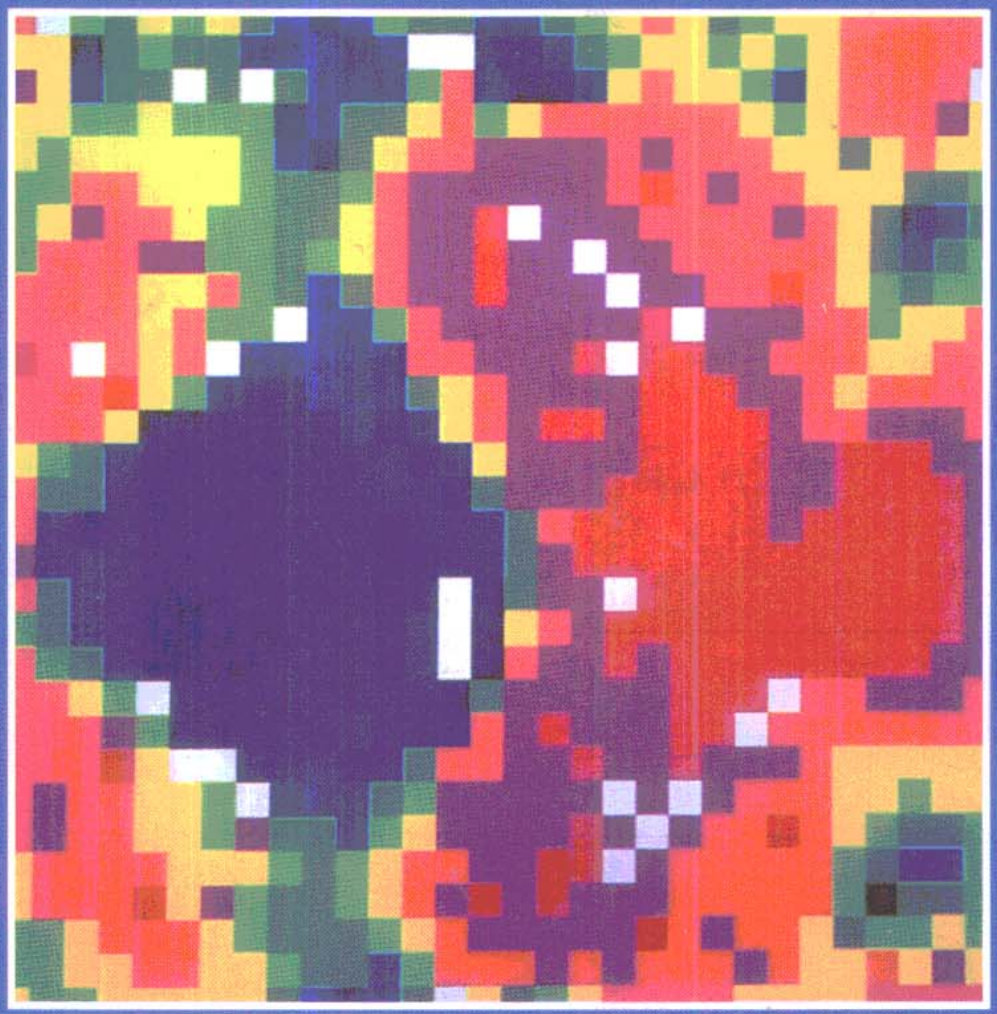
J. Zupan. J. Gasteiger 著 潘忠孝 陈玲然 译 张懋森 校

神经网络及其在化学中的应用



中国科学技术大学出版社

神经网络及其在化学中的应用

J. Zupan【斯洛文尼亚】 J. Gasteiger【德】 著

潘忠孝 陈玲然 译
张懋森 校

中国科学技术大学出版社
2000 · 合肥

图书在版编目(CIP)数据

神经网络及其在化学中的应用/(斯洛文尼亚) 多巴 (Zupan. J),(德)加斯特艾格(Gasteiger. J.)著;潘忠孝,陈玲然译. —合肥:中国科学技术大学出版社,2000.5
书名原文:Neural Networks for Chemists: an Introduction
ISBN 7-312-01107-1

Ⅰ. 神… Ⅱ. ①多… ②加… ③潘… ④陈… Ⅲ. ①神经网络-基本知识 ②人工神经网计算机-应用-化学-研究 Ⅳ. O6-05

中国版本图书馆 CIP 数据核字(2000)第 23641 号

中国科学技术大学出版社出版发行
(安徽省合肥市金寨路 96 号,邮编:230026)
中国科学技术大学印刷厂印刷
全国新华书店经销

开本:787×1092/16 印张:13.75 彩色插页:4 字数:330 千
2000 年 5 月第 1 版 2000 年 5 月第 1 次印刷
印数:1—1500 册
ISBN 7-312-01107-1/O・223 定价:25.00 元

第一篇

基 本 概 念

第一章　说　　明

学习目的：

神经网络是什么；
解决问题的途径；
术语

1.1　向信息学习

信息的应用价值在我们的社会中显得日益重要。事实上，我们正在进入一个信息社会。在我们的许多活动中，从业务管理到科学、信息都将成为最宝贵的财富之一。然而，同时也可以看到我们正处于被大量各种各样数据包围的危险之中，对某一特定问题获得正确的信息已越来越难。因此，分析数据、从各种各样的数据中提取知识以及从一些单次观察结果中总结出基础的原则和信息结构是极其有趣的。我们必须从各种观察结果中进行**学习**。

数据分析已进行了很多年，并不是什么新鲜事，绝大多数采用统计学的和模式识别的方法。然而，长期以来人们已弄清楚：人脑分析数据和信息与这类方法有相当大的区别，它处理大批量的数据并用不同的方式从这些数据中学习。人脑采集知识并不通过统计学方法进行。统计学方法和模式识别方法的固有局限已导致专家系统的发展。在一个专家系统中，针对问题的某一特定范畴的知识与推理机相分离。推理机根据知识引出结论（由此也做出决断）。然而，为专家系统获取知识的机制还远不完善。一个知识工程师通过咨询专家来建造知识库的方法有许多涉及到科学的、技术的和心理学方面的缺点。总之，这并不是大脑获取知识的方法。

在神经生理学和诸如电脑 X 光照相术（EEG）、计算机辅助层析 X 射线摄影法（CAT）、磁共振成像（MRI）、正电子发射层折 X 射线摄影法（PET）、超导量子干涉装置（SQID）及单光子发射计算机化层折 X 射线摄影法（SPECT）等新实验技术方面的进展，已大大的加强了我们对人脑组织以及出现在其中的物理和化学过程的了解。而且，已设计出数学模型和算法来模拟人类的信息处理和知识获取方法。这些模型叫做**神经网络**。

本书的目的有两个方面：第一，阐述**基本原理**和较重要的神经网络模型的范围及局限性；第二，（甚至是更重要的）明白如何**应用**这些神经网络来处理信息，通过弄清楚信息间的关系而获取需要的知识。

1.2 总体目的和概念

我们首先将一个神经网络当作一个黑箱来考虑，这个黑箱可以接收一系列的输入数据并由此产生一个或更多的输出（见图 1-1）。

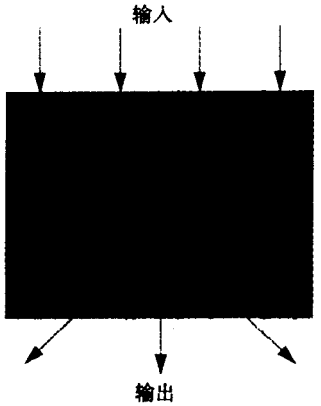


图 1-1 黑箱

输入的值可能来自于股票市场，其输出则可能是购入或抛出某种股票的建议。我们或者可输入一个病人的医学数据而获得对他（或她）所患疾病种类的预测；或从一个化合物的光谱可预测它的结构；或在一个自动化实验室中，输入一个目标物的运动情况而输出机器人手臂的反应。

对于许多神经网络使用者来讲，虽然不必准确知道在这个黑箱中发生的事情，但仍然能成功地把神经网络应用到他们的问题中去。不过，本书的目的还是想让读者逐步明白在这个黑箱中的运作过程。

在接下来的章节中，我们将会弄清楚在黑箱内部存在着基本的操作单元，它们以某种方式相连接（图 1-2）。输入沿着这些连接（称为**网络**的通道）通过，并被分配、转换且最终被重新结合而产生输出。数据的转换在许多基本处理单元中进行。这些基本处理单元，被称为**人工神经元**或简称为**神经元**，它们执行着同等的任务。

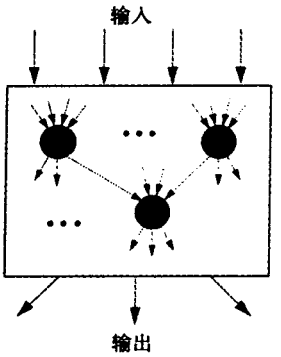


图 1-2 黑箱中的基本单元

因此，顾名思义，神经网络由连接成网络的神经元构成。我们将首先给出一个神经元的基本概念，然后察看连接它们的方式。

1.3 神经网络适宜做什么

如把神经网络看成一个黑箱，它将一个 m 变量输入转换成一个 n 变量输出。其输入或输出变量可以是：实数，最好是 0～1 或－1～＋1 之间的实数，如它们超出此范围，则输入数据必须加以处理使其回到该范围（但是，某些方法可处理较大或较小的实数）；二进制数，即 0 和 1；双极型数，即－1 和＋1。

输入和输出变量的个数受制于所采用的计算机硬件和计算所用的机时。输出变量的个数往往小于输入变量的个数，但这决不是强制性的。

用神经网络处理的问题可以是完全不同的。最一般地讲，它们可被分为四种基本类型：**联想**（自动联想或异联想）；分类；转换（不同的表示）；模型化。

自动联想指的是当系统所学习的是一个不完整的或是失真的模式时，系统能够重建一个准确的模式。图 1-3 说明了这种意思，图中的输入是由 26 个字母（在图的顶部）的栈构成的。如果一个系统能进行自动联想，那么，即使是在输入的字母不完整或失真的情况下，它也能产生出与任何已学习过的字母完美匹配的符号。

异联想指的是系统能在两套模式的成员之间进行一对一的联想。图 1-4 显示出 26 个大写字母与一套 26 个箱子之间的异联想，黑色的箱子确定输出的字母。因此，当输入一个完美的或模糊不清的字母 C 时，则用 26 个大写字母训练过且能进行异联想的系统将以一个 26 箱子模式（其中的第 3 个箱子已被占用，因为字母 C 在英文字母序中被定义为第 3 个字母）来响应。

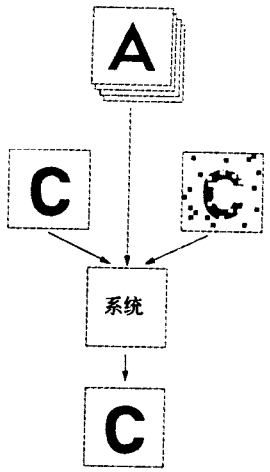


图 1-3　自动联想，根据不完整的或失真的输入来重构其原型

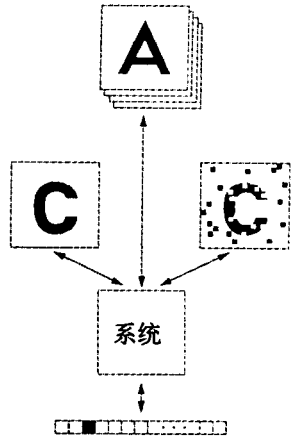


图 1-4　异联想，根据一个理想的或模糊不清的输入来重构被联想过的模式

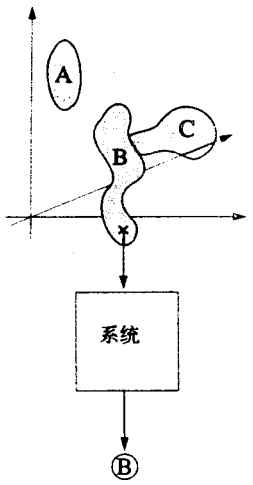


图 1-5　多元数据的分类

分类，或多或少是一个熟悉的概念，它的目的是根据一种或多种表征一个给定类别的性质，把全部给定的样本分配到合适的类别（**聚类**）中去（图 1-5）。神经网络主要用于一水平或两水平聚类中。神经网络方法的优点是多元样本仅有一小部份被用于训练，之后，网络便能预报出未知样本所归属的类（聚类）。其中某些应用与异联想应用非常相近。

分类过程可在**有监督**或**无监督**方式下进行。在一个有监督学习过程中，系统被强迫分配每一个样本到一个特定的类别中去，而在一个无监督学习中，在不需任何事先给定信息的情况下，聚类是自然形成的。

一个多元空间至另一个同维或低维空间的**转换**或**映射**是神经网络常见的一种应用。许多研究人员认为：思考、学习和推理的本质过程是来自我们感官的多元刺激至人脑神经元的一个二维平面的映射。这种映射可以从较低维至较高维来进行，而这并不常发生；在许多应用中，二维映射适用于对多维样本做简易、清晰的描述（图 1-6）。

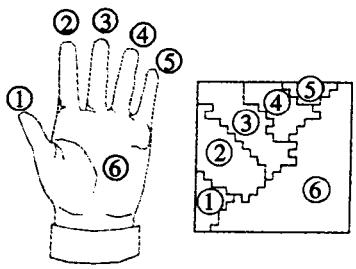
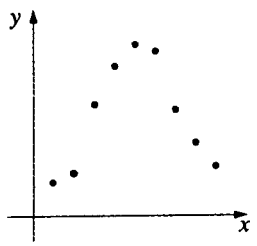


图 1-6　一个三维表面（手）至神经元矩形平面的映射

模型化，作为科学中最常被采用的一种数学应用，主要是寻找一个分析函数或一个程序（模型），这一函数或程序对于任何 m 变量输入将会给出一

个特定的 n 变量输出。从过程控制到专家系统设计，模型化在许多方面是很有用的。标准的模型化技术要求事先知道数学函数。在“拟合”过程中(见图 1-7)，根据实验数据(输入)与计算所得数据(输出)之间的最佳吻合来确定函数的参数。如实验数据均匀地覆盖变量空间并在整个范围内具有合适的密度，则预测可达到最佳。神经网络模型的优点是，它并不要求该数学函数方面的知识—单个的单元转换的非线性和充分大量的变量参数(权)保证足够的“自由度”，而使得神经网络适应于输入和输出数据之间的任何联系。



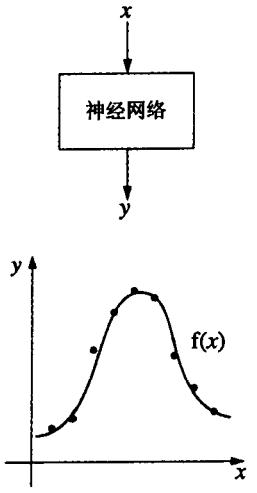


图 1-7　用神经网络来模型化实验数据

1.4　符号、惯例与缩写

论述神经网络的文献有许多符号，使初学者在把一种方法与另一种方法进行比较时发生困难。在本教科书中，我们将采用一致的命名法和符号。

标量值的名称用小斜体字母印刷：a。只有 Net 采用一个大写字母开头，以便不与“网络”或“网”相混淆。

向量和**矩阵**名称用大写黑斜体表示：$\boldsymbol{A}$。

一个**输入向量**(***Inp*** 或 $\boldsymbol{X}$)的单个值用小写字母 x 表示，用脚标 i 来表示其元素序，m 表示维数：inp_i 或 x_i，$(i=1,2,\cdots,m)$。

一批神经元的**输出向量**(***Out*** 或 $\boldsymbol{Y}$)的单个的值用小写字母表示，用脚标 j 表示元素序，n 表示维数：out_j 或 y_j，$(j=1,2\cdots,n)$。

一层神经元的**权重矩阵** $\boldsymbol{W}$ 包含有各单个的值 w_{ji}，第一个脚标 j 表示在考虑的神经元，第二个脚标 i 则表示输入它的单元(前面传输信号的神经元)：w_{ji}。

当不同层次的权重矩阵彼此进行比较时，属层次 l 的第一个权重矩阵，$\boldsymbol{W}^l$，带有脚标 i 和 j，而属下一层次的权重矩阵，$\boldsymbol{W}^{l+1}$，则带有脚标 j 和 k，k 值从 1 至 r：w_{jk}；

如有几个**输入样本**，它们用一个具有最大值为 p 的脚标 s 来鉴别，这样一来，输入样本用 $\boldsymbol{X}_s$ 来表示，它的各单个的组分(信号)则表示为：x_{si}

在一个多层网络中，其各个**层**用上标 l 来鉴别，于是，层 l 的输出向量为 $\boldsymbol{Out}^l$，它的各单个的值是 out_j^l。

贯穿一个网络的**迭代**用一个带括弧的上角标(t)来表示，于是，一个权重矩阵的初始值为 $\boldsymbol{W}^{(0)}$；在下一次迭代中它将变成 $\boldsymbol{W}^{(1)}$，改变值过程中的相连续步骤用上标(旧)和(新)来表示：$\boldsymbol{W}^{(\text{旧})}$，$\boldsymbol{W}^{(\text{新})}$。

在每一章的开头，我们给出主要论题和将要学习的**目的**。在一章的结尾处，最重要的问题和公式予以汇总，某些要点则加框以醒目。在每一章正文后，列出一批相关的文献，以便读者**更进一步的阅读**。

重要的新概念在本书中均用黑体印刷，使其醒目。想强调的词也用**黑体字**印刷。

第二章　神经元

学习目的：

生物学中的神经网络；神经元与突触

输入、权重、输出及输入-输出转换函数

线性学习机

δ-规则

人工神经元的图形描述

2.1　突触和输入信号

无论是人工的还是计算机模拟的，“神经元”都是被用来模仿生物组织神经细胞功能的。因此，我们应该考察一下生物神经元的简洁描述。对于进一步的细节，可参阅生理学或神经生理学方面的教科书（例如，生理学彩色图集）。

人类神经系统由约 10^{10} 个神经细胞（也称神经元）组成。虽然至少有 5 种不同类型的神经细胞，但只要介绍一种就足够了。运动组合体的一个典型的**神经元**由含一个细胞核的细胞体（soma）组成。细胞体有两类伸展部分：树突和轴突。图 2-1 就是这样一个神经元的极其简化的图示。树突接收信号并将其传递给细胞体。事实上，一个神经元上的树突比图 2-1 所显示的要多，树突的分枝也要多得多。因此，树突具有相当大的表面（高达 0.25 mm^2）可用以接收来自其它神经元的信号。轴突传输信号给其它神经元（或肌肉细胞），它分叉成几个“侧枝”。

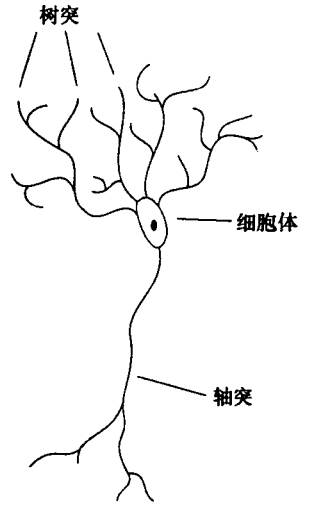


图 2-1　一个运动神经元的简化示意图

轴突与枝节的端部形成**突触**。这些突触与树突或其它神经元的细胞体相接触。一个运动神经元有成千上万个突触；高达 40% 的神经元表面被这些接触部位所覆盖。

在树突和轴突内部信号传递是电性的，通过离子移迁发生。而信号传越突触则是由化学物质完成的。轴突中的电信号释放出一种化学物质即**神经递质**（如氯乙酰），这种化学物质储存在突触前膜的小泡里，扩散穿过突触间隙，经突触后膜进入其它神经元的树突（图 2-2）。

在树突部分，神经递质产生一个新的电信号并被输送到第二个神经元。由于突触后膜不能释放神经递质，突触只能单一方向传递信号，因此，其功能就象一扇门，这对于信号传递是一个必不可少的先决条件。此外，其它神经元可通过突触来**调整**信号的传递。

由神经元产生的信号(不考虑产生这些信号的物种)非常相似,几乎不可区分,即使是由最原始的或最复杂的物种(从进化论的观点来看)产生的信号,情况也是如此。Kuffler 和 Nicholls 在它们的著作《From Neuron to Brain》第 4 页中写道:"…在体内神经细胞中,这些信号实际上是完全等同的。在不同的动物中,它们是如此地相似,以至于即使一个高级研究者也不能够确切地告知一个神经脉冲图形记录是否来自于鲸鱼、老鼠、猴子、毒蜘蛛或是教授的神经纤维"。

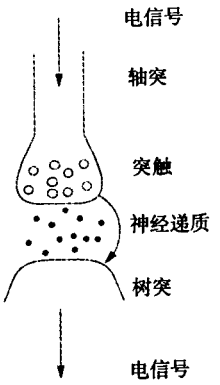


图 2-2 突触的图示描述

非常清楚,由神经元产生的信号强度(发放频率)会随刺激强度的不同而有差异。然而,不同信号的形状和轮廓外观却是非常相似的。

为什么神经网络研究的这一结论如此之重要呢?这一问题的简要(虽然不是最全面的)答案就是,信号的相似性清楚地表明大脑的真正功能并不太多依赖于单个神经元的作用,而主要取决于神经元的整体(即神经元的相互连接方式)的作用。

因此,"神经网络"这个词的重点在于"网络"而不在于"神经"。

对于信号由邻近的神经元通过突触进入一特定神经元来讲,突触代表着一种障碍,这种障碍肯定会对通过它们的信号进行调制。其改变的量取决于所谓的**突触强度**。在人工神经元中,突触强度被称为**权重** w。这种情况如图 2-3 所示。

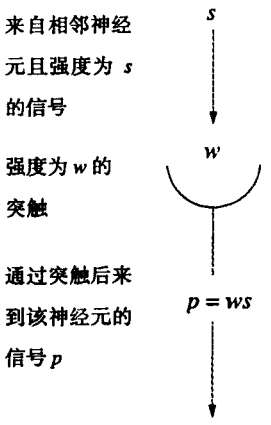


图 2-3 突触强度 w 改变进入信号 s 的强度

不涉及膜的物理和化学性质,我们可以说突触强度决定通过树突进入神经元体信号的相对量。突触强度的快速变化,即使是发生在两个连续脉冲之间的,被认为是大脑独特和有效功能的极其重要的机理。如何使突触强度适应于某个特定问题,这是我们学习的要点所在。

2.2 权 重

因为每个神经元都有大量的树突/突触(见图 2-1),故许多信号可以同时被神经元接收。各单个信号用 s_i 表示,相应的突触强度(权重)用 w_i 表示。假定在神经元诸多突触中的每一个突触上的权重于某一已知时刻可以有不同的值,我们可以认为输入信号会加和成一种**集合效应**,或称**净输入**。事实上,我们无法确切知道这种集合效应是如何形成的,也不知道对于所有输入信号来讲它到底有多大。因此,在建立一个神经元模型时,必须做一些非常粗略的简化:净输入(称为 Net)是在一给定时间间隔内到达的全部信号 s_i 和全部突触强度(权重 w_i)的函数,以及连接这些量的函数是送入的信号 s_i 及其相应权重 w_i 的乘积的简单加和。因此,可写成:

$$Net = w_1 s_1 + w_2 s_2 + \cdots + w_i s_i + \cdots + w_m s_m \tag{2.1}$$

这并不是表示神经元净输入的唯一方式，有些作者已提出了相当复杂的函数。但是，我们相信若要模拟一个**庞大**的神经元体系，则 *Net* 的表达式可能比较简单。

细心的读者们可能已注意到迄今我们都在避免谈及神经元的“输出”。由于我们不知道在神经元内发生了什么事，也无法说明在何种程度上净输入与发送信号的神经元的实际输出相等，它又会被接收者调制到何种程度。但以后将会明白的，现在的模型分两步计算出神经元的实际输出。当前，我们只关心计算的**第一步**，即所谓净输入 *Net* 的计算。

方程(2.1)的简单计算对应于图 2-4。人工神经元的描述是受真实神经元的结构启发而得到的。

举一个例子，我们将计算一个仅有 4 个突触的神经元的 *Net*，这些突触的权重分别为 0.1、0.2、−0.3 和 −0.02。因为已经指定这一人工神经元仅有 4 个突触，故它只能同时处理相连的 4 个神经元(在本例中，各信号的强度分别为 0.7、0.5、−0.1、1.0，见图 2-5)。因为 *Net* 不能用神经元的实际输出来表示，故只画出表示神经元体的上半圆。下半圆(虚线)提醒我们还需要另一步骤以根据净输入来获得输出。关于这一点，将在 2.4 节加以解释。

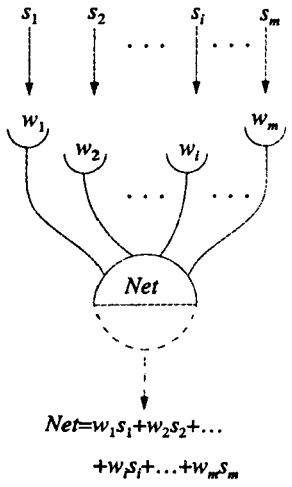


图 2-4 到达一个具有 m 个突触的人工神经元的净输入 *Net* 的计算

从这一例子可以见到，一个**人工**神经元的 *Net* 可以有正值或负值。因为没有理由对符号或权重的大小加以限制，故 *Net* 可以有范围很大的值。特别是当考虑到具有成千上万个与周围神经元相连接的突触的神经元时，情况的确如此。

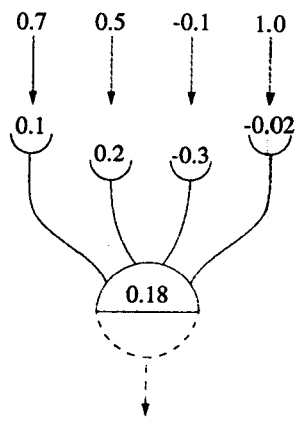


图 2-5 具有 4 个突触的神经元净输入的计算

除了用一组符号 $s_1, s_2, s_3, \cdots, s_i, \cdots, s_m$ 表示一已知神经元从相邻 m 个神经元接收的信号之外，将它们结合成一个多变量信号会显得更方便：一个多维向量 $\boldsymbol{X}$，其组元即为各单个信号：

$$\{s_1, s_2, s_3, \cdots, s_i, \cdots, s_m\} = \boldsymbol{X}(x_1, x_2, \cdots, x_m) \qquad (2.2)$$

采用这种表示法，我们前面的那个例子中的四维输入向量 $\boldsymbol{X}$ 可以写成：

$$\boldsymbol{X} = (0.7, 0.5, -0.1, 1.0) \qquad (2.3)$$

同理，一个神经元的所有突触强度可用一个多维**权重向量** $\boldsymbol{W}$ 来描述：

$$\boldsymbol{W} = (w_1, w_2, w_3, \cdots, w_i, \cdots, w_m) \qquad (2.4)$$

对于一个神经元，图 2-5 中的四维权重向量可写成：

$$\boldsymbol{W} = (0.1, 0.2, -0.3, -0.02) \qquad (2.5)$$

显然，每个神经元所具有的权重数目应至少与其所接触的神经元数目相同。在一个生物神经元中，当轴突和树突相连时，立即形成一个具有确定突触强度的突触；而在一个计算机

模拟的神经元中，程序员也应提供相应数目的权重。

2.3 线性学习机

所谓的**线性学习机**(在论及模式识别方法的标准教科书中出现的一个标题)引入了许多在以后更复杂的情形中可能会用到的有价值的概念和技术。线性学习机采用由权重向量 $\boldsymbol{W}$ 对一个多变量信号 $\boldsymbol{X}$ 进行线性变换(见式(2.1))而得到一个单变量信号，60 年代在包括化学在内的许多学科中非常流行(见 Nilsson：Linear Learning Machines)。学习机过程主要是用来判定一个已知多变量输入信号 $\boldsymbol{X}$ 是否属于一个确定的类(见 Jurs 和 Isenhour，Analytical Chemistry，1971 年 8 月；参考图 2-10)。

一个多维向量 $\boldsymbol{X}$ 可以代表许多东西，如一种声音信号、一个光学图像、任何种类的一条光谱、一种多组分化学分析、在某一时刻一个技术过程的状态、蛋白质中氨基酸的序列、某一地区的天气记录和一些股票的市场价值。因此，寻找一个合适的、会产生正确判定的权重向量 $\boldsymbol{W}$ 是非常吸引人的事情。

线性学习机的目标就是输入一套 m 维信号 $\boldsymbol{X}$ 并通过逐渐改进初始值 $\boldsymbol{W}^{(0)}$，来寻找合适的向量 $\boldsymbol{W}$。这个过程是通过将结果与事先已知的正确判定加以比较而获得的大量校正手续来完成的。为了保持与神经网络的相似性，这一过程的结果在这里称为 Net，用方程(2.6)表示：

$$Net = w_1 s_1 + w_2 s_2 + \cdots + w_i s_i + \cdots + w_m s_m \tag{2.6}$$

式(2.6)可被看成是两个向量(权重向量 $\boldsymbol{W}$ 和含有 m 个信号 s_i 的向量 $\boldsymbol{X}$)的点积。因为我们更喜欢向量符号，故各单个信号 s_i 被标记成为信号输入向量 $\boldsymbol{X}$ 的组元 x_i：

$$Net = w_1 x_1 + w_2 x_2 + \cdots + w_i x_i + \cdots + w_m x_m = \boldsymbol{WX} \tag{2.7}$$

输入向量与校正结果 Net 成线性关系，因此这一校正过程(2.7)被称为**线性**学习机。

现在，“输入向量 $\boldsymbol{X}$”是一个信号集：由几个参数描述的一个多变量样本。如果有多个输入样本，则用下标 s 以示区别，如 $\boldsymbol{X}_s$。因此传到各单个突触的实际单变量信号是这些多变量样本的具有两个下标的组元 x_{si}，第一个下标表示多变量样本，第二个下标表示与这一单个信号相连接的突触。

从现在起，组元 x_{si} 被称为**信号**，多变量输入称为**样本或向量** $\boldsymbol{X}_s$。多变量空间称为**测量空间**，在该空间中，样本被表示为指向点 $\boldsymbol{X}_s$ 的径向向量。在方程(2.6)和(2.7)中，符号 Net、$\boldsymbol{W}$ 和 $\boldsymbol{X}$ 都是如此命名，以便提醒我们注意它们在人工神经元中的相应特性。

Net 作为权重向量 $\boldsymbol{W}$ 和在测量空间中代表一个随意样本的多变量向量 $\boldsymbol{X}_s$ 的标量积，在做判定时是一个非常方便的量。它的正负号可以指示出在事先所选择的两个类，C_1 或 C_2 中，$\boldsymbol{X}_s$ 属于哪一个类。图 2-6 表示一个向量 $\boldsymbol{W}(0.2, 0.5, 0.6)$ 如何作为一个判定向量来对 3 个不同的三维向量 $\boldsymbol{X}_1 = (0.5, 0.7, -0.3)$，$\boldsymbol{X}_2 = (0.1, -0.4, 0.3)$，$\boldsymbol{X}_3 = (0.3, -0.1, -0.8)$ 进行分类。

根据 Net 是正或负来将样本分类不是唯一可行的分类方式；Net 可以很容易地分成 3 个或更多个间隔来定义 3 个或更多个类。例如，划分间隔为 $-\infty$ 到 $-a$、$-a$ 到 $+a$ 及 $+a$ 到 $+\infty$，可用来判定 $\boldsymbol{X}_s$ 属于这 3 类中的哪一类。绝大部分的判定都是二元的，即两个类间的判

定，因为所有复杂的判定可以由一系列二元判定来构成。

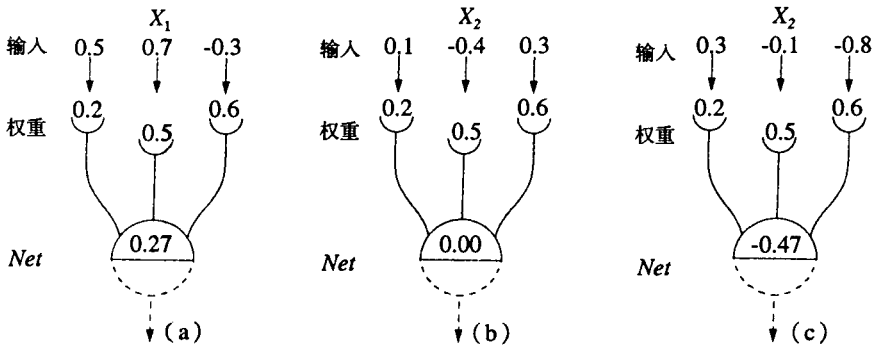


图 2-6　权重向量 $\boldsymbol{W}=(0.2,0.5,0.6)$和 3 个任意挑选的输入向量 $\boldsymbol{X}_1$、$\boldsymbol{X}_2$ 和 $\boldsymbol{X}_3$ 之间的点积 Net

让我们选择这样一个权重向量 $\boldsymbol{W}$，使得属于 C_1 类的所有样本 $\boldsymbol{X}_s$ 产生标量积 $Net=\boldsymbol{W}\boldsymbol{X}_s$ 为正，而属于 C_2 类的所有样本 $\boldsymbol{X}_s$ 产生 $Net=\boldsymbol{W}\boldsymbol{X}_s$ 为负；则此 $\boldsymbol{W}$ 就被称为**理想判定向量**。原则上，如果样本是线性可分的，理想判定向量是可以找到的；若数据不是线性可分的，则难以得到一个好的判定向量的实际途径。

通常 $\boldsymbol{W}$ 是由一个初始的(有可能是坏的)猜测 $\boldsymbol{W}^{(0)}$通过某种学习步骤而获得的。通过迭代来改善 $\boldsymbol{W}^{(0)}$。在这里，迭代过程是指如果 $\boldsymbol{W}^{(t+1)}$是由 $\boldsymbol{W}^{(t)}$得到的，那么 $\boldsymbol{W}^{(t+1)}$应该是一个比 $\boldsymbol{W}^{(t)}$略好一点的判定函数(图 2-7)。在文献中，除了线性学习机之外，还有许多不同的学习步骤。其中，来自神经网络领域的一些步骤稍后将在本书中予以讨论。

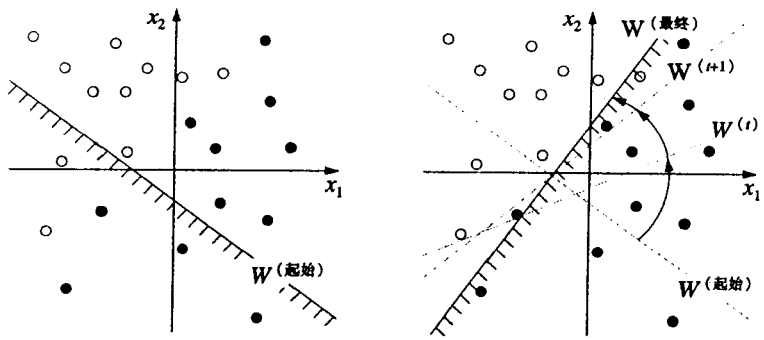


图 2-7　将判定向量 W 朝着更好的位置改变：$\boldsymbol{W}^{(t+1)}$是一个比 $\boldsymbol{W}^{(t)}$更好的判定向量，因为 $\boldsymbol{W}^{(t+1)}$分类时仅有两个样本失败；而 $\boldsymbol{W}^{(t)}$则有 5 个样本失败。理想判定向量 $\boldsymbol{W}$ 将全部 C_1 类的样本与全部 C_2 类样本(圆圈)分开。

图 2-8 表示 3 个判定向量 $\boldsymbol{W}_1$、$\boldsymbol{W}_2$、$\boldsymbol{W}_3$ 用于分层方式以便判定一个输入点应落在 x-y 平面的哪一个象限。

如果我们能设计一个产生可靠判定向量 $\boldsymbol{W}$ 的方法，则可根据这些 $\boldsymbol{W}$ 来建立任何复杂和任何规模的判定方案，从而进行二元判定(即所谓的分段线性分类器)。图 2-8 中的树结构是由三种判定构成的：要判定的点是落在纵坐标轴的左边还是右边以及(针对每一种情况)该点相对于横坐标的位置在何处。在这里，样本是 x-y 平面上的一些点，而权重向量由树的

判定结点的括号内的值表示。为了使判定树能正确地工作，每一个二元判定(在树中的任何点)都必须单独地**被生成和被检验**。

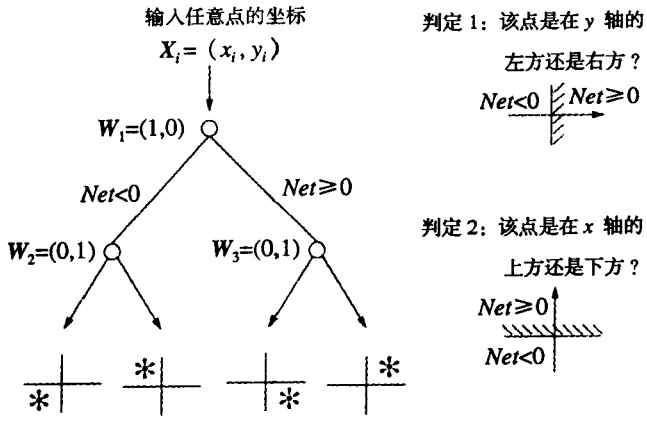


图 2-8　用于判定一已知点落在哪一象限的一个小型判定层状系统

可尝试对某些点计算其判定。例如：取点 $\boldsymbol{P}=(-3,4)$，求出其与权重向量 $\boldsymbol{W}_1$ 的点积：$Net=\boldsymbol{P}\boldsymbol{W}_1=(-3,4)(1,0)=-3$。$Net$ 的结果决定下一步将到达图 2-8 树中的哪一个分枝；现在，$Net<0$，指示为左手分枝，然后进行第二层次的判定。此时，采用 $\boldsymbol{W}_2$ 来重复上述步骤以获取最后的结果。

如上所述，我们可以把判定函数(记住，其结果是 Net)当作样本 $\boldsymbol{X}$ 与权重向量 $\boldsymbol{W}$ 表达式之间的点积。两个向量间的点积可被看作一个线性函数：

$$y=ax+b \tag{2.8}$$

式中，$y=Net$，a 和 x 分别为向量 $\boldsymbol{W}$ 和 $\boldsymbol{X}$ 的值，而 b 是一个标量常数，可被认为是这条直线的偏移量。为与神经网络表示法相一致，我们记这个偏移量为 ϑ：

$$\begin{aligned} Net&=\boldsymbol{W}\boldsymbol{X}+\vartheta=w_1x_1+w_2x_2\\ &\quad+w_3x_3+\cdots+w_mx_m+\\ &=\sum_{i=1}^{m}w_ix_i+\vartheta \end{aligned} \tag{2.9}$$

显然，向量 $\boldsymbol{W}$ 和 $\boldsymbol{X}$ 须具有相同的组元数。

观察方程(2.9)和(2.6)的相似性，它们的不同仅在常数 ϑ，这一常数实际上使前面讲到的线性形式广义化。因为我们在讨论神经元的转换函数(见 2.4 节)时将碰到这个常数，因此，让我们对它做更详细的解释。

由于 ϑ 是一个随意标量常数，可以写成两个其它常数(如 w_{m+1} 和 1)的积。如果这两个常数被看作是两个向量 $\boldsymbol{W}$ 和 $\boldsymbol{X}$ 的第$(m+1)$个组元，方程(2.9)可改写成两个$(m+1)$维向量的标量积。$\boldsymbol{X}$ 与被增大一维的样本的原表示式等同。增大一维的值总是 1(所谓增广特性向量)，而 w_{m+1} 等于 ϑ：

$$\begin{aligned} Net&=w_1x_1+w_2x_2+w_3x_3+\cdots+w_mx_m+\vartheta\\ &=w_1x_1+w_2x_2+w_3x_3+\cdots+w_mx_m+w_{m+1}\cdot 1\\ &=\boldsymbol{W}\boldsymbol{X} \end{aligned} \tag{2.10}$$

从现在起，请记住，$\boldsymbol{W}$ 和 $\boldsymbol{X}$ 都是$(m+1)$维。

这一新组元 w_{m+1} 在神经网络学习和适应中具有深远的意义(见 2.5 节“偏置量”概念的描述)。

如前所述，计算出的标量积 Net 的正负决定了样本 $\boldsymbol{X}$ 所属的类别。对某一个确定向量 $\boldsymbol{X}$，如果点积 $\boldsymbol{W}\boldsymbol{X}$ 的正负相对于 $\boldsymbol{X}$ 类样本的类别来讲是错误时，那么，我们对 $\boldsymbol{W}$ 计算一个增

量作为一些新的、经调整过的权重向量 $\boldsymbol{W}^{(新)}$ 与原来的、未经调整过的向量 $\boldsymbol{W}^{(旧)}$ 之间的差值：

$$\Delta \boldsymbol{W} = \boldsymbol{W}^{(新)} - \boldsymbol{W}^{(旧)} \tag{2.11}$$

因为权重的调整将在迭代步骤(0,1,2,…,t,$t+1$,…)中进行，故 $\boldsymbol{W}^{(新)}$ 和 $\boldsymbol{W}^{(旧)}$ 实际上分别为 $\boldsymbol{W}^{(t+1)}$ 和 $W^{(t)}$。

为了实现这种调整，我们采用所谓的 δ **规则**。该规则指出，为了改善判定向量，调整量 $\Delta \boldsymbol{W}$ 应该与某个参数 δ(它与误差成比例)成比例，**以及**与那个得到错误答案的输入 $\boldsymbol{X}$ 成比例。调整后，新的权重向量应该正确地将向量 $\boldsymbol{X}$ **分类**或至少得到的误差比前一次的误差小。

$$\Delta \boldsymbol{W} \sim \Delta \boldsymbol{X} \quad 或 \quad \delta \sim \Delta \boldsymbol{W}/\boldsymbol{X} \quad 或 \quad \delta = \eta(\Delta \boldsymbol{W}/\boldsymbol{X}) \tag{2.12}$$

式中，δ 为我们要找的调整常数，η 为比例常数，$\boldsymbol{X}$ 为被 $\boldsymbol{W}^{(旧)}$ **错误**分类的样本。我们的目标就是要找到一个参数 δ，通过它，新的权重向量 $\boldsymbol{W}^{(新)}$ 将正确地对 $\boldsymbol{X}$ 进行分类。

若点积 $Net = \boldsymbol{W}^{(旧)}\boldsymbol{X}$ 有一个错误的正负号，则点积 $Net = \boldsymbol{W}^{(新)}\boldsymbol{X}$ 必须有一个与之相反的正负号：

$$\boldsymbol{W}^{(新)}\boldsymbol{X} = -\boldsymbol{W}^{(旧)}\boldsymbol{X} \tag{2.13}$$

如将式(2.11)代入式(2.12)：

$$\delta = \eta(\boldsymbol{W}^{(新)} - \boldsymbol{W}^{(旧)})/\boldsymbol{X} \tag{2.14}$$

如将式(2.14)右边乘以 $\boldsymbol{X}/\boldsymbol{X}$，得到：

$$\delta = \eta(\boldsymbol{W}^{(新)} - \boldsymbol{W}^{(旧)})\boldsymbol{X}/(\boldsymbol{X} \cdot \boldsymbol{X}) \tag{2.15}$$

在式(2.15)的分母中，我们得到向量 $\boldsymbol{X}$ 与其自身的点积，称为向量 $\boldsymbol{X}$ 的**模**，写成 $\| \boldsymbol{X} \|^2$。式(2.15)写成：

$$\delta = \eta(\boldsymbol{W}^{(新)} - \boldsymbol{W}^{(旧)})\boldsymbol{X}/\| \boldsymbol{X} \|^2 \tag{2.16}$$

此式又可写成：

$$\delta = \eta(\boldsymbol{W}^{(新)}\boldsymbol{X} - \boldsymbol{W}^{(旧)}\boldsymbol{X})/\| \boldsymbol{X} \|^2 \tag{2.17}$$

将式(2.13)代入上式，则修正量 δ 的最终表达式为：

$$\begin{aligned}\delta &= \eta(-\boldsymbol{W}^{(旧)}\boldsymbol{X} - \boldsymbol{W}^{(旧)}\boldsymbol{X})/\| \boldsymbol{X} \|^2 \\ &= -2\eta \boldsymbol{W}^{(旧)}\boldsymbol{X}/\| \boldsymbol{X} \|^2\end{aligned} \tag{2.18}$$

应用方程式(2.18)，可容易得到调整量 $\Delta \boldsymbol{W} = \boldsymbol{W}^{(新)} - \boldsymbol{W}^{(旧)}$ 的表达式：

$$\begin{aligned}\Delta \boldsymbol{W} &= \boldsymbol{W}^{(新)} - \boldsymbol{W}^{(旧)} \\ &= -(2(\boldsymbol{W}^{(旧)}\boldsymbol{X}/\| \boldsymbol{X} \|^2)\boldsymbol{X}\end{aligned} \tag{2.19}$$

或以一个更广义的形式表示：

$$\boldsymbol{W}^{(新)} = \boldsymbol{W}^{(旧)} - (2\eta(\boldsymbol{W}^{(旧)}\boldsymbol{X}/\| \boldsymbol{X} \|^2)\boldsymbol{X} \tag{2.20}$$

$\boldsymbol{W}^{(旧)}$ 的 δ **规则调整**。式(2.19)和式(2.20)，**保证**了 $\boldsymbol{W}^{(新)}$ 将正确地对 $\boldsymbol{X}$ 进行分类。然而，δ 规则对还用 $\boldsymbol{W}^{(旧)}$ 分类的其它样本并无任何表示，它们之中的一些甚至全部此时可能被错误地进行分类。

如果比例常数 η 设为 1，则方程(2.20)给出的经调整过的判别向量 $\boldsymbol{W}^{(新)}$ 便会成为 $\boldsymbol{W}^{(旧)}$ 的**镜像**(见图 2-9)。在一定的情况下，由镜像所产生的调整量是小的，但镜像往往能明

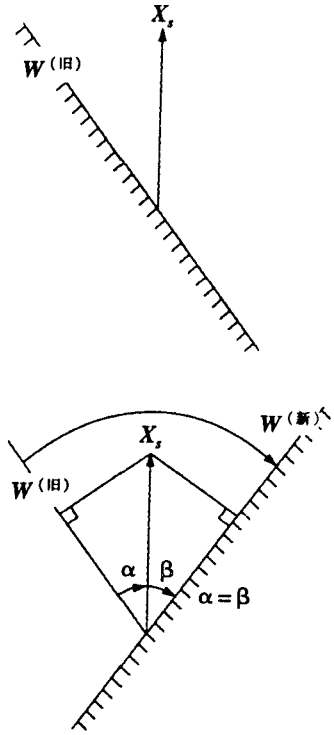


图 2-9 向量 $\boldsymbol{W}^{(新)}$(是向量 $\boldsymbol{W}^{(旧)}$ 的镜像)正确地将样本 $\boldsymbol{X}$ 归入另一个类中

显地改变 $W^{(旧)}$。这种大的改变是我们所不希望的，尤其是在迭代学习的末尾时更是如此，因为判定向量 W 的大变化意味着许多原先已被正确分类的样本可能会变成错误的分类。因此，适度调整有时会显得更合适。

在方程式(2.18)到(2.20)中，**常数** η 被一个小于 1 的**变量** η 所取代，即可获得适度的调整：

$$W^{(新)} = W^{(旧)} - (2\eta W^{(旧)}X / \| X \|^2)X$$

或

$$\Delta W = -(2\eta W^{(旧)}X / \| X \|^2)X \tag{2.21}$$

将校正量 $-2W^{(旧)}X / \| X \|^2$ 写成 δ 并记住 X 是一个输入样本，我们得到 δ 规则标准方程的最普遍已知形式：

$$\Delta W = \eta \delta X \tag{2.22}$$

由于样本 X 为输入，它也可称为 Inp：

$$\Delta W = \eta \delta Inp \tag{2.22a}$$

方程(2.22)代表用于自我学习过程的调整时 δ 规则的最一般形式；在第八章中它将被充分地用来描述误差反向传播算法。

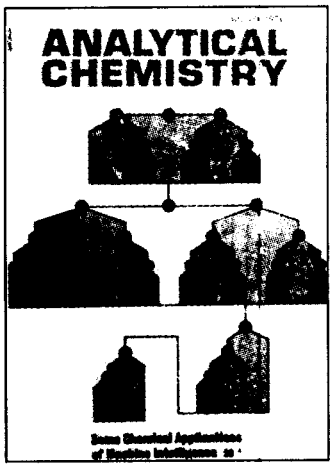


图 2-10　1971 年 8 月期 *Analytical Chemistry* 的封面图案，一棵从质谱预测分子式的判定树，该树有 26 个判定

一般地说，对任何一种子判定，一个好的二元判定向量 W 要求一个长的迭代过程。这个过程从一个随意值 $W^{(0)}$（可含有随机数值）和一系列样本 $\{X_i\}$ 开始，这些样本都同样被表示成 m 维样本，对它们应该进行的判定结果是早已知道了的。通过这些了解，我们就可以判断对当前的判定向量 W 的修正是否有必要。

学习开始后，按顺序检验所有样本 X 并随时依据需要调整 W。当对整个样本集完成一遍检验之后，判定向量 W 已发生了明显的改变；因此先前已被正确分类的样本在第二遍中可能被错误地分类。所以，只要还有一个错误，判定向量的检验和调整都应该反复地进行下去。由线性学习机方法获得的结果可用在许多复杂的判定过程中。Jurs 和 Isenhour 已使用该法根据质谱预测分子式，所用的判定树由 26 个二元判定构成。该判定树的示意图甚至被用作 1971 年 8 月期 *Analytical Chemistry* 杂志的封面图案（见图 2-10）。

尽管取得了某些成功，但由线性学习机获得的复杂判定方案总的来说并不很满意，特别是在处理庞大的实际问题时情况更是如此。因此，在经历了某些初始的兴奋之后，也发表了对该法的严厉批评（参见 Minsky 和 Papert 的文献）。目前，线性学习机主要用在非常有限的数据集上和教育方面。

2.4　神经元中的转换函数

在根据输入信号来获得输出方面，目前的神经元模型由两个不同的步骤组成。第一步是