**第5章**

# 反向传播

|  |  |
| --- | --- |
|  | 本章内容 |
|  | ◎ 深度学习历史简介  ◎ 深度学习原理 |

## 3.1 深度学习历史简介

神经网络诞生于人类对大脑的探索，现如今的神经网络已经是一个多学科交叉的科研领域。深度学习则是多层神经网络的组合形式，其内部结构灵活，可利用繁多的数学方法及不同的数学方法组合。

神经网络的发展历程可谓崎岖，1943年两位芝加哥大学的研究人员McCullough 和 Walter Pitts借鉴了神经细胞生物原理首次提出神经元的M-P模型如图3.1，该模型是历史上的第一个神经元模型即对人类大脑工作原理的一个建模，可以看到其原理和神经网络很类似，均是对输入信号的加权求和并与阈值进行比较。

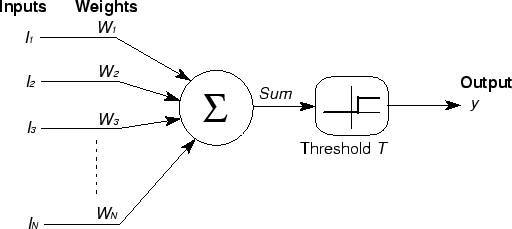


图3.1 M-P模型

1958年，Frank Rosenblatt提出了一种名为感知机的神经网络如图3.2，也就是目前神经网络中常用的全连接层，该网络的主要作用是做二元分类，该模型的提出也标志着神经网络正式走进了实际应用场景之中，也加快了神经网络的发展速度。

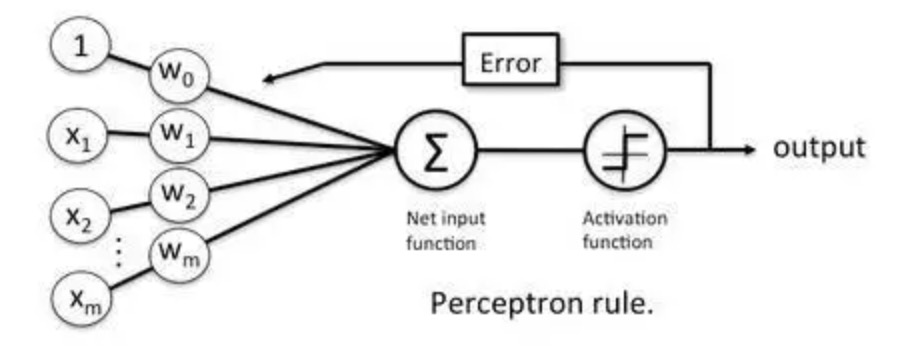


图3.2 感知机

1969年之前神经网络都一直都是计算机领域热门的研究内容，Marvin Minsky 和 Seymour Papert从数学的角度证明了神经网络的局限性，由于其简单的模型结构，使其连异或逻辑也无法表述，神经网络的发展也因此止步不前。直到1983年，Hinton, G. E. 和 Sejnowski, T. J.提出了隐单元的概念，即现如今的隐藏层，隐藏层进一步提高了神经网络的灵活性，并且反向传播算法也因此得到了重视，使得神经网络能有效的进行训练与参数更新。但其后的10多年的时间内，由于统计学模型的理论快速发展与其可解释性，并且神经网络的不可解释的黑盒性质，再加上当时提出了支持向量机这样的通用性较好的模型，神经网络再一次淡出了人们的视野。直到近年来，硬件设备的快速发展，GPU的强大处理能力，为神经网络铺平了一条大道路。随后神经网络在ImageNet图像分类的竞赛中，打破了传统算法的记录，并且准确率逐年提升。

现今，深度学习已经不仅在图像方面独具一头，在很多其它领域也都取得了突破性进展，包括语音识别、自然语言处理、人脸识别、机器翻译等等。在未来，基于统计的模型还会有什么样的突飞猛进我们不得而知，但是深度学习一定会是这些模型中不可小觑的中坚力量。

## 3.2 深度学习原理

上文中，我们提到过神经网络最大的缺点即是其结构过于简单，没法表述复杂的模型，从理论上来说，模型的参数如果越多，那么其能力就会越强，这就意味着模型能学习更复杂的任务，而参数多了，模型训练效率却会降低，并且存在过拟合的可能。随着大数据时代的到来，硬件设备的快速发展，高效的计算能力解决了训练低效的问题，大量的数据也一定概率的避免了过拟合的发生，深度学习因此也应运而生。

深度学习与神经网络的不同主要在于“深”，其由于隐层数量的增加，模型参数也因此增加，并且深度学习不仅仅只增加了线性情况的神经元，非线性的激活函数也随着隐层的增加而增加，模型能力自然会有提升。如果从另外一个角度看，也可以理解为把输入信号进行逐层的加工，从而把输入信号从与目标没有关联的数据分布转变为与目标联系更加紧密的分布，换言之即把深度学习中浅层的神经网络理解为特征提取器，而后面的层则是对特征的进一步的处理，把浅显的低层特征转变为了高层特征，之后用简单的模型即可完成复杂的机器学习任务。对比传统的机器学习算法来说，通常需要人为的收集整理特征，即特征工程，特征的好坏又对模型结果有巨大的影响并且可能会引入一些人为误差，深度学习也变相解决了特征提取问题，从而成为了一种端到端的模型。