人工神经网络的多维空间几何分析及其理论

王守觉 ,王柏南
(中国科学院半导体研究所神经网络实验室 ,北京 100083)

摘 要: 本文系统地讨论了作为用于分析人工神经网络的一种方法即多维空间几何方法 ,并对多维空间几何学进行系统的研究 ,推导了必要的定理 ,为人工神经网络的分析提供了新的手段.

关键词: 神经网络; 多维空间; 学习算法; 多维几何
中图分类号: TP183 **文献标识码**: A **文章编号**: 0372-2112 (2002) 01-0001-04

Analysis and Theory of High-Dimension Space Geometry for Artificial Neural Networks

WANG Shou-jue ,WANG Bai-nan
(Lab of Artificial Neural Networks , Institute of Semiconductors , CAS , Beijing 100083 , China)

Abstract : We systematically discuss a high-dimensional space geometry method used for analyzing artificial neural networks ,do some research work about the high-dimensional space geometry and prove the necessary theorems. It is a novel tool to analyze artificial neural networks.

Key words : neural networks ;high-dimensional space ;learning algorithm ;high-dimensional geometry

1 引言

人工神经网络一方面由于它极强的函数拟合和自学习能力、很好的鲁棒性以及高超的分类划分能力等优点已经广泛地被工程技术界用来解决各类实际问题[1]. 但另一方面 ,由于每个神经元行为的数学描述都是一个多变量的非线性函数方程 ,组成网络以后的庞大多变量非线性方程组更难以进行数学的分析. 虽然不少作者曾提出过多种数学分析方法[2] ,但复杂多变量非线性方程组的处理与分析问题始终是个难题 ,它限制着人工神经网络从分析角度的深入发展.

人类生活在自然界中最早对物体形状的概念是个几何概念 ,几千年以后由于代数学的深入发展才产生了解析几何、微积分、微分方程以及各种抽象空间的几何. 这就把人脑中朴素的形状概念上升到用许多抽象的数学公式来表达. 一般说人们对物体形状的几何概念很清楚 ,但解析几何、微积分用方程式来分析计算则更加精确.

对于人工神经网络而言 ,由于对庞大的复杂多变量非线性方程组目前数学上还缺乏通用的解析方法 ,因而对神经网络行为用数学解析全面进行精确分析的途径还难以实现.

为改善对神经网络行为的认识和研究中的"黑匣子"式难以处理的状态 ,本文提出了一种用多维空间几何推理的认识、分析方法. 在作者已发表的有关神经网络模型与算法的论文[3-7]的实验结果中 ,初步反映了这种认识与分析方法的优越性.

2 神经元的多维空间几何对应

广义地看待一个神经元可以理解为一个如下的基本运算操作:

$$Y = f[\Phi(x_1, x_2, \cdots x_n) - \theta] \quad (1)$$

神经元的基本运算规则. 如对于 BP 网络的神经元

$$\Phi(x_1, x_2, \cdots x_n) = \sum_{i=1}^{n} w_i x_i \quad (2)$$

而对于 RBF 网络的神经元 ,则可取

$$\Phi(x_1, x_2, \cdots x_n) = [\sum_{i=1}^{n} (x_i - w_i)^2]^{\frac{1}{2}} \quad (3)$$

因而可以把一个神经元对应于多维空间中一个超平面或超曲面 ,其方程式为

$$\Phi(x_1, x_2, \cdots x_n) - \theta = 0 \quad (4)$$

而神经元输出函数 f 的基是输入空间的输入点偏离该超平面或超曲面的程度.

显然超平面或超曲面的几何概念对于帮助人们对神经网络行为的认识与分析是十分有效的.

3 用几何分析方法发展神经网络新模型新算法的实例

把模式识别用神经网络中每个神经元看作是在多维空间

收稿日期:2000-09-25 ;修回日期:2000-11-20
基金项目:国家自然科学基金(No. 60076020)

中作的一个超平面或超曲面,依此来划分输入空间的样本点.而如果把先分离的某些样本点挖掉不再参加后面的划分,就象一只苹果中有许多不同颜色的污点,先靠边切一刀把若干个同一颜色的污点切掉,并把这切下的一片苹果拿走,则依次一刀刀地切下去总能把所有的各色污点都分离在许多片切下的苹果中.每切一片至少能切去一个或几个同色的污点.这一最简单的几何常识,就可以得出以下对应于模式识别用神经网络训练算法中的几点结论:

(1)如果神经元编有序号而能使某一个神经元激活后就能抑制所有以后的神经元,则就能相当于把切下的一片苹果拿走.

(2)如果神经元编有序号而在前面的神经元的输出结果优先于后面的神经元,则同样相当于把它切下的一片苹果拿走.

(3)如果能实现上述二点中任意一点,则这个训练算法必定收敛,而且所有训练样本能全部正确识别.

正是根据上述第1点,作者发展了通用前馈网络与排序学习前向掩蔽模型[3,4]即SLAM模型.实验证明它的训练速度较BP模型快几十倍[3,4].

根据上述第2点,作者提出了优先度排序人工神经网络(PONN)[5,6]的概念,不仅可用于SLAM模型,而且可用于任何其它神经元的网络.如用于作者在另一论文中所提出的"方向基函数"(DBF)[7]神经元,用优先度排序的方向基函数(PODBF)[5]神经网络,其训练时间更缩短了一个数量级,而且识别率大为提高.

正如上述第3点所述,这些神经网络模型与算法都有很好的收敛性.

上述这些事实正是作者希望进一步深入研究多维空间几何的动机.

4 多维空间几何的基本分析方法与定理证明

原苏联科学院院士和通讯院士A.Д亚历山大洛夫等的一部数学名著[8]中对多维空间几何学的内容及其发展史作出了简单的介绍:德国数学家格拉斯曼和英国的数学家凯利通过与普通的解析几何作形式的类比的途径得到多维空间解析几何原理的表达.瑞士数学家史雷夫里讨论了多维空间的正多面体的问题.但多维空间几何学虽历经多年的发展,而目前尚未形成系统而实用的几何分析方法.本论文试图在这方面做一些初步的补充,给出了如下的多维空间几何学的公理、定理及其证明,以便应用于人工神经网络行为的分析研究中.

4.1 基本定义

自由度:自由度等于维数.n维空间有n个自由度.

约束:对于n维空间的点$P(x_1,x_2,\cdots,x_n)$,若存在$f(x_1,x_2,\cdots,x_n)=c$,则称f为p的一个完全约束.若存在$f(x_1,x_2,\cdots,x_n)\le c$或$f(x_1,x_2,\cdots,x_n)\ge c$,则称$f$为$p$的一个不完全约束.(若无特殊说明,本文提及的约束均指完全约束)

超平面:n维空间中的k维超平面定义为满足下面的$n-k$个各自独立的线性方程的点的轨迹:

$$a_{11}x_1+a_{12}x_2+\cdots a_{1n}x_n+b_1=0$$

$$a_{21}x_1+a_{22}x_2+\cdots a_{2n}x_n+b_2=0$$

$$\vdots$$

$$a_{n-k,1}x_1+a_{n-k,2}x_2+\cdots a_{n-k,n}x_n+b_{n-k}=0$$

这些方程中的每一个都表示一个$n-1$维超平面,而它们全体则共同决定$n-k$个这种超平面的公共点.每个方程中x的系数$a_1,a_2,\cdots,a_n$就是该$n-1$维超平面的法线向量的系数.

超球面:到定点距离等于定长的点集.

超球:到定点距离小于或等于定长的点集.

平行:同一平面内永不相交的两条直线平行.n维空间中在$n-1$维超平面外一条直线和该超平面永不相交,则该直线与该超平面平行.若同一超平面内两个低一维的超平面永不相交,则二者平行.

垂直:夹角为直角的两直线垂直.若直线垂直于某一超平面内的任意与它相交的直线,则该直线垂直于该超平面.若一超平面内某一直线垂直于另一超平面,则二者垂直.

点到超平面的距离:已知超平面α和α外找一点p,过p作一直线与a垂直相交于点q,称p和q两点之间的距离为点p到超平面α的距离.

两个平行超平面之间的距离:在两个平行超平面外找一点使能作一直线与二者垂直相交,两个交点之间的距离就是两个超平面之间的距离.

投影:直线外一点p在直线l上的投影是指经过p点且垂直于l的直线与l的交点.平面外一点p在平面α上的投影是指经过点p且垂直于α的直线与α的交点.超平面外一点p在超平面α上的投影是指经过点p且垂直于α的直线与α的交点.相应地,其它几何构形在直线、平面、超平面上的投影是指该几何构形中的每一点在直线、平面、超平面上的投影的集合.

直线与超平面的夹角:是指直线与其在该超平面上的投影的夹角.(本文中平面、球面维数为二维)

4.2 点、线、超平面的相互关系

公理1:对于n维空间的任一点,作用于该点的每个独立的完全约束都减少它的一个自由度.作用于该点的不完全约束不减少它的自由度,只是限制它的尺度.

公理2:通过不在一个$k-1$维超平面上的每$k+1$个点,有且仅有一个k维超平面.

公理3:在每个k维超平面上至少有$k+1$个不在同一较低维数超平面上的点.在n维空间中至少有$n+1$个不在任何同一个超平面上的点.

公理4:如果直线与任意维数的超平面有两个公共点,则该直线位于这个超平面上.一般地,如果l维超平面与k维超平面有$l+1$个不在同一$l-1$维超平面上的公共点,则它就整个处在该k维超平面上.

如果n维空间中的l维和k维超平面相交(至少有一个公共点),则它们交集的维数x满足下面的关系:

$$max(l+k-n,0)\le x\le min(l,k) \quad (5)$$

证明:l维超平面有$n-l$个约束;k维超平面有$n-k$个约束;交集同时满足所有的$(n-l)+(n-k)=n-(l+k-$

n)个约束.如果所有的约束互相独立,则交集为 $l+k-n$ 维超平面;反之得到更大维数的超平面.又,交集的维数显然不大于 l 和 k.证毕.

定理 1:n 维空间中,过 k 维超平面 α 外一点 p 引直线 l 与 α 相交,有且仅有一个交点.

证明:由题设,l 与 α 相交,所以至少有一个交点.如果交点多于一个,由公理 4 可知 $l\subset\alpha$,与题设矛盾.证毕.

定理 2:n 维空间中,两个不重合的平面相交,交集是一点或一条直线.

证明:由公式 5,交集的维数 $0\le x\le 2$.由题设,两个平面不重合,所以交集只能是一点或一条直线.证毕.

讨论:在 $n(n>4)$ 维空间中,两个平面若相交,交集是点的几率大于交集是直线的几率.

定理 3:n 维空间中,在 k 维超平面 α 上一点 p 作直线 l 垂直于 α,l 的轨迹是 $n-k$ 维超平面.

证明:设 l 的轨迹 β 的自由度为 x,β 与 α 有且仅有一个交点,它的自由度为 0.由式(5),$k+x-n\le 0$,所以 $x=n-k$.证毕.

讨论:本定理同时说明,对于 k 维超平面,共有 $n-k$ 个与之垂直且相互垂直的方向.

定理 4:n 维空间中,通过 k 条两两垂直且交于一点的直线,有且仅有一个 k 维超平面.

证明:每条直线有一个自由度,每增加一条垂直于所有现存直线的直线,则增加一个自由度.所以这 k 条直线共有 k 个自由度.在这 k 条直线上,分别任取除公共点之外的 k 点,连同公共点,共有 $k+1$ 点,这 $k+1$ 点不可能处于同一个 $k-1$ 维超平面上.于是由公理 2,这 $k+1$ 点唯一确定一个 k 维超平面.证毕.

定理 5:n 维空间中,过 k 维超平面 α 外一点作直线 l 垂直于 α 内 k 条两两垂直且交于一点的直线,则 l 垂直于 α.

证明:由定理 4,这 k 条两两垂直且交于一点的直线唯一确定 α,l 垂直于这些直线,故 l 垂直于 α.证毕.

定理 6:$k+1$ 维空间中,如果 k 维超平面 α 外一条直线 l 平行于 α 内一条直线 m,则 l 平行于 α.

证明:用反证法.假设 l 不平行于 α,那么 l 必然与 α 相交,设交点为 p.显然 $p\not\subset m$,否则 l 不平行于 m.这样,由直线 m 以及 m 外一点 p 可以唯一确定一个平面.该平面既与 l 平行(l 平行于 m),又与之相交(交点 p),矛盾.所以 l 平行于 a.证毕.

定理 7:$k+1$ 维空间中,如果 k 维超平面 α 平行于 l 维超平面 β 内 l 条两两垂直且交于一点的直线,那么 α 平行于 β.

证明:由定理 4,β 内 l 条两两垂直且交于一点的直线唯一确定 β.α 平行于这些直线,所以 α 平行于 β.证毕.

两个平面之间的夹角:在 n 维空间中,两个平面经过平移后如果相交且不重合,交集可能是一点或一条直线.如果交集是一点,这时两个平面垂直,夹角为直角;如果交集是直线,这时两个平面实际上处于一个三维空间中,在两个平面上分别作直线与两平面的交线垂直,这两条直线之间的夹角称为两个平面之间的夹角.

两个超平面之间的夹角:在 n 维空间中,两个超平面经过平移后如果相交且任何一个超平面都不是另一个超平面的子集,那么交集有多种情况.设两个超平面 α、β 的维数分别为 k、l,由式(5),交集的维数 x 满足如下关系

$$\max(k+l-n,0)\le x\le\min(k,l)-1 \tag{6}$$

如果 $k=l=n-1$,那么 $x=n-2$,即 α、β 的交集是 $n-2$ 维超平面 γ,在 γ 上任取一点 O,过点 O 在 α 内 γ 外作直线 OA 垂直于 γ,再过点 O 在 β 内 γ 外作直线 OB 垂直于 γ,称 $\angle AOB$ 为 α、β 的夹角.

4.3　超平面和超曲面的相互关系

公理 5:n 维空间中,任一 k 维超平面中都包含无穷多较低维数的超曲面.

定理 8:n 维空间中,平面 α 与超球面 σ 相交,交集是一点或一个圆.

证明:交点在 σ 上,所以交点到超球面的球心 O 的距离为定值.O 到 α 的距离为定值,所以交点到 O 在 α 的投影点的距离也是定值.所以交集是圆.证毕.

讨论:一般地,交集是圆.如果圆退化成一点,则称平面 α 与超球面 σ 相切.

定理 9:n 维空间中,k 维超平面 α 与 l 维超球面 σ($k<l$,$k+l>n$)相交,一般地交集是超球面.并且交集的维数 x 满足如下关系:

$$k+l-n\le x\le k-1 \tag{7}$$

证明:α 受到 $n-k$ 个约束,σ 受到 $n-l$ 个约束,如果这些约束相互独立,那么交集的维数为 $n-(n-k)-(n-l)=k+l-n$,否则交集的维数要更大.又由题设,α 与 σ 相交,这就至少有一个约束,所以,$k+l-n\le x\le k-1$.证毕.

讨论:考虑到极限情况,交集可能是一点或一个圆.如果交集是一点,称 α 与 σ 相切.

定理 10:n 维空间中,k 维超球面 α 与 l 维超球面 β($k<l$,$k+l>n$)相交,一般地交集是超球面.并且交集的维数 x 满足如下关系:

$$k+l-n\le x\le k-1 \tag{8}$$

证明及讨论同定理 9.

4.4　常见几何体的体积公式

边长为 a 的 k 维立方体的体积 $V=a^k$.

在推导体积公式之前,首先给出两个公理.

公理 6:n 维空间中,若几何体沿某个方向伸长 a 倍,则其体积扩大 a 倍.

公理 7:n 维空间中,对于夹在两个平行的 k 维超平面 α 和 β 之间的两个 $k+1$ 维几何体 ε_1 和 ε_2,若以平行于 α 和 β 的任意 k 维超平面 δ 截 ε_1 和 ε_2 所得到的两个 k 维超平面的面积恒相等,则 ε_1 和 ε_2 的体积相等.

定理 11:n 维空间中,若 k 维长方体的边长分别为 $a_1,a_2\cdots a_k$,则其体积为 $\prod_{i=1}^{k}a_i$.

证明:由公理 6 即可推出.

定理 12:n 维空间中,k 维超棱锥的体积是与它等底等高的超棱柱的体积的 $1/k$.

证明:首先考虑底边和高相等的情况.以超棱柱的任一顶点为顶点,以包含该顶点的棱为高,以对应面为底,可以做出 k 个超棱锥,显然这 k 个超棱锥全等并且将超棱柱填满.所以每个超棱锥的体积是超棱柱的 $1/k$.再由公理 6,即可得证.

推论:n 维空间中,k 维超圆锥的体积是与它等底等高的超圆柱的体积的 $1/k$.

定理 13:n 维空间中,k 维超圆柱的体积等于底面积乘高,其底面积等于同样半径的 $k-1$ 维超球的体积.

证明:k 维超圆柱的底面显然是 $k-1$ 的超球.构造一个和 k 维超圆柱等底等高的超棱柱,由公理 7 可知,二者具有相同的体积.由定理 11,k 维超棱柱的体积等于底面积乘高,从而得证.

k 维超球的截面是 $k-1$ 维超球,设 $k-1$ 维超球的体积为 V_{k-1},则 $V_k = \int V_{k-1}(x)\,dx$.已知 $V_3 = \frac{3}{4}\pi R^3$,按照递推公式,有:

$$V_k = \begin{cases} \dfrac{\pi^{\frac{k}{2}}}{\left(\frac{k}{2}\right)!} R^k, & k = 2m \\ \dfrac{2^{\frac{k+1}{2}}\pi^{\frac{k-1}{2}}}{k!!} R^k, & k = 2m+1 \end{cases} \tag{9}$$

k 维超球面的面积是对应超球沿半径方向的微分,即 $S = dV/dR$,所以,

$$S_k = \begin{cases} \dfrac{2\pi^{\frac{k}{2}}}{\left(\frac{k}{2}-1\right)!} R^{k-1}, & k = 2m \\ \dfrac{2^{\frac{k+1}{2}}\pi^{\frac{k-1}{2}}}{(k-2)!!} R^{k-1}, & k = 2m+1 \end{cases} \tag{10}$$

定理 14:n 维空间中,k 维超球的体积等于半径与其面积乘积的 $1/k$.

几何体沿某个方向均匀增大 λ 倍,则体积随之增大 λ 倍.由此不难得出超椭球的体积公式.设 k 维超椭球的半轴分别为 $a_1, a_2, \cdots, a_k$,则体积为:

$$V_k = \begin{cases} \dfrac{\pi^{\frac{k}{2}}}{\left(\frac{k}{2}\right)!} \prod_{i=1}^{k} a_i, & k = 2m \\ \dfrac{2^{\frac{k+1}{2}}\pi^{\frac{k-1}{2}}}{k!!} \prod_{i=1}^{k} a_i, & k = 2m+1 \end{cases} \tag{11}$$

5 结论

多维空间几何方法用于定性地分析人工神经网络的行为,在提出新的模型和算法方面,以其简单直观的特点,较传统的分析方法具有独特的优越性.本文初步系统地发展了多维空间几何的分析、推理和证明的方法,为人工神经网络的分析与发展提供了一种新的手段.

参考文献:

[1] 周其节,等.神经网络控制系统的研究与展望[J].控制理论与应用,1992,9(6):569.

[2] J J Hopfield. Neural networks and physical systems with emergent collective computational abilities [A]. Proc. Natl. Acad. Sci., U. S. A, 1982,79:2554 - 2558.

[3] 王守觉,等.通用前馈网络及排序学习前向掩蔽模型在模式识别中的应用[J],电子学报,1998:26(8):1 - 6.

[4] Wang Shoujue etc. The Sequential Learning Ahead Masking (SLAM) Model of Neural Networks for Pattern Classification [A], Proceedings of JCIS'98, RTP, North Carolina, USA October 23 - 28, 1998, IV:199 - 202.

[5] Wang Shoujue. Priority Ordered Neural Networks With Better Similarity to Human Knowledge Representation [J]. Chinese Journal of Electronics, Jan. 1999, 8(1):1 - 4.

[6] Wang Shoujue etc. Priority Ordered Architecture of Neural Networks [A]. #58 Session:4.2, Proceedings of IJCNN'99 [C], Washington, DC, USA. (Electronic Version), July 10 - 16, 1999.

[7] Wang Shoujue etc. Direction-Basis-Function Neural Networks [A]. #89 Session:5.4, Proceedings of IJCNN'99 [C], Washington, DC, USA. (Electronic Version), July 10 - 16, 1999.

[8] A Д亚历山大洛夫等著,王元等译.数学——它的内容、方法和意义[M].第三卷,北京:科学出版社,1962.

作者简介:

王守觉 男.1925 年生于上海,早年就读于西南联大和同济大学,毕业后在北平研究院镭学研究所从事氧化亚铜研究,解放后改为中国科学院应用物理所结晶学室.1960 年成立半导体所后历任器件室主任、副所长、所长等职务.1980 年当选为学部委员(院士).现为中国电子学会副理事长,《电子学报》主编.中国半导体学科奠基人之一,现从事半导体超高速电路与人工神经网络算法、模型、硬件和应用的研究.

王柏南 男.1973 年出生,1997 年获北京大学电子学系理学学士学位,2000 年获中科院半导体所电路与系统专业工学硕士学位.攻读学位期间,从事人工神经网络的模型与算法工作.