

前馈神经网络及其应用

邢红杰 哈明虎 著



科学出版社

013028303

TP183
122

前馈神经网络及其应用

邢红杰 哈明虎 著



科学出版社

北京



北航

C1634756

TP183
122

内 容 简 介

本书较系统地介绍了前馈神经网络的理论与应用. 本书共7章, 主要内容
包括前馈神经网络的模型选择、单个前馈神经网络、混合前馈神经网络和前
馈神经网络的应用.

本书可作为应用数学、计算机科学与技术、信息与通信工程、电气工程、
控制科学与技术等专业高年级本科生、研究生的教材或教学参考书, 也可供
相关领域的科研人员和工程技术人员阅读参考.

图书在版编目(CIP)数据

前馈神经网络及其应用/邢红杰, 哈明虎著. —北京: 科学出版社, 2013

ISBN 978-7-03-037174-4

I. ①前… II. ①邢… ②哈… III. ①前馈-人工神经网络-研究
IV. ①TP183

中国版本图书馆 CIP 数据核字 (2013) 第 051963 号

责任编辑: 王丽平 / 责任校对: 林青梅

责任印制: 钱玉芬 / 封面设计: 陈 敬

科学出版社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

骏杰印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2013 年 3 月第 一 版 开本: B5(720 × 1000)

2013 年 3 月第一次印刷 印张: 12 1/2

字数: 240 000

定价: 56.00 元

(如有印装质量问题, 我社负责调换)

前 言

神经网络是由大量简单的神经元按照一定连接方式形成的智能仿生网。它以非线性神经元作为处理单元,通过广泛连接构成大规模分布式并行处理系统。神经网络不需预知其训练数据中输入输出之间的函数关系,而以数据驱动的方式解决问题。由于神经网络具有强大的模式识别能力和灵活的非线性建模能力,它引起了越来越多的学者及工程技术人员的关注。

神经网络主要有两种:前馈神经网络(feedforward neural network)和反馈(递归)神经网络(recurrent neural network)。相对而言,前馈神经网络主要适用于处理静态数据,而反馈神经网络则更适合处理动态数据。本书主要讨论前馈神经网络。尽管前馈神经网络在实际中得到了广泛的应用,但由于其模型选择至今仍无严格的理论支撑,故模型选择仍是神经网络领域亟待解决的问题。前馈神经网络的模型选择主要包括:对隐含层的个数、输入层和隐含层中的节点个数、隐含层激活函数类型及连接权重的选取。前馈神经网络模型选择的目标是对于一个特定的学习任务,选择合适的网络规模。如果网络规模过大,则网络容易产生“过拟合”(overfitting)现象,即虽然对训练数据暗含的输入输出映射关系可以产生较好的拟合,而网络的泛化能力却较差。如果网络的规模过小,则网络容易出现“欠拟合”(underfitting)现象,即网络学习的时间很长或者根本无法学习。因此,对前馈神经网络模型选择问题的研究具有非常重要的意义。

此外,尽管前馈神经网络的理论和算法不断完善,但是单个前馈神经网络在分类(或函数逼近)性能上难以得到很大的提高。为了追求更高的分类准确率(或逼近精度),对不同类型的单个前馈神经网络进行混合使用便自然是一种切实可行的办法,由此产生了混合前馈神经网络。混合前馈神经网络的目的是“综合利用各种类型的单个前馈神经网络的优点”。虽然混合前馈神经网络目前尚无严格统一的定义,但是它主要有两种混合方式:串联和并联,其中串联方式是指在混合网络的两个或者多个隐含层中分别使用不同类型的激活函数,而同一隐含层中的激活函数类型则完全相同。并联方式具有多种构造形式,如在混合网络的同一隐含层中使用不同类型的激活函数,或者将不同类型的单个前馈神经网络通过特定的方式直接并联在一起。与单个前馈神经网络相比,混合前馈神经网络具有更高的灵活度和学习能力,从而它一直是前馈神经网络领域的研究热点。

本书主要介绍作者已公开发表和尚未发表的研究工作。全书共7章,其内容安排如下:第1章为绪论。第2章为有监督学习前馈神经网络。第3章为无监督学习

前馈神经网络. 第 4 章为前馈神经网络的模型选择. 第 5 章为单个前馈神经网络. 第 6 章为混合前馈神经网络. 第 7 章为前馈神经网络的应用.

本书的部分内容选自于第一作者邢红杰的博士论文, 在此特向悉心指导博士论文的指导教师、中国科学院自动化研究所胡包钢研究员表示衷心感谢. 陈继强讲师、张振辉博士和王超博士生参加了本书的讨论, 硕士生王新美和刘李飞参与了本书的排版和校对, 在此一并向他们表示感谢.

本书的部分研究内容得到了中国博士后科学基金 (项目编号: 20080440820), 国家自然科学基金 (项目编号: 60903089, 61075051, 61073121), 河北省自然科学基金 (项目编号: F2012402037), 河北省教育厅自然科学基金青年基金项目 (项目编号: Q2012046) 和河北大学杰出青年基金 (项目编号: 3504020) 的资助, 特此致谢.

由于作者学识和水平所限, 书中不足之处在所难免, 敬请同仁及读者批评指正.

作 者

2012 年 12 月 28 日

符号说明

$\mathcal{R}$	实数域
$\mathcal{R}^d$	d 维实数域
$\mathbf{x}$	输入向量
t	目标输出向量
y_k	多层感知器第 k 个输出节点的输出
z_j	第 j 个隐含节点的输出
w_{ji}	连接第 j 个隐含节点和第 i 个输入节点的权重
w_{j0}	第 j 个隐含节点的偏差项
v_{kj}	连接第 k 个输出节点和第 j 个隐含节点的权重
v_{k0}	第 k 个输出节点的偏差项
E	误差函数
η	学习率
α	动量常数
ε	停止阈值
$\mathbf{c}_j$	径向基函数神经网络的第 j 个隐含节点的中心向量
σ_j	宽度参数
$T_j(\cdot)$	切比雪夫多项式正交函数
N_{train}	训练样本个数
$K(\cdot, \cdot)$	核函数
$\mathbf{K}$	核矩阵
N_{SV}	支持向量个数
$\mathbf{D}$	训练样本集
$\ \cdot\ $	欧氏范数
τ	迭代次数
$h_{ji}(\mathbf{x})$	自组织映射神经网络的近邻函数
T_{max}	最大迭代次数
$\phi(\cdot)$	从输入空间到特征空间的映射
∇J	目标函数的梯度
$\mathbf{X}$	随机向量
$\Sigma_{\mathbf{X}}$	随机向量 $\mathbf{X}$ 的协方差矩阵
CV	交叉验证误差
$H(C)$	熵

$H(C X)$	条件熵
$I(C; X)$	互信息
$\Gamma(\cdot)$	Gamma 函数
$\mathbf{I}$	单位矩阵
μ_{ji}	第 j 个样本属于第 i 个聚类的隶属度
$\text{sgn}(\cdot)$	符号函数
RMSE	均方根误差
$p(\mathbf{x})$	概率密度函数
π_k	高斯混合模型的混合权重
$\text{tr}[\cdot]$	矩阵的迹
Φ_{θ_i}	参数 θ_i 的重要性
$\mathbf{S}_{y\theta}$	输出敏感度矩阵
$g_i(\cdot)$	门网函数
$\ \cdot\ _1$	$L1$ 范数
β	特征权重向量

目 录

前言

符号说明

第 1 章 绪论	1
1.1 有监督学习和无监督学习	1
1.1.1 有监督学习	1
1.1.2 无监督学习	3
1.2 神经网络的分类	3
1.2.1 前馈神经网络	4
1.2.2 反馈神经网络	6
1.3 前馈神经网络的模型选择与混合策略	6
1.3.1 前馈神经网络的模型选择	6
1.3.2 前馈神经网络的混合策略	9
参考文献	10
第 2 章 有监督学习前馈神经网络	14
2.1 多层感知器神经网络	14
2.1.1 网络结构	14
2.1.2 学习算法	15
2.1.3 逼近理论	17
2.2 径向基函数神经网络	18
2.2.1 网络结构	18
2.2.2 学习算法	19
2.2.3 逼近理论	21
2.3 切比雪夫神经网络	22
2.3.1 网络结构	22
2.3.2 学习算法	23
2.3.3 逼近理论	24
2.4 支持向量机	25
2.4.1 网络结构	25
2.4.2 学习算法	27

2.4.3 逼近理论	28
参考文献	29
第 3 章 无监督学习前馈神经网络	33
3.1 自组织映射神经网络	33
3.1.1 网络结构	33
3.1.2 学习算法	34
3.1.3 核自组织映射神经网络	35
3.2 神经气网络	37
3.2.1 学习算法	38
3.2.2 核神经气网络	39
3.2.3 生长型神经气网络	40
3.3 主成分分析及其改进方法	41
3.3.1 主成分分析	42
3.3.2 核主成分分析	45
3.3.3 二维主成分分析	46
参考文献	47
第 4 章 前馈神经网络的模型选择	49
4.1 基于假设检验的方法	49
4.1.1 Wald-检验	49
4.1.2 LM-检验	50
4.2 基于信息准则的方法	51
4.2.1 AIC 准则和 BIC 准则	51
4.2.2 最小描述长度和交叉验证	52
4.3 基于敏感度分析的方法	53
4.3.1 基于偏导数的敏感度分析方法	53
4.3.2 基于随机分析的敏感度分析方法	58
4.4 基于互信息的方法	64
4.4.1 互信息及其估计	64
4.4.2 基于互信息的多层感知器两阶段构造方法	67
参考文献	78
第 5 章 单个前馈神经网络	81
5.1 基于正则化相关熵的径向基函数神经网络学习方法	81
5.1.1 正则化相关熵准则	82
5.1.2 数值实验	87
5.2 椭圆基函数神经网络的混合学习方法	91

5.2.1 椭球基函数神经网络	92
5.2.2 椭球基函数神经网络的混合学习策略	93
5.2.3 数值实验	96
5.3 基于互信息的特征加权支持向量机	100
5.3.1 基于互信息的特征权重估计	101
5.3.2 特征加权支持向量机	103
5.3.3 数值实验	105
参考文献	111
第 6 章 混合前馈神经网络	116
6.1 高斯、Sigmoid、切比雪夫混合前馈神经网络	116
6.1.1 Gauss-Sigmoid 神经网络	116
6.1.2 高斯-切比雪夫神经网络	117
6.1.3 数值实验	120
6.2 基于自适应模糊 c 均值的混合专家模型	125
6.2.1 基于 PBMF-index 的模糊 c 均值聚类算法	129
6.2.2 结构描述和实现方法	130
6.2.3 数值实验	131
参考文献	142
第 7 章 前馈神经网络的应用	146
7.1 前馈神经网络在人脸识别中的应用	146
7.2 前馈神经网络在非线性时间序列预测中的应用	150
7.3 前馈神经网络在图像分割中的应用	154
7.4 前馈神经网络在异常检测中的应用	157
参考文献	162
附录 部分前馈神经网络的 Matlab 源代码	165
附录 1 基本模型	165
附录 2 模型选择	172
附录 3 改进模型	176
索引	188

第1章 绪 论

神经网络是由简单处理单元构成的大规模并行分布式处理系统, 具有联想记忆和存储知识的特性. 神经网络是从人脑的结构出发模拟人的智能行为. 它与人脑具有相似之处, 主要表现在两个方面: 一是通过学习过程从外部环境中获取知识; 二是利用内部神经元存储知识.

神经网络为逼近连续值、离散值及向量值的目标函数提供了鲁棒的学习方法. 它善于联想、概括、类比和推广. 对于学习、解释复杂的传感器数据等问题, 神经网络是一种高效的学习方法. 另外, 神经网络具有很强的自学习能力, 可在学习过程中不断完善.

神经网络涉及应用数学、电子科学与技术、信息与通信工程、计算机科学与技术、电气工程、控制科学与技术等诸多学科, 其应用已扩展到模式识别与图像处理、时间序列分析、控制与优化、系统辨识、通信等众多领域.

1.1 有监督学习和无监督学习

在学习神经网络参数时, 如果训练样本的目标输出已知, 可以统计出训练样本不同的描述量, 如概率分布或在特征空间分布的区域等. 利用这些信息进行神经网络设计的过程称为有监督学习. 在实际应用中, 有时训练样本的目标输出未知, 仅能利用训练样本的输入特征进行神经网络设计, 这种学习过程就是无监督学习.

1.1.1 有监督学习

图 1.1 给出了有监督学习系统的框图. 有监督学习的学习形式是误差-修正学习 (error-correction learning) 的基础. 注意在图 1.1 中, 未知环境不包含在框图中. 通常可以采用训练样本的均方误差或平方误差和作为有监督学习系统性能的目标函数, 该函数可以定义为学习系统关于自由参数 (如连接权重) 的函数. 随着时间的变化, 学习系统的性能会逐渐提高, 同时其目标函数值也会越来越小, 直至达到误差极小点, 该极小点可能是局部最小点, 也可能是全局最小点.

对于有监督学习, 需要给出训练样本的目标输出. 若目标输出为连续值, 则该学习问题为回归问题. 若目标输出仅能在一个包含有限个元素的集合中取值, 则该学习问题为分类问题. 下面分别举例说明.

对于回归问题, 考虑一维 Sinc 函数:

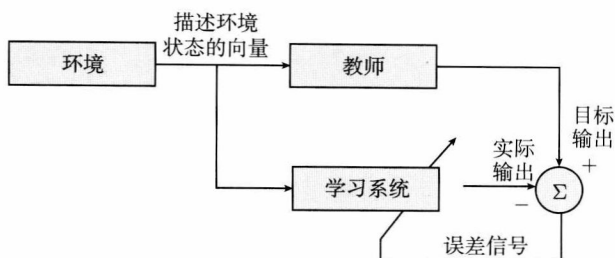


图 1.1 有监督学习系统的框图

$$y = \frac{\sin(x)}{x} + \rho \quad (1.1)$$

其中 $\rho \sim N(0, \sigma^2)$ 为服从高斯分布的噪声。按式 (1.1) 生成 100 个训练数据 $\{(x_i, y_i)\}_{i=1}^{100}$, x_i 在区间 $[-10, 10]$ 上均匀抽取, 标准差 $\sigma = 0.2$ 。所产生的训练数据如图 1.2 所示。利用这些训练数据学习神经网络, 得到回归函数。由神经网络产生的回归曲线如图 1.2 所示。对于待测数据 x , 将其输入神经网络, 即可得到相应的网络输出。

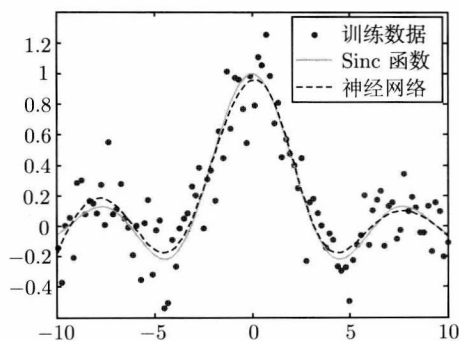


图 1.2 神经网络在 Sinc 数据集上的回归效果

对于分类问题, 考虑图 1.3 中的香蕉型数据集, 该数据集是二维空间中线性不可分的人工数据集, 它由 100 个样本组成, 其中正类和负类样本各 50 个, 两类样本

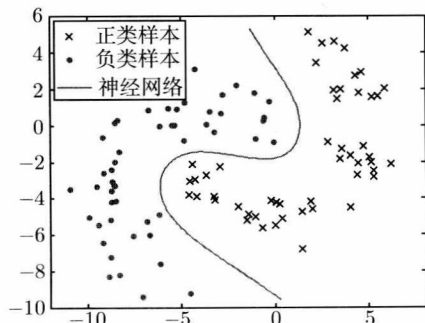


图 1.3 神经网络在香蕉型数据集上的分类效果

均呈现为香蕉形状. 利用训练数据学习神经网络, 得到决策函数, 所得分类边界如图 1.3 所示. 对于待测特征向量 $x \in \mathbb{R}^2$, 将之输入神经网络产生的决策函数, 即可得到 x 的类别标号.

1.1.2 无监督学习

在无监督学习中, 没有用于监督学习过程的目标输出. 无监督学习系统的框图如图 1.4 所示. 无监督学习必须提供任务独立度量 (task-independent measure) 来评价神经网络的表达质量, 让网络学习该度量并根据该度量优化网络参数. 对一个特定的任务独立度量, 一旦神经网络能够和输入数据的统计规律一致, 则网络将具有描述输入数据编码特征的内部表示能力, 从而自动产生新的类别^[1].

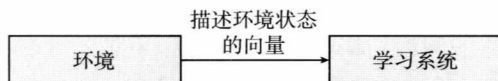


图 1.4 无监督学习系统的框图

为了完成无监督学习, 可以使用竞争学习规则. 例如, 可以采用包含两层神经网络: 输入层和竞争层. 输入层接收训练数据, 竞争层由相互竞争 (按一定的学习规则) 的神经元组成, 它们试图获得响应包含在训练数据中的特征的“机会”. 最简单的形式就是神经网络采用“胜者全得”(winner-take-all) 的学习策略.

考虑图 1.5(a) 中的人工数据集, 该数据集包含 100 个二维样本点, 分布在 5 组中. 这些样本的类别标号是未知的. 采用上述竞争学习规则对神经网络进行训练, 即可得到图 1.5(b) 的分组结果, 从而得到每个样本的类别标号.

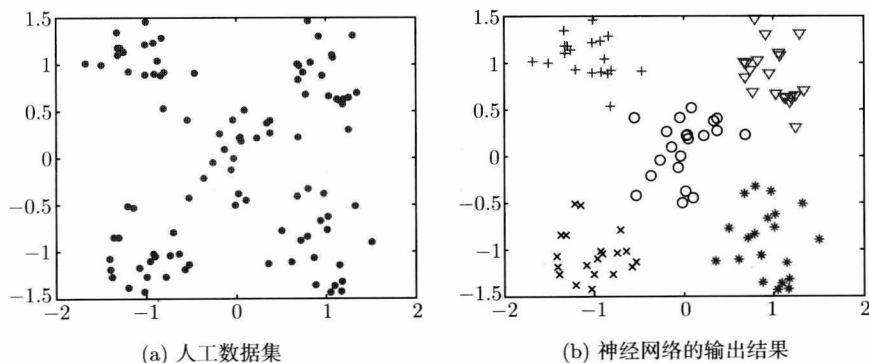


图 1.5 神经网络在人工数据集上的分组结果

1.2 神经网络的分类

可以从不同的角度对神经网络进行分类, 如①从网络性能角度可将神经网络分为连续型神经网络与离散型神经网络、确定性神经网络与随机性神经网络; ②从网

络结构角度可将神经网络分为前馈神经网络与反馈神经网络; ③从学习方式角度可将神经网络分为有监督学习神经网络和无监督学习神经网络; ④从突触连接性质角度可将神经网络分为一阶线性关联神经网络和高阶非线性关联神经网络。

下面简要介绍几种常用的前馈神经网络和反馈神经网络。

1.2.1 前馈神经网络

本书中的前馈神经网络包括两大类: 单个前馈神经网络和混合前馈神经网络。混合前馈神经网络由两个或两个以上单个前馈神经网络按照某种混合方式组合而成。下面分别介绍几种常用的前馈神经网络。

1.2.1.1 单层前馈神经网络

单层前馈神经网络的网络结构如图 1.6 所示, 所谓的“单层”是指前馈神经网络拥有的计算节点仅有一层, 输入层中的节点不具有执行计算的功能。单层感知器和自适应线性元件均属于单层前馈神经网络。

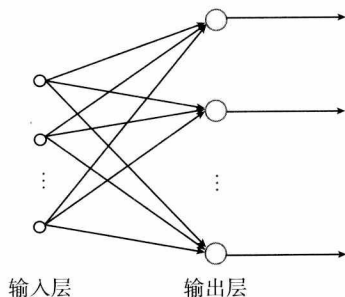


图 1.6 单层前馈神经网络

1.2.1.2 多层前馈神经网络

多层前馈神经网络是单层前馈神经网络的推广, 两者的区别在于: 多层前馈神经网络含有一个或更多的隐含层, 其计算节点被称为隐含神经元或隐含节点, 一个四层前馈神经网络的网络结构如图 1.7 所示。

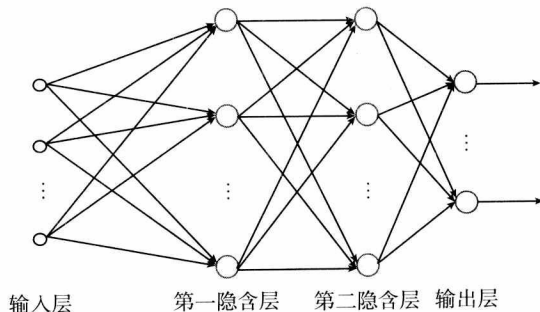


图 1.7 四层前馈神经网络

多层前馈神经网络输入层中的每个源节点的激励模式 (输入向量) 作为第二层 (第一隐含层) 的输入信号, 第二层输出信号成为第三层的输入, 其余层类似. 网络各隐含层中的神经元将它们前一层神经元的输出信号用作输入. 最终, 网络输出层神经元的输出信号组成了对网络中输入层源节点产生的激励模式的全部响应.

本书中的多层感知器和径向基函数神经网络均属于多层前馈神经网络.

1.2.1.3 竞争前馈神经网络

竞争前馈神经网络的主要特点是其输出神经元相互竞争以确定胜者, 胜者会指出哪个原型向量 (中心向量) 最能代表输入向量. Hamming 网络是一种最简单的竞争前馈神经网络, 其网络结构如图 1.8 所示. 该神经网络有一个单层的输出神经元, 每个输出神经元都与输入节点全连接, 输出神经元之间也相互连接且相互横向侧抑制.

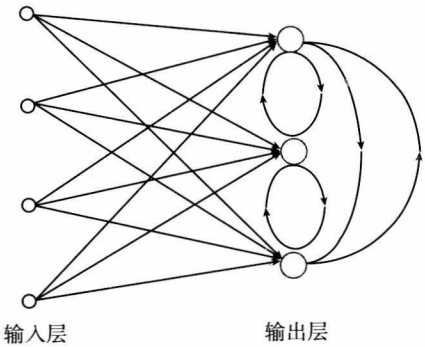


图 1.8 Hamming 网络

本书中的自组织映射神经网络属于竞争前馈神经网络.

1.2.1.4 混合前馈神经网络

混合前馈神经网络由两个或两个以上单个前馈神经网络混合形成. 图 1.9 所示

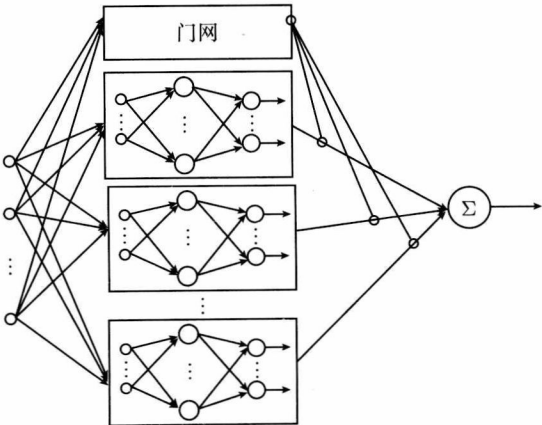


图 1.9 并联型混合前馈神经网络

的是并联型混合前馈神经网络, 它由多个单个前馈神经网络和一个门网组成, 门网充当协调者的角色, 按贡献量的大小为各单个前馈神经网络赋予不同的权重, 使不同的单个前馈神经网络在输入空间的不同区域上工作得最好。

1.2.2 反馈神经网络

在反馈神经网络中至少含有一个反馈回路。反馈神经网络可以包含一个单层神经元, 其中每个神经元将自己的输出信号反馈给其他所有神经元的输入, 如图 1.10(a) 所示的 Hopfield 网络, 图 1.10(b) 所示的是含有隐含层的反馈神经网络, 图 1.10(a) 和图 1.10(b) 中的反馈连接始于隐含神经元和输出神经元, 且反馈连接没有自反馈回路。

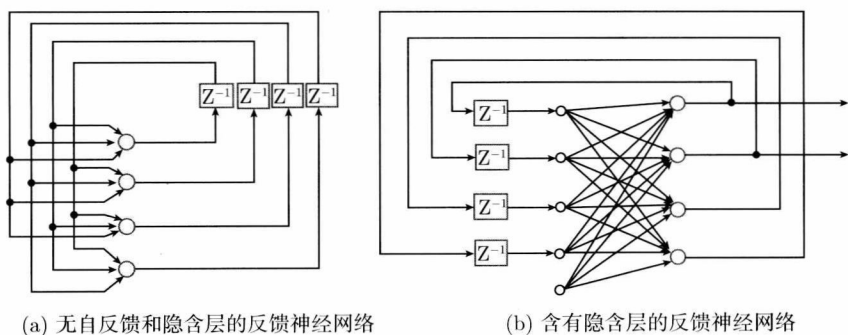


图 1.10 反馈神经网络

1.3 前馈神经网络的模型选择与混合策略

由于具有强大的模式识别能力和灵活的非线性建模能力, 前馈神经网络一直是模式识别和机器学习领域关注的焦点。但是, 我们仍需面对前馈神经网络的两大挑战: 网络结构选取和最优权重训练^[2]; 混合前馈神经网络的混合策略^[3]。下面分别对这两方面的研究成果进行简要介绍。

1.3.1 前馈神经网络的模型选择

前馈神经网络的模型选择主要包括: 隐含层个数、输入层和隐含层中的节点个数、隐含层激活函数类型及连接权重的选取。本节主要介绍前馈神经网络模型选择的研究进展, 并介绍与选取最优连接权重相关的前馈神经网络学习方法。

1.3.1.1 结构选取

前馈神经网络的泛化能力取决于三个主要因素, 即问题本身的复杂程度、网络结构及样本量大小^[4]。前馈神经网络结构选取的任务与模型选择完全相同, 在本

书中对两者不加区分. 下面仅讨论多层感知器和径向基函数神经网络的结构选取问题.

对于早期的前馈神经网络结构选取, Andres 和 Kron^[5] 将已有方法分为三类: 正则化方法、剪枝法和停止训练法.

在正则化方法中, 网络权重的选取需要满足条件: 最小化以网络复杂度为惩罚项的目标函数. 于是训练网络的目标函数可以表示为

$$E = E_T + \lambda E_C, \quad (1.2)$$

其中, E_T 为误差函数 (如误差平方和), E_C 为惩罚项^[6,7]. 常数 λ 控制惩罚项的影响程度. 为了自动选取最优的 λ , MacKay^[8,9] 提出了贝叶斯正则化方法.

剪枝法是从一个较大的网络开始逐渐删去一些不重要的权重或节点^[10]. 被删去的权重或节点往往对整个网络的性能不能提供“显著性”的贡献. 然而, 这种“显著性”通常不能在统计测试的基础上得到. 剪枝的方法为: 将某个权重设为零, 则这个权重的“显著性”定义为网络模型误差的增加, 从而移除那些“显著性”较小的权重.

在停止训练法中, 训练数据集被分为训练集 (training set) 和验证集 (validation set). 训练集用来优化拟合度规则, 而验证集则用来反映模型的性能^[11]. 如果在训练过程中模型对于验证集的误差开始增加则停止训练.

虽然这些传统方法可以取得满意的效果, 但是它们却偏向于启发式. 因此针对不同的传统结构选取方法, 出现了多种改进策略, 如基于敏感度分析的方法和基于信息理论的方法.

基于敏感度分析的方法属于剪枝方法, 它利用网络输出 (或目标函数) 关于参数的导数来度量参数与输出的相关程度. 用敏感度分析方法删减多层感知器不相关输入节点的代表性工作可参见文献^[12]. Engelbrecht^[13] 将输入节点剪枝和隐含节点剪枝合并为一个问题, 提出了基于统计检验的剪枝方法并取得了较好的实验结果. Zeng 和 Yeung^[14] 提出了基于随机敏感度分析的多层感知器隐含节点剪枝方法, 在所提方法中, 隐含节点的重要性由其随机敏感度和连接权重进行度量. 在互信息的基础上, Xing 和 Hu^[15] 提出了基于信息理论准则的多层感知器结构选取方法. 在该方法中, 隐含节点的相关度由互信息和权重的贡献量确定, 实验结果表明该方法能够产生优于敏感度分析^[13] 的网络结构和泛化性能.

近年来, 又出现了许多新的前馈神经网络结构选取方法. Xiang 等^[16] 从几何角度对多层感知器进行了解释, 为多层感知器提出了有效的结构选取方法, 并指出如果在目标函数图像的中间部位有平整曲面 (flat surface), 则使用四层感知器代替三层感知器会取得更优的结果. Yeung 等^[17] 将局部泛化误差界用于确定径向基函数神经网络的最优网络结构. Eğrioglu 等^[18] 提出了基于多种信息准则的神经网络结