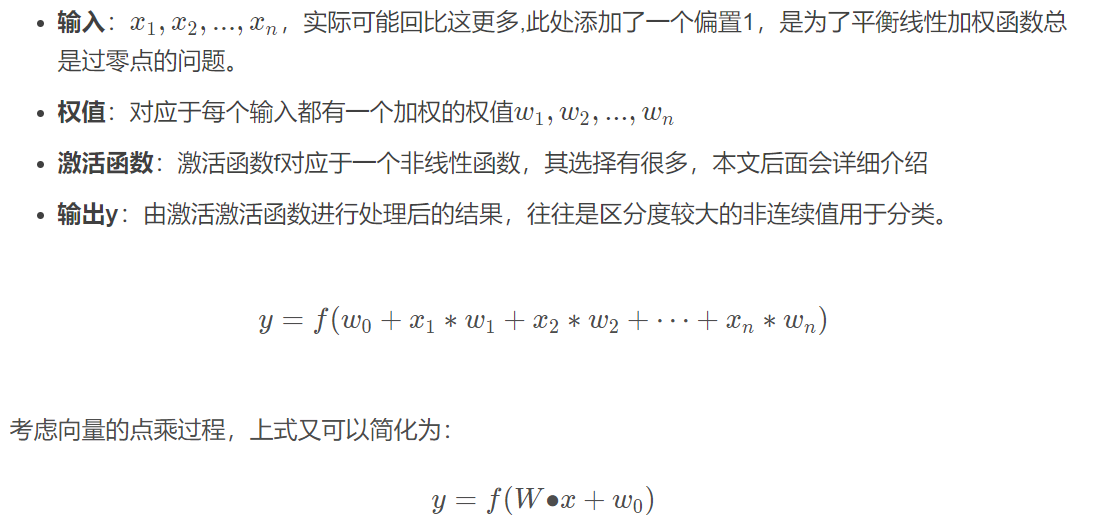
人工智能是最早出现的，也是最大、最外侧的同心圆；其次是机器学习，稍晚一点；最内侧，是深度学习，当今人工智能大爆炸的核心驱动。

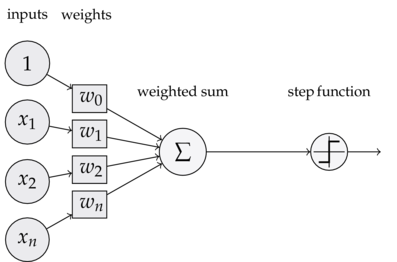
1. 神经元，单层感知机，多层感知机

神经元：一个神经元通常具有多个树突，主要用来接受传入信息；而轴突只有一条，轴突尾端有许多轴突末梢可以给其他多个神经元传递信息。轴突末梢跟其他神经元的树突产生连接，从而传递信号。这个连接的位置在生物学上叫做“突触”。突触之间的交流通过神经递质实现。

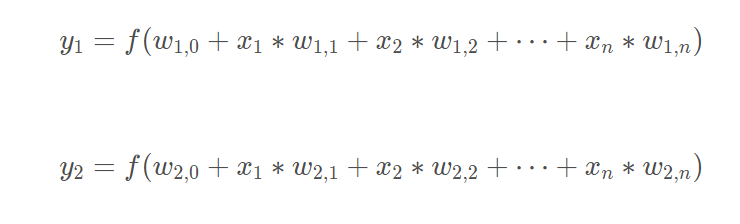


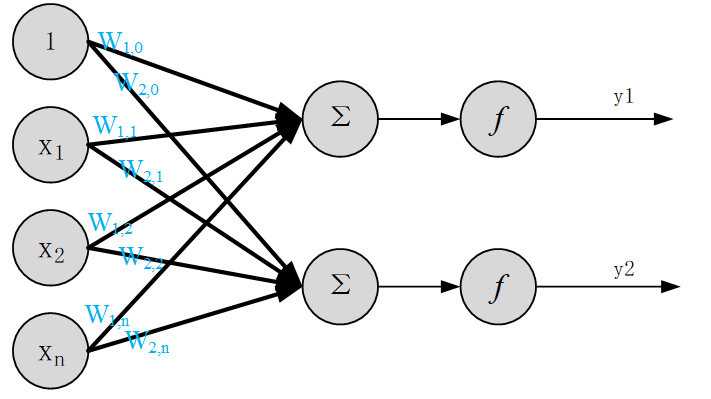


单层感知机：



多层感知机：

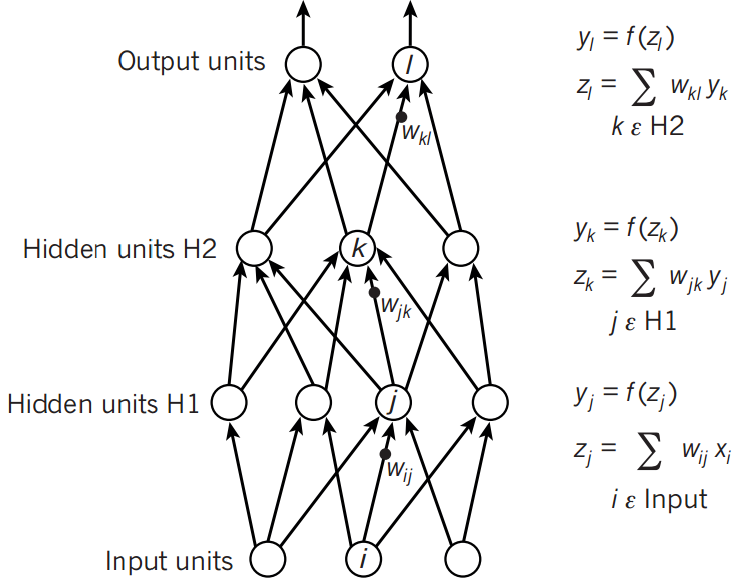




1. 什么是前向传播

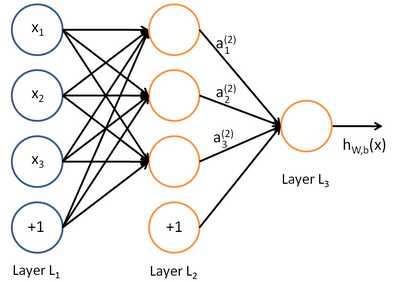
所谓前向传播(Forward propagation)，指的就是:数据从输入层开始，依次经过隐藏层(如果有)最终到达输出层的过程。

其中，数据每经过一层传播，其节点输出的值所代表的信息层次就越高阶和概括。节点中输出的值是通过与其相连的前一层中所有的节点输出值的加权求和处理后的结果。

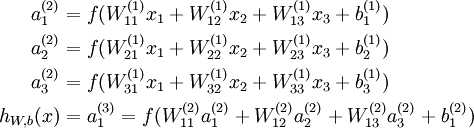


1. 什么是反向传播

其实反向传播的核心就是复合函数的链式求导法则。我们知道神经元就是一个函数映射，那么多层神经网络，本质上就是一个多层的复合函数。



其对应表达式为：



1. Hebb D. O., The organization of behaviour.New York: Wiley & Sons. [↑](#endnote-ref-0)
2. Rosenblatt, Frank. “The perceptron: a probabilistic model for information storage and organization in the brain.” Psychological review65.6 (1958): 386 [↑](#endnote-ref-1)