

深度学习基本知识

• 1.深度学习发展历史

1940年代：由神经科学家麦卡洛克和数学家皮兹在《数学生物物理学公告》上发表论文《神经活动中内在思想的逻辑演算》（A Logical Calculus of the Ideas Immanent in Nervous Activity）。建立了神经网络和数学模型，称为MCP模型。所谓MCP模型，其实是按照生物神经元的结构和工作原理构造出来的一个抽象和简化了的模型，也就诞生了所谓的“模拟大脑”，人工神经网络的大门由此开启。首次提出神经元的结构，但权重是不可学的。

1958年：计算机科学家罗森布拉特提出了两层神经元组成的神经网络，称之为“感知器”。第一次将MCP用于机器学习（machine learning）分类(classification)**。“感知器”算法使用MCP模型对输入的多维数据进行二分类，且能够使用梯度下降法从训练样本中自动学习更新权值。1962年,该方法被证明为能够收敛，理论与实践效果引起第一次神经网络的浪潮。

1969年：美国数学家及人工智能先驱 Marvin Minsky在其著作中证明了感知器本质上是一种线性模型（linear model），只能处理线性分类问题，就连最简单的XOR（亦或）问题都无法正确分类。这等于直接宣判了感知器的死刑，神经网络的研究也陷入了将近20年的停滞。

1986年：由神经网络之父 Geoffrey Hinton 在1986年发明了适用于多层感知器（MLP）的BP（Backpropagation）算法，并采用Sigmoid进行非线性映射，有效解决了非线性分类和学习的问题。该方法引起了神经网络的第二次热潮。

90年代时期：1991年BP算法被指出存在梯度消失问题，也就是说在误差梯度后项传递的过程中，后层梯度以乘性方式叠加到前层，由于Sigmoid函数的饱和特性，后层梯度本来就小，误差梯度传到前层时几乎为0，因此无法对前层进行有效的学习，该问题直接阻碍了深度学习的进一步发展。此外90年代中期，支持向量机算法诞生（SVM算法）等各种浅层机器学习模型被提出，SVM也是一种有监督的学习模型，应用于模式识别，分类以及回归分析等。支持向量机

以统计学为基础，和神经网络有明显的差异，支持向量机等算法的提出再次阻碍了深度学习的发展。

发展期 2006年 - 2012年：2006年，加拿大多伦多大学教授、机器学习领域泰斗、神经网络之父——Geoffrey Hinton 和他的学生 Ruslan Salakhutdinov 在顶尖学术刊物《科学》上发表了一篇文章，该文章提出了深层网络训练中梯度消失问题的解决方案：无监督预训练对权值进行初始化+有监督训练微调。斯坦福大学、纽约大学、加拿大蒙特利尔大学等成为研究深度学习的重镇，至此开启了深度学习在学术界和工业界的浪潮。

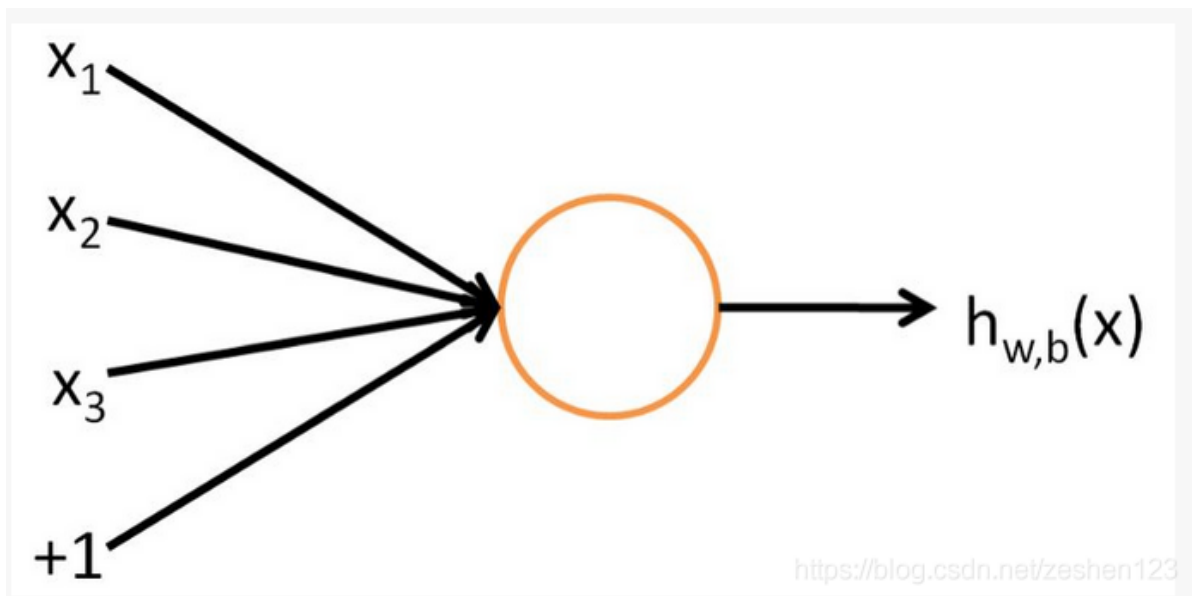
爆发期 2012 - 2017：2012年，Hinton课题组为了证明深度学习的潜力，首次参加ImageNet图像识别比赛，其通过构建的CNN网络AlexNet一举夺得冠军，且碾压第二名（SVM方法）的分类性能。也正是由于该比赛，CNN吸引到了众多研究者的注意。

• 2.人工智能、机器学习、深度学习有什么区别联系

概括来说，人工智能、机器学习和深度学习覆盖的技术范畴是逐层递减的。人工智能是最宽泛的概念。机器学习是当前比较有效的一种实现人工智能的方式。深度学习是机器学习算法中最热门的一个分支，近些年取得了显著的进展，并替代了大多数传统机器学习算法。即：人工智能 > 机器学习 > 深度学习。

• 3.神经元，单层感知机，多层感知机

神经元：神经元是一种处理单元，是对人脑组织的神经元的某种抽象、简化和模拟。是人工神经网络的关键部分。通过神经元，人工神经网络可以以数学模型模拟人脑神经元活动，继而进行高效的计算以及其他处理。

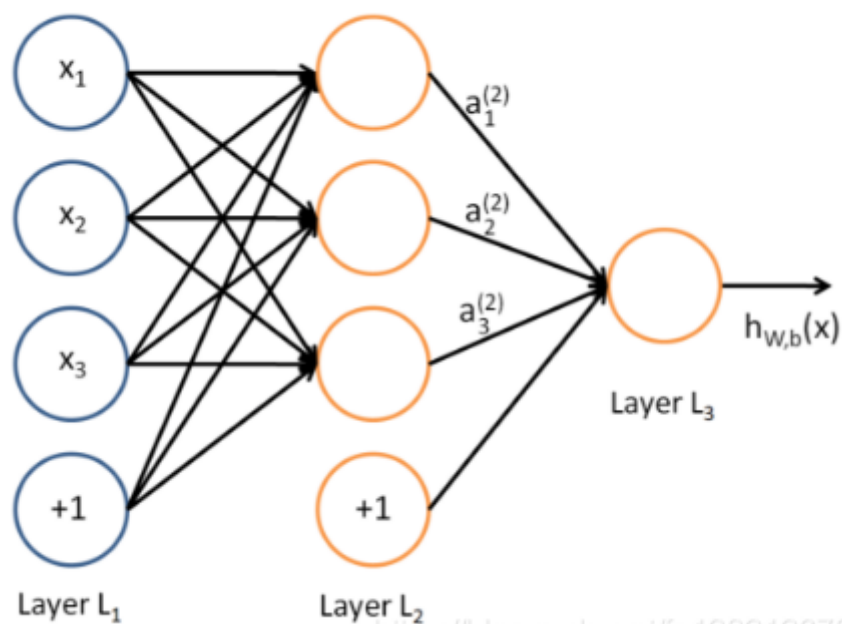


单层感知机：单层感知机目标是将被感知数据集划分为两类的分离超平面，并计算出该超平面。单层感知机是二分类的线性分类模型，输入是被感知数据集的特征向量，输出时数据集的类别 $\{+1, -1\}$ 。感知器的模型可以简单表示为：

$$f(x) = \text{sign}(wx + b) \quad (1)$$

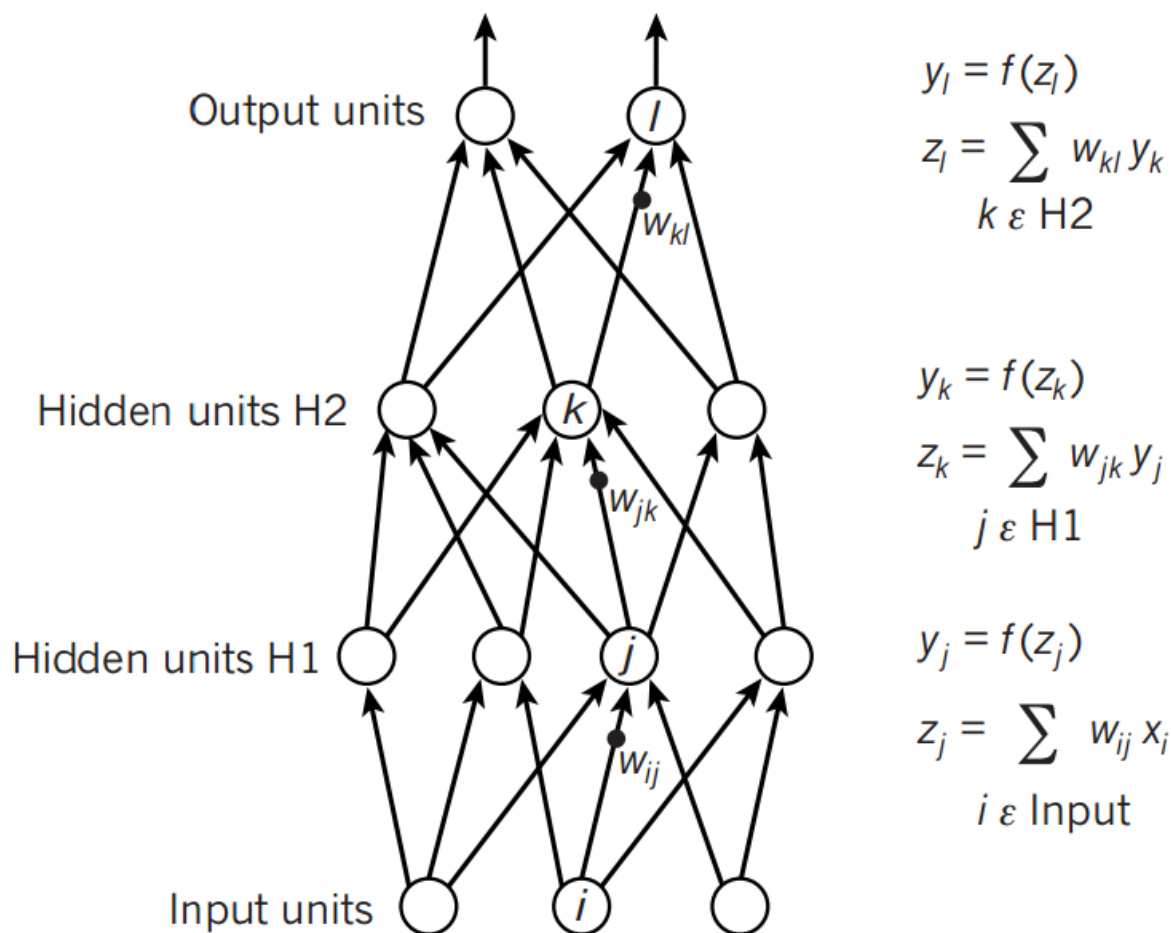
其中 w 是网络的 N 维权重向量， b 是网络的 N 维偏置向量， $w \cdot x$ 是 w 和 x 的内积， w 和 b 的 N 维向量取值要求在实数域。

多层感知机：多层感知机（MLP, Multilayer Perceptron）也叫人工神经网络（ANN, Artificial Neural Network），除了输入输出层，它中间可以有多个隐层，最简单的MLP只含一个隐层，即三层的结构，如下图：



• 4.什么是前向传播

通常，当我们使用神经网络时，我们输入某个向量 x ，然后网络产生一个输出 y ，这个输入向量通过每一层隐含层，直到输出层。这个方向的流动叫做前向传播。



• 5.什么是反向传播

误差反向传播算法简称反向传播算法(Back Propagation)。使用反向传播算法的多层感知器又称为BP神经网络。

BP算法是一个迭代算法，它的基本思想如下：将训练集数据输入到神经网络的输入层，经过隐藏层，最后达到输出层并输出结果，这就是前向传播过程。由于神经网络的输出结果与实际结果有误差，则计算估计值与实际值之间的误差，并将该误差从输出层向隐藏层反向传播，直至传播到输入层；在反向传播的过程中，根据误差调整各种参数的值（相连神经元的权重），使得总损失函数减小。

迭代上述三个步骤（即对数据进行反复训练），直到满足停止准则。

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial out_{h1}} * \frac{\partial out_{h1}}{\partial net_{h1}} * \frac{\partial net_{h1}}{\partial w_1}$$

$$\downarrow$$

$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial out_{h1}} + \frac{\partial E_{o2}}{\partial out_{h1}}$$

