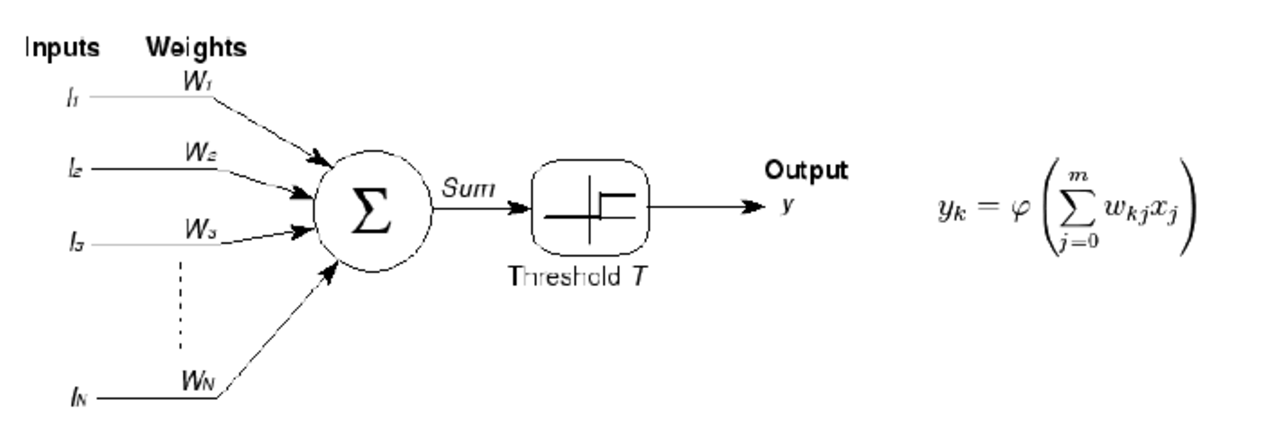
1. **深度学习发展历史**

1943年，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》。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。

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：输入信号线性加权，求和，非线性激活（阈值法）。

如下图所示



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用于机器学习分类。“感知器”算法算法使用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。

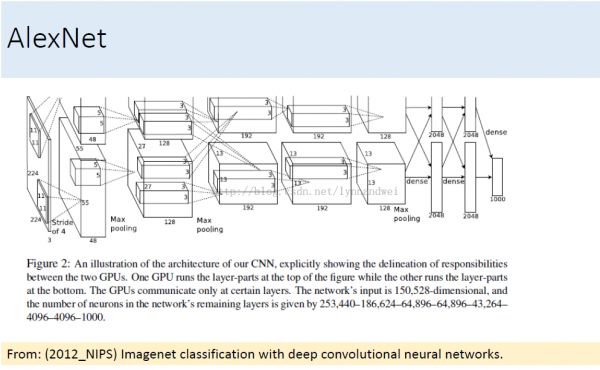
1969年，纵观科学发展史，无疑都是充满曲折的，深度学习也毫不例外。 1969年，美国数学家及人工智能先驱 Marvin Minsky 在其著作中证明了感知器本质上是一种线性模型，只能处理线性分类问题，就连最简单的XOR（亦或）问题都无法正确分类。这等于直接宣判了感知器的死刑，神经网络的研究也陷入了将近20年的停滞。

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。

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。

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。

2011年，ReLU激活函数被提出，该激活函数能够有效的抑制梯度消失问题。2011年以来，微软首次将DL应用在语音识别上，取得了重大突破。微软研究院和Google的语音识别研究人员先后采用DNN技术降低语音识别错误率20％~30％，是语音识别领域十多年来最大的突破性进展。2012年，DNN技术在图像识别领域取得惊人的效果，在ImageNet评测上将错误率从26％降低到15％。在这一年，DNN还被应用于制药公司的DrugeActivity预测问题，并获得世界最好成绩。



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。

1. **人工智能、机器学习、深度学习的区别与联系**



如上图，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。

人工智能（为机器赋予人的智能）：它是研究、开发用于模拟、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。人工智能是计算机科学的一个分支，当前各项人工智能技术不断得到突破，并找到相对明确的应用场景。

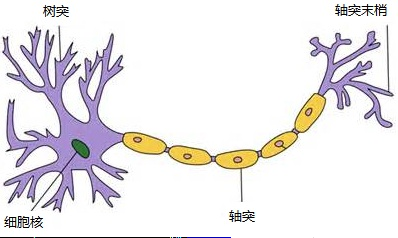
机器学习（一种实现人工智能的方法）：机器学习属于人工智能的一个分支，也是人工智能的核心，是使计算机具有智能的根本途径。

深度学习（一种实现机器学习的技术）：是机器学习领域中一个新的研究方向，它被引入机器学习使其更接近于最初的目标——人工智能。深度学习是学习样本数据的内在规律和表示层次，它的最终目标是让机器能够像人一样具有分析学习能力，能够识别文字、图像和声音等数据。

综上所述，三者之间的区别不小，但是联系非常的紧密。

1. **神经元、单层感知机、多层感知机**
2. 神经元：

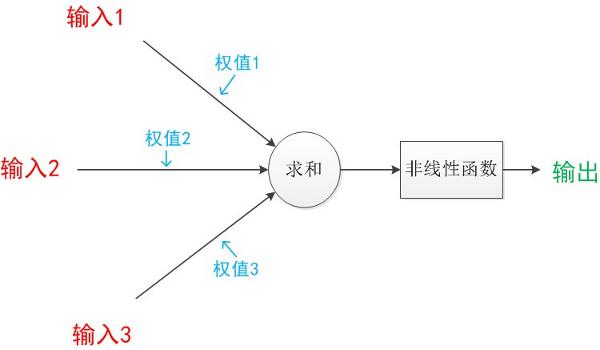
一个神经元通常具有多个树突，主要用来接受传入信息；而轴突只有一条，轴突尾端有许多轴突末梢可以给其他多个神经元传递信息。轴突末梢跟其他神经元的树突产生连接，从而传递信号。这个连接的位置在生物学上叫做“突触”。突触之间的交流通过神经递质实现。



对该模型进行抽象处理。首先考虑到神经元结构有多个树突，一个轴突可将其抽象为下图的黑箱结构：

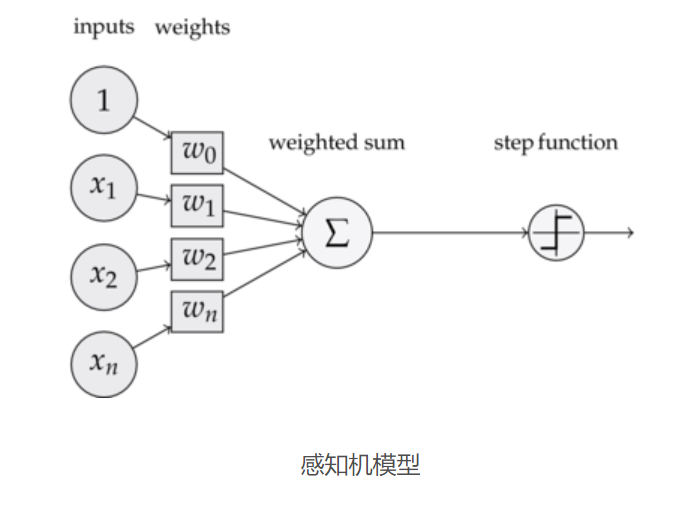


但是黑箱结构有诸多不便，首先是不知道黑箱中的函数结构就不能为我们所用，其次是输入输出与黑箱的关系也无法量化。因此考虑将上述结构简化，首先把树突到细胞核的阶段简化为线性加权的过程（当然了，该过程也有可能是非线性的，但是我们可以把其非线性过程施加到后面的非线性函数以及多层网络结构中），其次把突触之间的信号传递简化为对求和结果的非线性变换，那么上述模型就变得清晰了：



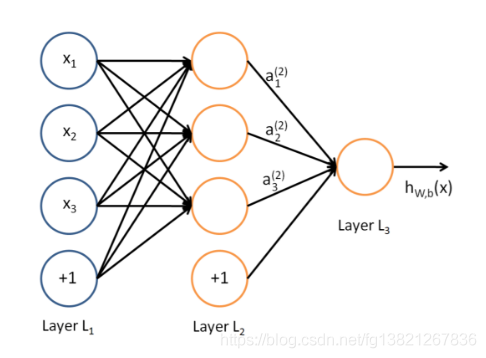
1. 单层感知机：

事实上，单个神经元就是我们常常所说的单层感知机模型，将其进行符号化所的如图所示：



1. 多层感知机：

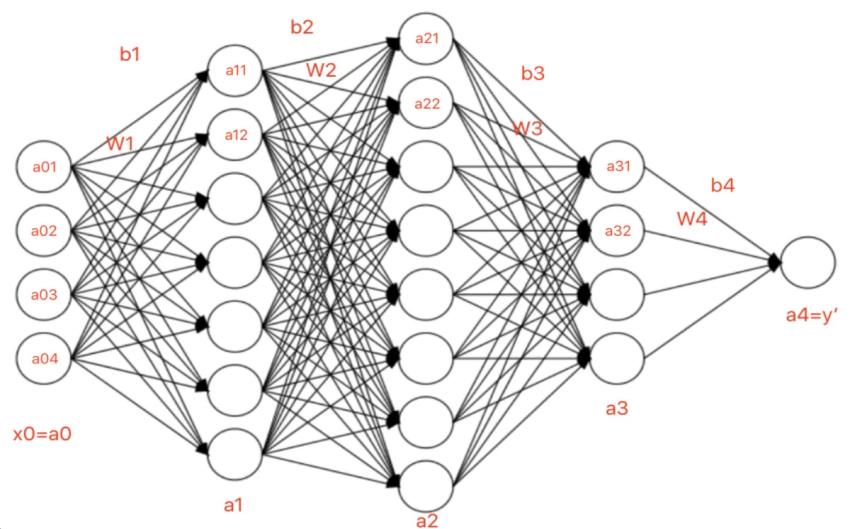
多层感知机也叫人工神经网络，除了输入输出层，它中间可以有多个隐层，最简单的MLP只含一个隐层，即三层的结构，如下图：



从上图可以看到，多层感知机层与层之间是全连接的。多层感知机最底层是输入层，中间是隐藏层，最后是输出层。

隐藏层的神经元与输入层是全连接的，假设输入层用向量X表示，则隐藏层的输出就是 f (W1X+b1)，W1是权重（也叫连接系数），b1是偏置，函数f 可以是常用的sigmoid函数或者tanh函数：

1. **前向传播**



如上图，

表示第一层即输入层，接收一个样本的特征,他有个神经元。

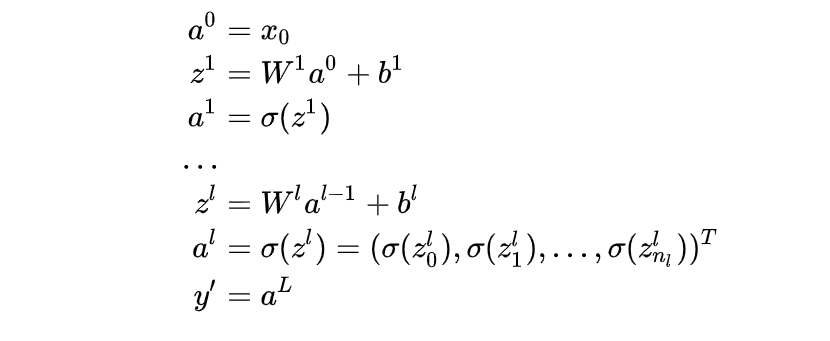
表示隐层每一层的结果是上一层的输出也是下一层的输入，分别有个神经元。

表示输出层，当然，对于二分类，输出层是一个神经元，如果是多分类，就可以是多个。有个神经元。

表示第l-1层与l层之间的权重矩阵。

表示第l层的偏至值。

除第一层外其他层中每个圆即神经元，都是一个LR模型。



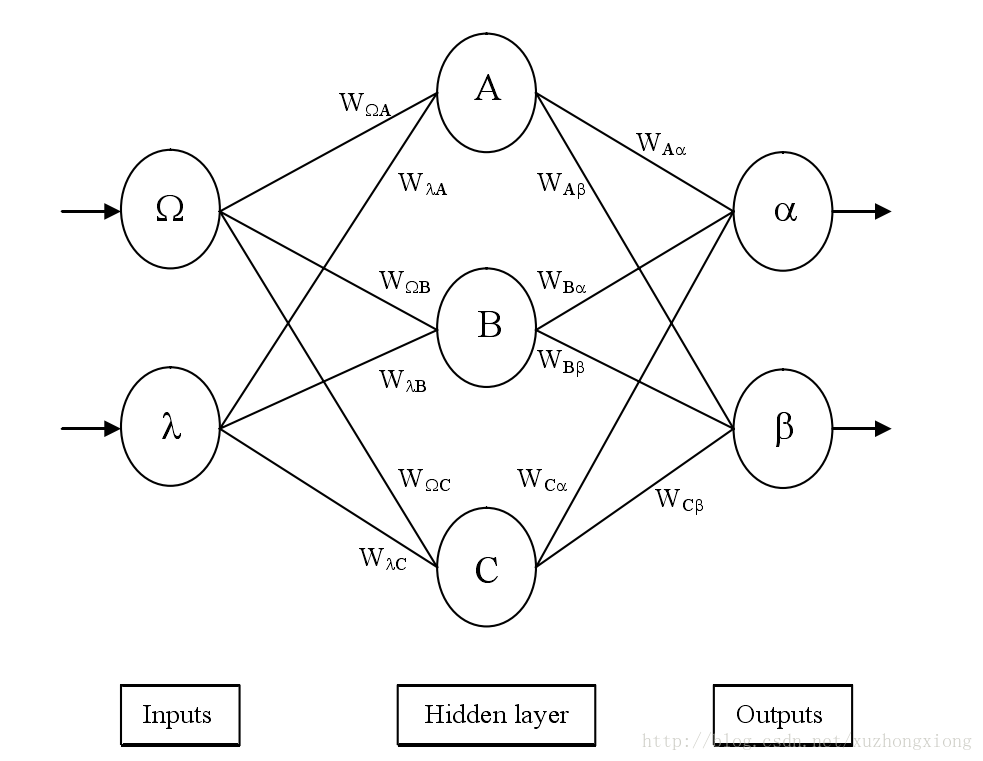
(这里的输出是通用输出，没按图1来。输出层神经元有个,模型共L层)

上述的计算值，从输入层开始，逐层向前传播，经过隐层到达输出层。所以叫做前向传播

1. **后向传播**

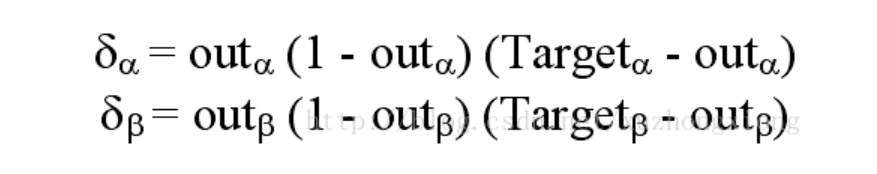
后向传播是深度学习中一种训练与学习方法，用来调整深度网络中各个节点之间的权重（即权值更新），使得网络输出的样本标签与实际标签相一致。本文将对后向传播算法的实现细节进行总结。

后向传播算法是将输出层的误差向后进行传播来实现权重的调整。如图1所示，一个含有1个隐藏层的网络，输入层有2个节点，隐藏层有3个节点，输出层有2个节点。



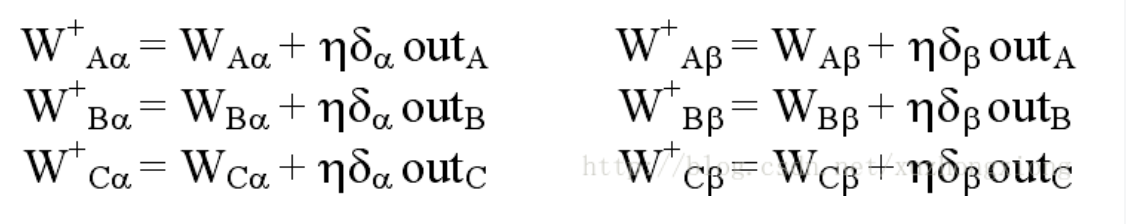
则后向传播的实现细节如下：

1.计算输出层各个节点的误差



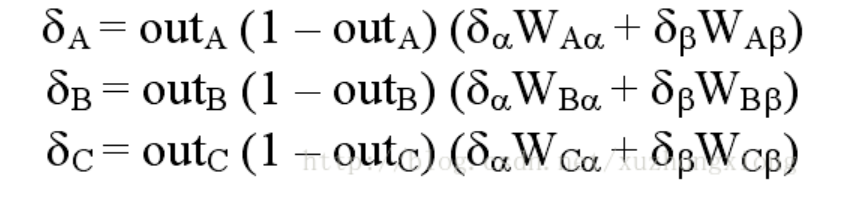
其中 δα表示 α节点的输出误差； outα表示经过前向传播后的输出； Targetα表示输出节点的真值，即样本对应的标签。式中 outα(1−outα)是由于激活函数采用sigmoid所造成的。

2.更新输出层的权重

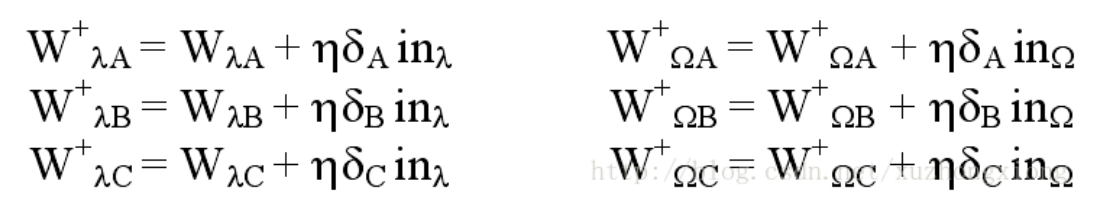


其中 W+W+为更新后的权重； WW为更新前的权重；ηη为学习率，控制权重更新的强度。

3.将输出层的误差传递给隐藏层



1. 更新隐藏层的权重



后向传播为深度学习中的调参操作提供很高效的数据依据。