①深度学习发展历史

（1）第一代神经网络（1958-1969)

1、 1958年Rosenblatt发明的感知算法可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。

2、1962年，该方法被证明为能够收敛，它的理论与实践效果引发了第一次神经网络的浪潮；

3、 1969年，美国数学家及AI先驱Minsky在著作中证明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问题都无法正确分类。这等于直接宣判了感知器的死弄，神经网络的研究也陷入了近20年的停滞。

（2）第二代神经网络（1986-1998）

1、 1986年Hiton发明的MLP的BP算法，并采用Sigmoid进行非线性映射，有效解决了非线性公类和学习的问题。该方法引发了神经网络的第二次热潮。

2、1989年，Robert Hecht-Nielsen证明了MLP的万能逼近定理，即对于任何闭区间内的一个连续函数f，都可以用含有一个隐含层的BP网络来逼近。

3、1989年，LeCun发明了CNN–LeNet，并将其用于数字识别，且取得了较好的成绩，不过当时并没有引起足够的注意。

4、1989年以后由于没有提出特别突出的方法，且神经网络一直缺少相应的严格数学理论支持，神经网络的热潮渐渐冷淡下去。冰点发生在1991年，BP算法被指出存在梯度消失问题，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。

5、1997年，LSTM模型被发明，尽管该模型在序列建模上表现出的特性非常突出，但由于处于神经网络的下坡期，也没有引起足够的重视。

（3）统计学习方法的春天（1986-2006）

1、1986年，决策方法被提出，很快ID3、ID4、CART等改进的决策树方法相继出现，这些方法也是符号学习方法的代表。正是由于这些方法的出现，使得统计学习开始进入人们的视野，迎来统计学习方法的春天。

2、1995年，统计学家Vapnik提出线性SVM。由于它有非常完美的数学理论推导做支撑（统计学与凸优化等），并且非常符合人的直观感受（最大间隔），更重要的是它在线性分类的问题上取得了当时最好的成绩，这使得神经网络更陷入无人问津的境地。

3、1997年，AdaBoost被提出，该方法是PAC理论在机器学习实践上的代表，也催生了集成学习这一类方法的诞生，在回归和分类任务上取得了非常好的效果。该方法通过一系列的弱分类器集成，达到强分类器的效果。

4、2000年，Kernel SVM被提出，核化的SVM通过一种巧妙的方法将原空间线性不可分的问题，通过Kernel映射成高维空间的线性可分的问题，成功解决了非线性分类的问题，且分类效果非常好。至此也更加终结了神经网络时代。

5、 2001年，随机森林被提出，这是集成方法的另一个代表，该方法比AdaBoost能更好地抑制过拟合问题。

6、2001年，一种新的统一框架—-图模型被提出，该方法试图统一机器 学习混的方法，如朴素贝叶斯、SVM、隐马尔可夫模型等，为各种学习方法提供一个统一的描述框架

（4）第三代神经网络深度学习（2006-至今）

快速发展期（2006-2012）

1、2006年，深度学习元年。Hiton提出了深层网络训练中梯度消失问题的解决方案：“无监督训练对权值进行初始化+有监督训练微调”。其主要思想是先通过自学习的方法学习到训练数据的结构，然后在该结构上进行有监督训练微调。

2、2011年，ReLU激活函数被提出，有效抑制梯度消失的问题。

3、2011年，微软首次将深度学习应用在语音识别上，取得了重大突破。

（5）爆发期（2012-至今）

2012年，Hiton课题组首次参加ImageNet图像识别比赛，AlexNet夺得冠军，并碾压了第二名（SVM）的分类性能。

②人工智能、机器学习、深度学习有什么区别和联系？

答：机器学习是一种实现人工智能的方法，深度学习是一种实现机器学习的技术。目前，业界有一种错误的较为普遍的意识，即“深度学习最终可能会淘汰掉其他所有机器学习算法”。这种意识的产生主要是因为，当下深度学习在计算机视觉、自然语言处理领域的应用远超过传统的机器学习方法，并且媒体对深度学习进行了大肆夸大的报道。深度学习，作为目前最热的机器学习方法，但并不意味着是机器学习的终点。起码目前存在以下问题：

深度学习模型需要大量的训练数据，才能展现出神奇的效果，但现实生活中往往会遇到小样本问题，此时深度学习方法无法入手，传统的机器学习方法就可以处理；

有些领域，采用传统的简单的机器学习方法，可以很好地解决了，没必要非得用复杂的深度学习方法；

深度学习的思想，来源于人脑的启发，但绝不是人脑的模拟，举个例子，给一个三四岁的小孩看一辆自行车之后，再见到哪怕外观完全不同的自行车，小孩也十有八九能做出那是一辆自行车的判断，也就是说，人类的学习过程往往不需要大规模的训练数据，而现在的深度学习方法显然不是对人脑的模拟。

③神经元、单层感知机、多层感知机

（1）神经元：

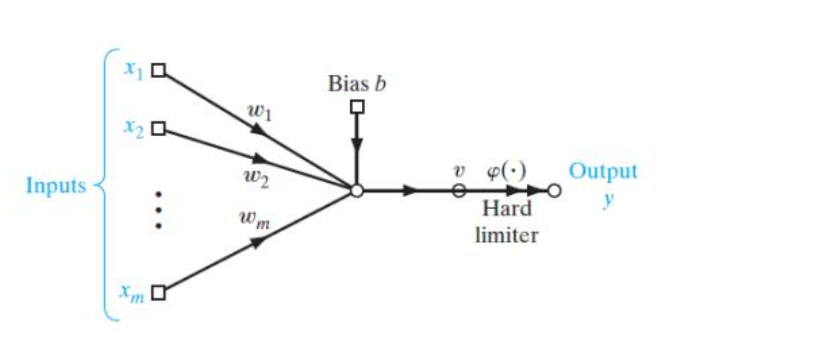
人工神经元（Artificial Neuron），简称神经元（Neuron），是构成神经网络的基本单元，其主要是模拟生物神经元的结构和特性，接受一组输入信号并产出输出。

一个生物神经元通常具有多个树突和一条轴突。树突用来接受信息，轴突用来发送信息。当神经元所获得的输入信号的积累超过某个阈值时，它就处于兴奋状态，产生电脉冲。轴突尾端有许多末梢可以给其他个神经元的树突产生连接（突触），并将电脉冲信号传递给其它神经元。

MP神经元[McCulloch and Pitts, 1943]。现代神经网络中的神经元和M-P 神经元的结构并无太多变化。不同的是，MP神经元中的激活函数f 为0或1的阶跃函数，而现代神经元中的激活函数通常要求是连续可导的函数。

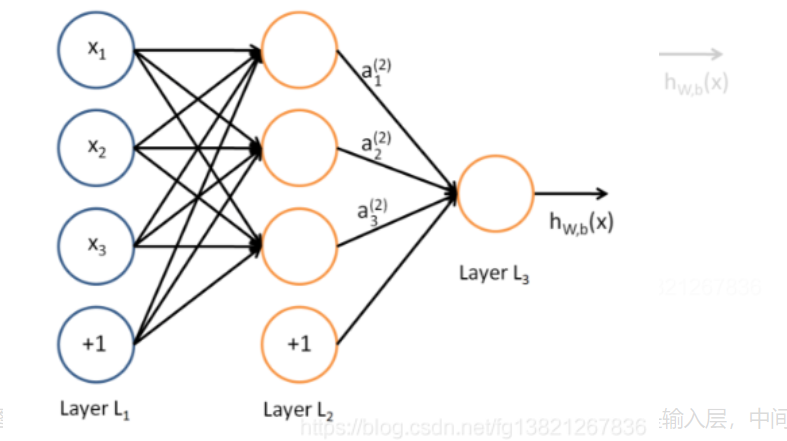
（2）单层感知机：

感知机是美国学者F.Rosenblatt提出的，他对最早提出的MP不一样，他的所有维度 的权值是可以改变的，通过对权值的迭代可以快速的解决线性可分的二分类问题。

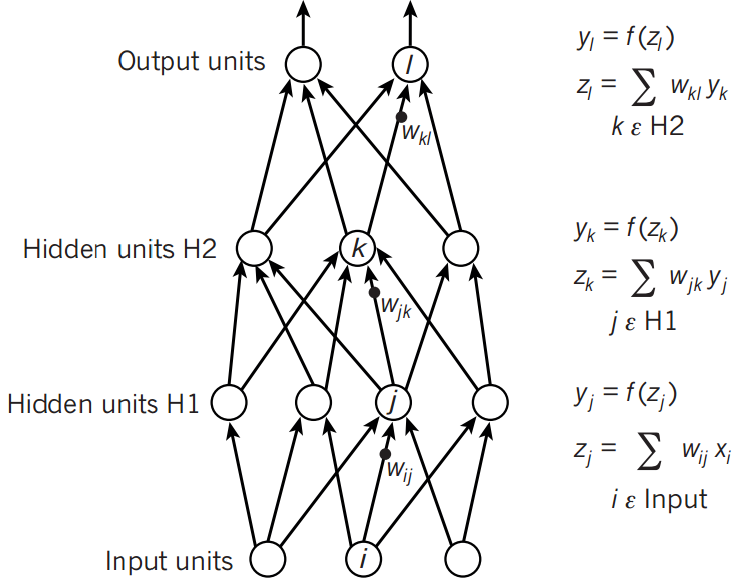


（3）多层感知机：

   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？



如图所示，这里讲得已经很清楚了，前向传播的思想比较简单。

举个例子，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：

a 2 = σ ( z 2 ) = σ ( a 1 ∗ W 2 + b 2 ) a^2 = \sigma(z^2) = \sigma(a^1 \* W^2 + b^2)a2=σ(z2)=σ(a1∗W2+b2)

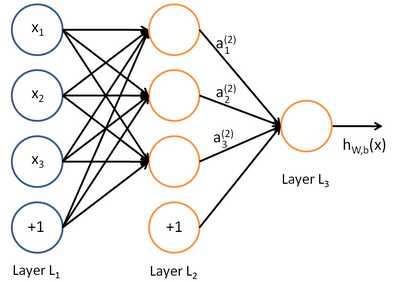
其中，上标代表层数，星号表示卷积，b表示偏置项bias，σ \sigmaσ表示激活函数。

⑤什么是后向传播？

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。借用网上找到的一幅图[1]，来直观描绘一下这种复合关系。



其对应的表达式如下：

