

神经网络和深度学习笔记

张晋

北京航空航天大学
数学与系统科学学院

最后更新于: May 26, 2017

目录 | 0

1	神经网络简介	2
1.1	生物神经网络 (Biological Neural Networks)	3
1.2	人工神经网络 (Artificial Neural Networks)	4
1.2.1	人工神经元模型 (Artificial Neuron Model)	4
1.2.2	人工神经网络的构成 (Structure of ANN)	6
1.2.3	人工神经网络的学习 (Training of ANN)	8
1.3	神经网络应用现状	9
2	反向传播算法	10

神经网络简介

1

1.1 生物神经网络 (Biological Neural Networks).....	3
1.2 人工神经网络 (Artificial Neural Networks)	4
1.2.1 人工神经元模型 (Artificial Neuron Model)	4
1.2.2 人工神经网络的构成 (Structure of ANN)	6
1.2.3 人工神经网络的学习 (Training of ANN)	8
1.3 神经网络应用现状.....	9

“神经网络是由具有适应性的简单单元组成的广泛并行互连的网络，它的组织能够模拟生物神经系统对真实世界物体所作出的交互反应” —[Kohonen,1988]

1.1 生物神经网络 (Biological Neural Networks)

大脑可视作为 1000 多亿神经元组成的神经网络。

神经元的信息传递和处理是一种电化学活动。树突由于电化学作用接受外界的刺激；通过胞体内的活动体现为轴突电位，**当轴突电位达到一定的值则形成神经脉冲或动作电位**；再通过轴突末梢传递给其它的神经元。从控制论的观点来看；这一过程可以看作一个**多输入单输出非线性**系统的动态过程。

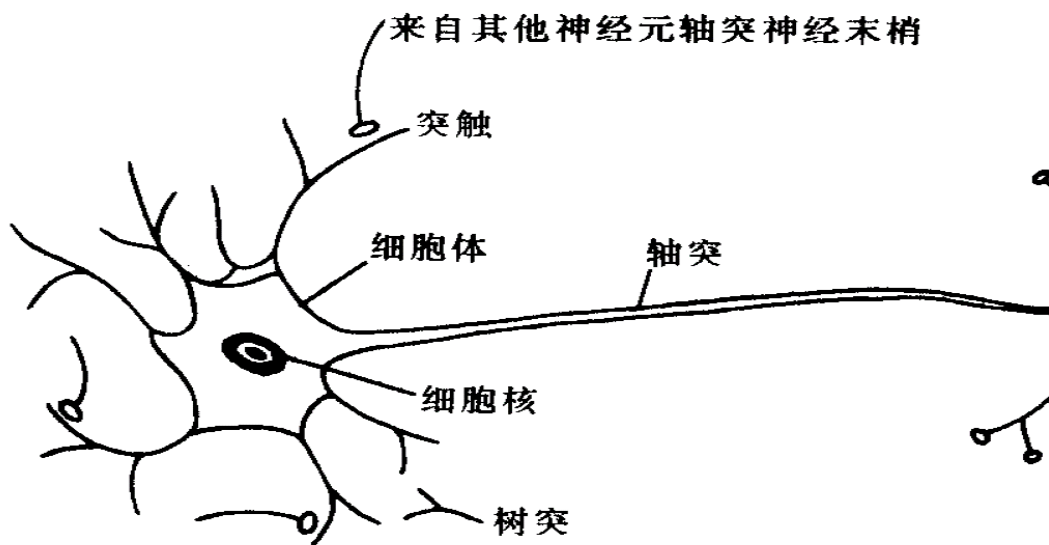


图 1.1: 神经元信息的传递

大脑处理信息的特点

- 分布存储与冗余性: 记忆在大量元中，每个元存在许多信息的部分内容，信息在神经网络中的记忆反映在神经元间的突触连接强度上
- 并行处理: 神经网络既是处理器又是存储器 (并行处理不同于并行机)
- 信息处理与存储合一: 每个元兼有二者功能
- 可塑性与自组织性: 可塑性是学习记忆的基础
- 鲁棒性: 高连接度导致一定的误差和噪声不会使网络性能恶化。是智能演化的重要因素

1.2 人工神经网络 (Artificial Neural Networks)

神经网络是一个并行和分布式的信息处理网络结构，它一般由许多个神经元组成，每个神经元只有一个输出，它可以连接到很多其他的神经元，每个神经元输入有多个连接通道，每个连接通道对应于一个连接权系数。

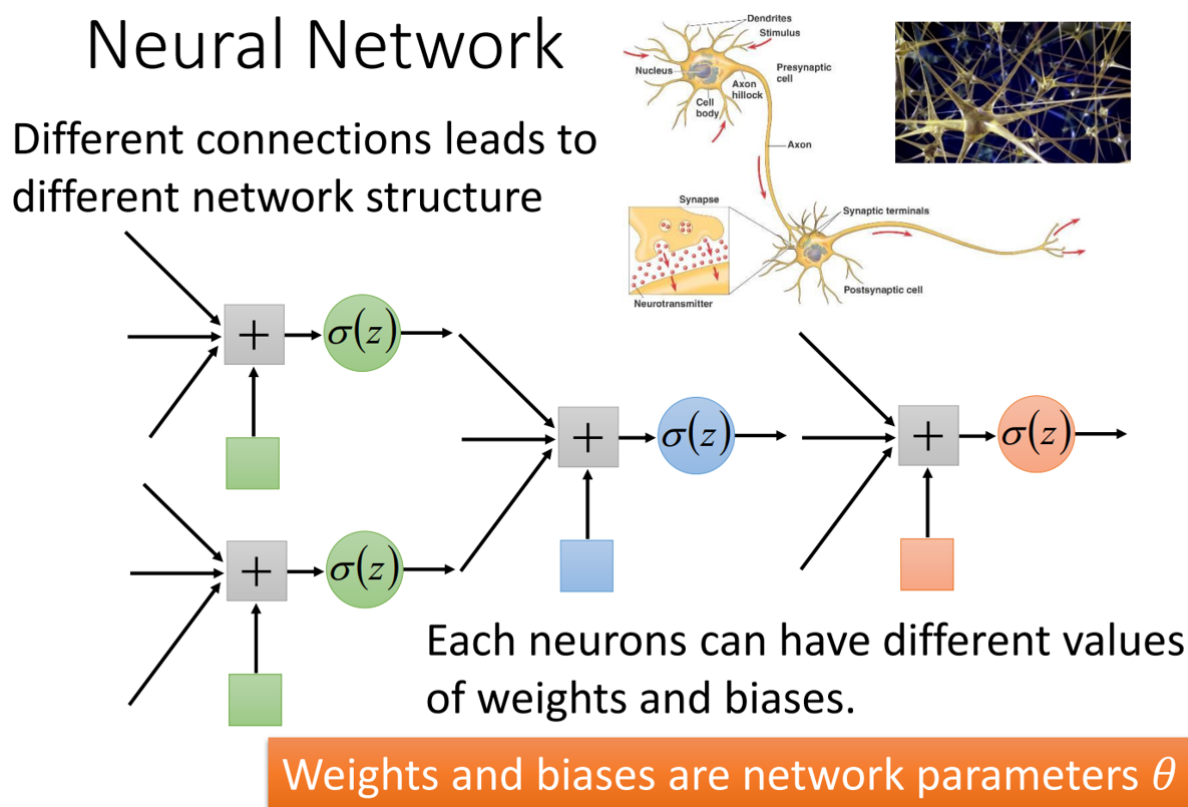


图 1.2: Artificial Neural Networks

1.2.1 人工神经元模型 (Artificial Neuron Model)

神经网络中最基本的成分是**神经元 (neuron) 模型**，即上述定义中的“简单单元”。在生物神经网络中，每个神经元与其他神经元相连，当它“兴奋”时，就会向相连的神经元发送化学物质，从而改变这些神经元内的电位；如果某神经元的电位超过了一个“**阈值**” (threshold)，那么它就会被激活，即“兴奋”起来，向其他神经元发送化学物质。

1943 年，[MccullochandPitts,1943] 将上述情形抽象为图所示的简单模型，这就是一直沿用至今的“M-P 神经元模型”在这个模型中，神经元接收到来自 n 个其他神经元传递过来的输入信号，这些输入信号通过带权重的连接 (connection) 进行传递，神经元接收到的总输入值将与神经元的阈值进行比较然后通过“**激活函数**” (activation function) 处理已产生神经元的输出。

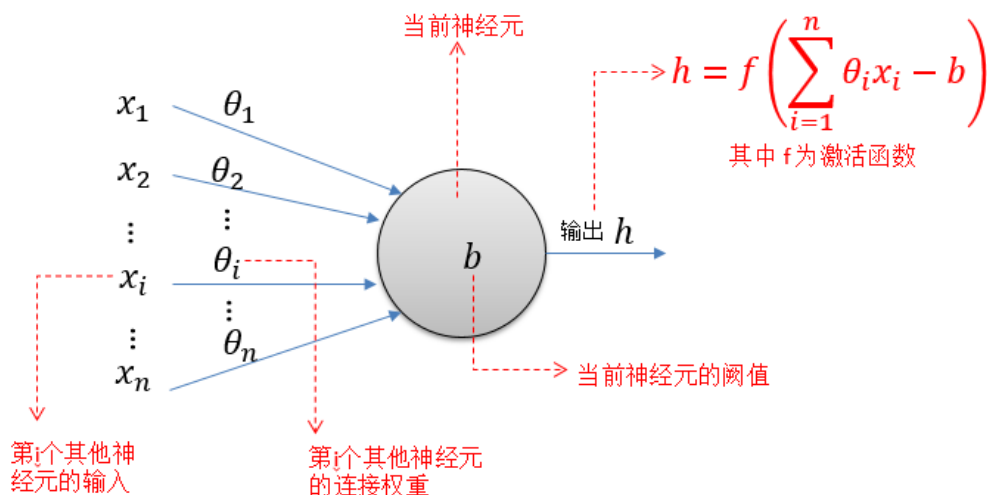


图 1.3: Artificial Neuron Model

激活函数

理想中的激活函数 $f(x)$ 是图 (a) 所示的阶跃函数，它将输入值映射为输出值“0”或“1”，“1”对应神经元兴奋，“0”对应神经元抑制。但是，阶跃函数具有不连续，不光滑（不连续可导）等不太好的性质¹，因此实际中常用 Logistic 回归中应用到的 **sigmoid 函数** 作为激活函数。典型的 sigmoid 函数如图 (b) 所示，它把可能在较大范围内变化的输入值挤压到 (0,1) 输出值范围内，因此有时又称之为“挤压函数” (squashing function)。

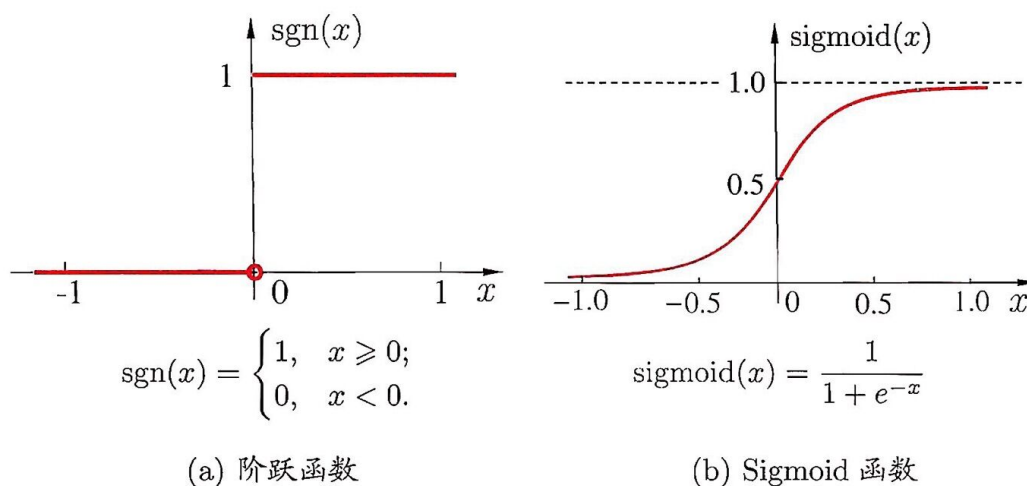


图 1.4: 激活函数

¹事实上在使用阶跃函数时，神经网络中的权值或偏置的微小改变可能会造成输出结果天翻地覆的变化，这使得网络的行为变得复杂且难以控制。

1.2.2 人工神经网络的构成 (Structure of ANN)

■ 感知机 (Perceptron)

感知机 (Perceptron) 由两层神经元组成, 如图所示, 输入层接收外界输入信号后传递给输出层, 输出层是 M-P 神经元, 亦称“阈值逻辑单元”(threshold logic unit)

感知器处理单元对 n 个二进制输入 x_1, x_2, \dots , 进行加权和操作并产生一个二进制输出:

$$y = f\left(\sum_i w_i x_i - \theta\right)$$

感知机能容易地实现逻辑与、或、非运算:

- 先定义 $f(x)$ 为阶跃函数
- “与” ($x_1 \wedge x_2$): 令 $w_1 = w_2 = 1, \theta = 2$; 则仅在 $x_1 = x_2 = 1$ 时, $y = 1$;
- “或” ($x_1 \vee x_2$): 令 $w_1 = w_2 = 1, \theta = 0.5$; 则仅在 $x_1 = x_2 = 1$ 时, $y = 1$;
- “非” ($\neg x_1$): 令 $w_1 = -0.6, w_2 = 0, \theta = -0.5$; 那么当 $x_1 = 1$ 时, $y = 0$;

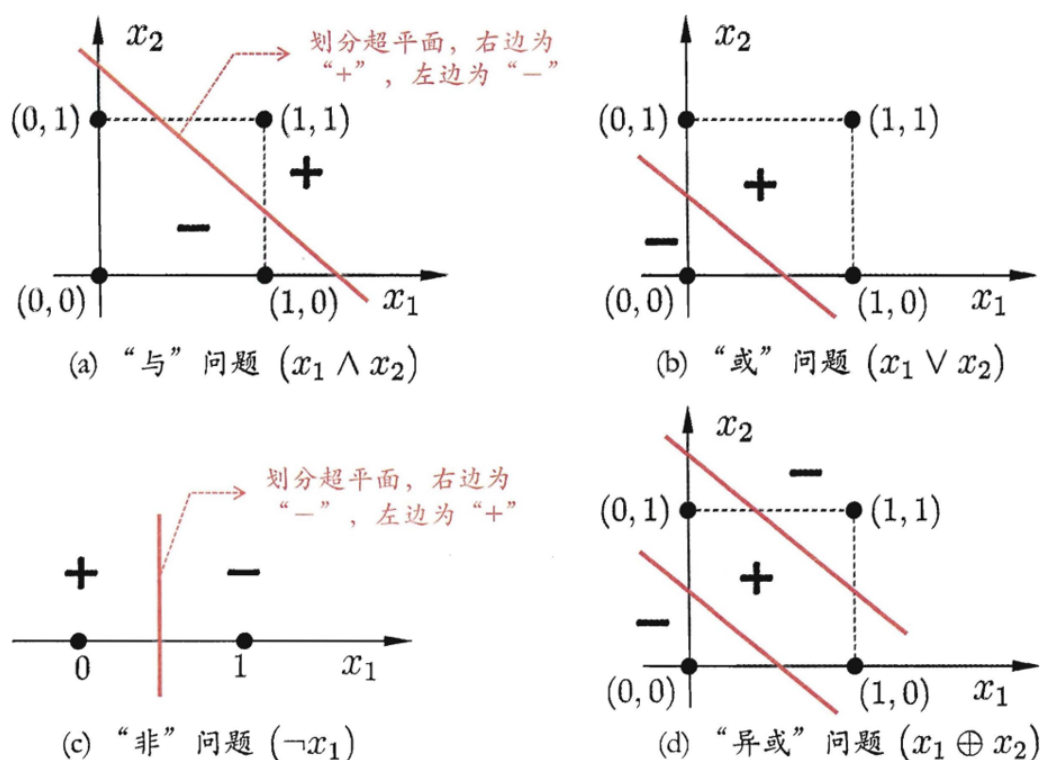


图 1.5: 与、或、非运算的实现

需注意的是, 感知机只有输出层神经元进行激活函数处理, 即只拥有一层功能神经元 (functional neuron), 其学习能力非常有限, 只能处理线性可分 (linearly separable) 的问题。

从图中可以看出, 与、或、非问题的分类样本是可以用一条直线分开的, 即具有“线性可分”性质。

■ 多层感知机 (MLP)

要解决非线性可分问题, 就需考虑多层功能神经元。

而“异或”问题的分类样本需要两条线才能将其分开, 故需要两层感知机才能实现, 如图 1.6

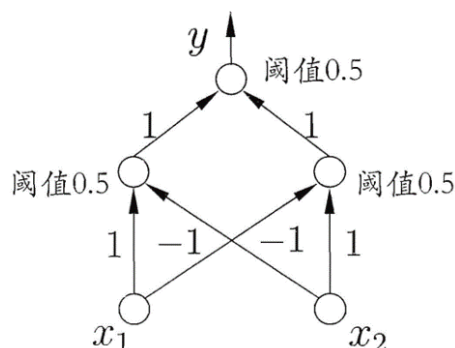
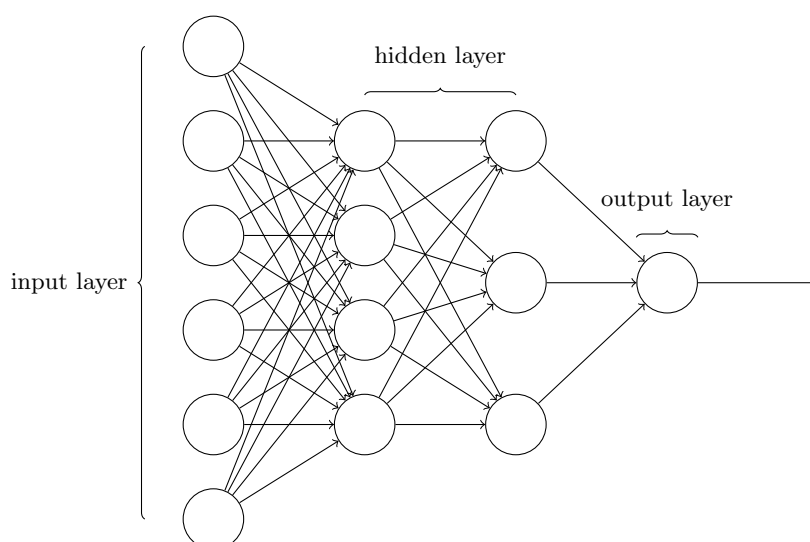


图 1.6: 能解决异或问题的两层感知机

多层感知机 (MLP) 的结构如下:



- 这个网络中最左边的称为**输入层**, 其中的神经元称为**输入神经元**。
- 最右边的, 即**输出层**包含有**输出神经元**。
- 中间层, 这层中的神经元既不是输入也不是输出, 被称为**隐藏层**。
- 隐含层和输出层神经元都是拥有**激活函数**的功能神经元。
- 只需包含隐层, 即可称为多层网络. 神经网络的学习过程, 就是根据训练数据来调整神经元之间的“**连接权**” (connection weight) 以及每个功能神经元的**阈值**;
- 换言之, 神经网络“学”到的东西, 蕴涵在**连接权**与**阈值**中。

1.2.3 人工神经网络的学习 (Training of ANN)

反向传播算法 (error Back Propagation, 简称 BP) 是迄今最成功的神经网络学习算法. 它可用来学习由一系列确定的单元互连形成的多层网络的权值. 它采用**梯度下降方法**试图最小化网络输出值和目标值之间的**误差平方**. 其主要思想是:

1. 将训练集数据输入到 ANN 的输入层, 经过隐藏层, 最后达到输出层并输出结果, 这是 ANN 的前向传播过程;
2. 由于 ANN 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;
3. 在反向传播的过程中, 根据误差调整各种参数的值; 不断迭代上述过程, 直至收敛。

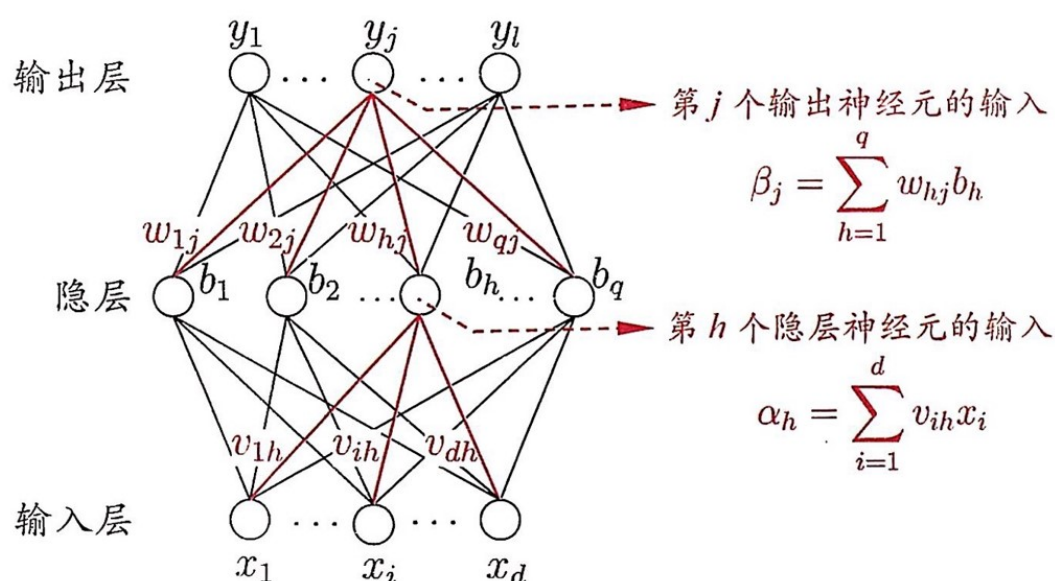


图 1.7: 神经网络

Algorithm 1 BP Algorithm

Input:

训练集 $D = (x_k, y_k)_{k=1}^m$, 学习率 η

Output:

连接权与阈值确定的多层前馈神经网络

- 1: 在 (0,1) 范围内随机初始化网络中所有连接权和阈值
 - 2: **while** 未达到停止条件 **do**
 - 3: 计算当前样本的输出 \hat{y}_k
 - 4: 计算输出层神经元的梯度项 g_j
 - 5: 计算隐层神经元的梯度项 e_h
 - 6: 更新连接权 w_{hj}, v_{ih} 与阈值 θ_j, γ_h
 - 7: **end while**
-

1.3 神经网络应用现状

人工神经网络 (ANN) 以其独特的结构和处理信息的方法, 在许多实际应用领域中取得了显著的成效, 主要应用如下:

- **模式识别**

在模式识别实践中通常存在着噪声干扰和输入输出信息缺失的情况, 神经网络的分布式信息存储特性使得其具有良好的容错性和鲁棒性, 因此在解决模式识别问题时具有较大的优越性。事实上, 模式识别技术的广泛应用也是人工神经网络技术发展的重要方面。神经网络在这个领域中有许多成功的应用, 例如目前市场上被人们广为熟知的手写输入和语音输入等智能系统。

- **自动控制**

神经网络的诸多特点决定了其应用于自动控制领域的优越性。如神经网络本质上就是并行计算、分布结构的非线性动力学系统, 可应用于那些难以用模型描述的过程或系统及应用与实时性要求较高的自动控制领域。另外, 神经网络具有较强的信息综合能力, 在多变量复杂系统的控制方案设计上具有一定明显的优势。

- **信号处理**

神经网络的自学习和自适应特性使其成为一种对各类信号进行多用途综合加工处理的天然工具, 被广泛地应用于自适应信号处理和非线性信号处理。

- **人工智能**

人们在应用专家系统解决类似于人脑形象思维的问题时遇到了较大的困难, 而人工神经网络的发展则为本已遇到“瓶颈”问题的人工智能带来了新的出路。这主要得益于人工神经网络的自学习、自适应能力和良好的容错特性。这方面的主要应用有自然语言处理、逻辑推理、模糊判断、系统诊断、密码破译、智能机器人等。

以上只是大概地简单列举了四个方面的应用。事实上, 人工神经网络的应用领域是十分广泛的。例如, 人工神经网络还可以用来解决最优化问题 (包括组合优化、条件约束优化等这类问题) 或者开发出用于矩阵/向量运算的数字计算机协处理器, 在经济领域中也可以用来进行市场预测和信贷分析等方面的工作。此外, 在医学领域、军事领域、化学领域甚至娱乐视听领域, 人工神经网络都取得了一定的成功应用。这也正是科学技术转化为具体生产力的表现。相信随着人工神经网络研究的进一步深入和发展, 特别是当人工神经网络作为一种智能化方法同其他学科和领域更为紧密地结合起来的时候, 人工神经网络的应用前景也将更为广阔。

反向传播算法的核心是一个对代价函数 C 关于任何权重 w 和偏置 b 的偏导数 $\partial C/\partial w, \partial C/\partial b$ 的表达式。这个表达式告诉我们在改变权重和偏置时，代价函数变化的快慢，所以反向传播不仅仅是一种学习的快速算法。实际上它让我们细致领悟如何通过改变权重和偏置来改变整个网络的行为。