1. 深度学习基础知识
2. 深度学习发展历史

从 20 世纪 50 年代研究机器学习以来，不同时期的研究途径和目标并不相同，可以划分为四个阶段。

第一阶段是 20 世纪 50 年代中叶到 60 年代中叶，这个时期主要研究“有无知识的学习”。这类方法主要是研究系统的执行能力。这个时期，主要通过对机器的环境及其相应性能参数的改变来检测系统所反馈的数据，就好比给系统一个程序，通过改变它们的自由空间作用，系统将会受到程序的影响而改变自身的组织，最后这个系统将会选择一个最优的环境生存。在这个时期最具有代表性的研究就是 Samuet 的下棋程序。但这种机器学习的方法还远远不能满足人类的需要。

第二阶段从 20 世纪 60 年代中叶到 70 年代中叶，这个时期主要研究将各个领域的知识植入到系统里，在本阶段的目的是通过机器模拟人类学习的过程。同时还采用了图结构及其逻辑结构方面的知识进行系统描述，在这一研究阶段，主要是用各种符号来表示机器语言，研究人员在进行实验时意识到学习是一个长期的过程，从这种系统环境中无法学到更加深入的知识，因此研究人员将各专家学者的知识加入到系统里，经过实践证明这种方法取得了一定的成效。在这一阶段具有代表性的工作有 Hayes-Roth 和 Winson 的对结构学习系统方法。

第三阶段从 20 世纪 70 年代中叶到 80 年代中叶，称为复兴时期。在此期间，人们从学习单个概念扩展到学习多个概念，探索不同的学习策略和学习方法，且在本阶段已开始把学习系统与各种应用结合起来，并取得很大的成功。同时，专家系统在知识获取方面的需求也极大地刺激了机器学习的研究和发展。在出现第一个专家学习系统之后，示例归纳学习系统成为研究的主流，自动知识获取成为机器学习应用的研究目标。1980 年，在美国的卡内基梅隆（CMU）召开了第一届机器学习国际研讨会，标志着机器学习研究已在全世界兴起。此后，机器学习开始得到了大量的应用。1984 年，Simon 等 20 多位人工智能专家共同撰文编写的 MachineLearning 文集第二卷出版，国际性杂志Machine Learning 创刊，更加显示出机器学习突飞猛进的发展趋势。这一阶段代表性的工作有 Mostow 的指导式学习、Lenat 的数学概念发现程序、Langley 的BACON 程序及其改进程序。

第四阶段 20 世纪 80 年代中叶，是机器学习的最新阶段。这个时期的机器学习具有如下特点：

（1）机器学习已成为新的学科，它综合应用了心理学、生物学、神经生理学、数学、自动化和计算机科学等形成了机器学习理论基础。

（2）融合了各种学习方法，且形式多样的集成学习系统研究正在兴起。

（3）机器学习与人工智能各种基础问题的统一性观点正在形成。

（4）各种学习方法的应用范围不断扩大，部分应用研究成果已转化为产品。

（5）与机器学习有关的学术活动空前活跃。

1. 人工智能、机器学习、深度学习有什么区别和联系?

从定义出发，人工智能（Artificial Intelligence, AI）是机器，特别是计算机系统对人类智能过程的模拟。人工智能是一个愿景，目标就是让机器像我们人类一样思考与行动，能够代替我们人类去做各种各样的工作。机器学习（Machine Learning，ML）是机器从经验中自动学习和改进的过程，不需要人工编写程序指定规则和逻辑。深度学习（Deep Learn，DL）是机器学习的一种，主要特点是使用多层非线性处理单元进行特征提取和转换。每个连续的图层使用前一层的输出作为输入。

对比之下，人工智能是一个宏大的愿景，目标是让机器像我们人类一样思考和行动，既包括增强我们人类脑力也包括增强我们体力的研究领域。而学习只是实现人工智能的手段之一，并且，只是增强我们人类脑力的方法之一。所以，人工智能包含机器学习。机器学习又包含了深度学习。深度学习是机器学习的一种，是机器学习的子集。深度学习更强调模型结构的重要性、非线性处理以及特征提取和特征转换。

（3）神经元、单层感知机、多层感知机

神经元：

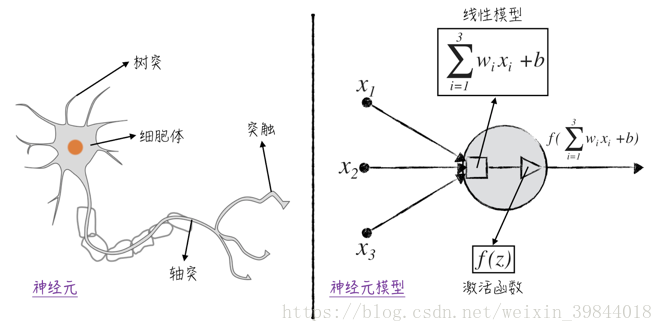
一个典型的神经元由4个部分组成。

树突：一个神经元有若干个树突，它们能接收来自其他神经元的信号，并将信号传递给细胞体。

细胞体：细胞体是神经元的核心，它把各个树突传递过来的信号加总起来，得到一个总的刺激信号。

轴突：当细胞体内的刺激信号超过一定阈值之后，神经元的轴突会对外发送信号。

突触：该神经元发送的信号（若有）将由突触向其他神经元或人体内的其他组织（对神经信号做出反应的组织）传递。需要注意的是，神经元通常有多个突触，但它们传递的信号都是一样的。

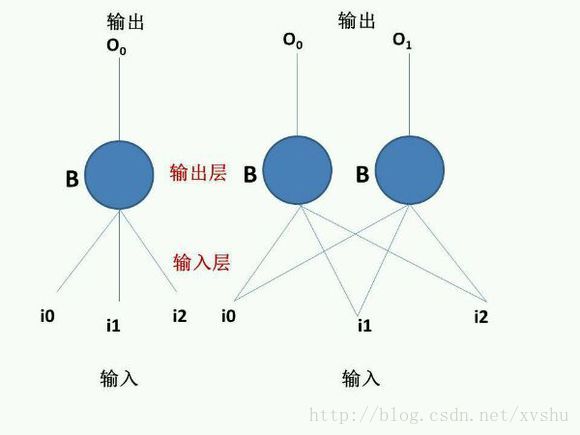


模仿神经元模型，我们提出了神经元模型。这个模型中可以有n个输入，，经过一个求和函数算出的值再经过激活函数，最终将激活函数的函数值作为输出。

单层感知机

单层感知器可以用来区分线性可分的数据，并且一定可以在有限的迭代次数中收敛。单层感知机与神经元模型不同。单层感知器能够用来模拟逻辑函数，例如逻辑非NOT、逻辑或非XOR、逻辑或OR、逻辑与AND和逻辑与非NAND等，但是不能用来模拟逻辑异或XOR（对于这个函数，必须用两层神经元）。

除了输入之外，偏置（bias）也经常被用于每个神经元，它在通过转换函数之前被加到输入的加权和上。权值也经常被应用于偏置上。偏置决定了多少输入激活（输入的加权和）才能激发神经元进入兴奋状态。偏置往往被设定为1，而偏置的权值可以通过学习算法加以调整。



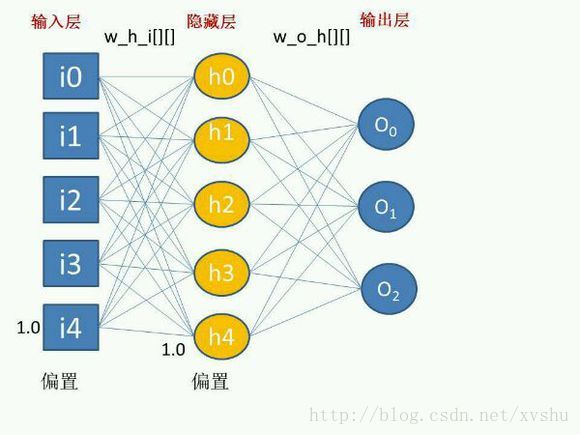
多层感知机

多层感知器中神经元的性质和单层感知器中的一样（例如偏置等）。然而由于存在多层，多层感知器中某一层的输出就是下一层的输入。因此多层感知器的实现比单层感知器略显复杂，但还是很直观的。注意，在这里对隐层结点和输出层结点使用相同的激励函数，即sigmoid函数。Sigmoid函数可以将神经元的输出值规格化到区间[0,1]。

该多层感知器有4个输入单元，4个隐层单元和3个输出单元。偏置作为输入单元和隐层单元来处理，但偏置具有常数值1（虽然可以通过调整权值来改变偏置的影响）。

程序的执行流程如下：首先，计算隐层单元的输出（利用输入单元和隐层单元到输入层单元的权值）；然后，继续计算下一层（这里指输出层）单元的输出。通常这个过程被称为前向传送数据，或简单的称之为前馈（feed forward）。

可以把神经网络看成并行计算系统。每个神经元是一个处理元件，该处理元件接受一个或多个输入并产生一个输出。输入和输出均可以看成信息。在多层感知器架构中，输出可以被送往多个其他处理单元，做进一步的处理。神经网络并行性的特性来自于同一层多个神经元之间的并行性。所有的神经元能够同时处理它们各自的输入，使得在多处理器系统中同一层次上具有大量神经元的神经网络的运行速度更快。



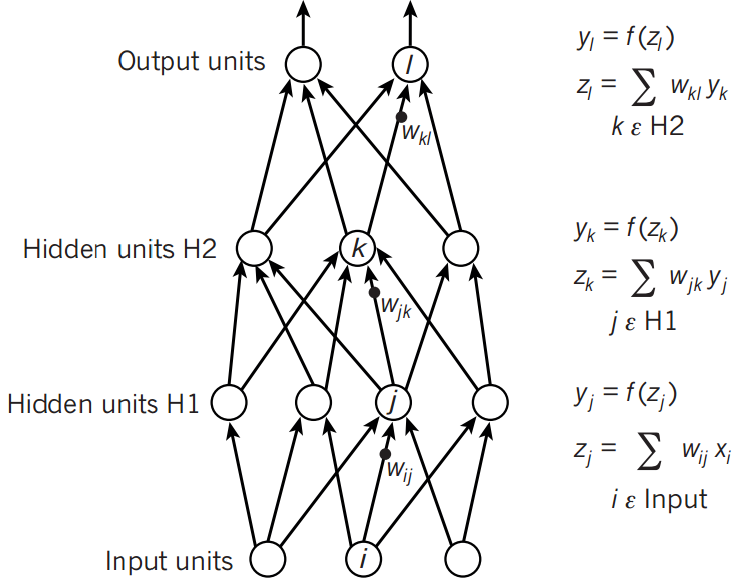
（4）前向传播

假设上一层结点i,j,k,…等一些结点与本层的结点w有连接，那么结点w的值怎么算呢？就是通过上一层的i,j,k等结点以及对应的连接权值进行加权和运算，最终结果再加上一个偏置项（图中为了简单省略了），最后在通过一个非线性函数（即激活函数），如ReLu，sigmoid等函数，最后得到的结果就是本层结点w的输出。

最终不断的通过这种方法一层层的运算，得到输出层结果。

对于前向传播来说，不管维度多高，其过程都可以用如下公式表示：

其中，上标代表层数，星号表示卷积，b表示偏置项bias，σ \sigmaσ表示激活函数。

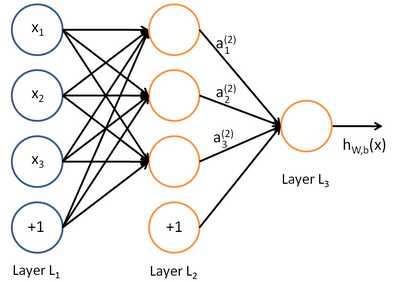


（5）反向传播

BackPropagation算法是多层神经网络的训练中举足轻重的算法。简单的理解，它的确就是复合函数的链式法则，但其在实际运算中的意义比链式法则要大的多。要回答题主这个问题“如何直观的解释back propagation算法？” 需要先直观理解多层神经网络的训练。

机器学习可以看做是数理统计的一个应用，在数理统计中一个常见的任务就是拟合，也就是给定一些样本点，用合适的曲线揭示这些样本点随着自变量的变化关系.

深度学习同样也是为了这个目的，只不过此时，样本点不再限定为(x, y)点对，而可以是由向量、矩阵等等组成的广义点对(X,Y)。而此时，(X,Y)之间的关系也变得十分复杂，不太可能用一个简单函数表示。然而，人们发现可以用多层神经网络来表示这样的关系，而多层神经网络的本质就是一个多层复合的函数。



2、