

机器读心术之神经网络与深度学习 第2周

【声明】 本视频和幻灯片为炼数成金网络课程的教学资料，所有资料只能在课程内使用，不得在课程以外范围散播，违者将可能被追究法律和经济责任。

课程详情访问炼数成金培训网站

<http://edu.dataguru.cn>

关注炼数成金企业微信



- 提供全面的数据价值资讯，涵盖商业智能与数据分析、大数据、企业信息化、数字化技术等，各种高性价比课程信息，赶紧掏出您的手机关注吧！



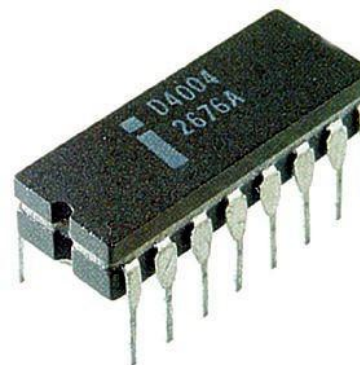
Bernard Widrow

- <http://www-isl.stanford.edu/~widrow/>



Marcian Hoff

- <http://www.computerhistory.org/fellowawards/hall/bios/Marcian,Hoff/>
- http://ethw.org/Ted_Hoff
- Widrow的合作者，共同提出“Widrow-Hoff学习规则”，但其最杰出的贡献是在英特尔公司发明了人类第一颗微处理器Intel 4004



- 与单层感知器结构相似
- 增加支持purelin传输函数
- 除了二值输出外还可以支持模拟输出，因此除了充当分类器还可以实现类似回归的效果
- 使用多个输出（Madaline网络）可以变相解决线性不可分问题
- 引入非线性成分（升维）可以一定程度上解决非线性问题

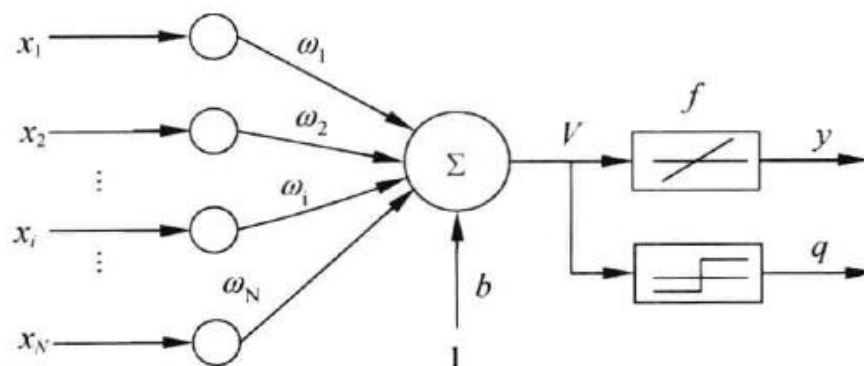


图 5-1 线性神经网络的结构

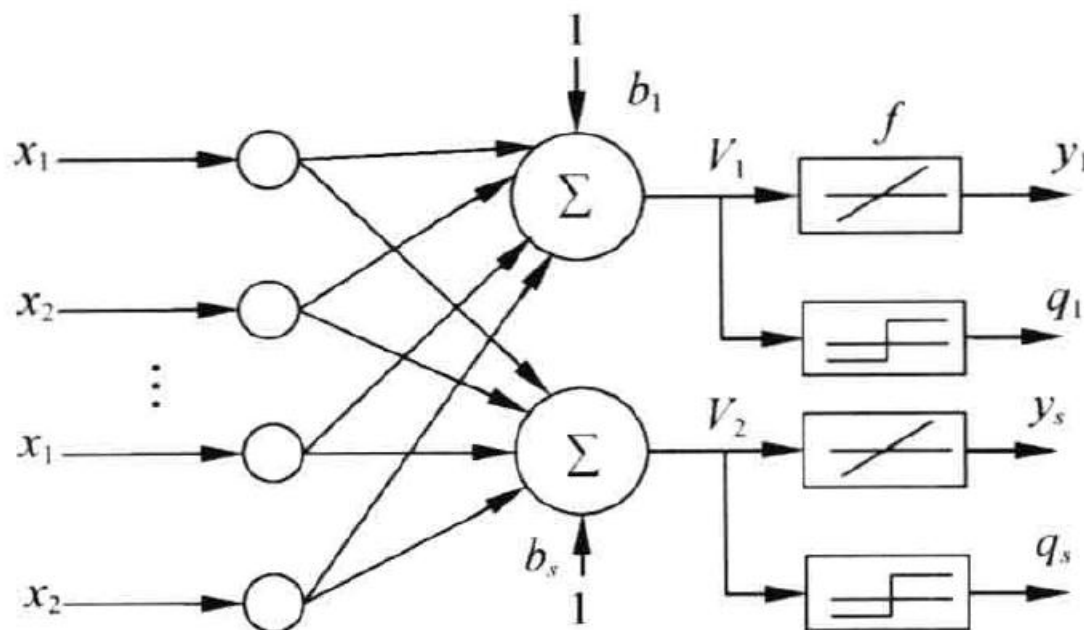


图 5-2 Madaline 结构图

线性网络能解决非线性问题吗

- 可以把变量的非线性项作为新输入变量来处理
- 《Matlab神经网络原理与实例精解》第130页

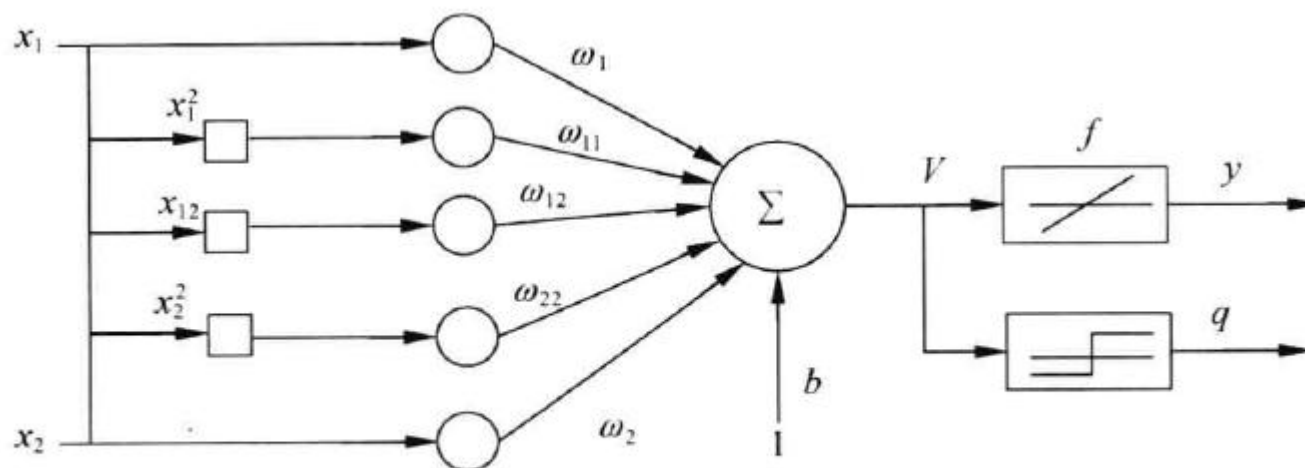


图 5-4 线性网络解决非线性问题

- 一种利用梯度下降法的一般性的学习规则
- 1986年，心理学家McClelland和Rumelhart引入
- 目标函数：最小平方误差条件（MSE）
- 学习规则：权值变化量（delta）正比于负梯度（比例系数即为学习率）
- 一个算例：《人工神经网络理论、设计及应用》第34页

- <http://zh.wikipedia.org/wiki/%E6%9C%80%E9%80%9F%E4%B8%8B%E9%99%8D%E6%B3%95>

梯度下降法，基于这样的观察：如果实值函数 $F(\mathbf{x})$ 在点 \mathbf{a} 处可微且有定义，那么函数 $F(\mathbf{x})$ 在 \mathbf{a} 点沿着梯度相反的方向 $-\nabla F(\mathbf{a})$ 下降最快。因而，如果

$$\mathbf{b} = \mathbf{a} - \gamma \nabla F(\mathbf{a})$$

对于 $\gamma > 0$ 为一个够小数值时成立，那么 $F(\mathbf{a}) \geq F(\mathbf{b})$ 。

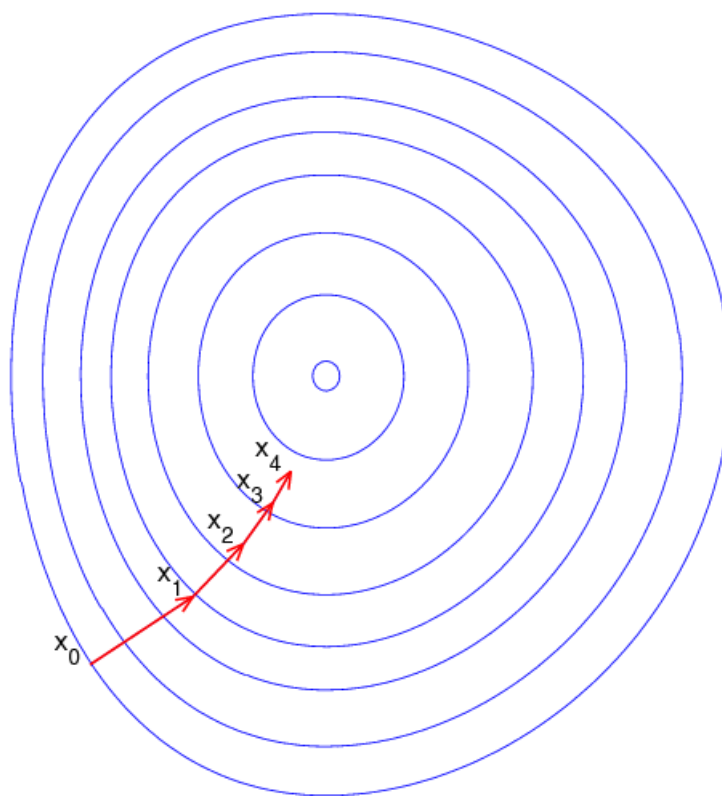
考虑到这一点，我们可以从函数 F 的局部极小值的初始估计 \mathbf{x}_0 出发，并考虑如下序列 $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots$ 使得

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \gamma_n \nabla F(\mathbf{x}_n), \quad n \geq 0.$$

因此可得到

$$F(\mathbf{x}_0) \geq F(\mathbf{x}_1) \geq F(\mathbf{x}_2) \geq \dots,$$

如果顺利的话序列 (\mathbf{x}_n) 收敛到期望的极值。注意每次迭代步长 γ 可以改变。

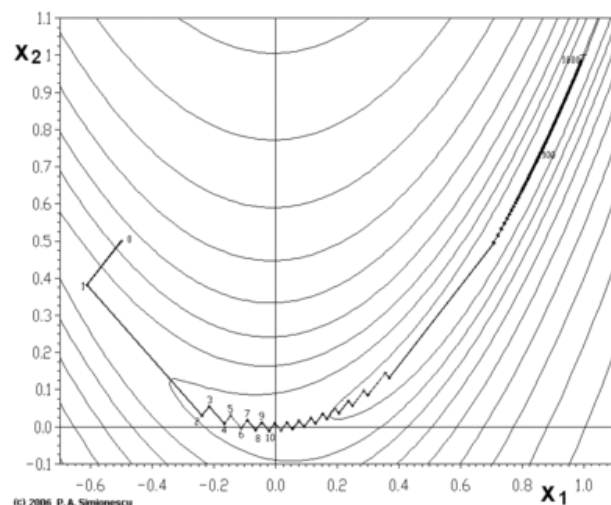


- 目标函数必须可微
- 学习率难以选取，太大会产生“之字形”震荡，太小迭代次数太多，前进很慢
- 容易陷入局部最优

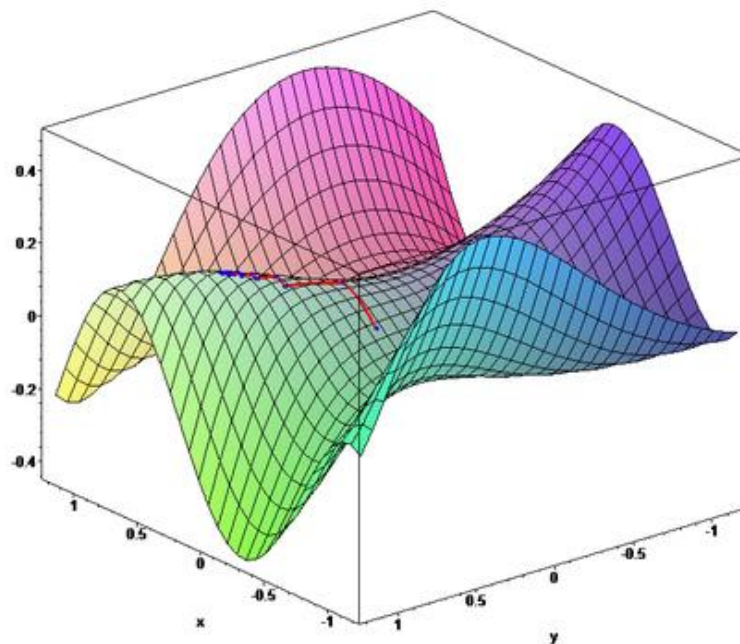
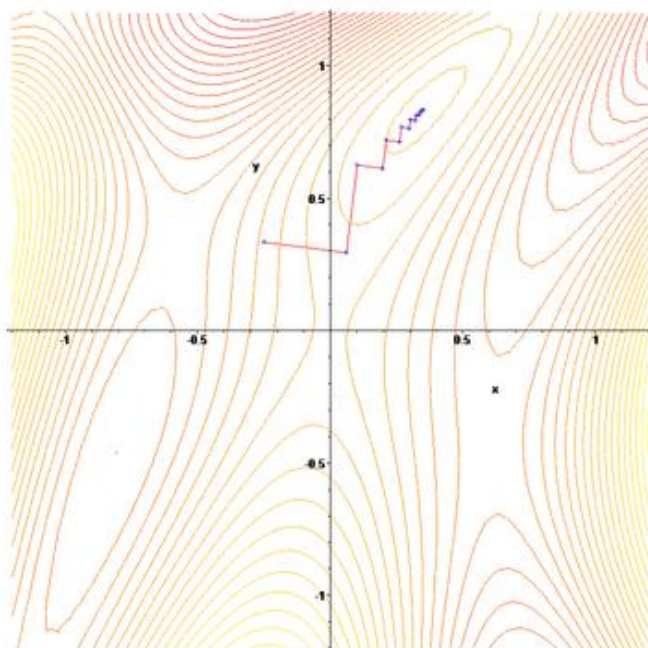
梯度下降法处理一些复杂的非线性函数会出现问题，例如Rosenbrock函数

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2.$$

其最小值在 $(x, y) = (1, 1)$ 处，数值为 $f(x, y) = 0$ 。但是此函数具有狭窄弯曲的山谷，最小值 $(x, y) = (1, 1)$ 就在这些山谷之中，并且谷底很平。优化过程是之字形的向极小值点靠近，速度非常缓慢。



下面这个例子也鲜明的示例了"之字"的下降，这个例子用梯度下降法求 $F(x, y) = \sin\left(\frac{1}{2}x^2 - \frac{1}{4}y^2 + 3\right) \cos(2x + 1 - e^y)$ 的极小值。



- Least Mean Square , Widrow和Hoff在1960提出
- 可以理解为Delta学习规则的特殊情况（激活函数使用purelin）
- 和感知器学习规则神似
- 仅用于训练单层神经网络（多层经过转化为单层，事实上也能应用）

以下是 LMS 算法的步骤。

(1) 定义变量和参数。

为方便处理，将偏置 b 与权值合并：

$$\boldsymbol{\omega}(n) = [b(n), \omega_1(n), \omega_2(n), \dots, \omega_N(n)]^T$$

相应地，训练样本为

$$\mathbf{x}(n) = [1, x_1(n), x_2(n), \dots, x_N(n)]^T$$

$b(n)$ 为偏置， $d(n)$ 为期望输出， $y(n)$ 为实际输出， η 为学习率， n 为迭代次数。

(2) 初始化。给向量 $\boldsymbol{\omega}(n)$ 赋一个较小的随机初值， $n = 0$ 。

(3) 输入样本，计算实际输出和误差。根据给定的期望输出 $d(n)$ ，计算

$$e(n) = d(n) - \mathbf{x}^T(n)\boldsymbol{\omega}(n)$$

(4) 调整权值向量。根据上一步算得的误差，计算

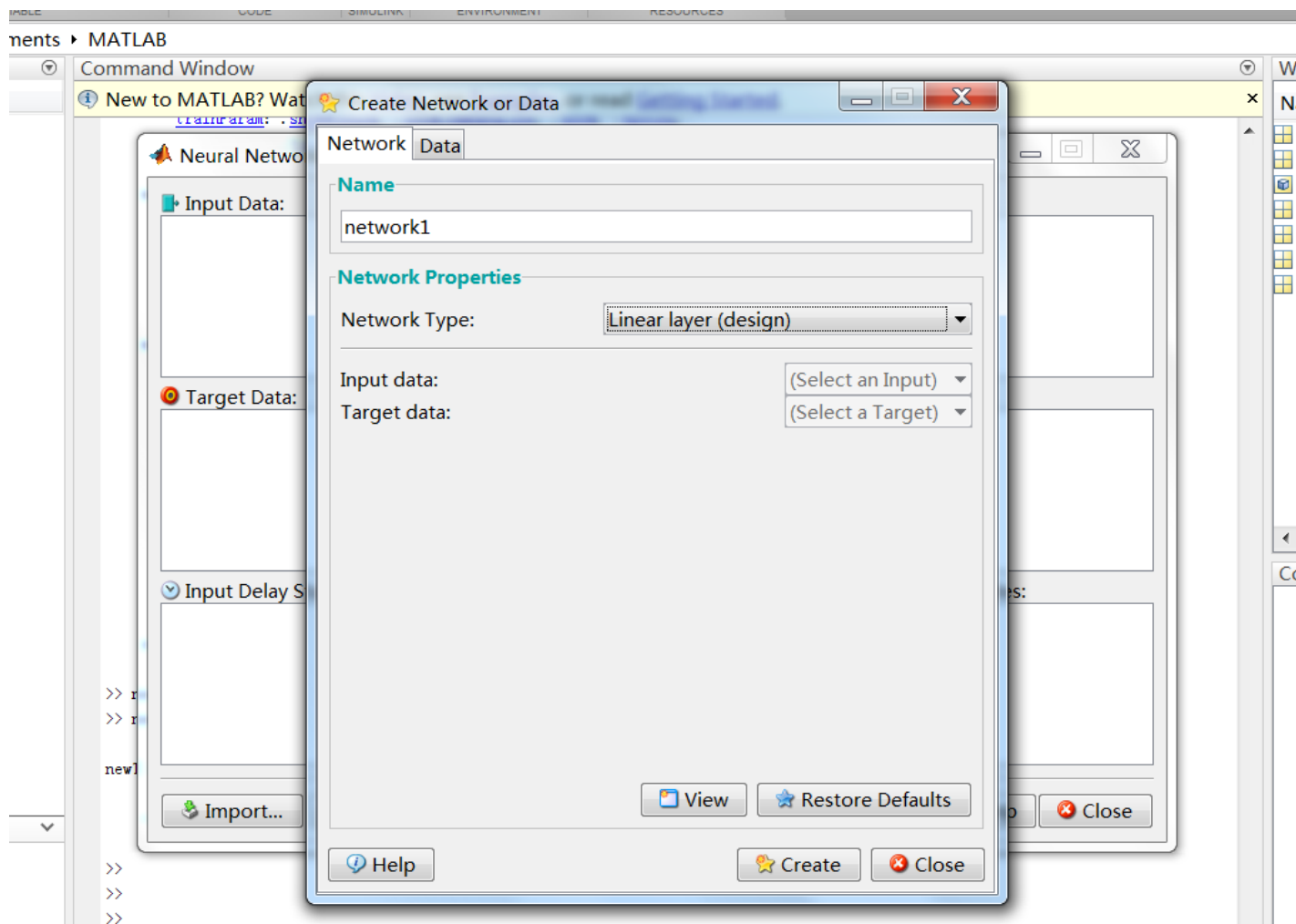
$$\boldsymbol{\omega}(n+1) = \boldsymbol{\omega}(n) + \eta \mathbf{x}^T(n)e(n)$$

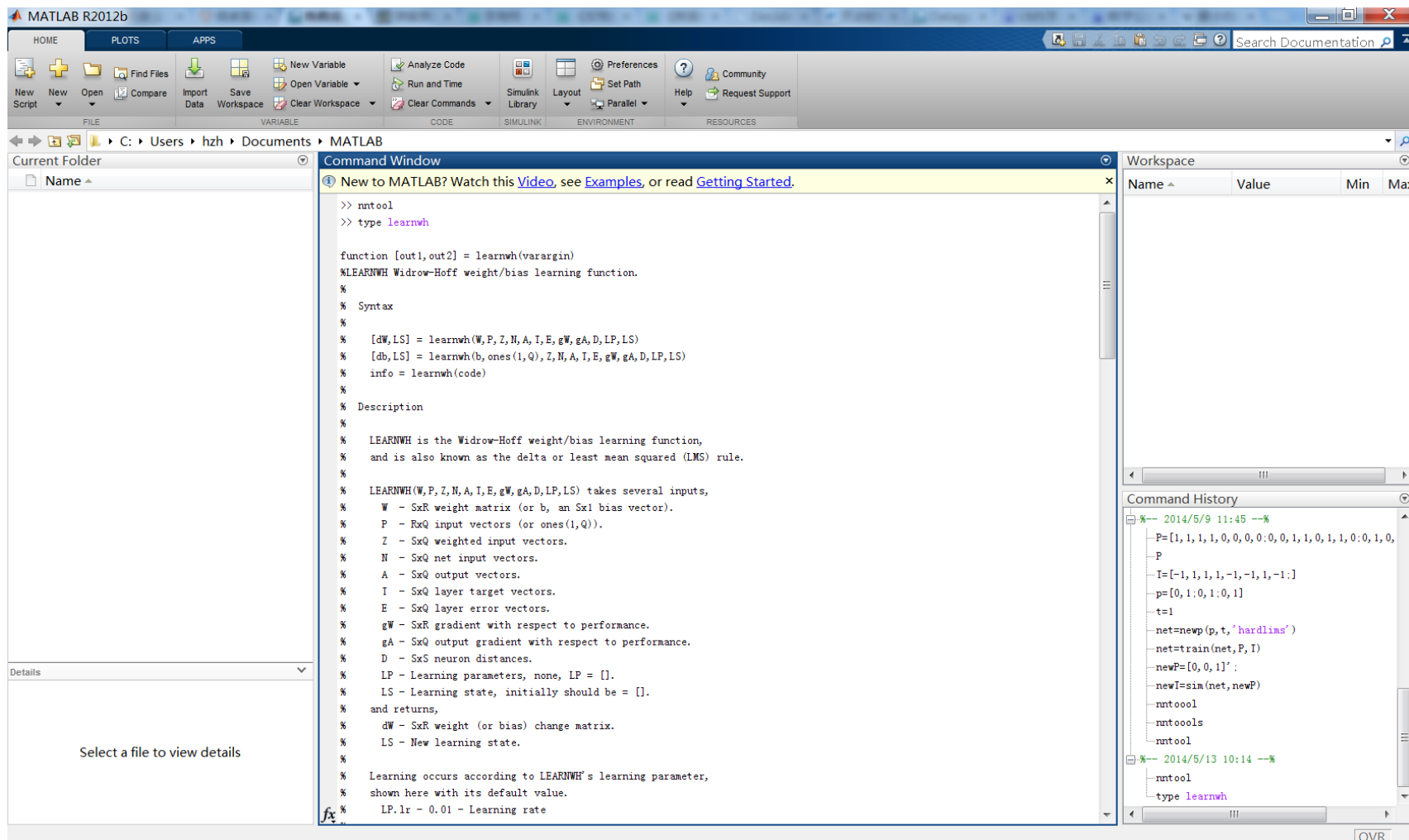
(5) 判断算法是否收敛。若满足收敛条件，则算法结束，否则 n 自增 1 ($n = n + 1$)，跳转到第 3 步重新计算。收敛条件的选择对算法有比较大的影响，常用的条件有：

- ❑ 误差等于零或者小于某个事先规定的较小的值，如 $|e(n)| < \varepsilon$ 或 $\text{mse} < \varepsilon$ ；
- ❑ 权值变化量已经很小，即 $|\boldsymbol{\omega}(n+1) - \boldsymbol{\omega}(n)| < \varepsilon$ ；
- ❑ 设置最大迭代次数，达到最大迭代次数 N 时，无论算法是否达到预期要求，都将

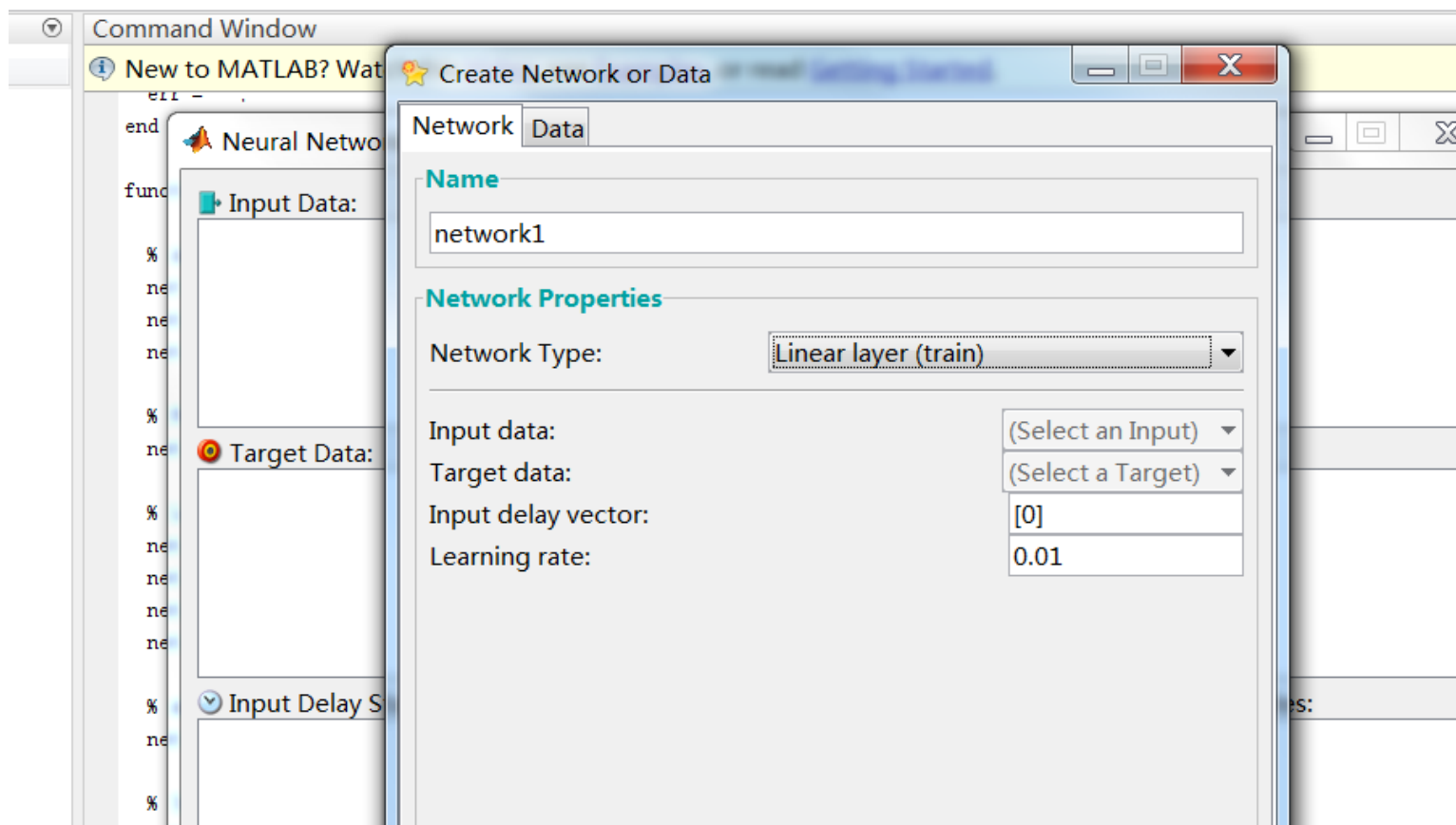
学习率怎样选择

- R 为输入样本 X 的自相关矩阵， a 是 R 的最大特征根
- 可以证明，只要选择学习率在0和 $2/a$ 之间，LMS算法收敛（Haykin，1996）
- 由于 a 一般难以计算，通常用 R 的迹 t 代替，因为 $2/t < 2/a$ ，所以学习率如果小于 $2/t$ ，那么它必定小于 $2/a$
- LMS也有类似多重共线性的情况，对 R 中的条件数敏感
- 学习率逐渐下降的算法（《Matlab神经网络原理与实例精解》第133页）





- 可以用于处理时间序列数据



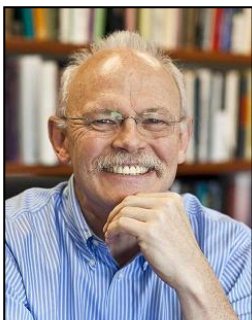
- 美国认知心理学家，主要从事认知神经科学和人工智能方面的研究。鲁姆哈特的研究从描述语义网络中长时记忆的特点开始，他在这方面所做的工作对70年代认知心理学的兴起作出了贡献。1991年当选为国家科学院院士，1996年获美国心理学会颁发的杰出科学贡献奖。
- Rumelhart (鲁梅尔哈特) 奖：<http://rumelhartprize.org/>



■ <http://www-psych.stanford.edu/~jlm/>

Last updated September 7, 2013

James L. (Jay) McClelland



Lucie Stern Professor in the Social Sciences
Director, Center for Mind, Brain and Computation
Department of Psychology
Stanford University

344 Jordan Hall, Bldg 420
450 Serra Mall
Stanford, CA 94305
+1-650-725-1232
mcclelland@stanford.edu

[Curriculum Vitae](#) and [Career Highlights](#).
[Publications](#) and [Resources for simulation and self-study](#).
[Interviews and news articles](#) and [Rumelhart prize lecture](#).
[PDP lab website](#) and [Heineken Prize Lecture](#).
This page in [Spanish](#), [Georgian](#), and [Ukrainian](#).

Welcome and Overview

Welcome and thank you for visiting my home page. I am a Professor in the Psychology Department and Director of the Center for Mind, Brain and Computation at Stanford. My research addresses a broad range of topics in cognitive science and cognitive neuroscience, including perception and perceptual decision making; learning and memory; language and reading; semantic and mathematical cognition; and cognitive development. I view cognitive functions as emerging from the parallel, distributed processing activity of neural populations, with learning occurring through the adaptation of connections among participating neurons, as discussed in [Parallel Distributed Processing](#) (Rumelhart, McClelland, and the PDP Research Group, 1986). Research in my lab revolves around efforts to develop explicit computational models based on these ideas; to test, refine, and extend the principles embodied in the models; and then to apply the models to substantive research questions through behavioral experiment, computer simulation, and mathematical analysis.

My [career highlights](#) are listed below. A full list of relevant publications is available on my [publications](#) page, and links to other resources are provided next to my photograph above.

Mathematical Cognition

Recently, I have begun what I expect will be a broad-ranging and long-term program of research in *mathematical cognition* (watch [this video](#) for a description of the approach). The work grows out of my long-standing interest in developmental transitions and in readiness to learn from new experiences as well as from the hope that a Parallel-Distributed Processing approach may shed light on some of the most awe-inspiring achievements of human thought --- the insights and structured reasoning systems that have been created by mathematicians. In this effort, we are combining experimental studies and computational modeling studies. The lab is seeking to recruit experimentally and/or computationally oriented students interested in contributing to this effort.

PARALLEL DISTRIBUTED PROCESSING

Explorations in the Microstructure of Cognition

Volume 1: Foundations

David E. Rumelhart James L. McClelland
and the PDP Research Group

Chisato Asanuma	Alan H. Kawamoto	Paul Smolensky
Francis H. C. Crick	Paul W. Munro	Gregory O. Stone
Jeffrey L. Elman	Donald A. Norman	Ronald J. Williams
Geoffrey E. Hinton	Daniel E. Rabin	David Zipser
Michael I. Jordan	Terrence J. Sejnowski	

Institute for Cognitive Science
University of California, San Diego

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

- error Back Propagation , 基于误差反向传播算法
- 1986年由Rumelhart和McClland为首的科学家小组提出
- 多层前馈网络
- 目前应用最广泛的神经网络模型之一
- 多层, 可以调整的权值很多, 可以使用多种激活函数
- 能学习和存贮大量的输入-输出模式映射关系, 而无需事前揭示描述这种映射关系的数学方程
- 学习规则使用最速下降法, 通过反向传播来不断调整网络的权值和阈值, 使网络的误差平方和最小
- BP神经网络模型拓扑结构包括输入层 (input)、隐层(hidden layer)和输出层(output layer)

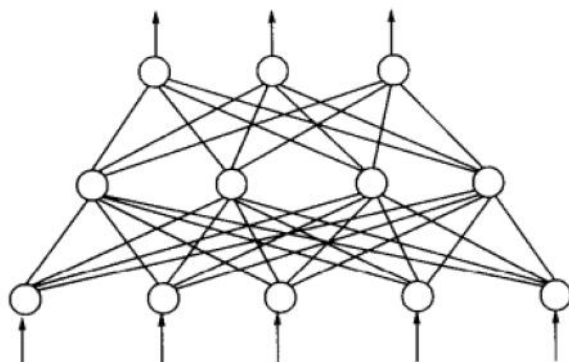


图 2.9 层次型网络结构示意图

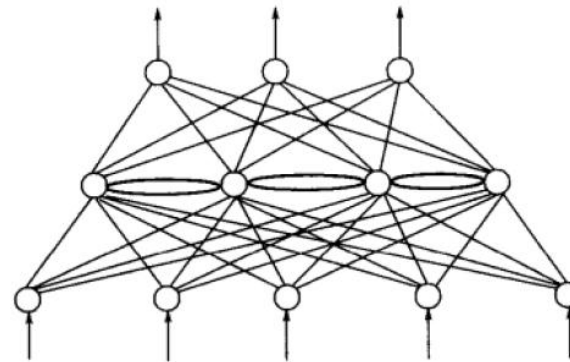


图 2.11 层内有互连的层次型网络结构示意图

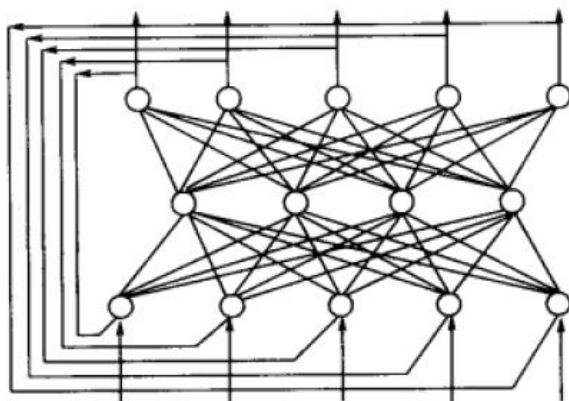


图 2.10 输出层到输入层有连接的
层次型网络结构示意图

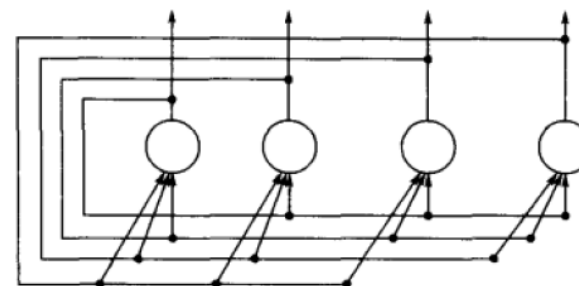
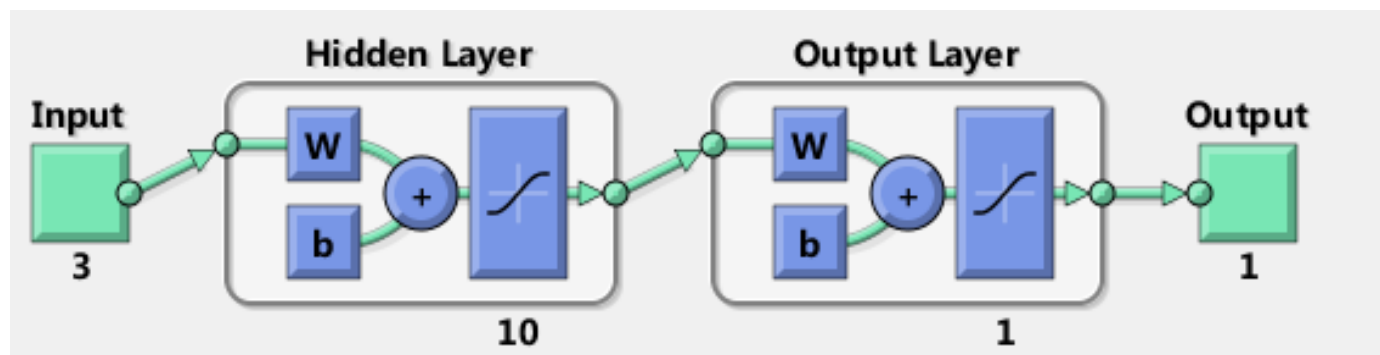


图 2.12 全互连型网络结构示意图



- 隐藏层与隐藏节点
- 前馈 —— 每一层的节点仅和下一层节点相连

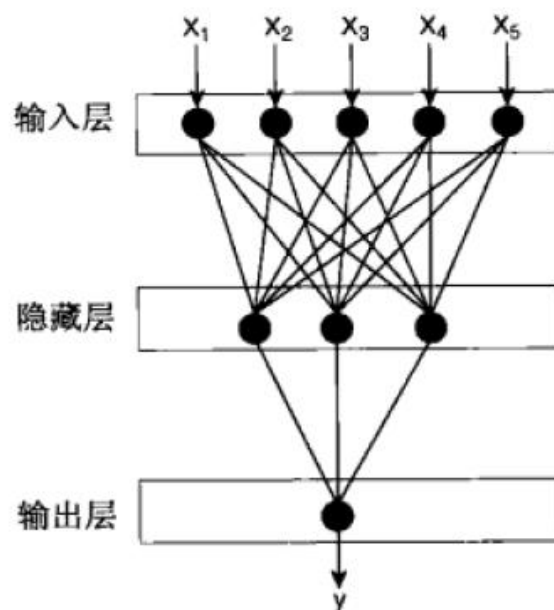
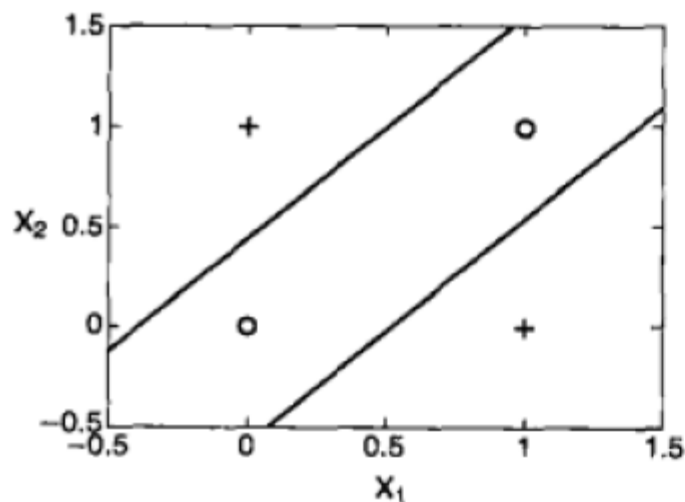
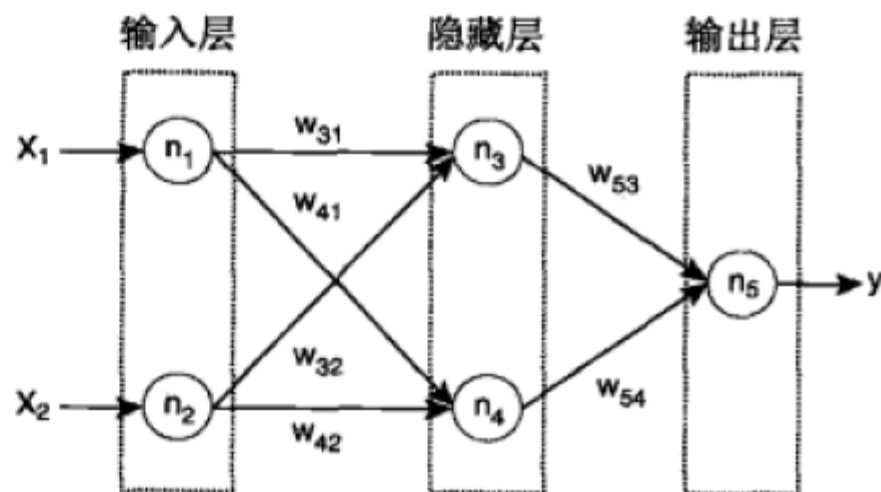


图 5-17 多层前馈人工神经网络（ANN）举例



(a) 决策边界



(b) 神经网络拓扑结构

图 5-19 XOR 问题的两层前馈神经网络

学习规则：最速下降BP法

- 很早就有多层神经网络的想法，但苦于没有合适的学习算法
- 误差反向传播算法催熟了人工神经网络
- 基本思想：学习过程由信号的正向传播和误差的反向传播两个过程组成
- 《人工神经网络理论、设计及应用》第47页

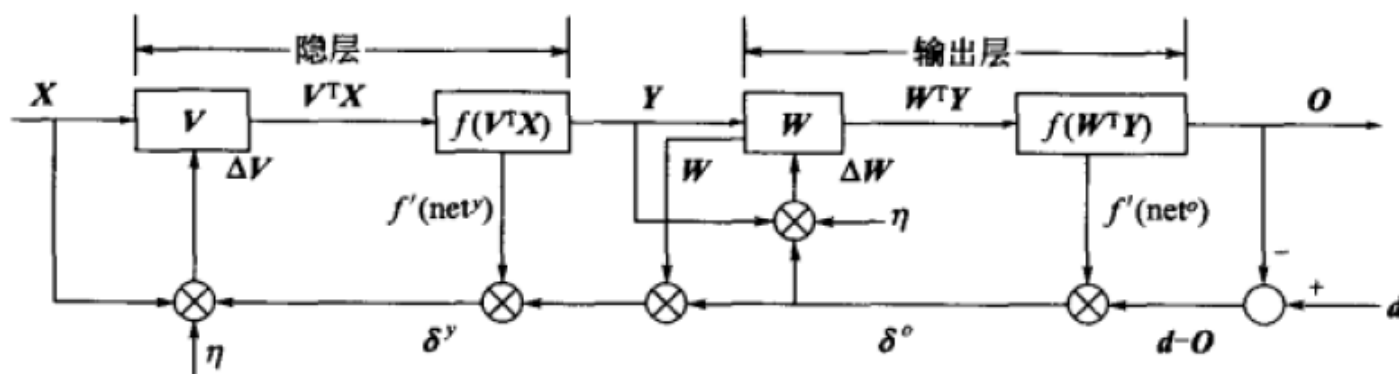


图 3.17 BP 算法的信号流向

- 选择适当的神经网络类型
- 对权重赋予初始值
- 选择一定的学习规则对模型进行迭代训练
- 最终收敛到合适的权重，确定模型
- 模型泛化
- 神经网络可以用于哪些场景？
- 神经网络的本质是一种通用的逼近器（类比于数学分析中的幂函数展开，傅里叶展开）

■ ftp://ftp.ics.uci.edu/pub/machine-learning-databases



← → ↻ ftp://ftp.ics.uci.edu/pub/machine-learning-databases		
README	35.7 kB	07-6-25 上午12:00:00
SUMMARY-TABLE	3.9 kB	98-9-1 上午12:00:00
abalone/		96-12-3 上午12:00:00
abscisic-acid/		08-6-17 上午12:00:00
access-lists/		96-12-3 上午12:00:00
acute/		09-9-14 上午12:00:00
adult/		01-1-31 上午12:00:00
annealing/		96-12-3 上午12:00:00
anonymous/		98-11-30 上午12:00:00
arcene/		08-3-25 上午12:00:00
arrhythmia/		98-3-20 上午12:00:00
artificial-characters/		97-12-9 上午12:00:00
audiology/		96-12-3 上午12:00:00
auto-mpg/		96-12-3 上午12:00:00
autos/		96-12-3 上午12:00:00
badges/		96-12-3 上午12:00:00
bag-of-words/		09-2-2 上午12:00:00
balance-scale/		96-12-3 上午12:00:00
balloons/		96-12-3 上午12:00:00
blood-transfusion/		08-11-30 上午12:00:00
breast-cancer/		00-1-6 上午12:00:00
breast-cancer-wisconsin/		96-12-3 上午12:00:00
bridges/		96-12-3 上午12:00:00
car/		97-6-19 上午12:00:00
census-income/		96-12-3 上午12:00:00
character-trajectories/		08-11-30 上午12:00:00
chess/		96-12-3 上午12:00:00
chorales/		96-12-3 上午12:00:00
cmc/		99-3-4 上午12:00:00
communities/		09-9-14 上午12:00:00
concrete/		09-9-14 上午12:00:00
connect-4/		96-12-3 上午12:00:00
contacts/		96-12-3 上午12:00:00
covtype/		99-12-14 上午12:00:00

← → ↺ <ftp://ftp.ics.uci.edu/pub/machine-learning-databases/statlog/german/german.data>

```
A11 6 A34 A43 1169 A65 A75 4 A93 A101 4 A121 67 A143 A152 2 A173 1 A192 A201 1
A12 48 A32 A43 5951 A61 A73 2 A92 A101 2 A121 22 A143 A152 1 A173 1 A191 A201 2
A14 12 A34 A46 2096 A61 A74 2 A93 A101 3 A121 49 A143 A152 1 A172 2 A191 A201 1
A11 42 A32 A42 7882 A61 A74 2 A93 A103 4 A122 45 A143 A153 1 A173 2 A191 A201 1
A11 24 A33 A40 4870 A61 A73 3 A93 A101 4 A124 53 A143 A153 2 A173 2 A191 A201 2
A14 36 A32 A46 9055 A65 A73 2 A93 A101 4 A124 35 A143 A153 1 A172 2 A192 A201 1
A14 24 A32 A42 2835 A63 A75 3 A93 A101 4 A122 53 A143 A152 1 A173 1 A191 A201 1
A12 36 A32 A41 6948 A61 A73 2 A93 A101 2 A123 35 A143 A151 1 A174 1 A192 A201 1
A14 12 A32 A43 3059 A64 A74 2 A91 A101 4 A121 61 A143 A152 1 A172 1 A191 A201 1
A12 30 A34 A40 5234 A61 A71 4 A94 A101 2 A123 28 A143 A152 2 A174 1 A191 A201 2
A12 12 A32 A40 1295 A61 A72 3 A92 A101 1 A123 25 A143 A151 1 A173 1 A191 A201 2
A11 48 A32 A49 4308 A61 A72 3 A92 A101 4 A122 24 A143 A151 1 A173 1 A191 A201 2
A12 12 A32 A43 1567 A61 A73 1 A92 A101 1 A123 22 A143 A152 1 A173 1 A192 A201 1
A11 24 A34 A40 1199 A61 A75 4 A93 A101 4 A123 60 A143 A152 2 A172 1 A191 A201 2
A11 15 A32 A40 1403 A61 A73 2 A92 A101 4 A123 28 A143 A151 1 A173 1 A191 A201 1
A11 24 A32 A43 1282 A62 A73 4 A92 A101 2 A123 32 A143 A152 1 A172 1 A191 A201 2
A14 24 A34 A43 2424 A65 A75 4 A93 A101 4 A122 53 A143 A152 2 A173 1 A191 A201 1
A11 30 A30 A49 8072 A65 A72 2 A93 A101 3 A123 25 A141 A152 3 A173 1 A191 A201 1
A12 24 A32 A41 12579 A61 A75 4 A92 A101 2 A124 44 A143 A153 1 A174 1 A192 A201 2
A14 24 A32 A43 3430 A63 A75 3 A93 A101 2 A123 31 A143 A152 1 A173 2 A192 A201 1
A14 9 A34 A40 2134 A61 A73 4 A93 A101 4 A123 48 A143 A152 3 A173 1 A192 A201 1
A11 6 A32 A43 2647 A63 A73 2 A93 A101 3 A121 44 A143 A151 1 A173 2 A191 A201 1
A11 10 A34 A40 2241 A61 A72 1 A93 A101 3 A121 48 A143 A151 2 A172 2 A191 A202 1
A12 12 A34 A41 1804 A62 A72 3 A93 A101 4 A122 44 A143 A152 1 A173 1 A191 A201 1
A14 10 A34 A42 2069 A65 A73 2 A94 A101 1 A123 26 A143 A152 2 A173 1 A191 A202 1
A11 6 A32 A42 1374 A61 A73 1 A93 A101 2 A121 36 A141 A152 1 A172 1 A192 A201 1
A14 6 A30 A43 426 A61 A75 4 A94 A101 4 A123 39 A143 A152 1 A172 1 A191 A201 1
A13 12 A31 A43 409 A64 A73 3 A92 A101 3 A121 42 A143 A151 2 A173 1 A191 A201 1
A12 7 A32 A43 2415 A61 A73 3 A93 A103 2 A121 34 A143 A152 1 A173 1 A191 A201 1
A11 60 A33 A49 6836 A61 A75 3 A93 A101 4 A124 63 A143 A152 2 A173 1 A192 A201 2
A12 18 A32 A49 1913 A64 A72 3 A94 A101 3 A121 36 A141 A152 1 A173 1 A192 A201 1
A11 24 A32 A42 4020 A61 A73 2 A93 A101 2 A123 27 A142 A152 1 A173 1 A191 A201 1
A12 18 A32 A40 5866 A62 A73 2 A93 A101 2 A123 30 A143 A152 2 A173 1 A192 A201 1
A14 12 A34 A49 1264 A65 A75 4 A93 A101 4 A124 57 A143 A151 1 A172 1 A191 A201 1
A13 12 A32 A42 1474 A61 A72 4 A92 A101 1 A122 33 A141 A152 1 A174 1 A192 A201 1
A12 45 A34 A43 4746 A61 A72 4 A93 A101 2 A122 25 A143 A152 2 A172 1 A191 A201 2
A14 48 A34 A46 6110 A61 A73 1 A93 A101 3 A124 31 A141 A153 1 A173 1 A192 A201 1
A13 18 A32 A43 2100 A61 A73 4 A93 A102 2 A121 37 A142 A152 1 A173 1 A191 A201 2
A13 10 A32 A44 1225 A61 A73 2 A93 A101 2 A123 37 A143 A152 1 A173 1 A192 A201 1
A12 9 A32 A43 458 A61 A73 4 A93 A101 3 A121 24 A143 A152 1 A173 1 A191 A201 1
A14 30 A32 A43 2333 A63 A75 4 A93 A101 2 A123 30 A141 A152 1 A174 1 A191 A201 1
A12 12 A32 A43 1158 A63 A73 3 A91 A101 1 A123 26 A143 A152 1 A173 1 A192 A201 1
A12 18 A33 A45 6204 A61 A73 2 A93 A101 4 A121 44 A143 A152 1 A172 2 A192 A201 1
A11 30 A34 A41 6187 A62 A74 1 A94 A101 4 A123 24 A143 A151 2 A173 1 A191 A201 1
```


基于BP网络的个人信贷信用评估

对个人信用的评估方法主要分为定性评估和定量评估两种，前者主要根据信贷人员的主观判断，后者则根据个人客户的资料，利用评分卡和信用评分模型等工具进行分析。

本实例采用BP神经网络，以已知用户信息及信用情况为训练样本，学习得出一个抽象模型，然后对新样本进行评估，正确率稳定在70%以上

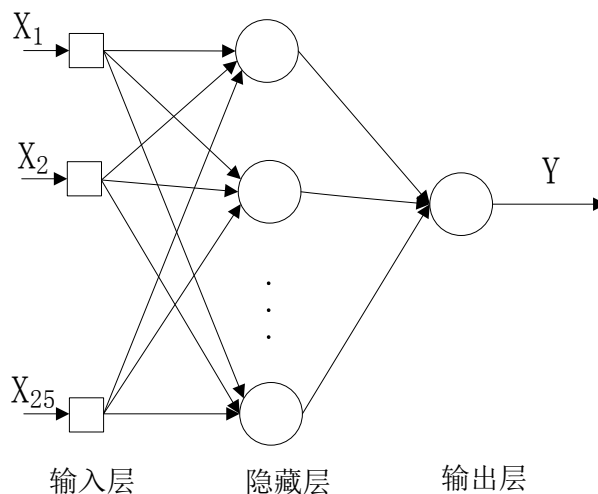
对所有客户做二分类，只区分**好和差**两种情况。数据采用德国信用数据库。德国信用数据库由Hans Hofmann教授整理，包含1000份客户资料，每位客户包含20条属性，并给出了信用好或差的标注。

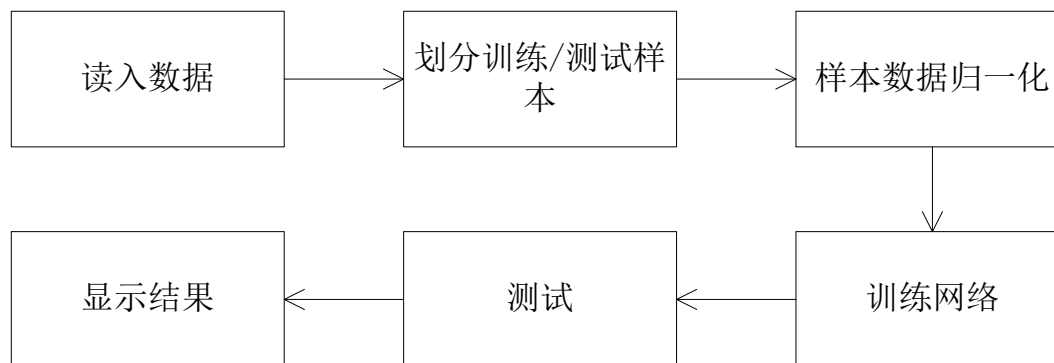
基于BP网络的个人信贷信用评估

原始数据保存在german.data文件中：

经常账户状况、账户持续时间、贷款历史状况、贷款用途、贷款数额。

使用MATLAB实现一个三层的BP神经网络。由于每个个人用户拥有24个属性，因此输入层包含24个神经元节点。该问题为针对信用好/差的二分类问题，因此输出层只包含一个神经元。隐含层的神经元个数与网络性能有关，需要通过实验确定





(1) 读入数据。

(2) 划分训练样本与测试样本。在全部1000份样本中，共有700份正例（信誉好），300份负例（信誉差）。划分时取前350份正例和前150份负例作为训练样本，后350份正例和后150份负例作为测试样本。

(3) 样本归一化。使用mapminmax函数对输入样本进行归一化

(4) 创建BP神经网络，并完成训练

(5) 测试。BP网络输出值并无限定为1或2，而是一个实数，因此还需要将输出转换为整数。取1.5为阈值，小于该阈值的输出判为1（信用好），否则判为2（信用差）

基于BP网络的个人信贷信用评估

运行脚本credit_class.m

测试20次的平均正确率为74.97%，最低正确率为73.4%，迭代次数均为3次

- 非线性映射能力：由于有多层，可以容纳更多节点，权值，能储存更多信息，善于处理没有业务逻辑和数学方法可以描述，缺乏专家经验的问题，天生是面对非线性问题的通用逼近工具
- 容错能力：个别样本的误差对权值的全局训练影响不大。正确的知识来自全体样本，是从大量样本中提取统计特性的过程
- 泛化能力：训练完毕后的模型可以直接投入使用，速度很快，既可以用软件实现，也可以用电子器件，甚至神经网络芯片实现

- 具有其它神经网络解决方案的同样弱点：神经网络设计具有较高技巧。同时它也是一个灰箱系统，容易掩盖某些业务背景细节。容易产生过度拟合。
- 事实上BP神经网络也不能够支持太多的隐层数。原因：学习信号反向传播过程会越来越弱，容易陷入局部极小值。
- 韩力群书第53页

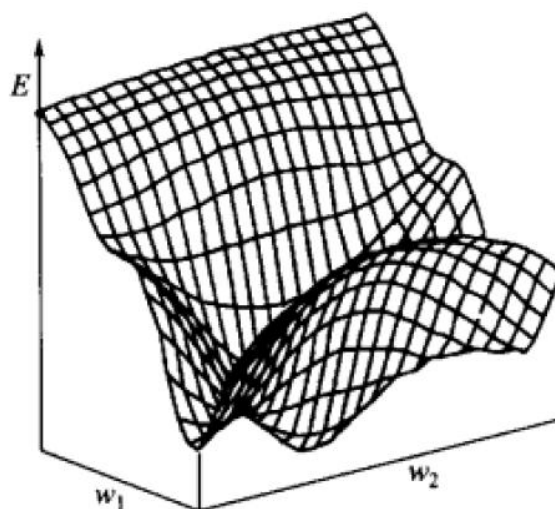


图 3.20 二维权空间的误差曲面

为什么BP网络不能支持太多的隐层数？

Understanding the difficulty of training deep feedforward neural networks

Xavier Glorot

DIRO, Université de Montréal, Montréal, Québec, Canada

Yoshua Bengio

Abstract

Whereas before 2006 it appears that deep multi-layer neural networks were not successfully trained, since then several algorithms have been shown to successfully train them, with experimental results showing the superiority of deeper vs less deep architectures. All these experimental results were obtained with new initialization or training mechanisms. Our objective here is to understand better why standard gradient descent from random initialization is doing so poorly with deep neural networks, to better understand these recent relative successes and help design better algorithms in the future. We first observe the influence of the non-linear activations functions. We find that the logistic sigmoid activation is unsuited for deep networks with random initialization because of its mean value, which can drive especially the top hidden layer into saturation. Surprisingly, we find that saturated units

learning methods for a wide array of *deep architectures*, including neural networks with many hidden layers (Vincent et al., 2008) and graphical models with many levels of hidden variables (Hinton et al., 2006), among others (Zhu et al., 2009; Weston et al., 2008). Much attention has recently been devoted to them (see (Bengio, 2009) for a review), because of their theoretical appeal, inspiration from biology and human cognition, and because of empirical success in vision (Ranzato et al., 2007; Larochelle et al., 2007; Vincent et al., 2008) and natural language processing (NLP) (Collobert & Weston, 2008; Mnih & Hinton, 2009). Theoretical results reviewed and discussed by Bengio (2009), suggest that in order to learn the kind of complicated functions that can represent high-level abstractions (e.g. in vision, language, and other AI-level tasks), one may need *deep architectures*.

Most of the recent experimental results with deep architecture are obtained with models that can be turned into deep supervised neural networks, but with initialization or training schemes different from the classical feedforward

Yoshua Bengio

- http://www.iro.umontreal.ca/~bengioy/yoshua_en/



- **Dataguru（炼数成金）是专业数据分析网站，提供教育，媒体，内容，社区，出版，数据分析业务等服务。我们的课程采用新兴的互联网教育形式，独创地发展了逆向收费式网络培训课程模式。既继承传统教育重学习氛围，重竞争压力的特点，同时又发挥互联网的威力打破时空限制，把天南地北志同道合的朋友组织在一起交流学习，使到原先孤立的学习个体组合成有组织的探索力量。并且把原先动辄成千上万的学习成本，直线下降至百元范围，造福大众。我们的目标是：低成本传播高价值知识，构架中国第一的网上知识流转阵地。**
- **关于逆向收费式网络的详情，请看我们的培训网站 <http://edu.dataguru.cn>**



Thanks

FAQ时间