



智能科学与技术本科专业系列教材

HINENG KEXUE YU JISHU BENKE ZHUANYE XILIE JIAOCAI

# 神经网络设计 方法与实例分析

施彦 韩力群 廉小亲 编著



北京邮电大学出版社  
[www.buptpress.com](http://www.buptpress.com)

智能科学与技术本科专业系

# 神经网络设计方法与实例分析

施彦 韩力群 廉小亲 编著

北京邮电大学出版社  
• 北京 •

## 内 容 提 要

本书从神经网络设计和应用实践出发,介绍了10种常见的人工神经网络的基本原理、设计方法,并从各个应用领域精选了丰富的典型应用实例进行剖析,旨在使读者对各类常用的人工神经网络的基本原理和学习算法进一步加深理解,熟悉其主要功能,掌握其设计方法,了解其主要应用,为设计各类神经网络和解决实际问题打下基础。主要内容包括神经网络模型评估与选择;10种典型网络的设计与应用,包括BP网络、RBF网络、SOFM网络、LVQ网络、CPN网络、ART网络、Hopfield网络、时序递归网络、CMAC网络、SVM网络;最后介绍了人工神经元网络设计开发平台。

本书可作为具有一定人工神经网络理论基础的科技工作者解决实际问题时的设计参考书,也可作为相关专业研究生及本科专业高年级学生的参考教材。

### 图书在版编目(CIP)数据

神经网络设计方法与实例分析/施彦,韩力群,廉小亲编著.--北京:北京邮电大学出版社,2009.12  
ISBN 978-7-5635-2102-9

I. 神… II. ①施…②韩…③廉… III. 人工神经—神经网络 IV. TP183

中国版本图书馆CIP数据核字(2009)第184573号

---

书 名:神经网络设计方法与实例分析  
作 者:施 彦 韩力群 廉小亲  
责任编辑:周 堃  
出版发行:北京邮电大学出版社  
社 址:北京市海淀区西土城路10号(邮编:100876)  
发 行 部:电话:010-62282185 传真:010-62283578  
E-mail:publish@bupt.edu.cn  
经 销:各地新华书店  
印 刷:北京源海印刷有限责任公司  
开 本:787 mm×960 mm 1/16  
印 张:23.5  
字 数:512千字  
印 数:1—3000册  
版 次:2009年12月第1版 2009年12月第1次印刷

---

ISBN 978-7-5635-2102-9

定 价:42.00元

· 如有印装质量问题,请与北京邮电大学出版社发行部联系 ·

# 前 言

人工神经网络自 20 世纪 80 年代复兴以来,已经经过了 20 余年。除了神经网络理论得到了进一步的发展以外,神经网络的应用成果也日益丰富。神经网络今后发展的一个方向是将其成功地应用于生产、生活的各个方面,发挥其信息处理能力,扩展其应用范围。作者在教学和科研中发现,即便掌握了一些神经网络的基本理论,在设计和应用中也需要一定的经验和指导。特别是初涉此领域的学生和工程技术人员,面对需要解决的问题,究竟选择哪一种网络和学习算法更为合适,如何进行设计和模型选择,都是比较困难的事情。本书编写的目的正是为已具有一定神经网络基础,希望应用神经网络解决实际问题的有关人员提供一定的指导和参考。

基于此目的,本书选择了近年应用较为普及的典型网络,其中包括静态网络和动态网络;既有监督式学习网络,也有无监督式学习网络;既有前馈网络,也有反馈网络;既有单个网络的应用,也有多个网络协同工作的应用。全书共 12 章,第 1 章介绍了通用的神经网络选择原则,常用的模型性能评价方法,样本的选择和使用等;第 2、3 章介绍了典型的前馈网络 BP 网络和 RBF 网络的基本理论、设计方法和应用实例;第 4、5、6 章分别介绍了基于竞争学习的 SOFM 网络、LVQ 网络以及 CPN 网络的基本理论、设计方法和应用实例;第 7 章介绍了自适应共振 ART 网络的基本理论、设计方法和应用实例;第 8 章介绍了具有联想记忆和优化计算功能的 Hopfield 网络的基本理论、设计方法和应用实例;第 9 章侧重介绍有反馈和时延环节的、适合处理时间序列问题的时序递归网络的基本理论、设计方法和应用实例;第 10 章介绍了智能控制中常用的局部逼近网络——CMAC 网络——的基本理论、设计方法和应用实例;第 11 章介绍了善于处理小样本问题的支持向量机的基本理论、设计方法和应用实例;第 12 章介绍了各种人工神经元网络设计开发平台。本书的 3 位作者分工如下:第 1 章及第 2、3、4、5、6、8、9、11 章的部分基本理论和实例分析由施彦撰写;第 7 章及第 2~11 章的基本理论和部分实例分析由韩力群撰写;第 12 章及第 2、4 章部分实例分析由廉小亲撰写。

本书重点介绍人工神经网络的主要设计方法,从各个应用领域精选了丰富的典型应用实例进行剖析,旨在使读者对各类常用人工神经网络的基本原理和学习算法进一步加深理解,熟悉其主要功能,掌握其设计方法,了解其主要应用,为设计各类神经网络和解决实际问题打下基础。

本书在编写过程中,特别注意以下几点:①实用性——理论联系实际,注意基本原理、设计方法和应用实例相结合,以便读者通过学习了解如何设计实用的神经网络,如何入手解决实际应用问题。②新颖性——体例、风格新颖,与常见的人工神经网络理论性教材有较大区别,特别重视实例分析、网络设计、模型适用性分析和模型选择。③实践性——作为科技人员解决实际问题时的设计参考书,具有较强的示范和指导意义;作为一种实践性教学环节的教材,则重在指导读者应用神经网络解决实际问题,积累分析和解决问题的能力。

本书可作为具有一定人工神经网络理论基础的科技工作者解决实际问题时的设计参考书,也可作为相关专业研究生及智能科学与技术本科专业的实践性环节的教材。

本书参考了许多文献,特别是在实例分析部分,除了本书作者的研究成果以外,还参考了许多其他作者的研究成果,在此一并表示感谢。

编 者

# 目 录

<b>第 1 章 神经网络模型评估及选择</b> .....	1
1.1 神经网络的泛化能力 <sup>[1-3]</sup> .....	1
1.2 神经网络预测模型的一般描述 <sup>[2,3]</sup> .....	2
1.3 研究神经网络泛化能力的理论基础 <sup>[2-6]</sup> .....	3
1.3.1 经验风险最小化原则.....	3
1.3.2 结构风险最小化原则.....	4
1.3.3 偏差-方差分解.....	5
1.4 影响神经网络的泛化能力的具体因素 <sup>[2-5]</sup> .....	6
1.4.1 神经网络具有泛化能力的基本必要条件.....	7
1.4.2 噪声的影响.....	7
1.4.3 “欠拟合”和“过拟合”.....	8
1.5 提高神经网络的泛化能力的方法 <sup>[4,6]</sup> .....	9
1.5.1 模型结构选择.....	9
1.5.2 训练集扩展方法.....	11
1.5.3 提前停止.....	12
1.5.4 权值衰减.....	12
1.5.5 贝叶斯学习.....	13
1.5.6 神经网络集成 <sup>[7-14]</sup> .....	14
1.6 神经网络模型的评估与选择 <sup>[4-6,15]</sup> .....	17
1.6.1 神经网络模型的评估.....	17
1.6.2 神经网络模型的选择.....	18
本章参考文献.....	22
<b>第 2 章 基于 BP 算法的多层感知器的设计与应用</b> .....	23
2.1 基于 BP 算法的多层感知器网络工作原理与主要特点 <sup>[1-6]</sup> .....	23
2.1.1 基于 BP 算法的多层前馈网络模型.....	23

2.1.2	BP 学习算法 .....	25
2.1.3	BP 算法的程序实现 .....	27
2.1.4	标准 BP 算法的改进 .....	28
2.2	基于 BP 算法的多层前馈网络设计基础 <sup>[1-6]</sup> .....	29
2.2.1	网络信息容量与训练样本数 .....	30
2.2.2	训练样本集的准备 .....	30
2.2.3	初始权值的设计 .....	34
2.2.4	多层前馈网结构设计 .....	34
2.2.5	网络训练与测试 .....	36
2.3	基于 BP 算法的多层前馈网络应用与设计实例 .....	37
2.3.1	基于 BP 算法的多层前馈网络在催化剂配方建模中的应用 <sup>[8]</sup> .....	37
2.3.2	基于 BP 算法的多层前馈网络在城市年用水量预测中的应用 <sup>[9]</sup> .....	38
2.3.3	基于 BP 算法的多层前馈网络在煤与瓦斯突出预测中的应用 <sup>[10]</sup> .....	41
2.3.4	基于 BP 算法的多层前馈网络在磨煤机料位监测中的应用 <sup>[11]</sup> .....	45
2.3.5	基于 BP 算法的多层前馈网络在道路安全评价中的应用 <sup>[12]</sup> .....	48
2.3.6	基于 BP 算法的多层前馈网络在项目投资风险评价中的应用 <sup>[13]</sup> .....	54
2.3.7	基于 BP 算法的多层前馈网络在大气环境质量评价中的应用 <sup>[14]</sup> .....	56
2.3.8	基于 BP 网络集成的除草剂定量构效关系模型 <sup>[15,16]</sup> .....	66
2.3.9	基于 BP 网络集成的物流中心选址模型 <sup>[17]</sup> .....	70
	本章参考文献 .....	75
<b>第 3 章</b>	<b>径向基函数神经网络的设计与应用 .....</b>	<b>77</b>
3.1	径向基函数网络原理与学习算法 <sup>[1]</sup> .....	77
3.1.1	正则化 RBF 网络原理与学习算法 .....	77
3.1.2	广义 RBF 网络原理与学习算法 .....	80
3.2	RBF 网络的设计要点 <sup>[2]</sup> .....	83
3.2.1	随机选取 RBF 中心 .....	83
3.2.2	自组织学习选取 RBF 中心及网络设计 .....	83
3.2.3	有监督学习选取 RBF 中心及网络设计 .....	84
3.2.4	正交最小二乘(OLS)法选取 RBF 中心及网络设计 .....	84
3.2.5	递归最小二乘(OLS)法选取 RBF 中心及网络设计 .....	84
3.2.6	其他方法 .....	85
3.2.7	RBF 网络与多层感知器的特点与设计比较 .....	85
3.3	RBF 网络的应用实例 .....	86
3.3.1	RBF 网络在液化气销售量预测中的应用 <sup>[3]</sup> .....	86

3.3.2	RBF 网络在地表水质评价中的应用 <sup>[4]</sup>	86
3.3.3	RBF 网络在汽油干点软测量中的应用 <sup>[5]</sup>	88
3.3.4	RBF 网络在地下温度预测中的应用 <sup>[6]</sup>	89
3.3.5	RBF 网络在工程车辆自动变速控制中的应用 <sup>[7]</sup>	92
3.3.6	RBF 网络在人脸年龄估计中的应用 <sup>[8]</sup>	95
3.3.7	RBF 网络在专利发展趋势预测中的应用 <sup>[9]</sup>	97
3.3.8	RBF 网络在图像融合中的应用 <sup>[10]</sup>	100
3.3.9	RBF 网络红外光谱法用于中药大黄样品的真伪分类 <sup>[11]</sup>	104
3.3.10	RBF 网络在船用柴油机智能诊断中的应用 <sup>[12]</sup>	106
3.3.11	RBF 网络在多级入侵检测中的应用 <sup>[13]</sup>	109
	本章参考文献	112
<b>第 4 章 SOFM 网络设计与应用</b>		<b>113</b>
4.1	SOFM 网络原理与学习算法 <sup>[1]</sup>	113
4.2	SOFM 网络的设计基础 <sup>[1]</sup>	115
4.2.1	输出层设计	115
4.2.2	权值初始化问题	116
4.2.3	优胜邻域 $N_{j^*}(t)$ 的设计	117
4.2.4	学习率 $\eta(t)$ 的设计	117
4.3	应用与设计实例	117
4.3.1	SOFM 网络用于皮革外观效果分类 <sup>[2]</sup>	117
4.3.2	SOFM 网络用于物流中心城市分类评价 <sup>[3]</sup>	119
4.3.3	SOFM 网络用于遥感影像分类 <sup>[4]</sup>	121
4.3.4	SOFM 网络用于火焰燃烧诊断 <sup>[5]</sup>	123
4.3.5	SOFM 网络在防火树种分类中的应用 <sup>[6]</sup>	125
4.3.6	SOFM 在基于汉字与部件聚类的汉字认知建模中的应用 <sup>[7]</sup>	128
4.3.7	SOFM 网络在膨胀土分类中的应用 <sup>[8]</sup>	132
4.3.8	SOFM 网络在水电厂技术供水系统故障诊断中的应用 <sup>[9]</sup>	136
4.3.9	SOFM 网络在不同地区人力资本构成分析中的应用 <sup>[10]</sup>	141
	本章参考文献	144
<b>第 5 章 LVQ 网络设计与应用</b>		<b>146</b>
5.1	LVQ 网络原理与学习算法 <sup>[1]</sup>	146
5.1.1	LVQ 网络工作原理	146
5.1.2	LVQ 网络的学习算法	147



5.2	LVQ 网络设计要点 <sup>[1]</sup> .....	149
5.2.1	竞争层设计 .....	149
5.2.2	权值初始化问题 .....	150
5.2.3	学习率 $\eta(t)$ 的设计 .....	151
5.3	应用与设计实例 .....	151
5.3.1	LVQ 网络在证券投资基金分类中的应用 <sup>[2]</sup> .....	151
5.3.2	LVQ 网络在探地雷达探雷中的应用 <sup>[3]</sup> .....	153
5.3.3	LVQ 网络在苹果等级判别中的应用 <sup>[4]</sup> .....	154
5.3.4	LVQ 网络在胃癌组织样品分类识别中的应用 <sup>[5]</sup> .....	157
5.3.5	LVQ 网络在液压系统故障诊断中的应用 <sup>[6]</sup> .....	159
5.3.6	LVQ 网络在汽车信贷客户分类中的应用研究 <sup>[7]</sup> .....	161
5.3.7	LVQ 网络在土地利用/覆盖变化探测中的应用 <sup>[8]</sup> .....	162
5.3.8	LVQ 网络在周期信号识别方面的扩展应用 <sup>[9]</sup> .....	165
	本章参考文献 .....	168
<b>第 6 章</b>	<b>对偶传播神经网络</b> .....	<b>169</b>
6.1	网络运行原理及学习算法 <sup>[1]</sup> .....	169
6.1.1	网络结构及运行原理 .....	169
6.1.2	CPN 的学习算法 .....	170
6.2	CPN 网的设计与改进 <sup>[1]</sup> .....	172
6.2.1	CPN 网的设计要点 .....	172
6.2.2	改进的 CPN 网 .....	172
6.3	CPN 网的应用 .....	173
6.3.1	CPN 网在烤烟烟叶颜色模式分类中的应用 <sup>[1]</sup> .....	173
6.3.2	CPN 网络的水电机组振动故障诊断中的应用 <sup>[2]</sup> .....	174
6.3.3	CPN 网络在博士学位论文质量评估中的应用 <sup>[3]</sup> .....	176
6.3.4	CPN 网在人脸识别中的应用 <sup>[4]</sup> .....	178
6.3.5	CPN 网络在时间序列预测问题中的应用 <sup>[5]</sup> .....	180
6.3.6	CPN 网络在棉花异性纤维识别中的应用 <sup>[6]</sup> .....	182
	本章参考文献 .....	184
<b>第 7 章</b>	<b>ART 网络设计及应用</b> .....	<b>185</b>
7.1	ART 网络原理与算法 <sup>[1-4]</sup> .....	185
7.1.1	ART 网络的主要特点 .....	185
7.1.2	ART I 型网络原理与算法 .....	185

7.1.3	ART II 型网络原理与算法 .....	188
7.2	ART 网络设计与改进 <sup>[1-4]</sup> .....	190
7.2.1	设计要点——警戒门限的选择 .....	190
7.2.2	几种改进算法 .....	191
7.3	ART 网络应用实例 .....	193
7.3.1	ART I 型网络在故障诊断中的应用 .....	193
7.3.2	ART I 型网络在垃圾邮件过滤中的应用 <sup>[7]</sup> .....	195
7.3.3	ART I 型网络在中文文档分类中的应用 <sup>[8]</sup> .....	197
7.3.4	ART II A-E 型网络在指纹识别中的应用 <sup>[9]</sup> .....	197
7.3.5	ART II 型网络在压力容器焊缝超声波探伤中的应用 <sup>[10]</sup> .....	200
7.3.6	变形 ART 网络的应用 .....	201
	本章参考文献 .....	205
<b>第 8 章</b>	<b>Hopfield 网络的设计与应用</b> .....	<b>207</b>
8.1	离散型 Hopfield 网络 <sup>[1]</sup> .....	207
8.1.1	离散型 Hopfield 网络工作原理 .....	207
8.1.2	网络的权值设计 .....	209
8.2	连续型 Hopfield 网络 <sup>[1]</sup> .....	210
8.3	Hopfield 网络应用实例 .....	212
8.3.1	DHNN 网络在杭州西湖富营养化分类中的应用 <sup>[2]</sup> .....	212
8.3.2	DHNN 在数字水印算法中的应用 <sup>[3,4]</sup> .....	215
8.3.3	DHNN 在信用风险评价中的应用 <sup>[5]</sup> .....	218
8.3.4	CHNN 网解决 TSP 优化计算问题 <sup>[1]</sup> .....	221
8.3.5	CHNN 网络在原料混匀中的应用 <sup>[6]</sup> .....	225
8.3.6	CHNN 在经济负荷分配中的应用 <sup>[7]</sup> .....	228
	本章参考文献 .....	231
<b>第 9 章</b>	<b>时序递归网络的设计与应用</b> .....	<b>232</b>
9.1	递归网络常用模型 .....	232
9.1.1	输入输出递归网络通用模型 <sup>[1]</sup> .....	232
9.1.2	Elman 网络模型 .....	233
9.1.3	递归多层感知器模型 .....	234
9.2	递归网络的学习算法 <sup>[1]</sup> .....	234
9.2.1	历时反向传播学习算法 .....	235
9.2.2	实时递归学习算法 .....	237

9.3 递归网络的应用实例 .....	239
9.3.1 NARX 网络在系统辨识中的应用 <sup>[2]</sup> .....	239
9.3.2 Elman 网络在股票价格预测中的应用 <sup>[3]</sup> .....	241
9.3.3 Elman 网络在故障诊断中的应用 <sup>[4]</sup> .....	243
9.3.4 Elman 网络集成在诺西肽发酵过程建模中的应用 <sup>[5]</sup> .....	245
9.3.5 递归神经网络在化工动态系统建模中的应用 <sup>[6,7]</sup> .....	247
9.3.6 递归神经网络在非线性预测语音编码中的应用 <sup>[8]</sup> .....	250
本章参考文献 .....	252
<b>第 10 章 CMAC 网络的设计与应用 .....</b>	<b>253</b>
10.1 CMAC 网络工作原理与设计要点 <sup>[1]</sup> .....	253
10.2 CMAC 网络的学习算法 <sup>[1]</sup> .....	258
10.3 CMAC 网络应用实例 .....	258
10.3.1 CMAC 网络在板厚控制系统中的应用 <sup>[2]</sup> .....	259
10.3.2 CMAC 网络在机器人的手臂协调控制中的应用 <sup>[3]</sup> .....	261
10.3.3 CMAC 网络在短期电价预测中的应用 <sup>[4]</sup> .....	263
10.3.4 CMAC 网络在系统辨识中的应用 <sup>[5]</sup> .....	266
10.3.5 CMAC 网络在温度传感器校正中的应用 <sup>[6]</sup> .....	268
10.3.6 CMAC 网络在变压器故障诊断中的应用 <sup>[7]</sup> .....	271
本章参考文献 .....	274
<b>第 11 章 支持向量机的设计与应用 .....</b>	<b>275</b>
11.1 支持向量机的基本思想和构建 <sup>[1,2]</sup> .....	275
11.1.1 最优超平面的概念和构建 .....	275
11.1.2 支持向量机的基本原理 .....	278
11.1.3 支持向量机用于多分类问题 .....	280
11.1.4 支持向量机用于回归问题 .....	281
11.2 支持向量机的学习算法 <sup>[1,2]</sup> .....	282
11.3 支持向量机的特点及影响因素 <sup>[2,3]</sup> .....	284
11.3.1 支持向量机的特点 .....	284
11.3.2 支持向量机的影响因素 .....	285
11.4 支持向量机设计的应用实例 .....	286
11.4.1 用 SVM 解决 XOR 问题 <sup>[1]</sup> .....	286
11.4.2 SVM 在手写体数字识别中的应用 <sup>[3]</sup> .....	289
11.4.3 支持向量机在混合气体种类光谱识别中的应用 <sup>[4]</sup> .....	292

11.4.4	支持向量机在高速公路车流量预测中的应用 <sup>[5]</sup> .....	296
11.4.5	基于 SVM 的沙尘暴预测模型 <sup>[6]</sup> .....	298
11.4.6	SVM 在人耳识别中的应用 <sup>[7]</sup> .....	300
11.4.7	SVM 在锅炉燃烧工况识别中的应用 <sup>[8]</sup> .....	302
11.4.8	SVM 在烤烟烟叶生长部分判别中的应用 <sup>[9]</sup> .....	304
	本章参考文献 .....	305
<b>第 12 章</b>	<b>人工神经元网络设计开发平台</b> .....	<b>307</b>
12.1	MATLAB 神经网络工具箱函数及其应用 <sup>[1,2]</sup> .....	307
12.1.1	感知器神经网络工具箱函数 .....	307
12.1.2	线性神经网络工具箱函数 .....	312
12.1.3	BP 神经网络工具箱函数 .....	317
12.1.4	径向基神经网络工具箱函数 .....	325
12.1.5	自组织神经网络工具箱函数 .....	329
12.1.6	学习向量量化(LVQ)神经网络工具箱函数 .....	337
12.1.7	Elman 神经网络工具箱函数 .....	340
12.1.8	Hopfield 神经网络工具箱函数 .....	345
12.2	基于 MATLAB Simulink 的神经网络控制 <sup>[3]</sup> .....	348
12.2.1	控制系统模型分类 .....	348
12.2.2	系统设计过程 .....	349
12.2.3	神经网络预测控制系统实例 .....	349
12.3	其他几种神经网络开发环境介绍 <sup>[4-13]</sup> .....	358
12.3.1	神经网络软件 NeuroSolutions .....	358
12.3.2	神经网络软件 NeuralWorks Professional II /PLUS V 5.5 .....	359
12.3.3	神经网络软件 LmNet PF .....	360
12.3.4	Plexi 神经网络开发环境 .....	360
12.3.5	Neuroshell 神经网络开发环境 .....	360
12.3.6	NETSET II 神经网络开发环境 .....	361
12.3.7	N-NET210 神经网络开发环境 .....	361
12.3.8	CaseNet 神经网络开发环境 .....	361
12.3.9	FS-Neuro 神经网络开发平台 .....	361
12.3.10	金自天正人工神经元网络开发平台软件 V1.0 .....	362
	本章参考文献 .....	362

# 第 1 章 神经网络模型评估及选择

---

神经网络的种类很多,每一种类型的网络适合解决不同的问题,无论对于哪一种网络,都需要在最终结束训练和设计时评估网络的性能。以什么指标作为评估网络性能的主要指标,如何对该指标进行量化估计,如何提高网络的性能,这是每个设计者都要遇到和需要解决的问题。即设计者在解决实际问题时,需要考虑以什么标准设置网络的各项参数,控制其训练过程,确定最终的模型等一系列问题。大多数情况下,我们不仅要求神经网络对现有训练集具有较好的拟合能力,也要求它能够对训练集以外的同分布数据做出较好的预测,这就与神经网络的泛化能力有关。本章从神经网络的泛化能力出发,讨论神经网络模型性能评估问题,分析提高泛化能力的原理和途径,给出估计泛化误差的方法,关于模型选择的问题则贯穿在各节当中。

## 1.1 神经网络的泛化能力<sup>[1-3]</sup>

神经网络模型设计常常需要满足多种不同的要求。例如,具有较好的泛化(推广)能力、易于硬件实现、训练速度快等。其中泛化能力最为重要,它是衡量神经网络性能优劣的一个重要方面,这是因为建立神经网络模型的一个重要目标是通过学习对已知环境信息的学习,掌握其中的规律,从而对新的环境信息做出正确的预测。

泛化能力的定义如下:它是指经训练(学习)后的预测模型对未在训练集中出现(但具有统一规律性)的样本做出正确反映的能力,学习不是简单地记忆已经学过的输入,而是通过对有限个训练样本的学习,学到隐含在样本中的有关环境本身的内在规律性。例如,对有导师学习的网络,通过对已有样本的学习,将所提取的样本对中的非线性映射关系存储在权值矩阵中,在其后的工作阶段,当向网络输入训练时未曾见过的非样本数据(与训练集同分布)时,网络也能完成由输入空间向输出空间的正确映射。

神经网络的泛化能力涉及其在独立的检验数据(Test Data)上的预测能力。在实践中,性能的评估非常重要,它指导了学习算法和模型的选择,并为最终选定模型提供质的

度量。神经网络的泛化能力与其预测功能密切相关,下面给出神经网络预测模型的一般描述、影响泛化能力的因素和提高泛化能力的途径方法。

## 1.2 神经网络预测模型的一般描述<sup>[2,3]</sup>

神经网络具有泛化能力通常表现在能够进行预测,即掌握已有数据的内在规律后,对新的情况做出预测,如天气预报、地震预报、价格预测、不同表情下人脸的识别等。

建立预测模型的目的是通过已有的样本,给出输入与输出之间的计算关系(公式或算法表示),以便据此用输入来确定(估计、预测)输出。如图 1.1 所示,被研究的系统由一些可被观测的变量描述,其中一组变量  $x$  为输入变量;另一组变量  $y$  为输出变量。 $G$  是示例发生器,它以某一未知(但固定的)的概率分布函数  $P(X)$ ,独立同分布产生变量  $X$ , $f$

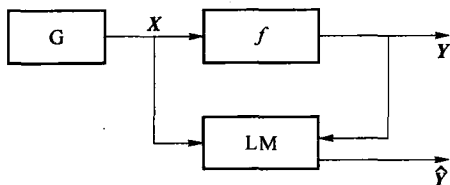


图 1.1 预测建模

是  $X$  和  $Y$  之间存在的映射关系。在理想条件下, $Y=f(X)$ ,但实际问题中,有些影响  $Y$  的因素可能观测不到,即  $Y$  中可能含有噪声。 $LM$  代表能够学习样本规律的某种模型,例如神经网络。

预测模型的优化就是选定某种类型的模型(函数集)类型,用相应的学习算法,依

据  $X$  和  $Y$  来确定模型的参数,从而给出一个最优的估计  $\hat{Y}$ 。尽管在不同的模型框架下,学习方法从形式上有所不同,但从学习过程的本质上来讲,学习过程是一个从给定的函数集中选择一个适当函数的过程,它以最佳可能方式满足给定的品质准则。

具体而言,观测到的数据集  $D$  中的训练对为  $z_i(x_i, y_i)(i=1, 2, \dots, N)$ ,满足概率分布  $P(z)$ ;所要优化的预测模型为  $\{g(z, \alpha), \alpha \in \Lambda\}$  ( $\alpha$  是集合  $\Lambda$  中的自由参数,对神经网络模型而言可以是激活函数参数、节点数、权值等)时,定义模型品质准则函数(统计学中称为风险泛函):

$$R(\alpha) = \int L(z, g(z, \alpha)) dP(z) \alpha \in \Lambda \quad (1.1)$$

其中:

$$Q(z, \alpha) = L(z, g(z, \alpha)) \quad (1.2)$$

为误差损失函数。 $R(\alpha)$ 在此处是误差损失函数的数学期望。对定量预测问题,误差损失函数典型的选择是均方误差或绝对误差:

$$L(z, g(z, \alpha)) = \begin{cases} (Y - g(z, \alpha))^2 & \text{均方误差} \\ |Y - g(z, \alpha)| & \text{绝对误差} \end{cases} \quad (1.3)$$

学习的目的就是在函数集  $\{g(z, \alpha), \alpha \in \Lambda\}$  中寻找一个使  $R(\alpha)$  值最小的函数(实际上

是确定  $\alpha$  的值)。对神经网络而言,如果风险泛函的最小值是在  $\alpha_0$  取得的,那么该神经网络  $\{g(z, \alpha_0), \alpha_0 \in \Lambda\}$  即是优化好的预测模型,这时它就可以根据  $X$  的值给出输出  $Y$  的估计值  $\hat{Y} = g(X, \alpha_0)$ 。

然而现在存在的问题是,式(1.1)中概率分布  $P(z)$  未知,因此由有限个样本点来恢复未知函数关系不可能完全精确,即便风险值相同,存在的解也不唯一,这在统计学上被认为是一个“不适定”(Ill-posed)问题。因此,根据已有数据是否可以建立符合设计要求的神经网络模型,模型的泛化能力能达到什么程度,如何提高模型的泛化能力是设计者在应用神经网络解决实际问题前需要考虑的问题。统计学界和机器学习界对这些问题进行了大量的理论和实践研究,以下章节将介绍一些较为成熟的理论。其中,在什么条件下可以根据经验数据来优化  $\alpha_0$ ,这是经验风险最小化原则解决的问题。另外,由于经验风险一致收敛的条件是在样本量无穷大的条件下成立的,然而在实际问题中这一条件是无法满足的,过分追求经验风险最小可能导致预测模型泛化能力下降;为了控制模型的泛化能力,还需要在经验风险最小化的基础上考虑样本容量,这就是结构风险最小化原则要解决的问题。除此之外,对误差损失函数的“偏差-方差分解”也是研究神经网络泛化能力的一个重要方向。

### 1.3 研究神经网络泛化能力的理论基础<sup>[2-6]</sup>

一般说来,神经网络模型的泛化能力取决于3个主要因素,即问题本身的复杂程度、参数  $\alpha$  的范围以及样本量的大小。经验风险最小化原则侧重于样本,结构风险最小化原则侧重于模型的复杂度,即应从不同的角度研究神经网络的泛化能力。而误差损失函数的“偏差-方差分解”是解释许多方法能够提高或降低神经网络泛化能力的重要依据。

#### 1.3.1 经验风险最小化原则

经验风险最小化原则解决在样本概率分布  $P(z)$  未知时,在什么条件下可以根据经验数据来优化  $\alpha_0$ 。现在  $P(z)$  未知,但得到了根据  $P(z)$  独立地随机抽取出的观测样本  $z_i (i = 1, 2, \dots, N)$ ,所以要构造经验风险泛函:

$$R_{\text{emp}}(\alpha) = \frac{1}{N} \sum_{i=1}^N Q(z_i, \alpha) \quad \alpha \in \Lambda \quad (1.4)$$

式(1.4)将各个样本出现的概率设为相同,这时就无须知道  $P(z)$ 。

假定风险泛函  $R(\alpha) = \int L(z, g(z, \alpha)) dP(z)$  的最小值是在  $Q(z, \alpha_0)$  上取得的,即参数  $\alpha$  为  $\alpha_0$  时  $R(\alpha)$  最小;经验风险泛函是在  $Q(z, \alpha_{\text{emp}})$  上取得的,即参数  $\alpha$  为  $\alpha_{\text{emp}}$  时  $R_{\text{emp}}(\alpha)$  最小。此时,需要解决的问题是:在什么条件下,可以将  $Q(z, \alpha_{\text{emp}})$  作为  $Q(z, \alpha_0)$  的一个近似,即  $\alpha_{\text{emp}}$  作为  $\alpha_0$  的一个近似。通过证明可以得到经验风险最小化的基本原理,即在经验

风险  $R_{\text{emp}}(\alpha)$  一致收敛于真正的风险  $R(\alpha)$  的条件下,  $R(\alpha_{\text{emp}})$  依概率收敛于真正的风险  $R(\alpha)$  的可能最小值。一致收敛是指:

当  $N \rightarrow \infty$  时,

$$P\{\sup_{\alpha \in \Lambda} |R(\alpha) - R_{\text{emp}}(\alpha)| > \epsilon\} \rightarrow 0 \quad (1.5)$$

此条件是经验风险具有一致收敛的充要条件。所谓经验风险最小化的一致性可以这样理解:当样本趋于无穷时,使经验风险最小化的  $\alpha_{\text{emp}}$  对应的期望风险,也差不多是最小。

这个原则在理论上奠定了经验风险最小化的可行性,如果不知道样本的概率分布,而可以获得的样本任意多(同时也有足够的计算资源来处理这些样本),则可计算任意点  $x$  对应的  $y$  (由  $x$  在任意的体积周围处对应  $y$  的均值,体积允许趋于 0)。对于一个实际问题来说,  $N$  不可能无穷大,这就要求神经网络预测模型能够通过有限样本的学习,可以对未知样本进行较好的预测,即具有较好的泛化能力。

这一原则指出,一般而言,神经网络若具有较好的泛化能力应该有较多的样本可供学习,当样本量不足的时候,可以根据样本特性进行虚拟构造以增加样本数。但有些时候样本量非常小且很难扩展,如何能够提高模型的泛化能力,可以用结构风险最小化(SRM)原则作为指导。

### 1.3.2 结构风险最小化原则

结构风险最小化(SRM)原则在最小化经验风险的基础上,通过函数集的容量(Vapnik-Chervonenkis 维, VC 维),找到一个函数,它对于固定数量的数据可以达到保证风险的最小值。对神经网络模型简单地说,就是要控制模型的复杂度。Vapnik-Chervonenkis 理论提供了复杂度的一般度量,并给出乐观性的相关界限。

如果  $N/M$  (训练样本数目与神经网络模型的 VC 维之比)较小(比如小于 20),则称大小为  $N$  的样本集为小样本集。可以证明,对于来自具有有限 VC 维  $M$  的完全有界函数集  $0 \leq Q(z, \alpha) \leq B$ ,  $\alpha \in \Lambda$  的所有函数 ( $B$  是  $Q(z, \alpha) = L(z, g(z, \alpha))$  的上界), (加性)不等式

$$R(\alpha) \leq R_{\text{emp}}(\alpha) + \frac{B\epsilon(N)}{2} \left( 1 + \sqrt{1 + \frac{4R_{\text{emp}}(\alpha)}{B\epsilon(N)}} \right) \quad (1.6)$$

至少依  $1 - \eta$  的概率成立,其中

$$\epsilon(N) = 4 \frac{M(\ln \frac{2N}{M} + 1) - \ln \eta / 4}{N} \quad (1.7)$$

利用上述不等式可以控制基于固定数量经验数据最小化风险泛化的过程。控制这一过程的最简单的方法是极小化经验风险值。不等式(1.6)的右端为风险的上界(即保证风险  $R_b(\alpha)$ ),它由两项组成,其中一项是经验风险,因此它随着经验风险值的减小而减小。这就是经验风险最小化原则对于大样本集经常能给出好结果的原因。

如果  $N/M$  较大,实际风险值由经验风险值来决定。因此,为了最小化实际风险,只需最小



化经验风险。然而,如果  $N/M$  较小,小的经验风险值并不能保证有小的实际风险值。在这种情况下,必须同时考虑不等式的两项,以最小化不等式的右端。第一项取决于函数集的某一个特定函数,而对于固定数量的观测数据来说,第二项主要取决于整个函数集的 VC 维。

SRM 方案:将结构  $S$  施加到函数  $Q(z, \alpha)$  ( $\alpha \in \Lambda$ ) 的集合  $S$  上,使之具有一个结构  $S$ , 即  $S_k = \{Q(z, \alpha), \alpha \in \Lambda_k\}, k=1, 2, \dots, n, S_1 \subset S_2 \subset \dots \subset S_n$ , 其 VC 维满足  $M_1 \leq M_2 \leq \dots \leq M_n$ 。用  $Q(z, \alpha_N^k)$  表示函数集  $S_k$  中最小化经验风险的函数。 $R_b(\alpha_N^k)$  为其保证风险上界。对于给定的观测数据,SRM 方法选择达到最小  $R_b(\alpha_N^k)$  的结构元素  $S_k$ , 即在经验风险值增加尽可能小的条件下降低 VC 维数。

这一方法成功实现的例子是支持向量机,它也可以作为模型选择的一个准则,但其主要缺点是计算函数的 VC 维比较困难。

这一原则从另一方面验证了,为什么在误差函数项中加入对模型复杂度控制的惩罚项(比如权值衰减),以及其他控制模型复杂度的方法有效。

### 1.3.3 偏差-方差分解

对模型复杂度和泛化能力的研究还可以从误差函数的“偏差-方差分解”的角度进行研究。下面以回归和曲线拟合为例,给出预测误差函数表达式的偏差-方差分解。假定  $Y=f(X)+\epsilon$ , 即期望输出中有噪声  $\epsilon$ , 满足均值  $E(\epsilon)=0$ , 方差  $\text{Var}(\epsilon)=\sigma_\epsilon^2$ , 使用均方误差损失函数,可以导出在任意输入点  $X$  上,神经网络进行拟合的输出  $\hat{Y}$  的误差:

$$\begin{aligned} \text{Err}(X; D) &= E_D [(g(X; D) - f(X))^2] \\ &= \sigma_\epsilon^2 + (E_D [g(X; D)] - f(X))^2 + E_D [g(X; D) - E_D [g(X; D)]]^2 \\ &= \sigma_\epsilon^2 + \text{Bias}^2(g(X; D)) + \text{Var}(g(X; D)) \\ &= \text{不可约的误差} + \text{偏差}^2 + \text{方差} \end{aligned} \quad (1.8)$$

第一项是神经网络目标输出在真正值  $f(X)$  附近的方差,这是噪声所引起的,除非  $\sigma_\epsilon^2=0$ , 否则无论神经网络对  $f(X)$  估计得多好,也无法避免;第二项是“偏差”(平方),代表的是神经网络估计值与真实值之间的差异,一个小的偏差意味着可以从数据集  $D$  中较准确地估计出  $f(X)$ , 这与经验风险最小原则是一致的;第三项是方差项,代表的是神经网络估计值在其本身期望均值附近的平方差,它反映了一个模型对数据的敏感度,即一个小的“方差”意味着神经网络的估计值不随训练集的波动而发生较大的波动。

式(1.8)表明均方误差可以用偏差项和方差项的和的形式表示。“偏差”项度量的是模型与真实系统之间匹配的“准确性”和“质量”:一个高的偏差意味着坏的匹配;而“方差”项度量的是模型与真实系统之间匹配的“精确性”和“特定性”:一个高的方差意味着弱匹配。设计者需要注意的是,“偏差和方差两难”或者“偏差和方差折中”是一个普遍的现象:一个模型学习算法如果逐渐提高对训练数据的适应性(如设计更多的权值),那么它将趋向于更小的偏差,但可能导致更大的方差。反之,如果一个模型的参数较少(通常对应较大的偏差),那么数据拟合的性能就不会太好,但拟合的程度不会随数据集的变化而变化太大(较低的方差)。换言之,随着模型复杂度的增加,方差趋于增加,偏差趋于减小;随着