

机器学习

Machine Learning

北京航空航天大学计算机学院

School of Computer Science and Engineering, Beihang University

黄迪 张永飞 陈佳鑫

2023年秋季学期

Fall 2023

第7章：人工神经网络

Chapter 7: Artificial Neural Networks

发展概述

● 萌芽期(20世纪40年代)

- 上世纪40年代初，美国McCulloch和Pitts从信息处理的角度，研究神经细胞行为的数学模型表达，提出了MP二值神经元模型。**MP模型**的提出开始了对神经网络的研究进程。
- 1949年，心理学家Hebb提出著名的**Hebb学习规则**，即由神经元之间结合强度的改变来实现神经学习的方法。



W. S. McCulloch
(1898-1969)



W. Pitts
(1923-1969)



Donald O. Hebb
(1904-1985)

发展概述

- 第一次高潮期(20世纪50-60年代)
 - 现代串行计算机的奠基人Von Neumann注意到计算机与人脑的差异，对类似于神经网络的分布系统做了许多研究。
 - 上世纪50年代末期，Rosenblatt提出**感知机模型**，它基本符合神经生理学的原理。在60代掀起了神经网络研究的第一次高潮。



Von Neumann
(1903-1957)



Frank Rosenblatt
(1928-1971)

发展概述

● 反思期(1969年-1980年)

- 1969年，人工智能创始人之一的Minsky和Papert出版了《感知器》一书，从数学上深入分析了感知器的原理，指出其局限性。Minsky持悲观态度。
- 串行计算机正处于全盛发展时期，早期人工智能的研究成果显著，掩盖了发展新计算模型的迫切性。使对神经网络的研究进入了低潮期。不少科学家坚持这一研究，为神经网络的发展奠定了理论基础。



Marvin Lee Minsky
(1921-)



Seymour Papert
(1928-)

发展概述

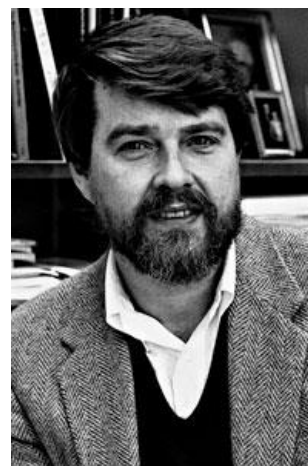
- 第二高潮期(1981年~1990年)
 - 1982年，Hopfield提出了人工神经网络的一种数学模型，引入了能量函数的概念，研究了网络的动力学性质；设计出用电子线路实现这一网络的方案，同时开拓了神经网络用于联想记忆和优化计算的新途径，大大促进了人工神经网络的研究。



John J. Hopfield
(1933-)

发展概述

- **第二高潮期(1981年~1990年)**
 - 1986年，Rumelhart及LeCun等学者提出了多层感知器的反向传播算法，解决了多层前向神经网络的学习问题，克服了当初阻碍感知机发展的重要障碍，并证明了多层神经网络具有很强的学习能力。



David Rumelhart
(1942-2011)



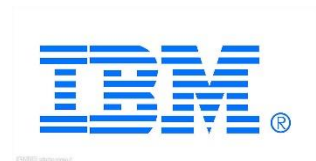
Yann Lecun
(1960-)

发展概述

- 再认识与应用研究期(1991年~2006年)
 - 问题
 - 应用面不够宽
 - 结果不够精确
 - 研究
 - 开发现有模型的应用，并在应用中根据实际运行情况对模型、算法加以改进，以提高网络的训练速度和运行准确度。
 - 在理论上寻找新的突破，建立新的专用/通用模型和算法。
 - 进一步对生物神经网络进行研究，不断丰富对人脑的认识。

发展概述

- 深度神经网络时期(2006年~)
- 2006年, Hinton提出了更有效的训练神经元的新方法, 通过逐层初始化(Layer-Wise Pre-Training)来克服多层神经网络在训练上的难度。
- 2012年, Alex Krizhevsky在ImageNet图像分类竞赛中取得了里程碑式的突破, Top5分类准确率由74%提升至85%.

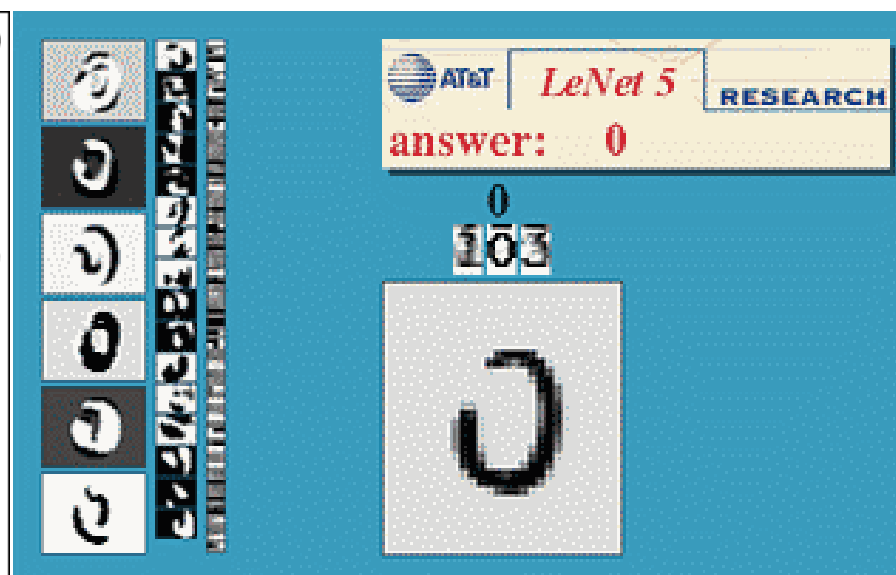


发展概述



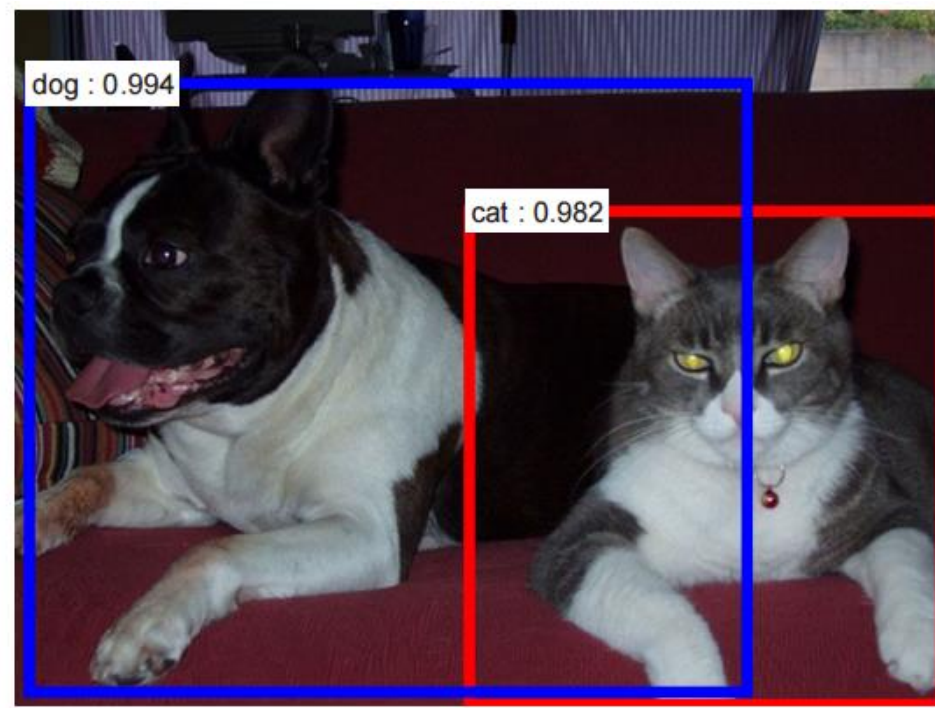
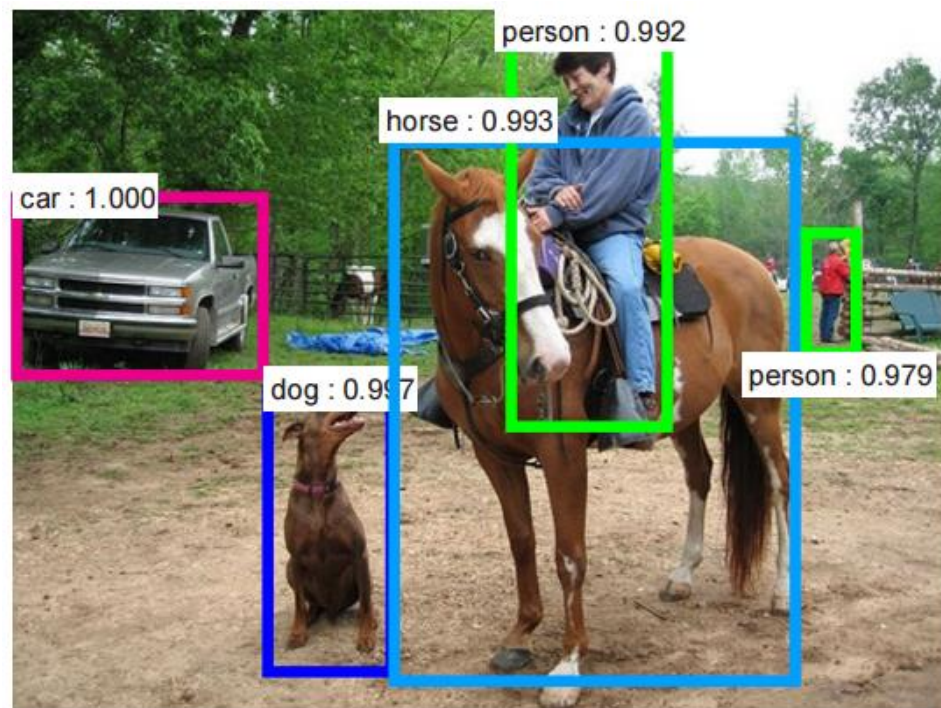
典型应用

● 手写体数字识别



典型应用

● 物体检测



典型应用

● 画风迁移



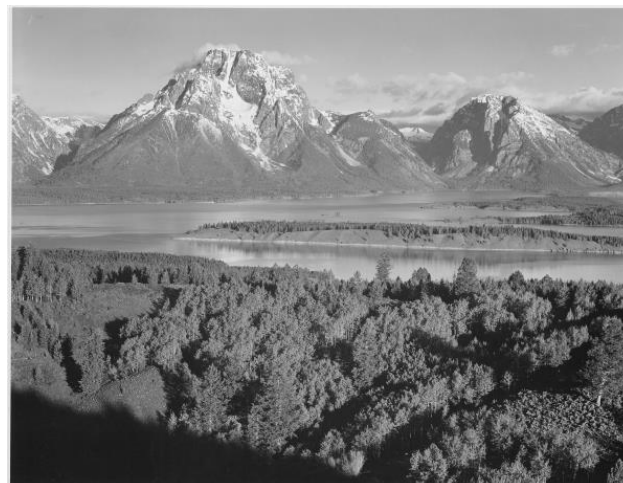
典型应用

● 图像着色

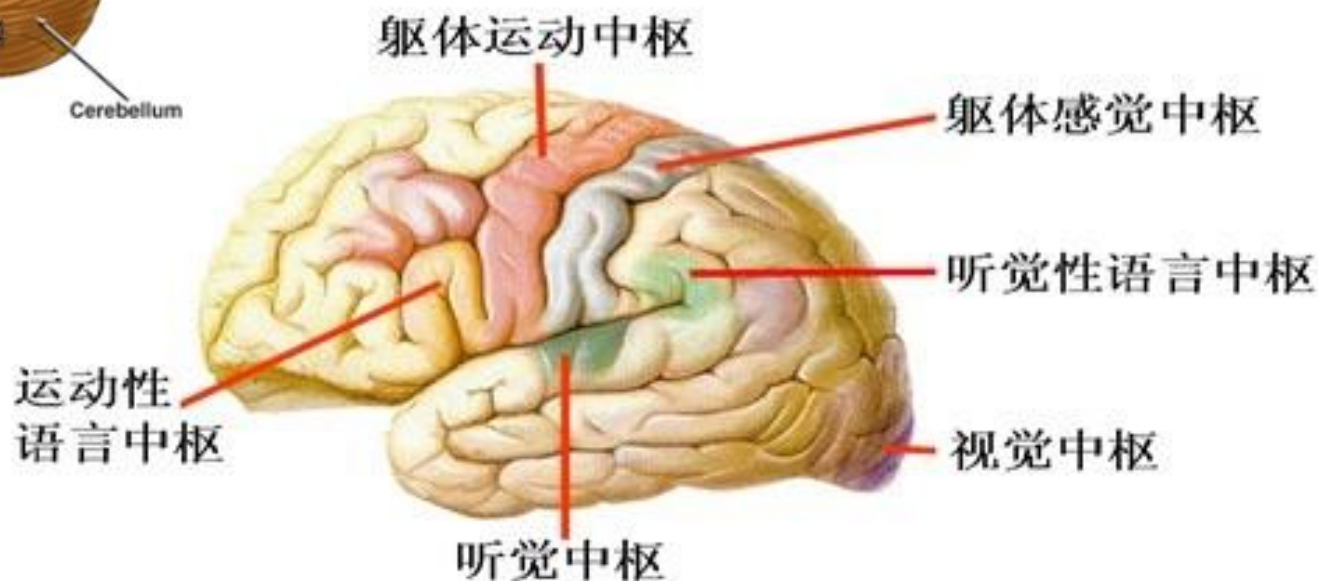
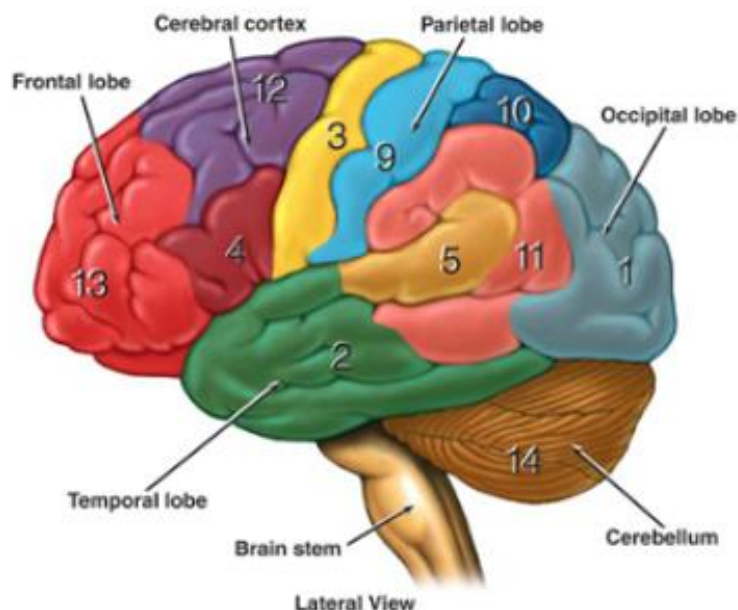
S. J. Barley, 1933



Mount Moran, 1941



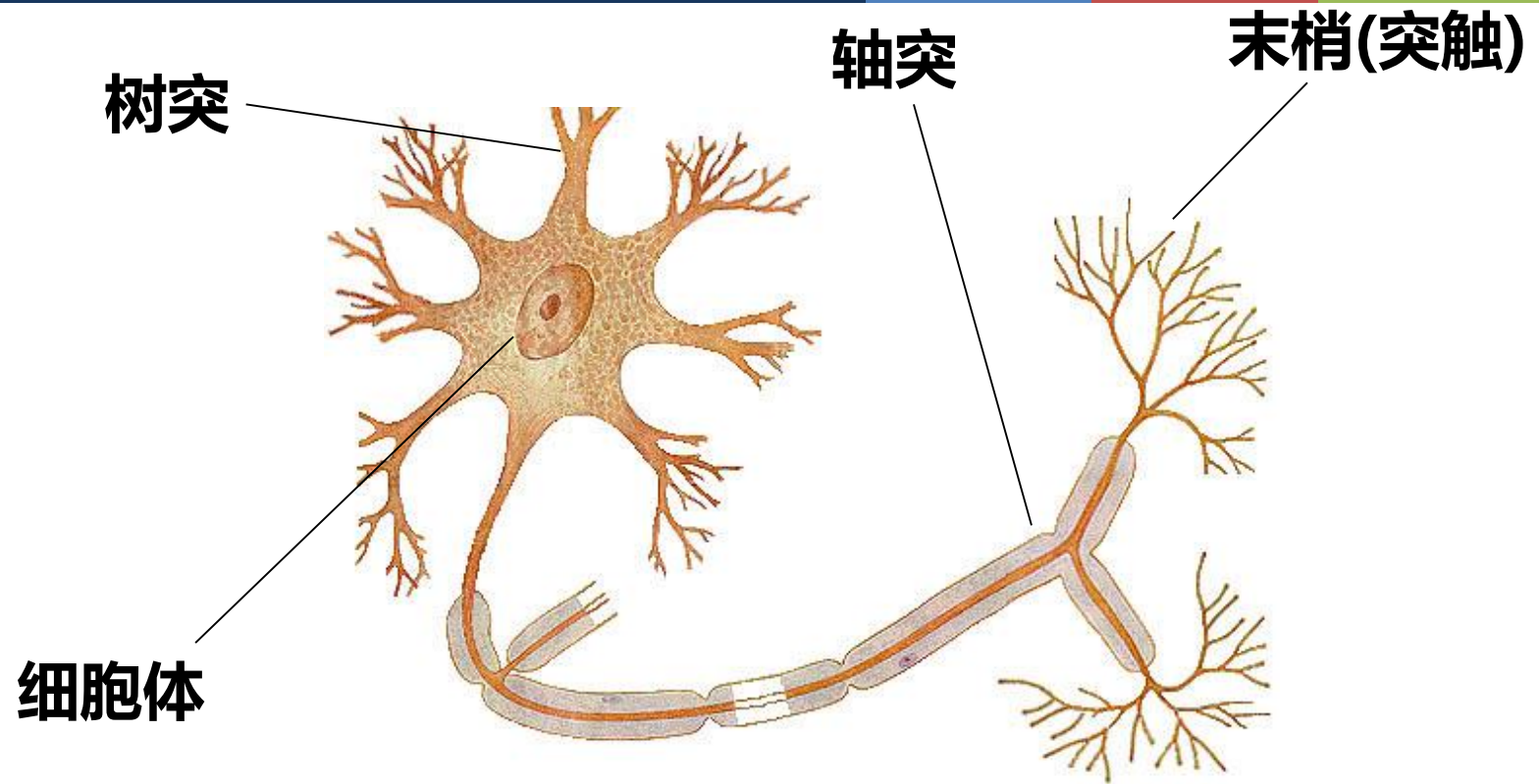
生物神经网络



生物神经网络



神经元结构



人类大脑皮层大约有**14,000,000,000**个神经细胞(神经元)。每个神经元有数以千计的通道同其它神经元相互连接，形成复杂的生物神经网络。

神经元结构

- **细胞体(Soma)**

神经元的主体，是神经元的代谢和营养中心

- **轴突(Axon)**

传导神经冲动，细胞体→其他神经元或效应细胞

- **树突(Dendrites)**

接受刺激并将冲动传入细胞体

神经元结构

● 工作机制

- 一个神经元有两种状态：兴奋和抑制；
- 平时处于抑制状态的神经元，其树突和胞体接收其他神经元由突触传来的兴奋电位，多个输入在神经元中以代数和的方式叠加；
- 如果输入兴奋电位总量超过某个阈值，神经元会被激发进入兴奋状态，发出输出脉冲，并由突触传递给其他神经元。
- 神经元被触发后进入不应期，在不应期不能被触发，然后阈值逐渐下降，恢复兴奋性。

神经元结构

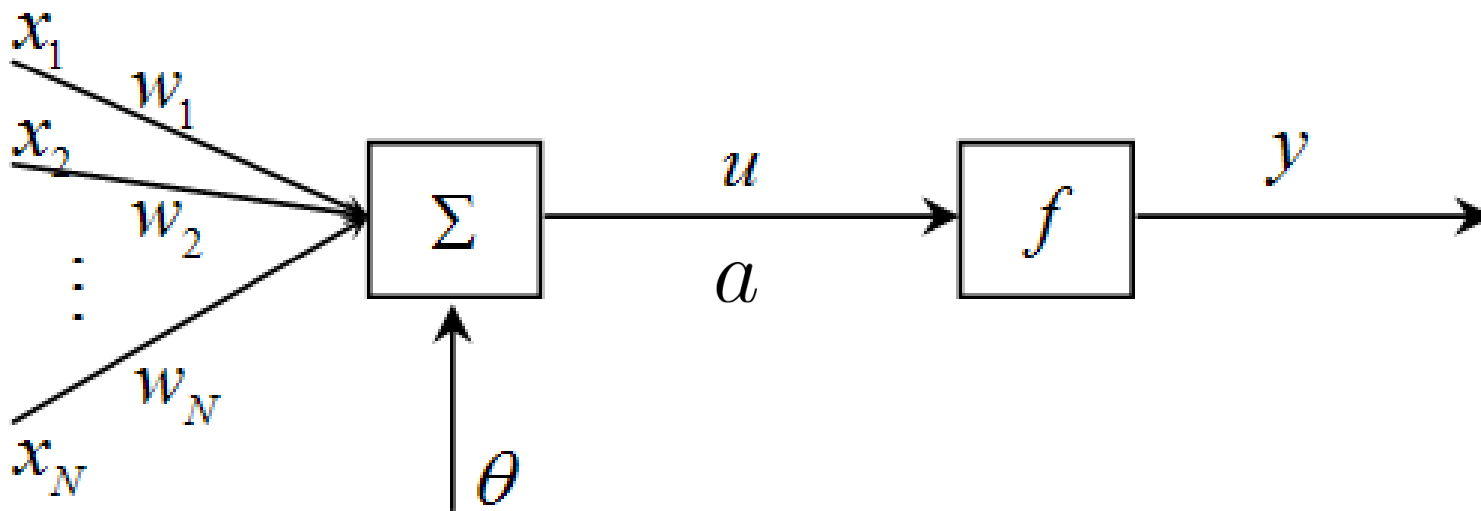
● 基本特征

- 神经元及其联接；
- 神经元之间的联接强度决定信号传递的强弱；
- 神经元之间的联接强度是可以随训练改变的；
- 信号可以是起刺激作用的，也可以是起抑制作用的；
- 一个神经元接受信号的累积效果决定该神经元的状态；
- 每个神经元可以有一个“阈值”。

人工神经元的MP模型

● MP模型

一种人工神经元的数学模型，它最早由美国的McCulloch和Pitts提出的神经元模型，是大多数神经网络模型的基础。



人工神经元的MP模型

人工神经元是仿照生物神经元提出的，神经元可以有 N 个**输入**：

$$x_1, x_2, \dots, x_N$$

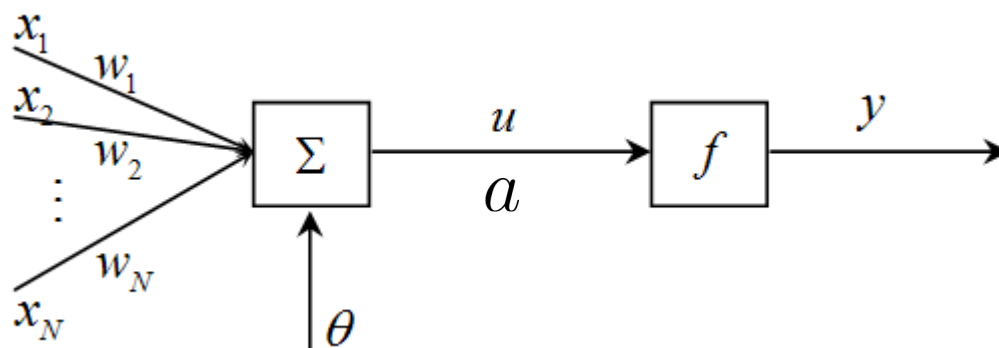
每个输入端与神经元之间有一定的**联接权值**：

$$w_1, w_2, \dots, w_N$$

神经元总输入为对每个输入的加权求和，同时减去**阈值** θ 。 a 代表神经元的活跃值，即**神经元状态**：

$$a = \sum_{i=1}^N w_i x_i - \theta = \sum_{i=0}^N w_i x_i$$

人工神经元的MP模型



神经元的输出 y 是对 a 的映射：

$$y = f(a) = f \left(\sum_{i=0}^N w_i x_i \right)$$

f 称为激活函数(激励函数，输出函数)。

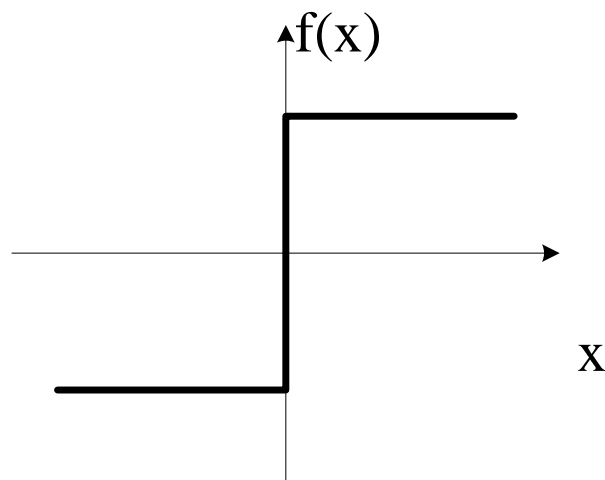
人工神经元的MP模型

MP模型采用**阶跃函数**作为激活函数

$$f(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}$$



$$y = \text{sgn} \left(\sum_{i=0}^N w_i x_i \right)$$

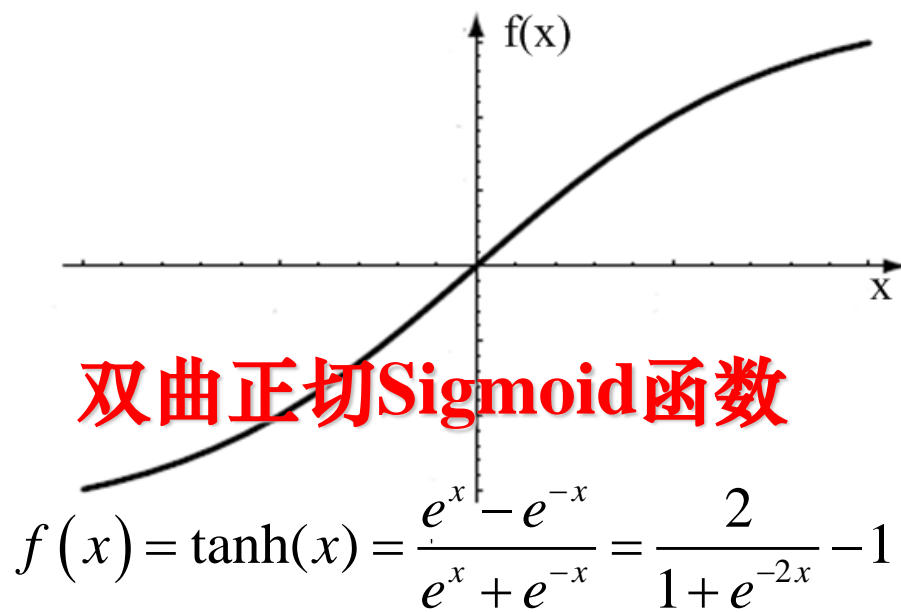
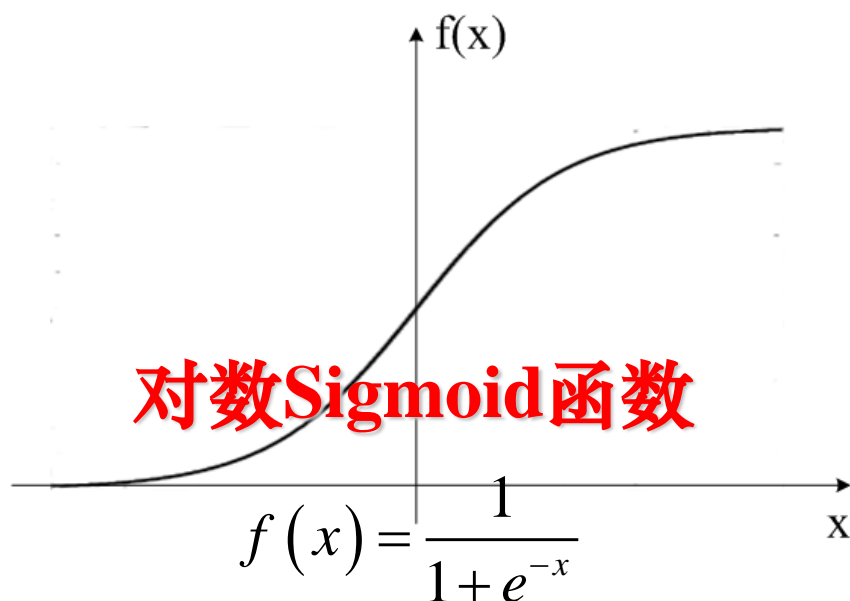


当激活函数 f 为阈值(阶跃)函数时，神经元就可以看作是一个**线性分类器**。

人工神经元的MP模型

● 激活函数

- 线性函数
- 非线性斜面函数(Ramp Function)
- Sigmoid输出函数(S型函数)
-

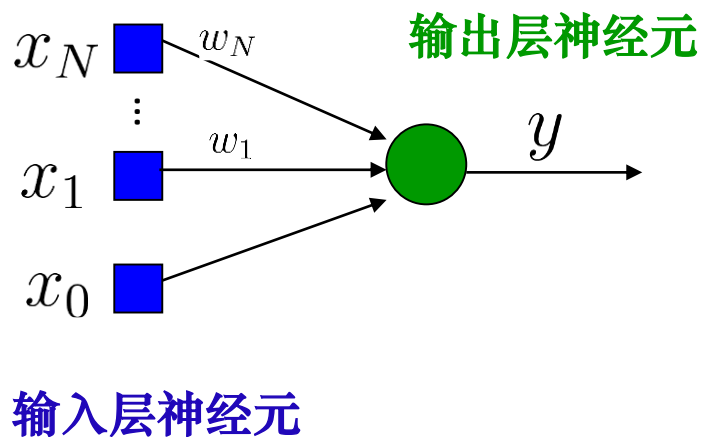


感知器

- 感知器是一种早期的神经网络模型，由美国学者F. Rosenblatt于1957年提出。
- 感知器中第一次引入了学习的概念，使人脑所具备的学习功能在基于符号处理的数学到了一定程度的模拟，所以引起了广泛的关注。
- 简单感知器模型实际上仍然是MP模型的结构。它是一种双层神经网络模型，一层为输入层(有时不把输入层计入网络层数之中)，另一层具有计算单元，可以通过监督学习来逐步增强模式划分的能力，达到学习的目的。

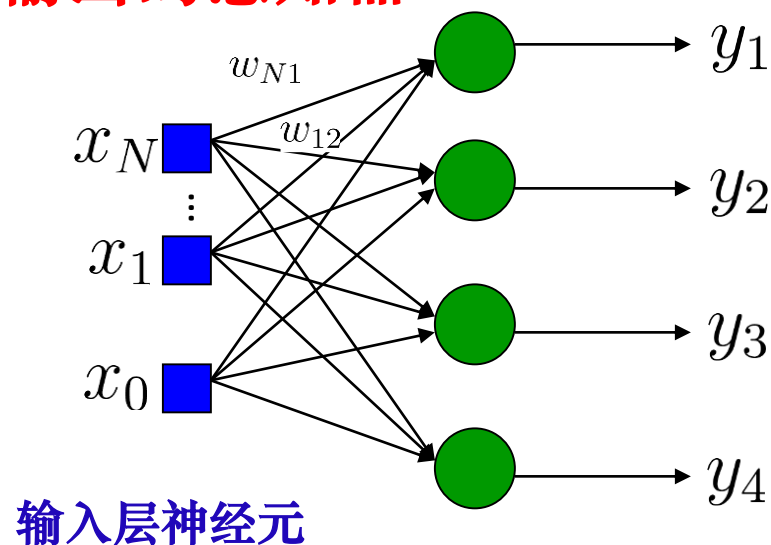
感知器

单输出的感知器(M-P模型)



$$y = f(a) = f\left(\sum_{i=0}^N w_i x_i\right)$$

多输出的感知器



$$y_j = f(a_j) = f\left(\sum_{i=0}^N w_{ij} x_i\right)$$

感知器

- 单个神经元可以实现两类问题的线性分类，多个感知器则可以实现多类别问题的线性分类。

例子中的网络就可以实现四类问题的分类。

在训练时，理想输出为：

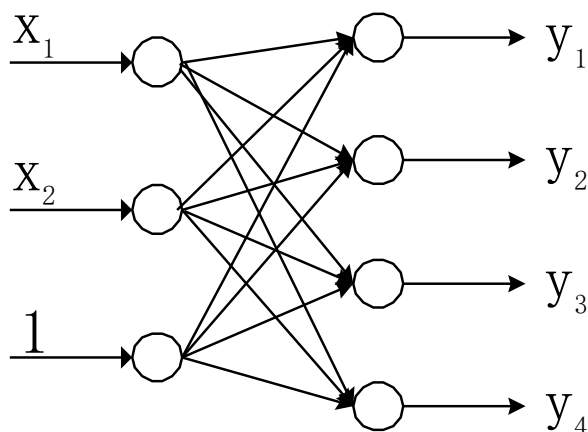
第1类训练样本：(1, 0, 0, 0) (1, 0)

第2类训练样本：(0, 1, 0, 0) (0, 1)

第3类训练样本：(0, 0, 1, 0) (0, 0)

第4类训练样本：(0, 0, 0, 1) (1, 1)

也就是每个神经元输出为1代表某一类别。

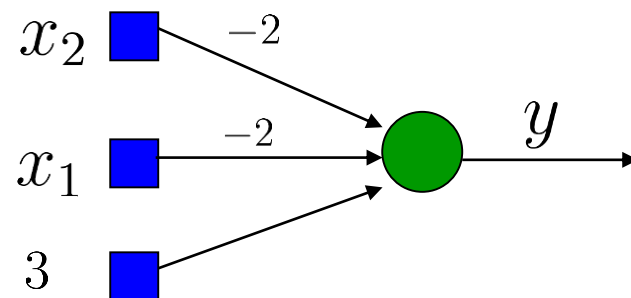


感知器

● 示例：与非门

真值表

		y	
		0	1
x	0	1	1
	1	1	0



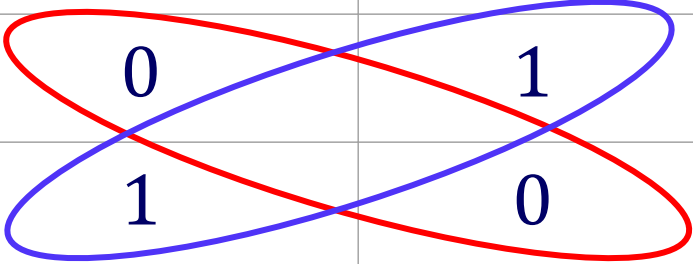
x_1	x_2	a	y
1	1	-1	0
1	0	1	1
0	1	1	1
0	0	3	1

感知器

- **线性不可分**问题：感知器模型的局限

例：异或(XOR)问题

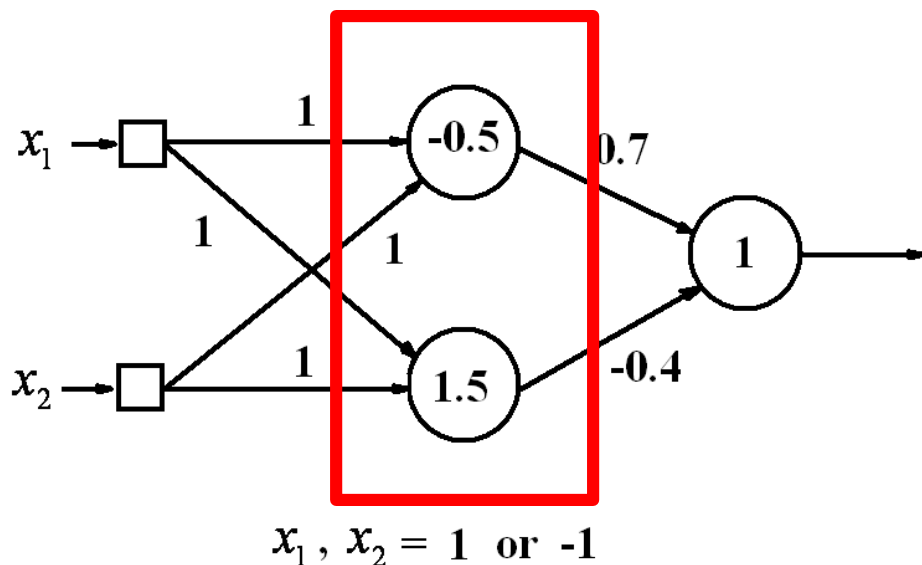
		y	
		0	1
x	0	0	1
	1	1	0



解决方法：三层感知器

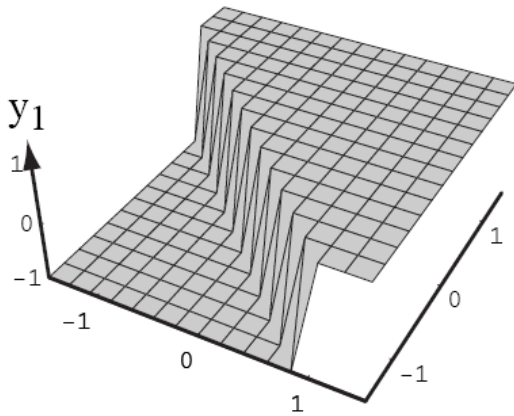
多层感知器

- 对于异或问题，用一个简单的三层感知器就可得到解决。

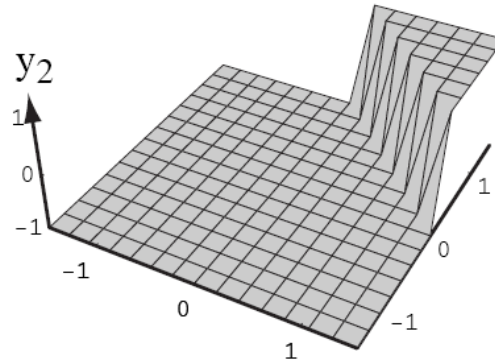


x_1	x_2	y
-1	-1	-1
-1	1	1
1	-1	1
1	1	-1

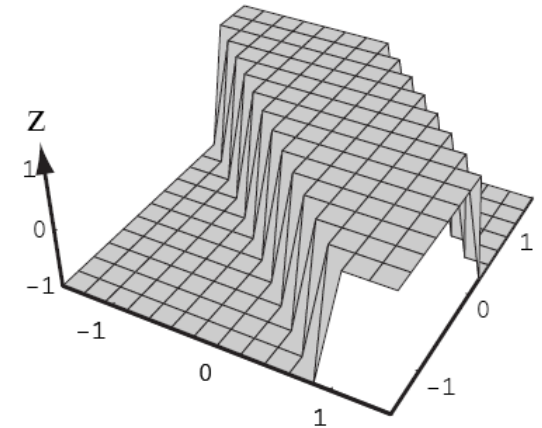
多层感知器



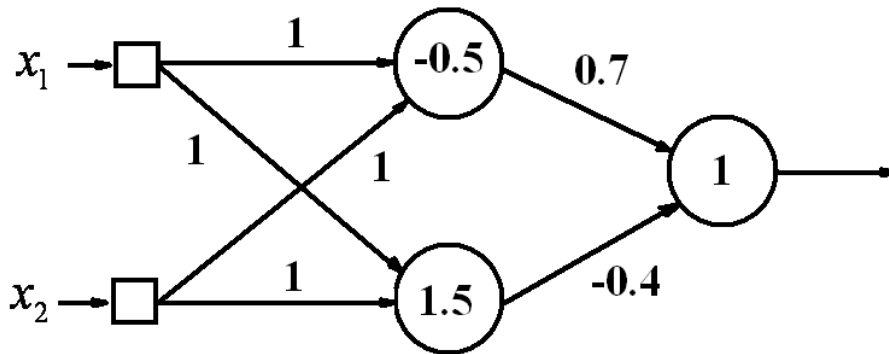
$$x_1 + x_2 + 0.5 = 0$$



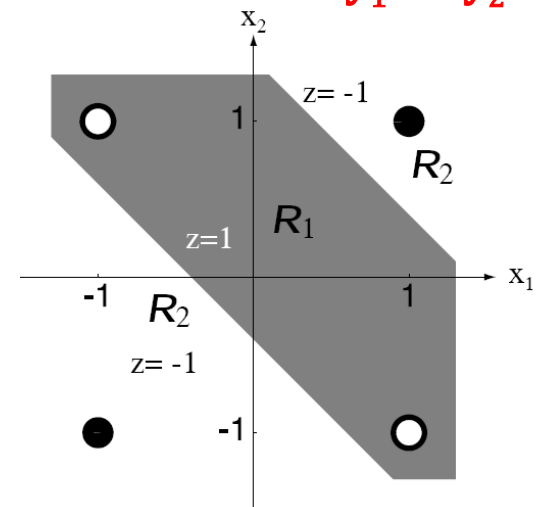
$$x_1 + x_2 - 1.5 = 0$$



$$0.7y_1 - 0.4y_2 - 1 = 0$$



$$x_1, x_2 = 1 \text{ or } -1$$

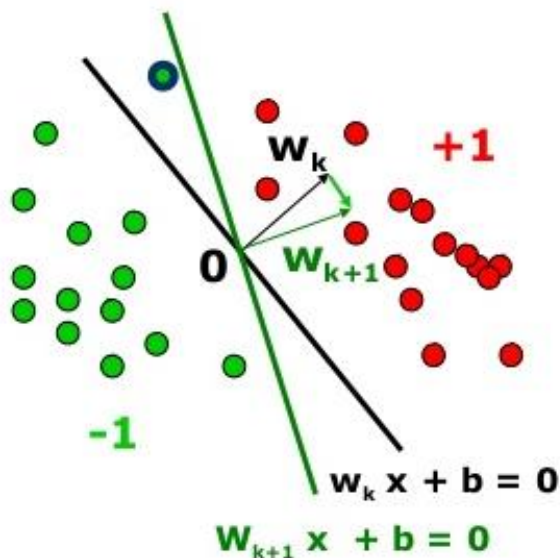


感知器

- 感知器训练算法的基本原理来源于著名的Hebb学习律(1949年提出)

- 基本思想

逐步地将样本集中的样本输入到网络中，根据输出结果和理想输出之间的差别来调整网络中的权重。



感知器

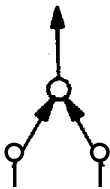
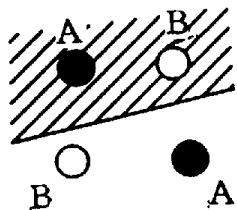
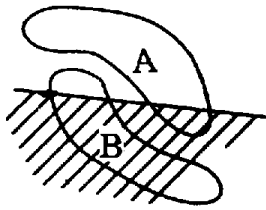

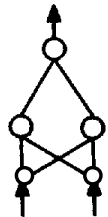
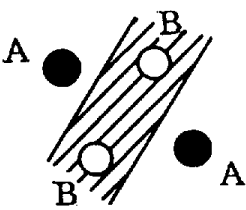
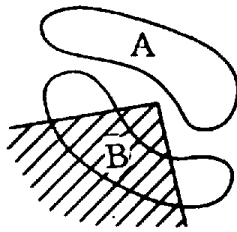

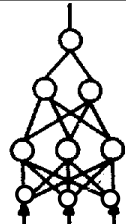
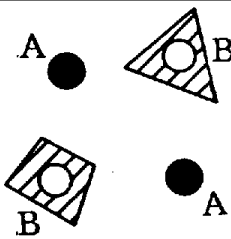
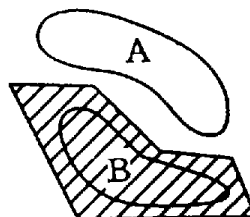
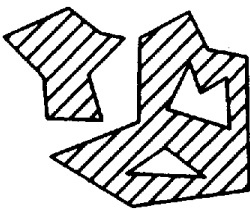
● Hebb学习规则

- 神经网络具有学习功能。对于人工神经网络而言，这种学习归结为神经元连接权的变化。
- Hebb学习规则调整 w_{ij} 的原则：若第 i 个神经元 u_i 和第 j 个神经元 u_j 同时处于兴奋状态，则它们之间的连接 w_{ij} 应当加强，即：

$$\Delta w_{ij} = \eta x_i y_j$$

这一规则与“条件反射”学说一致，并已得到神经细胞学说的证实。 η 是表示学习速率的比例常数(学习率)。

感知器小结

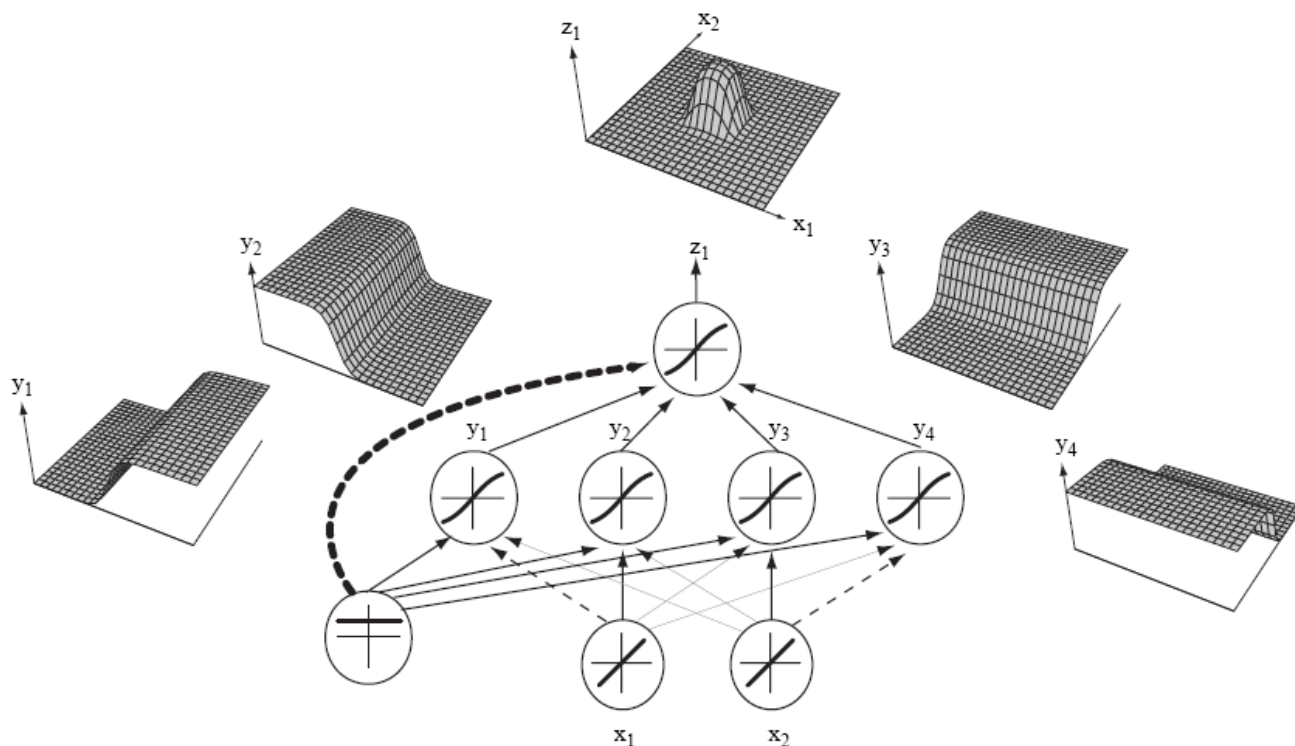
感知器结构	异或问题	复杂问题	判决域形状	判决域
无隐层 				半平面
单隐层 				凸 域
双隐层 				任意复杂 形状域

理论证明，三层感知器可以实现任意的逻辑运算，在激活函数为Sigmoid函数的情况下，可以逼近任何非线性多元函数。

感知器小结

● 输出函数：S型函数

若给出足够多隐单元，任何从输入到输出的连续函数都可用三层神经网络以任意精度近似。



感知器

● 小结

- 感知器的学习过程与求解线性判决函数(单样本感知器算法)的过程等价；
- 两层感知器(一个输入层和一个输出层)只能解决线性可分问题；
- 只要隐层和隐层单元数足够多，多层感知器网络可实现任何模式分类；
- 确定多层网络的权值，即网络如何进行学习，在感知器上没有得到解决。