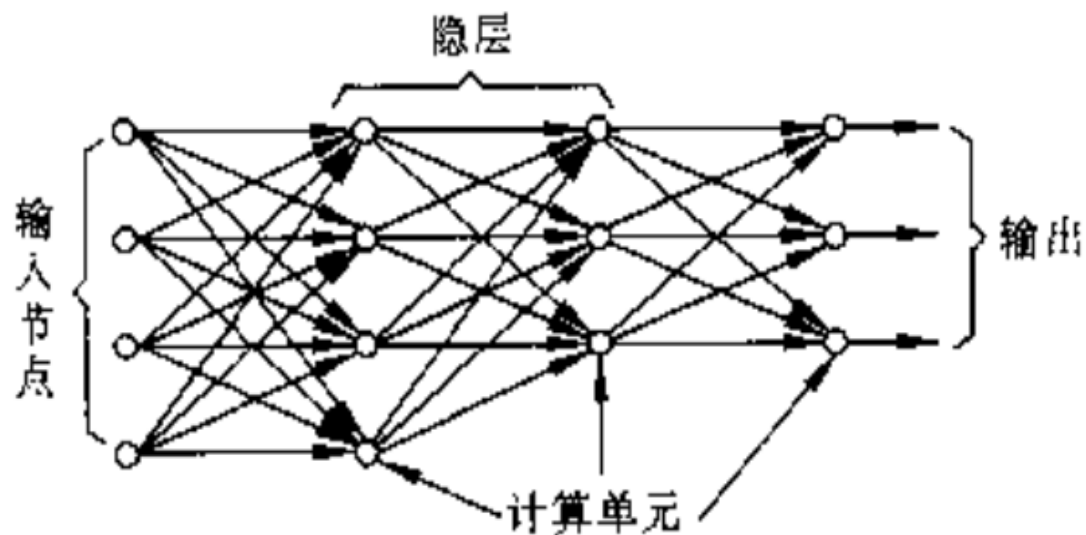


# 前馈神经网络及其主要算法

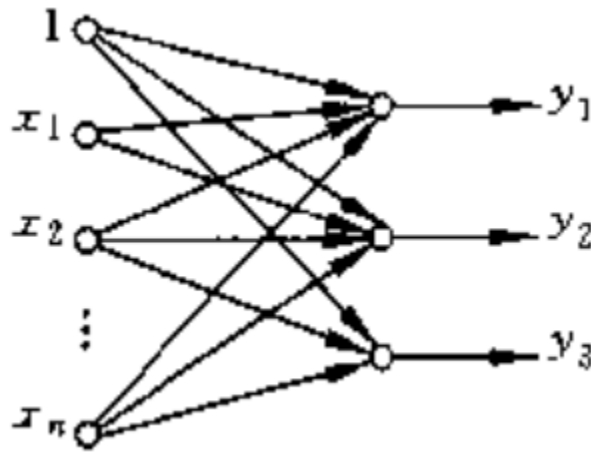
什么叫前馈神经网络 (feed-forward NN) ?

构成前馈网络的各神经元接受前一级输入,并输出到下一级,无反馈,可用一有向无环图表示。图的节点分为两类,即输入节点与计算单元。每个计算单元可有任意个输入,但只有一个输出,而输出可耦合到任意多个其他节点的输入。前馈网络通常分为不同的层,第 $i$ 层的输入只与第 $i-1$ 层的输出相联,这里认为输入节点为第一层,因此所谓具有单层计算单元的网络实际上是一个两层网络<sup>①</sup>。输入和输出节点由于可与外界相连,直接受环境影响,称为可见层,而其他的中间层则称为隐层,如图 11.3 所示。



## 淘宝店铺-酷流科技 掌柜：我是雷锋的朋友

前面我们在线性判别函数角度学习了感知器，现在我们从神经网络的角度来回顾感知器



感知器

$$y_i \in \{0, 1\}$$

$$Y = [y_1, y_2 \dots y_m]^T$$

$$\hat{Y} = [\hat{y}_1, \hat{y}_2 \dots \hat{y}_m]^T$$

$$W_{ij}(t+1) = W_{ij}(t) + \Delta W_{ij}(t)$$

$$\Delta W_{ij} = \eta (y_j - \hat{y}_j) x_i$$

如果使用MSE准则

$$\Delta W_{ij} = \frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial \hat{y}_j} \cdot \frac{\partial \hat{y}_j}{\partial net_j} \cdot \frac{\partial net_j}{\partial W_{ij}}$$

$$= \frac{\partial}{\partial \hat{y}_j} \sum_i \frac{1}{2} (y_i - \hat{y}_i)^2 \cdot f'(net_j)$$

$$\cdot x_i$$

$$= -(y_j - \hat{y}_j) \cdot 1 \cdot x_i$$

感知器的学习过程与求取线性判别函数的过程是等价的,此处只指出感知器的一些特性:

- ①两层感知器只能用于解决线性可分问题。
- ②学习过程收敛很快,且与初始值无关。

淘宝店铺-酷流科技 掌柜:我是雷锋的朋友

## 三层前馈网络

利用人工神经元的非线性特性,可以实现各种逻辑门。

①任何逻辑函数都可以用前馈网络实现。

②单个阈值神经元可以实现任意多输入的与、或及与非、或非门;由于任何逻辑函数都可以化为析取(或合取)形式,所以任何逻辑函数都可用一个三层(只有两层计算单元)的前馈网络实现。

当神经元的输出函数为Sigmoid函数时,上述结论可以推广到连续的非线性函数,在很宽松的条件下,三层前馈网络可以逼近任意的多元非线性函数,突破了二层前馈网络线性可分的限制。这种三层或三层以上的前馈网络通常又被叫做多层感知器(Multi-Layer Perceptron,简称MLP)。

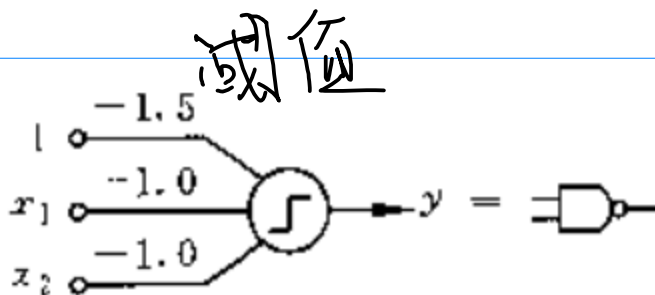


图 11.5 用感知器模型实现  
“与非”逻辑

$$-x_1 - x_2 + 1.5 \geq 0$$

1	1	-0.5
1	-1	1.5
-1	-1	3.5

## 用多个感知器实现非线性分类

正如在第4章中看到的，单个感知器神经元能够完成线性可分数据的分类问题，是一种最简单的可学习机器。但是，它无法解决非线性问题。比如，在图5-8的四个点中，如果(1,1)点和(-1,-1)点同属于第一类，而(1,-1)点和(-1,1)点同属于第二类，这一问题在逻辑学里称作异或(XOR)问题。对于这样的问题，单个感知器神经元是无法正确分类的。

异或

单个阈值逻辑单元神经元作为分类器的这一局限早在1969年Minsky和Papert的专著《感知器》里就进行了透彻的分析<sup>①</sup>。他们证明了感知器只能解决所谓一阶谓词逻辑问题，如与

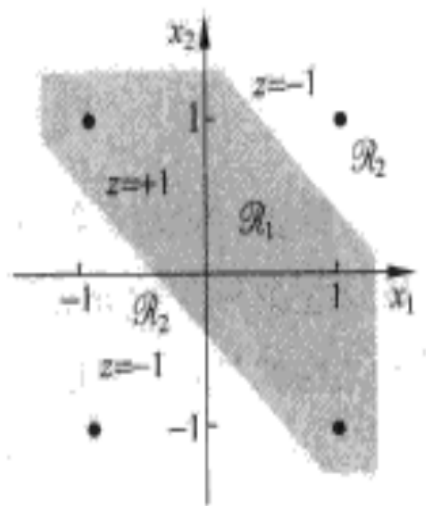


图5-8 异或问题

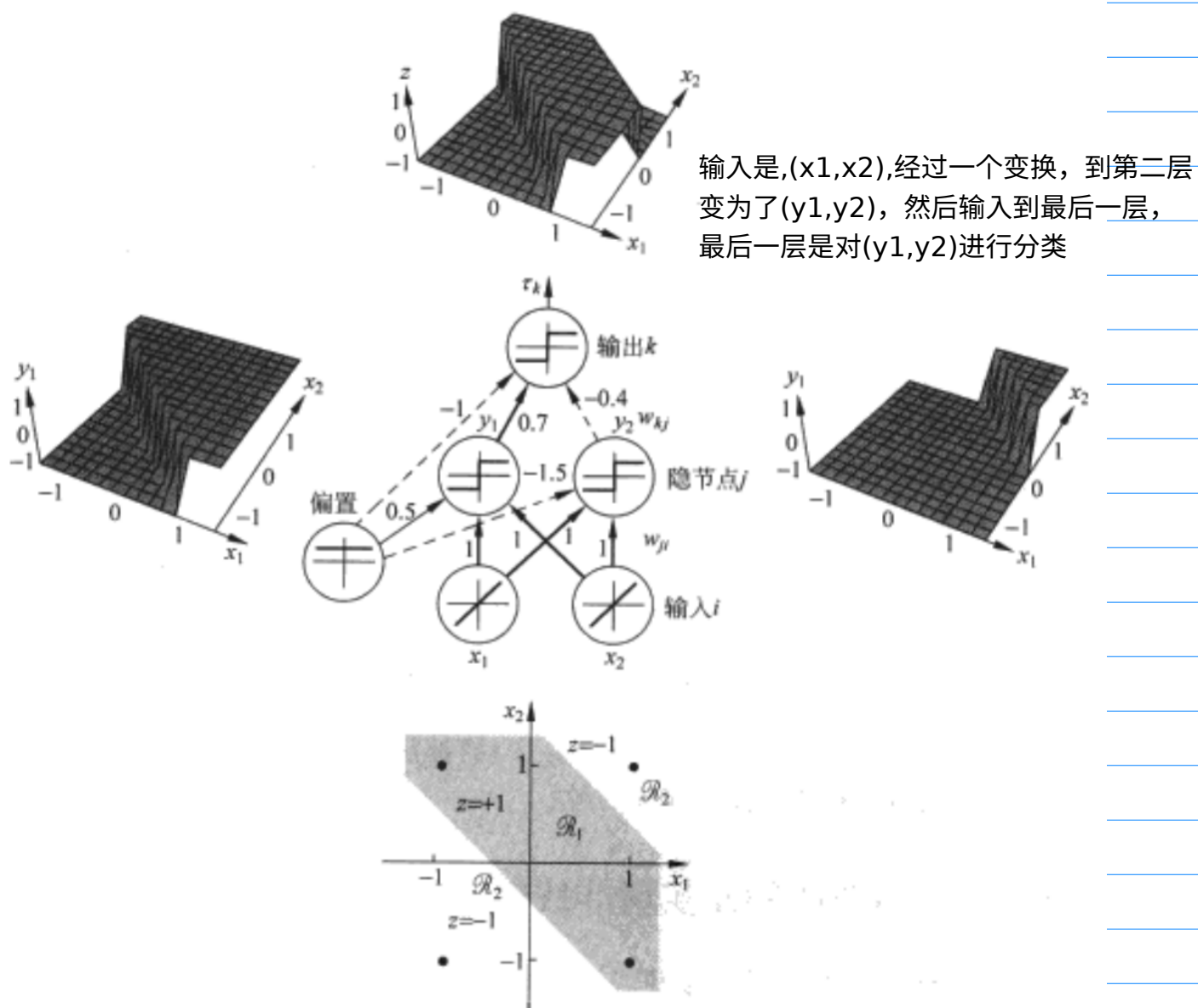
(AND)、或(OR)等，而不能解决异或(XOR)之类的高阶谓词逻辑问题。

很快人们就发现，虽然一个神经元无法实现诸如异或这样的高阶谓词逻辑问题，但是可以通过将多个神经元分层组合起来实现复杂的空间现状的分割。比如，如果按图5-9所示，将两层神经元按照一定的结构和系数进行组合，用第一层神经元分别实现两个线性分类器，把特征空间分割，而在这两个神经元节点的输出之上再加一层感知器节点，就可以实现异或运算。

每个  $W^T x$  是一个感知器，其中  $w$  是个向量

淘宝店铺-酷流科技 掌柜：我是雷锋的朋友

## 参数搜索

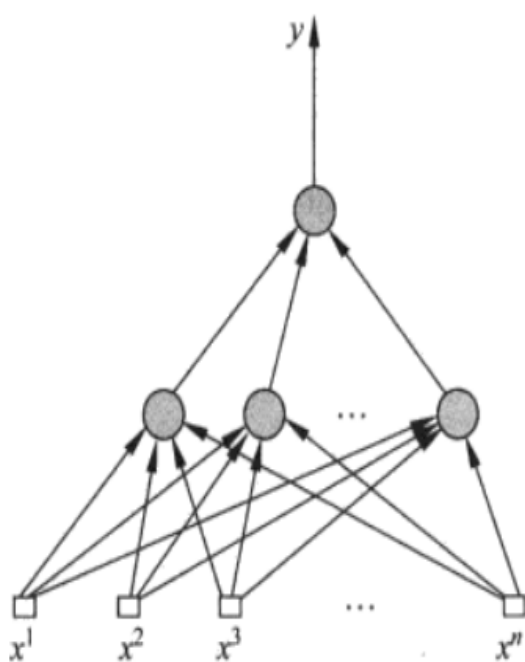


在认识到单个感知器的局限后, Rosenblatt 提出了这种多层的学习模型: 前一层神经元的输出是后一层神经元的输入, 最后一层只有一个神经元, 它接收来自前一层的  $n$  个输入, 给出作为决策的一个输出。

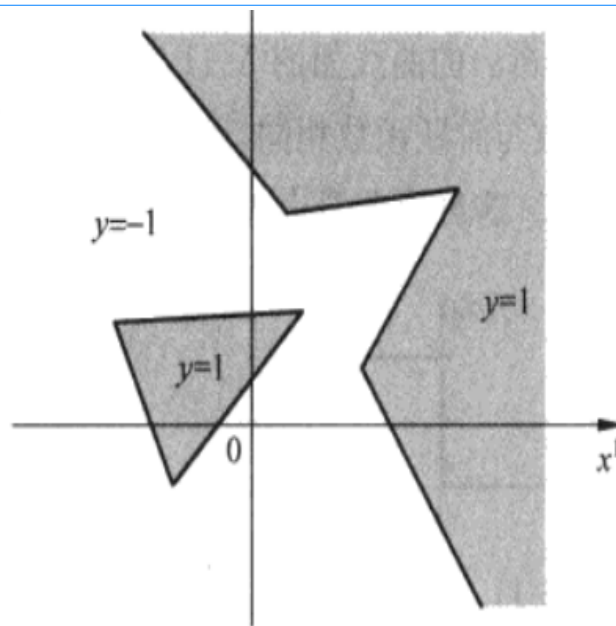
从第二层向后的每个神经元都相当于一个感知器classifier

淘宝店铺-酷流科技 掌柜: 我是雷锋的朋友





(a) 感知器模型



(b) 输入空间中的两上区域

图 5-10 用多个感知器组成两层结构实现分段线性分类

遗憾的是,在 20 世纪 60 年代,人们发现感知器学习算法无法直接应用到这种多层模型的学习上,因此,Rosenblatt 提出了这样的方案<sup>①</sup>:除了最后一个神经元之外,事先固定其他所有神经元的权值,学习过程只是用感知器学习算法来寻找最后一个神经元的权系数。实际上,这样做就相当于通过第一层神经元把原始的特征空间变换到了一个新的空间,第一层的每个神经元构成新空间的一维,每一维取值都为二值( $\{0,1\}$ 或 $\{-1,1\}$ ),然后再在这个新空间里用感知器学习算法构造一个线性分类器。显然,由于第一层神经元的权值是需要人为给定的,模型的性能很大程度上取决于能否设计出恰当的第一层神经元模型,而这又取决于对所面临的数据和问题的了解。人们当时没有找到能够针对任意问题求解第一层神经元参数的方法,所以这方面没有进一步进展,人们对感知器的研究就此停滞了大约 25 年。

这里体现了核函数思想,核SVM