关于神经网络详解与实例

引　言

- 利用机器模仿人类的智能是长期以来人们认识自然、改造自然和认识自身的理想。
- 研究ANN目的：
- （1）探索和模拟人的感觉、思维和行为的规律，设计具有人类智能的计算机系统。
- （2）探讨人脑的智能活动，用物化了的智能来考察和研究人脑智能的物质过程及其规律。

ANN的研究内容

（1）理论研究：ANN模型及其学习算法，试图从数学上描述ANN的动力学过程，建立相应的ANN模型，在该模型的基础上，对于给定的学习样本，找出一种能以较快的速度和较高的精度调整神经元间互连权值，使系统达到稳定状态，满足学习要求的算法。

（2）实现技术的研究：探讨利用电子、光学、生物等技术实现神经计算机的途径。

（3）应用的研究：探讨如何应用ANN解决实际问题，如模式识别、故障检测、智能机器人等。

研究ANN方法

（1）生理结构的模拟：

用仿生学观点，探索人脑的生理结构，把对人脑的微观结构及其智能行为的研究结合起来即人工神经网络（Artificial Neural Netwroks，简称ANN）方法。

（2）宏观功能的模拟：

从人的思维活动和智能行为的心理学特性出发，利用计算机系统来对人脑智能进行宏观功能的模拟，即符号处理方法。

ANN研究的目的和意义

(1)通过揭示物理平面与认知平面之间的映射，了解它们相互联系和相互作用的机理，从而揭示思维的本质，探索智能的本源。

(2)争取构造出尽可能与人脑具有相似功能的计算机，即ANN计算机。

(3)研究仿照脑神经系统的人工神经网络，将在模式识别、组合优化和决策判断等方面取得传统计算机所难以达到的效果。

神经网络研究的发展

(1)第一次热潮(40-60年代末)

1943年,美国心理学家W.McCulloch和数学家W.Pitts在提出了一个简单的神经元模型，即MP模型。1958年，F.Rosenblatt等研制出了感知机(Perceptron）。

(2)低潮(70-80年代初):

(3)第二次热潮

1982年，美国物理学家J.J.Hopfield提出Hopfield模型，它是一个互联的非线性动力学网络.他解决问题的方法是一种反复运算的动态过程,这是符号逻辑处理方法所不具备的性质. 1987年首届国际ANN大会在圣地亚哥召开，国际ANN联合会成立，创办了多种ANN国际刊物。1990年12月，北京召开首届学术会议。

人工神经网络研究的局限性

（1）ANN研究受到脑科学研究成果的限制。

（2）ANN缺少一个完整、成熟的理论体系。

（3）ANN研究带有浓厚的策略和经验色彩。

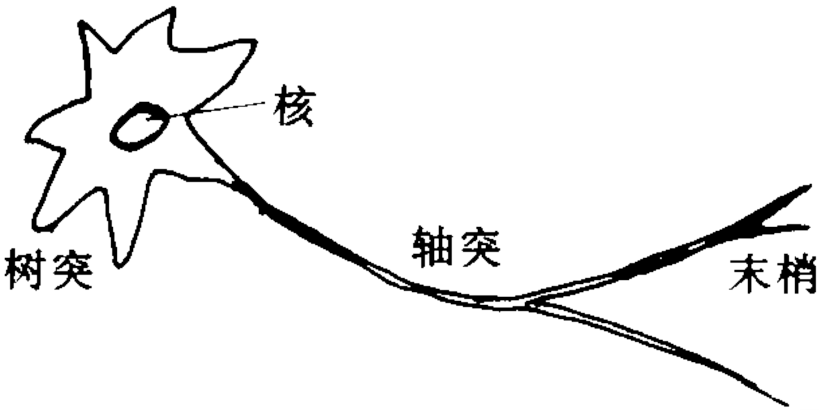
（4）**ANN**与传统技术的接口不成熟。

人工神经网络概述

- 什么是人工神经网络？
- **T.Koholen**的定义：“人工神经网络是由 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。”

二、神经元与神经网络

- 大脑可视作为1000多亿神经元组成的神经网络



- 图3 神经元的解剖图

- 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

神经网络研究的两个方面

- 从生理上、解剖学上进行研究
- 从工程技术上、算法上进行研究

脑神经信息活动的特征

(1)巨量并行性。

(2)信息处理和存储单元结合在一起。

(3)自组织自学习功能。

神经网络基本模型

- 神经元的数学模型

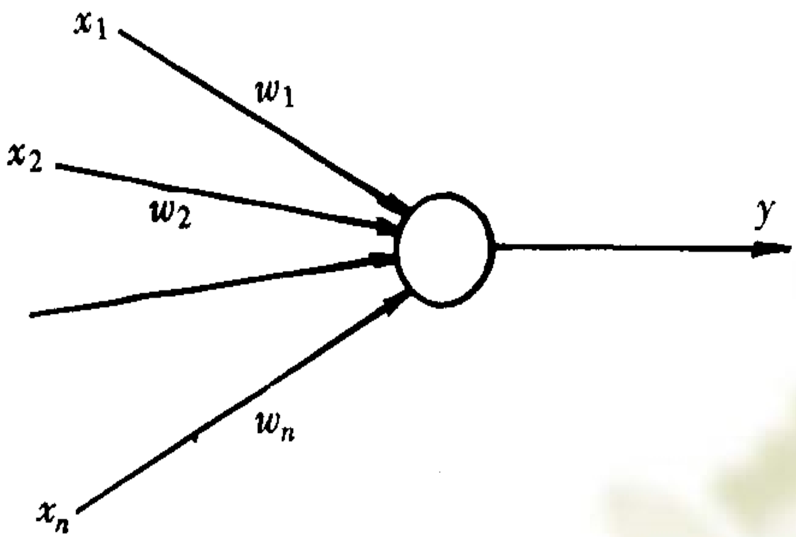


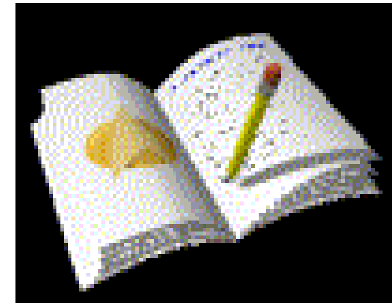
图4神经元的数学模型

- 其中$x=（x_1，…x_m）^T$输入向量，y为输出，w_i是权系数；输入与输出具有如下关系：

$$y = f\left(\sum_{i=1}^{m} w_i x_i - \theta\right)$$

- θ为阈值，f（X）是激发函数；它可以是线性函数，也可以是非线性函数.

例如，若记



$$z=\sum_{i=1}^{m} w_i x_i-\theta$$

取激发函数为符号函数

$$\operatorname{sgn}(x)=\begin{cases}1, & x>0,\\ 0, & x\le 0.\end{cases}$$

则 $$y=f(z)=\begin{cases}1, & \sum_{i=1}^{m} w_i x_i>\theta,\\ 0, & \sum_{i=1}^{m} w_i x_i\le\theta,\end{cases}$$

S型激发函数：

$$f(x)=\frac{1}{1+e^{-x}},\qquad 0\le f(x)\le 1;$$

或 $$f(x)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}, \qquad -1<f(x)<1.$$

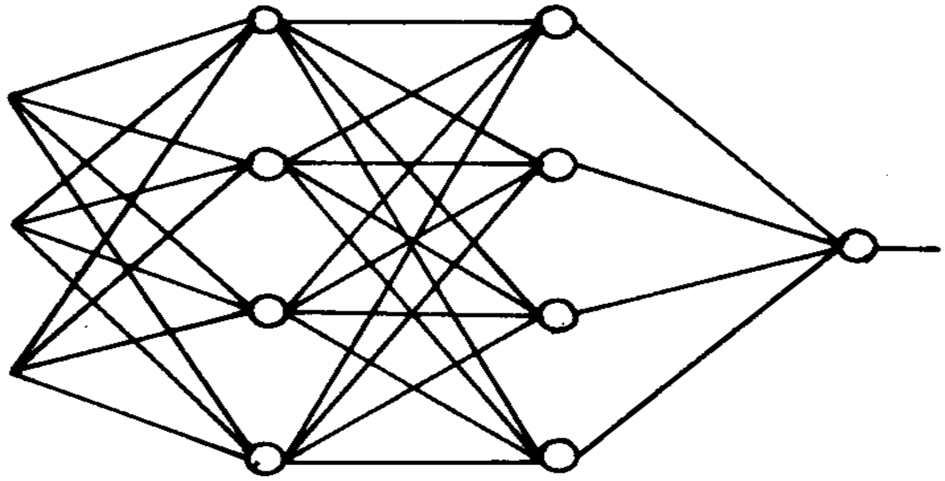
- 注：若将阈值看作是一个权系数，-1是一个固定的输入，另有m-1个正常的输入，则（1）式也可表示为：

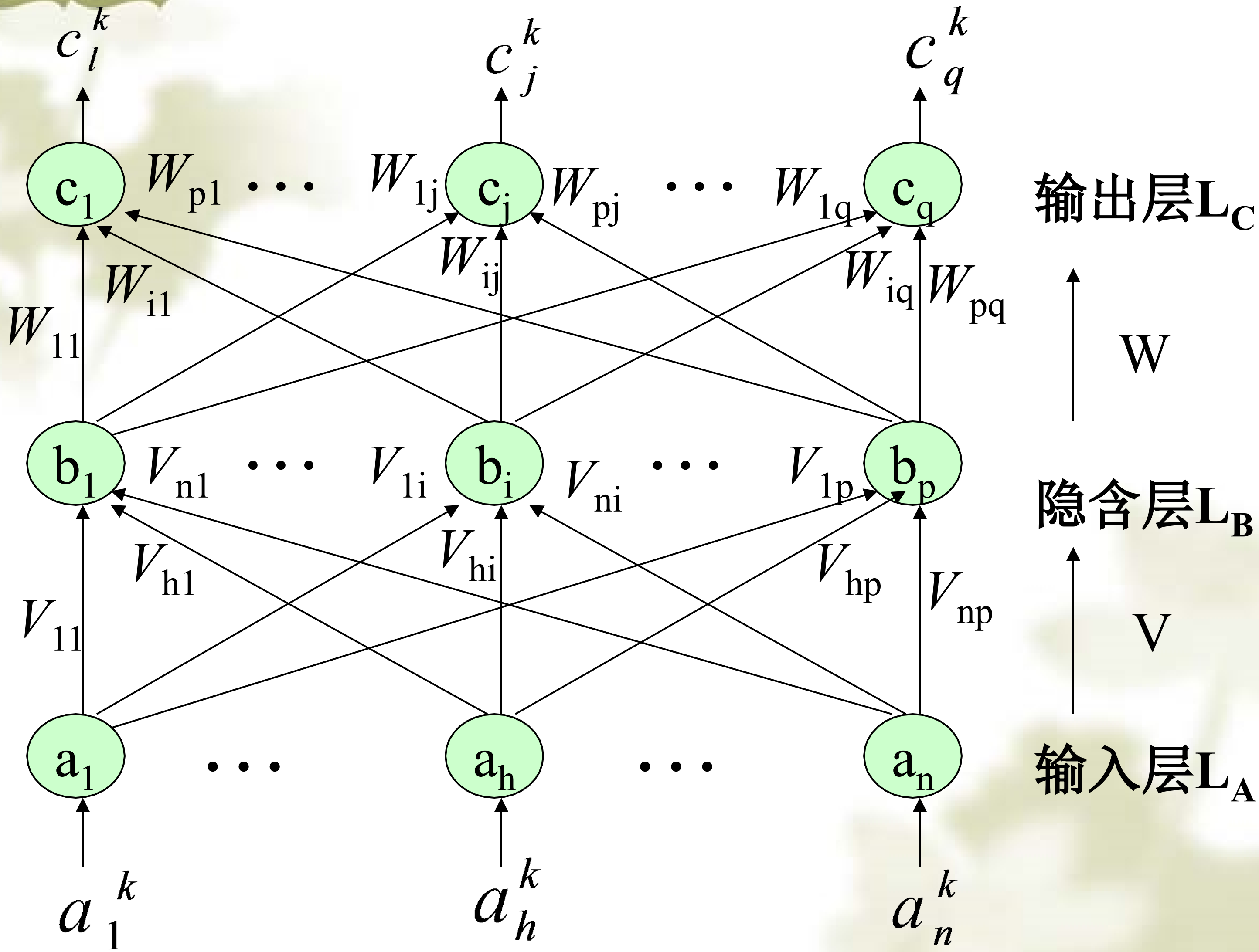
$$y=f\left(\sum_{i=1}^{m} w_i x_i\right) \qquad (1)$$

- 参数识别：假设函数形式已知，则可以从已有的输入输出数据确定出权系数及阈值。

2、神经网络的数学模型

- 众多神经元之间组合形成神经网络，例如下图的含有中间层（隐层）的B-P网络





基本BP网络的拓扑结构

ANN类型与功能

计算功能	ANN模型代表
数学近似映射，如模式识别、分类、函数逼近	BP·CPN·RBF·Elman
概率密度函数估计	SOM·CPN
从二进制数据基中提取相关信息	BSB(脑中盒模型)
形成拓扑连续及统计意义上的同构映射	SOM·Kohonen
最近相邻模式分类	BP·BM(玻尔兹曼机)·CPN·Hopfield·BAM·ART(自适应共振理论模型)·Kohonen
数据聚类	ART
最优化	BM·Hopfield

一般而言, ANN与经典计算方法相比并非优越, 只有当常规方法解决不了或效果不佳时ANN方法才能显示出其优越性。尤其对问题的机理不甚了解或不能用数学模型表示的系统,如故障诊断、特征提取和预测等问题,ANN往往是最有利的工具。另一方面, ANN对处理大量原始数据而不能用规则或公式描述的问题, 表现出极大的灵活性和自适应性。

人工神经网络
（Artificial Neuron Nets=ANN）

例

- 1981年生物学家格若根（W． Grogan）和维什（W． Wirth）发现了两类蚊子(或飞蠓midges)．他们测量了这两类蚊子每个个体的翼长和触角长，数据如下：

翼长	触角长	类别
1.78	1.14	Apf
1.96	1.18	Apf
1.86	1.20	Apf
1.72	1.24	Af
2.00	1.26	Apf
2.00	1.28	Apf
1.96	1.30	Apf
1.74	1.36	Af

翼长	触角长	类别
1.64	1.38	Af
1.82	1.38	Af
1.90	1.38	Af
1.70	1.40	Af
1.82	1.48	Af
1.82	1.54	Af
2.08	1.56	Af

❖ 问：如果抓到三只新的蚊子，它们的触角长和翼长分别为(l.24,1.80); (l.28，1.84)；（1.40，2.04）．问它们应分别属于哪一个种类？

解法一：

- 把翼长作纵坐标，触角长作横坐标；那么每个蚊子的翼长和触角决定了坐标平面的一个点.其中 6个蚊子属于 APf类；用黑点“·”表示；9个蚊子属 Af类；用小圆圈“。”表示.
- 得到的结果见图1

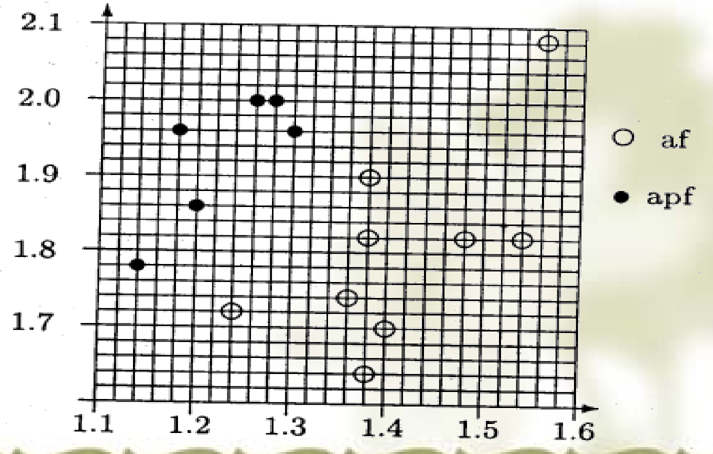


图1飞蠓的触角长和翼长

- **思路：** 作一直线将两类飞蠓分开

- 例如；取A＝（1.44，2.10）和 B＝(1.10，1.16)，过A B两点作一条直线：
- $y = 1.47x - 0.017$
- 其中X表示触角长；y表示翼长.

- **分类规则：** 设一个蚊子的数据为（x, y）
- 如果$y \geq 1.47x - 0.017$，则判断蚊子属Apf类；
- 如果$y < 1.47x - 0.017$；则判断蚊子属Af类.

❖ 分类结果：(1.24， 1.80)，(1.28， 1.84)属于Af类；(1.40，2.04）属于 Apf类.

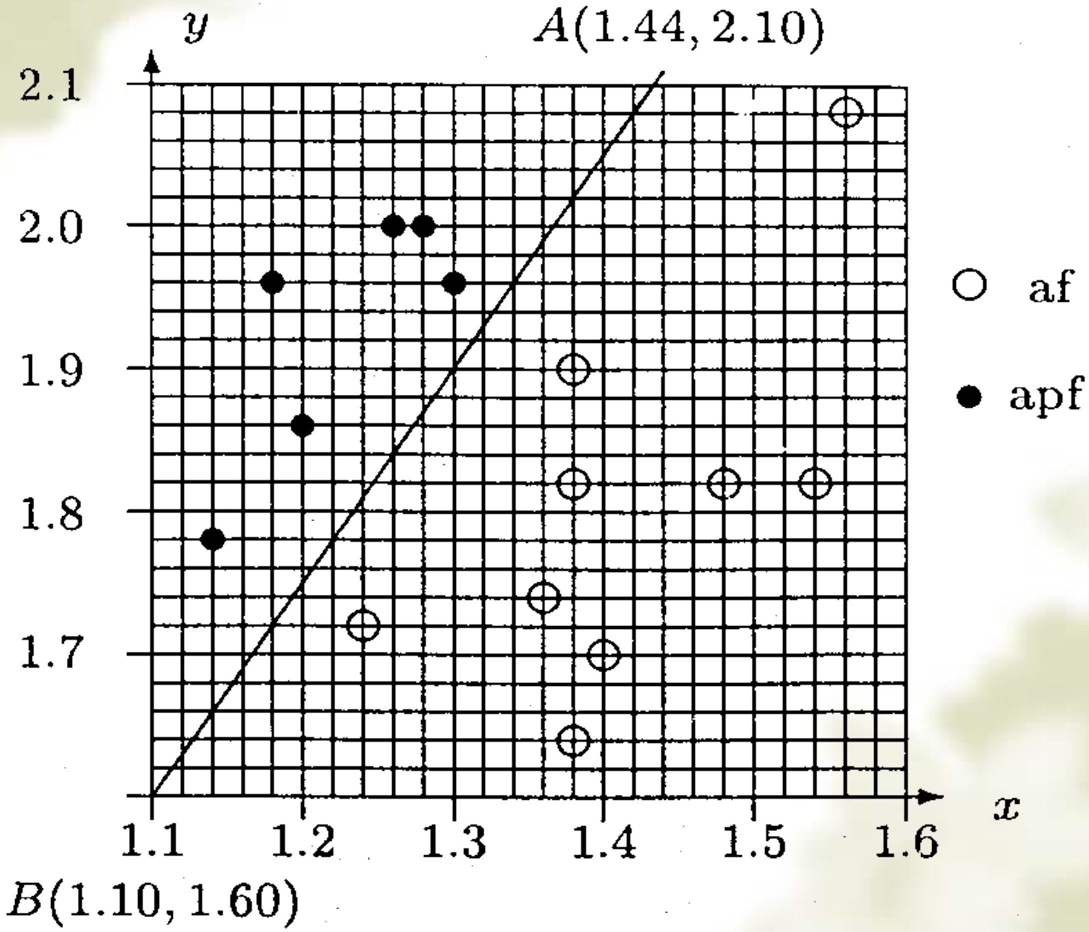


图2 分类直线图

• 缺陷：根据什么原则确定分类直线?

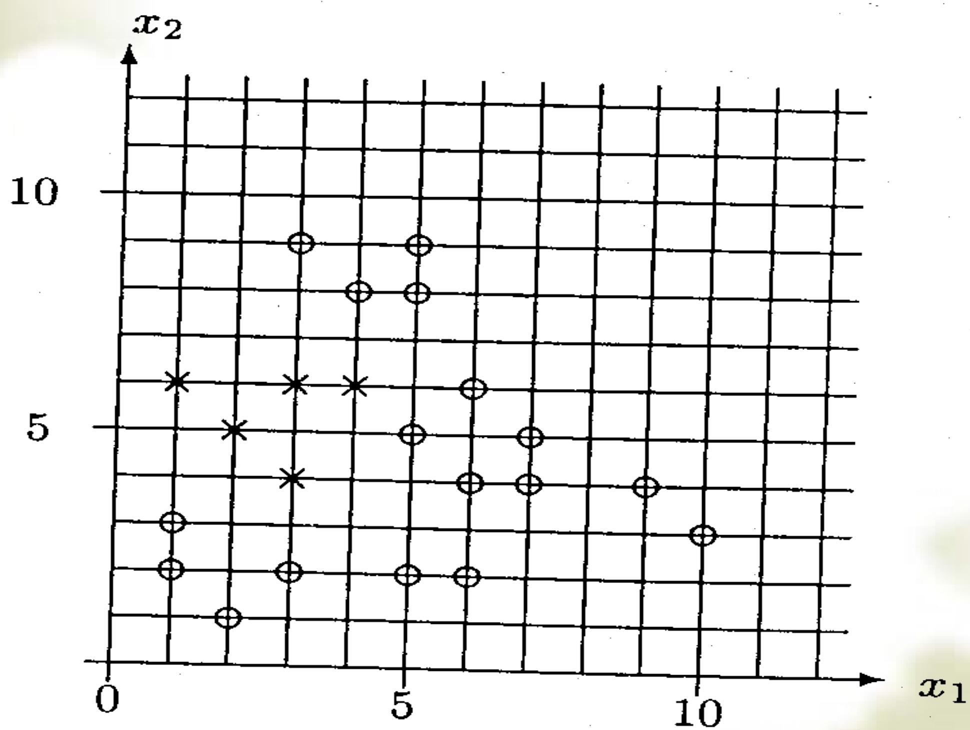
- 若取A=(1.46,2.10), B=(1.1,1.6)不变，则分类直线变为　　y=1.39x+0.071

分类结果变为：(1.24,1.80)，(1.40,2.04) 属于Apf类；(1.28,1.84)属于Af类

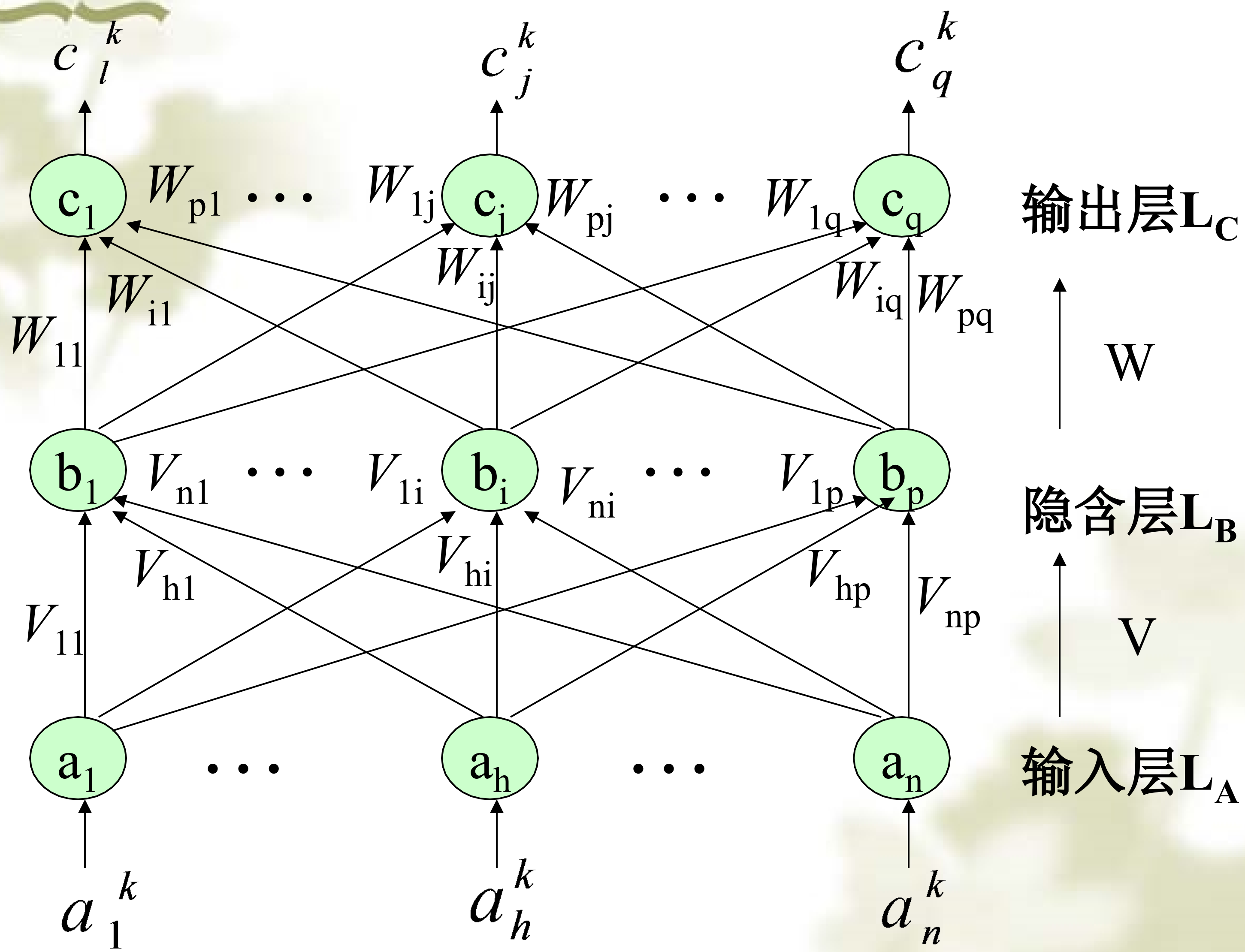
- 哪一分类直线才是正确的呢？

- 因此如何来确定这个判别直线是一个值得研究的问题．一般地讲，应该充分利用已知的数据信息来确定判别直线．

❖ 再如，如下的情形已经不能用分类直线的办法：



v **新思路：**将问题看作一个系统，飞蠓的数据作为输入，飞蠓的类型作为输出，研究输入与输出的关系。



基本BP网络的拓扑结构

四、反向传播算法（B-P算法）

Back propagation algorithm

算法的目的：根据实际的输入与输出数据，计算模型的参数（权系数）

1．简单网络的B-P算法

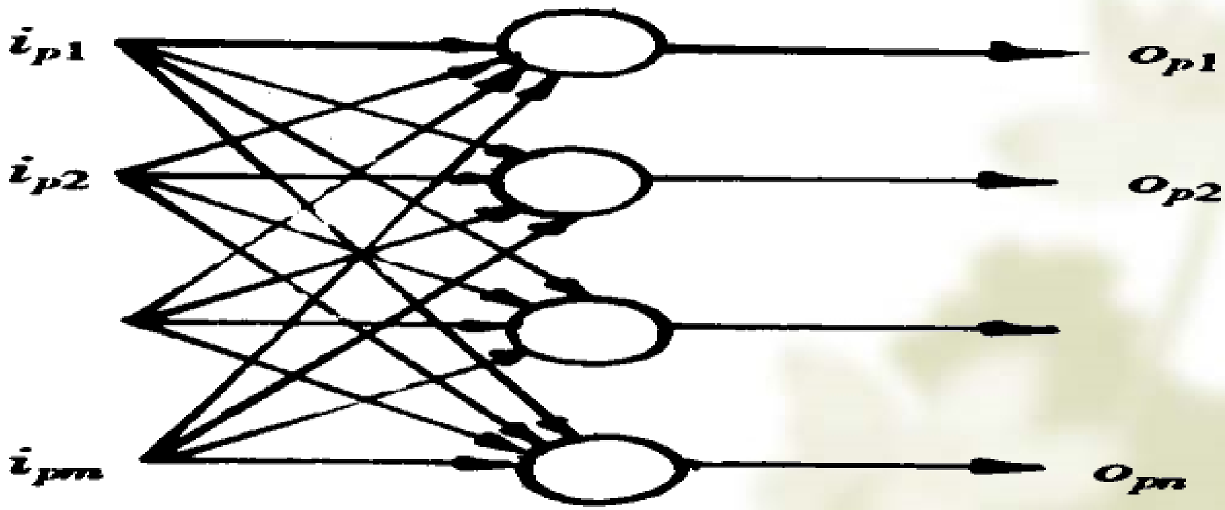


图6　简单网络

- 假设有P个训练样本，即有P个输入输出对
- （I_p, T_p）,p=1,…,P,

其中

输入向量为 ： $$I_p = (i_{p1},\dots, i_{pm})^T$$

目标输出向量为（实际上的）：

$$T_p = (t_{p1},\dots, t_{pn})^T$$

网络输出向量为（理论上的）

$$O_p = (o_{p1},\dots, o_{pn})^T$$

记wij为从输入向量的第j (j=1,…,m) 个分量到输出向量的第i (i=1,…,n)个分量的权重。通常理论值与实际值有一误差，网络学习则是指不断地把与比较，并根据极小原则修改参数wij，使误差平方和达最小：

$$\min \sum_{i=1}^{n} (t_{pi} - o_{pi})^2 \qquad （p=1,\ldots,P） \qquad （2）$$

- Delta学习规则:

记 Δw_{ij} 表示递推一次的修改量，则有

$$w_{ij} + \Delta w_{ij} \Rightarrow w_{ij} \qquad （3）$$

$$\Delta w_{ij} = \sum_{p=1}^{P} \eta (t_{pi} - o_{pi}) i_{pj} = \sum_{p=1}^{P} \eta \delta_{pi} i_{pj} \qquad （4）$$

$$\delta_{pi} = t_{pi} - o_{pi}$$

η 称为学习的速率

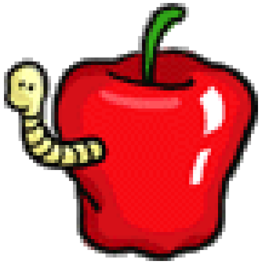
注：由（1）式，第i个神经元的输出可表示为

$$o_{pi} = f\left(\sum_{j=1}^{m} w_{ij} i_{pj}\right)$$

i_{pm}= -1 , w_{im}= (第i个神经元的阈值)　　（5）

特别当f是线性函数时

$$o_{pi} = a\left(\sum_{j=1}^{m} w_{ij} i_{pj}\right) + b \qquad (6)$$



定理 4.1 按上面描述的神经网络, 设其中每个神经元都是线性的, 取训练指标为

$$E=\sum_{p=1}^{P}E_p,$$

$$E_p=\frac{1}{2}\sum_{i=1}^{n}(t_{pi}-o_{pi})^2$$

时, 求E的最小值的梯度最速下降法就是Delta学习规则.

2. 多层前馈网络

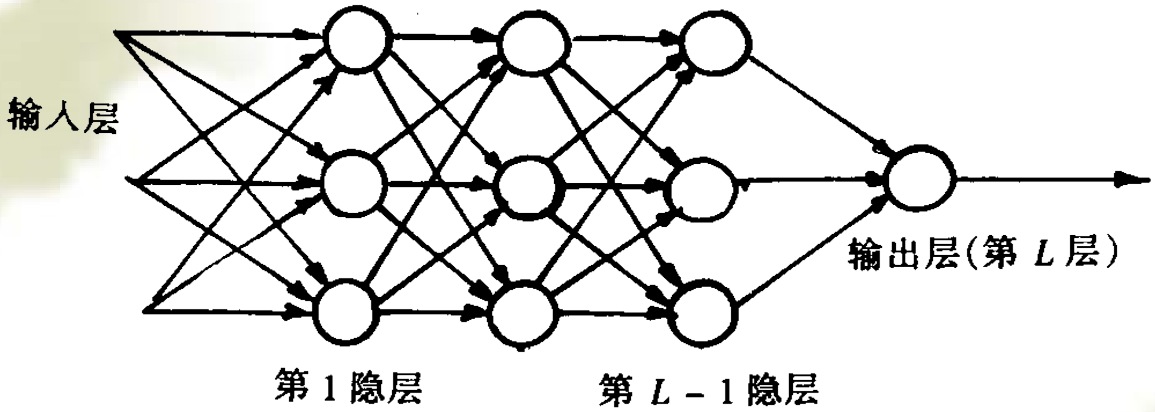


图7 多层前馈网络

假设：

（l）输入层不计在层数之内，它有N0个神经元．设网络共有L层；输出层为第L层；第 k层有Nk个神经元．

(2) 设 $u_k(i)$ 表示第k层第i神经元所接收的信息

$w_k(i,j)$ 表示从第k-1层第j个元到第k层第i个元的权重，

$a_k(i)$ 表第k层第i个元的输出

假设：

（3）设层与层间的神经元都有信息交换（否则，可设它们之间的权重为零）；但同一层的神经元之间无信息传输.

(4) 设信息传输的方向是从输入层到输出层方向；因此称为前向网络．没有反向传播信息.

（5）$a_0(j)$ 表示输入的第j个分量.

在上述假定下网络的输入输出关系可以表示为:

$$\begin{cases} u_1(i) = \sum_{j=1}^{N_0} w_1(i,j)a_0(j) + \theta_1(i), \\ a_1(i) = f(u_1(i)), & 1 \le i \le N_1, \\ u_2(i) = \sum_{j=1}^{N_1} w_2(i,j)a_1(j) + \theta_2(i), \\ a_2(i) = f(u_2(i)), & 1 \le i \le N_2, \\ \qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ u_L(i) = \sum_{j=1}^{N_{L-1}} w_L(i,j)a_{L-1}(j) + \theta_{L-1}(i), \\ a_L(i) = f(u_L(i)), & 1 \le i \le N_L, \end{cases} \quad （7）$$

v 其中表示第k层第i个元的阈值.

定理2 对于具有多个隐层的前馈神经网络；设激发函数为S函数；且指标函数取

$$E = \sum_{p=1}^{P} E_p \qquad （8）$$

其中 $$E_p = \frac{1}{2}\sum_{i=1}^{N_L}(t^{(p)}(i) - a_L^{(p)}(i))^2 \qquad （9）$$

则每个训练循环中按梯度下降时；其权重迭代公式为

$$w_l^{(p)}(i,j) = w_l^{(p-1)}(i,j) + \eta\delta_l^{(p)} a_{l-1}^{(p)}(j), \qquad （10）$$

$$l = 1,\dots,L,$$

$w_l^{(p)}(i,j)$ 表示第-1层第个元对第层第个元输入的第次迭代时的权重

$$\delta_L^{(p)}(i) = (t^{(p)}(i) - a_L^{(p)}(i)) f'(u_L^{(p)}(i)) \qquad （11）$$

$$\delta_l^{(p)}(i) = f'(u_l^{(p)}(i)) \sum_{j=1}^{N_{l+1}} \delta_{l+1}^{(p)}(j) w_{l+1}^{(p-1)}(j,i) \qquad （12）$$

$$1 \le l \le L-1.$$

BP算法

Step1

- 选定学习的数据,p=1,…,P, 随机确定初始权矩阵W（0）

Step2

- 用学习数据计算网络输出

Step3

- 用（10）式反向修正,直到用完所有学习数据.

$$w_l^{(p)}(i,j) = w_l^{(p-1)}(i,j) + \eta\delta_l^{(p)} a_{l-1}^{(p)}(j),$$

$$l = 1,\dots,L,$$

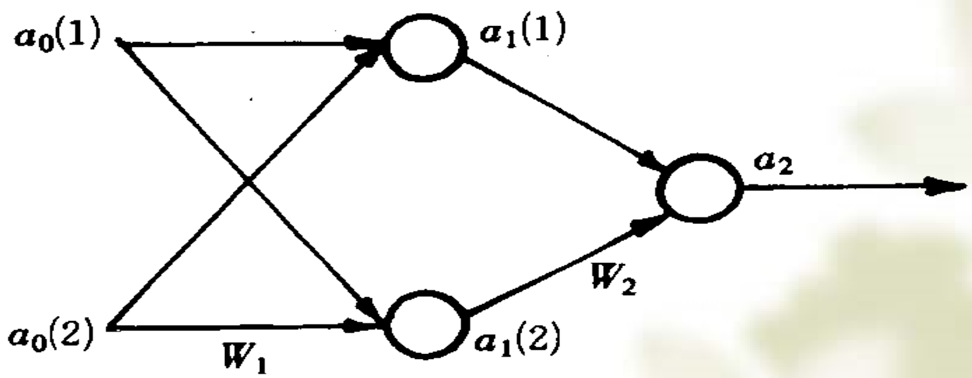
五．应用之例：蚊子的分类

v 已知的两类蚊子的数据如表1：

翼长	触角长	类别	目标值
1.78	1.14	Apf	0.9
1.96	1.18	Apf	0.9
1.86	1.20	Apf	0.9
1.72	1.24	Af	0.1
2.00	1.26	Apf	0.9
2.00	1.28	Apf	0.9
1.96	1.30	Apf	0.9
1.74	1.36	Af	0.1

翼长	触角长	类别	目标t
1.64	1.38	Af	0.1
1.82	1.38	Af	0.1
1.90	1.38	Af	0.1
1.70	1.40	Af	0.1
1.82	1.48	Af	0.1
1.82	1.54	Af	0.1
2.08	1.56	Af	0.1

- 输入数据有15个，即，p=1,…,15; j=1, 2; 对应15个输出。
- 建模：（输入层，中间层，输出层，每层的元素应取多少个？）
- 建立神经网络



- 规定目标为： 当t(1)=0.9 时表示属于Apf类，t(2)=0.1表示属于Af类。
- 设两个权重系数矩阵为：

$$W_1 = \begin{bmatrix} w_1(1,1) & w_1(1,2) & w_1(1,3) \\ w_1(2,1) & w_1(2,2) & w_1(2,3) \end{bmatrix}$$

$$W_2 = \begin{bmatrix} w_2(1,1) & w_2(1,2) & w_2(1,3) \end{bmatrix}$$

其中 $w_i(j,3) = \theta_i(j)$ 为阈值

分析如下：

$$u_1(1) = w_1(1,1)a_0(1) + w_1(1,2)a_0(2) - \theta_1(1)$$

$$u_1(2) = w_1(2,1)a_0(1) + w_1(2,2)a_0(2) - \theta_1(2)$$

$a_1(1) = f(u_1(1))$ $a_1(2) = f(u_1(2))$

为第一层的输出，同时作为第二层的输入。

其中,θ_i为阈值,f为激励函数

若令 $a_0(3) = -1$ （作为一固定输入）

$w_1(j,3) = \theta_j$ $j = 1,2$

（阈值作为固定输入神经元相应的权系数）

则有：

$$u_1(1)=w_1(1,1)a_0(1)+w_1(1,2)a_0(2)+w_1(1,3)a_0(3)=\sum_{j=1}^{3}w_1(1,j)a_0(j)$$

$$u_1(2)=w_1(2,1)a_0(1)+w_1(2,2)a_0(2)+w_1(2,3)a_0(3)=\sum_{j=1}^{3}w_1(2,j)a_0(j)$$

取激励函数为 $f(x)=\dfrac{1}{1+e^{-x}}$

则 $a_1(i)=f(u_1(i))=\dfrac{1}{1+\exp(-u_1(i))}$ $\quad i=1,2$

同样，取 $a_1(3)=-1,$ $\quad w_2(1,3)=\theta$

$$u_2(1)=\sum_{j=1}^{3}w_2(1,j)a_1(j)$$

$$a_2(1)=\frac{1}{1+\exp(-u_2(1))}$$