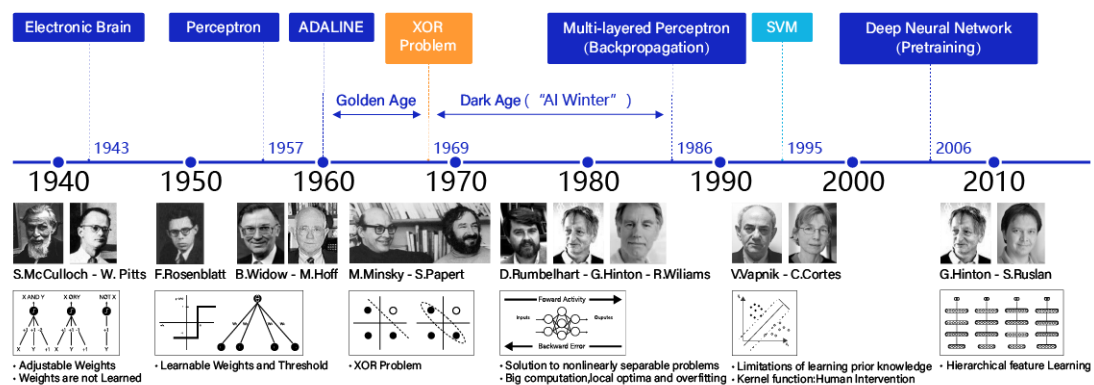


①深度学习发展历史

1943 年，神经科学家麦卡洛克(W. S. McCulloch)和数学家皮兹(W. Pitts)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。MCP 当时是希望能够用计算机来模拟人的神经元反应的过程，该模型将神经元简化为了三个过程：多路信号输入，线性加权求和，非线性激活。但是，这个模型的参数一开始并不是可学习的，而是人为设置好的。利用这样的结构来完成一定的逻辑运算。

1958 年，计算机科学家罗森布拉特(Rosenblatt)提出了两层神经元组成的神经网络，称之为“感知器”(Perceptrons)。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 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，神经网络之父 Geoffrey Hinton 发明了适用于多层感知器(MLP)的 BP 算法，并采用 Sigmoid 进行非线性映射，有效地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。

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 年代中后期，除了神经网络模型之外，还有大量的传统机器学习模型也被陆续提出，其中最经典的就是支持向量机(SVM)。由于 SVM 的理论非常完备，而且在当时的数据集上的表现效果极佳。相比之下，深度学习网络模型并没有展现出很好的效果。因此，当时虽然很多人用深度学习网络，但没有把它放在一个突出的位置，这再次阻碍了深度学习的发展。

直到 21 世纪，在互联网的兴起、数据量的增大、GPU 计算的出现等条件下，人们发现神经网络模型在自然语言处理、计算机视觉等方面有非常惊艳的表现，而且要比传统机器学习模型的效果好得多。从这个时间点开始，我们真正进入了深度学习的时代。

②人工智能、机器学习、深度学习有什么区别和联系？

人工智能是研发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。它是最宽泛的概念，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。三者的关系如图 1 所示，即：人工智能 > 机器学习 > 深度学习。

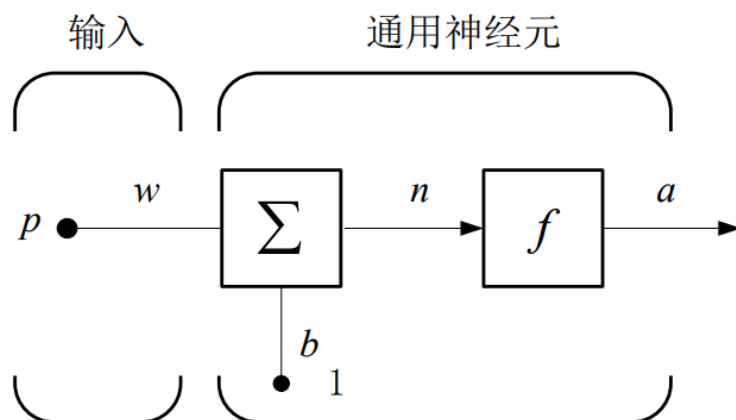


图 1 人工智能、机器学习和深度学习三者关系示意

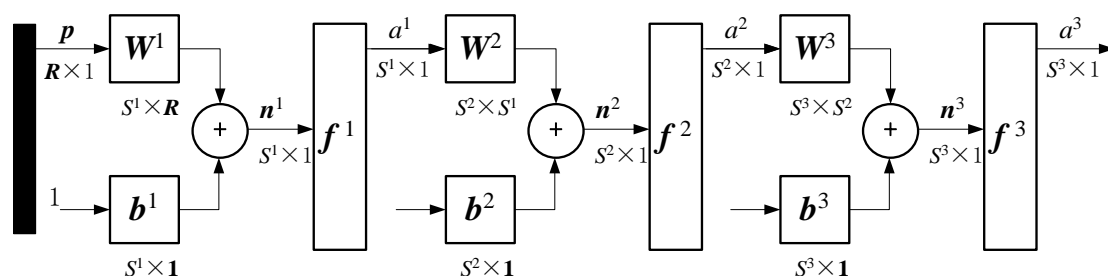
③神经元、单层感知机、多层感知机

神经元指神经网络中的每个节点，由加权和、非线性变换（激活函数）组成。

单层感知器属于单层前向网络，即除输入层和输出层之外，只拥有一层神经元节点。特点是输入数据从输入层经过隐藏层向输出层逐层传播，相邻两层的神经元之间相互连接，同一层的神经元之间没有连接。

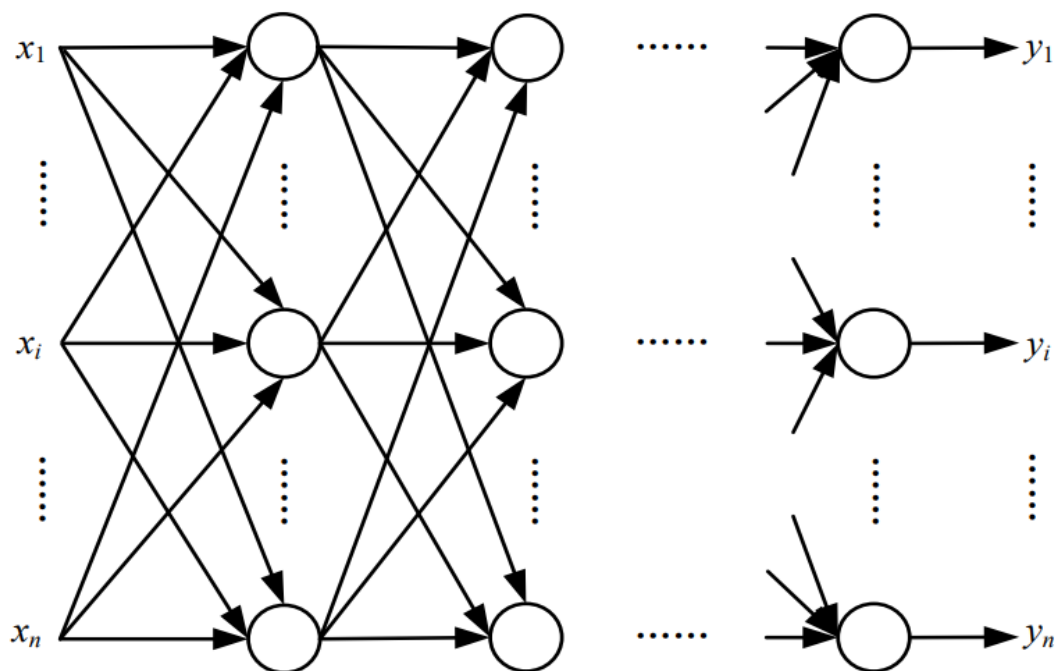


多层感知器，除了输入输出层，它中间可以有多个隐层，最简单的 MLP 只含一个隐层，即三层的结构。



④什么是前向传播？

前向传播指由输入层到输出层的单向传播，每层神经元仅与其前一层的神经元相连，仅接受前一层传输来的信息。



⑤什么是反向传播?

梯度下降可以应对带有明确求导函数的情况，或者说可以应对那些可以求出误差的情况，我们可以把它看做没有隐层的网络。但是，对于多隐层的神经网络，输出层可以直接求出误差来更新参数，但其中隐层的误差是不存在的，因此不能对它直接应用梯度下降，而是先将误差反向传播至隐层，然后再应用梯度下降，其中将误差从末层往前传递的过程需要链式法则的帮助，因此反向传播算法可以说是梯度下降在链式法则中的应用。

反向传播算法的思想：从后向前反向逐层传播输出层的误差，以间接计算隐层的误差。算法可以分为两个阶段：

正向过程：从输入层经隐层逐层正向计算各单元的输出；

反向过程：由输出误差逐层反向计算隐层各单元的误差，并用此误差修正前层的权值。

