東莞理工學院
DONGGUAN UNIVERSITY OF TECHNOLOGY

人工智能概论

第五章：神经网络

丁烨，计算机科学与技术学院
dingye@dgut.edu.cn

目录

- 神经元模型
- 感知机与多层网络
- 误差逆传播算法
- 全局最小与局部极小

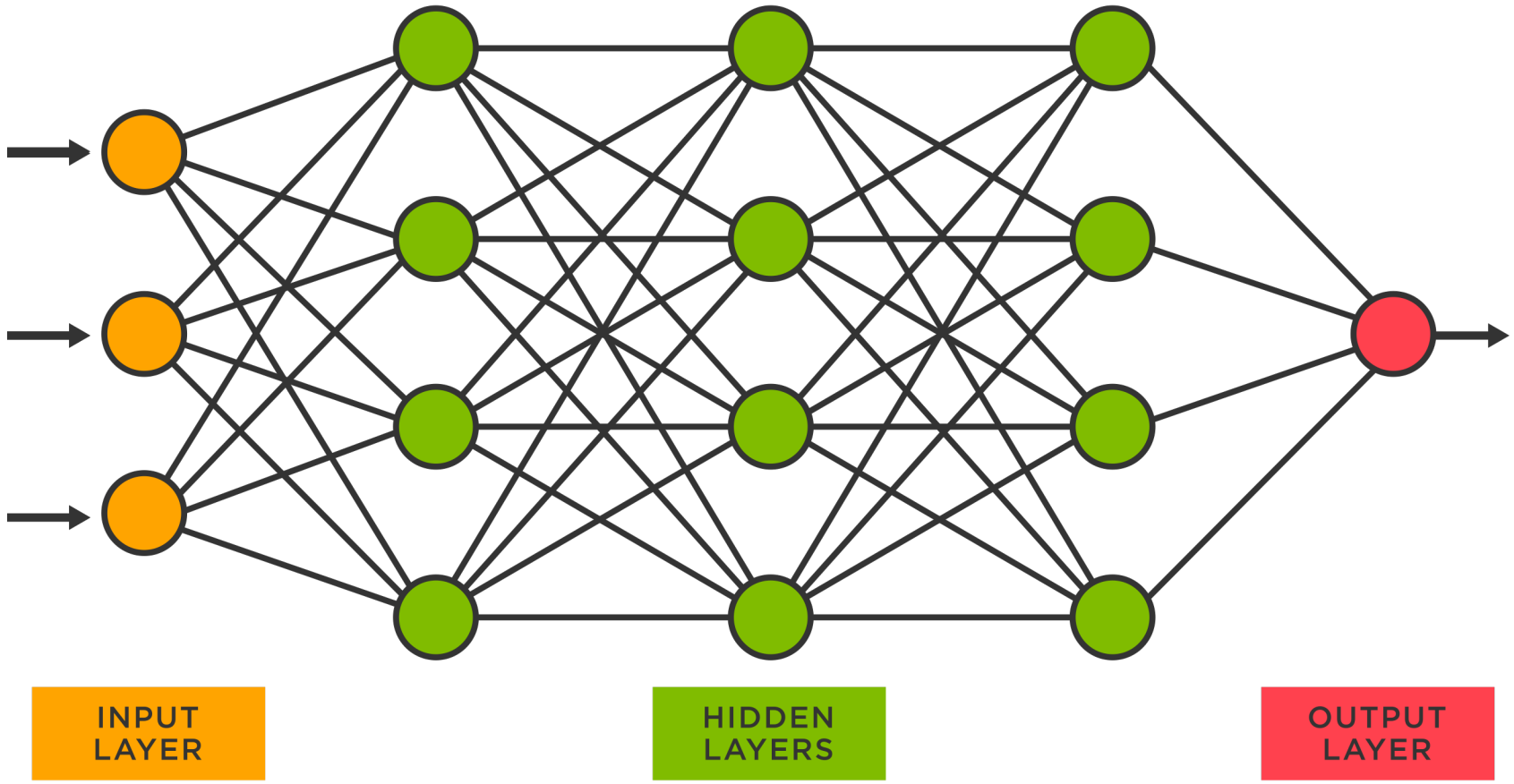
神经元模型

基本概念

- 人工神经网络（Artificial Neural Network，ANN）
- 一种模仿生物神经网络的结构和功能的机器学习算法
- 神经网络由大量的人工神经元（Neuron）联结进行计算
- 人工神经网络能在外界信息的基础上改变内部结构，是一种自适应系统
- 通俗地讲就是具备学习功能
- 和其他机器学习方法一样，神经网络已经被用于解决各种各样的问题，例如机器视觉和语音识别
- 这些问题都是很难被传统基于规则的算法所解决的

神经元模型

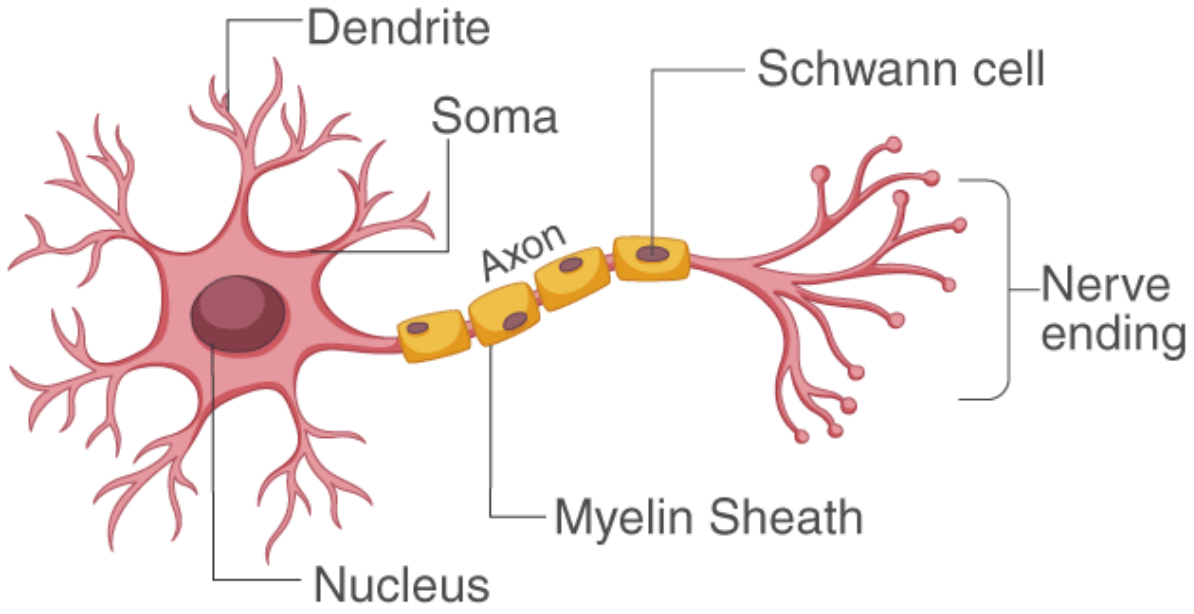
基本概念



神经元模型

基本概念

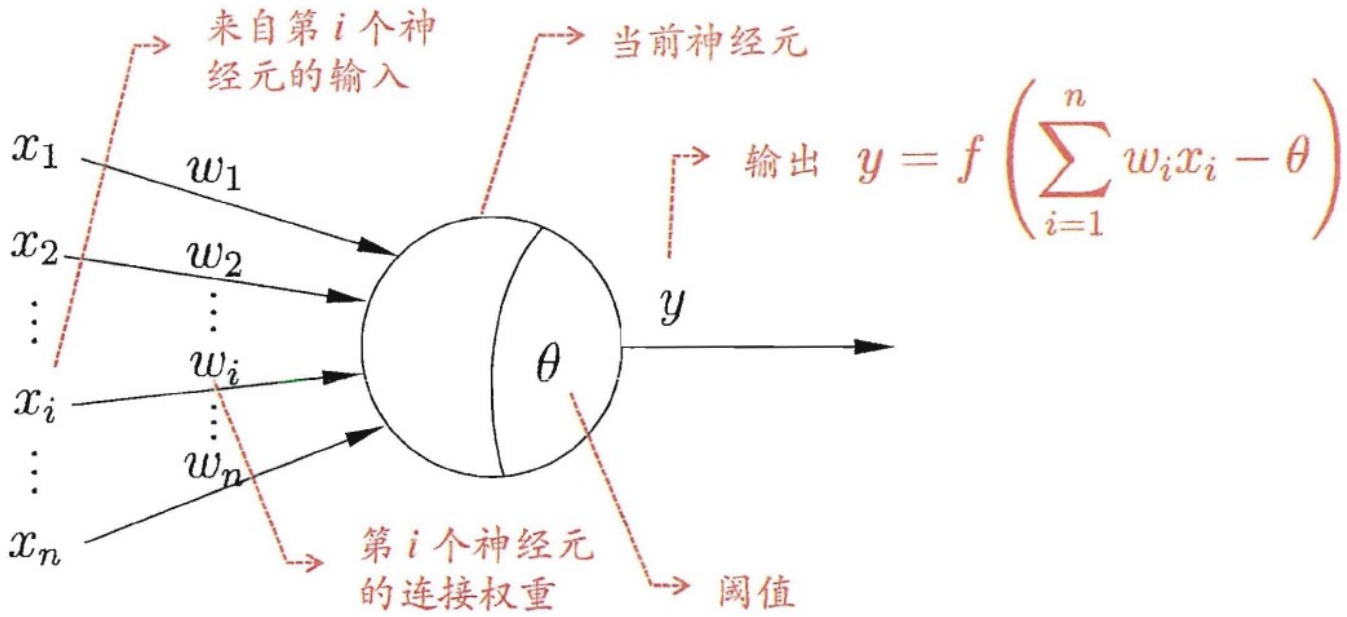
- 神经元（Neuron）
- 在生物神经网络中，每个神经元与其他神经元相连，当它被激活（“兴奋”）时，就会向相连的神经元发送化学物质，从而改变这些神经元内的电位
- 当神经元的电位超过了一个“阈值（Threshold）”，它就会被激活



神经元模型

基本概念

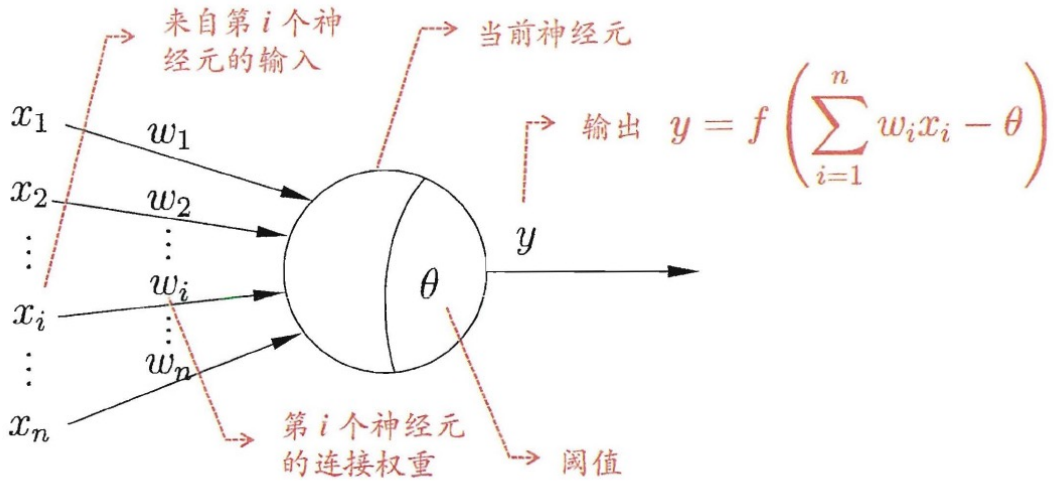
- 1943 年，McCulloch 和 Pitts 将上述情形抽象下图的数学模型
- 这就是一直沿用至今的“M-P 神经元模型”



神经元模型

基本概念

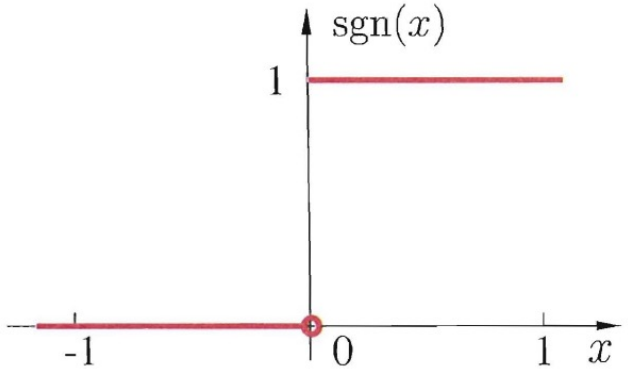
- 在这个模型中，神经元接收到来自其他神经元传递过来的输入信号
- 这些输入信号通过带权重的连接进行传递
- 神经元接收到的总输入值将与神经元的阈值进行比较
- 然后通过“激活函数（Activation Function）”处理
- 并产生神经元的输出



神经元模型

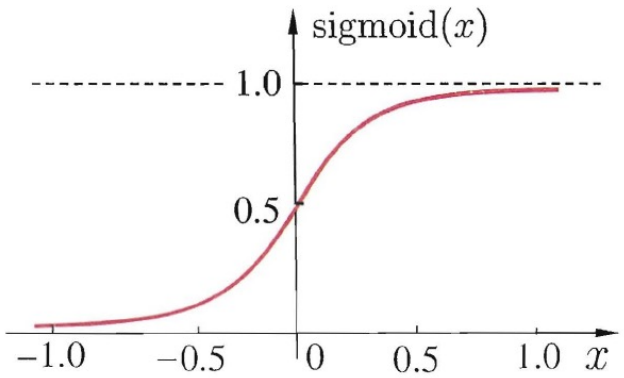
基本概念

- 理想中的激活函数是上图所示的阶跃函数
- 但阶跃函数具有不连续、不光滑等不太好的性质
- 因此实际常用 Sigmoid 函数作为激活函数



$$\mathrm{sgn}(x) = \begin{cases} 1, & x \geqslant 0; \\ 0, & x < 0. \end{cases}$$

(a) 阶跃函数



$$\mathrm{sigmoid}(x) = \frac{1}{1+e^{-x}}$$

(b) Sigmoid 函数

神经元模型

基本概念

- 把许多个这样的神经元按一定的层次结构连接起来，就得到了神经网络
- 事实上，从计算机科学的角度看，我们可以先不考虑神经网络是否真的模拟了生物神经网络
- 只需将一个神经网络视为包含了许多参数的数学模型，这个模型是若干个函数相互嵌套代入形成，例如：

$$y_i = f\left(\sum_i w_i x_i - \theta_j\right)$$

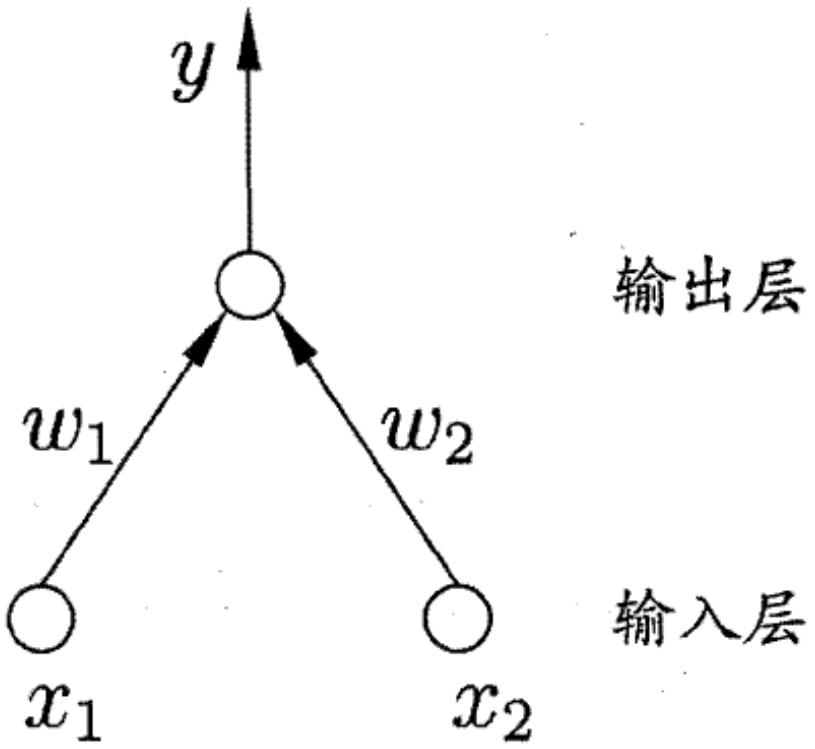
目录

- 神经元模型
- 感知机与多层网络
- 误差逆传播算法
- 全局最小与局部极小

感知机与多层网络

感知机

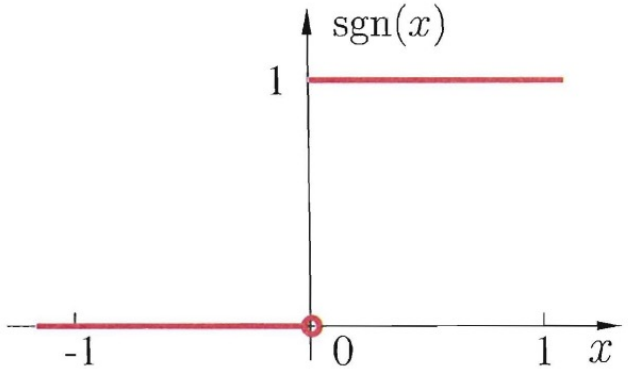
- 感知机（Perceptron）
- 由两层神经元组成
- 输入层接收外界输入信号后传递给输出层
- 输出层是 M-P 神经元
- 感知机能容易地实现逻辑与、或、非运算



感知机与多层网络

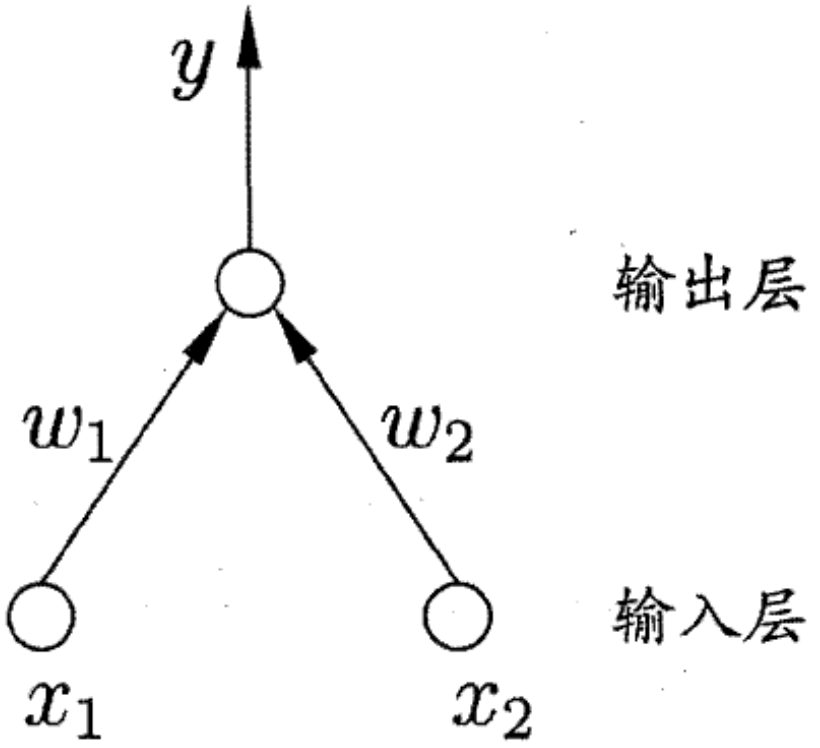
感知机

$$y = f\left(\sum_i w_i x_i - \theta\right)$$



$$\mathrm{sgn}(x) = \begin{cases} 1, & x \geqslant 0; \\ 0, & x < 0. \end{cases}$$

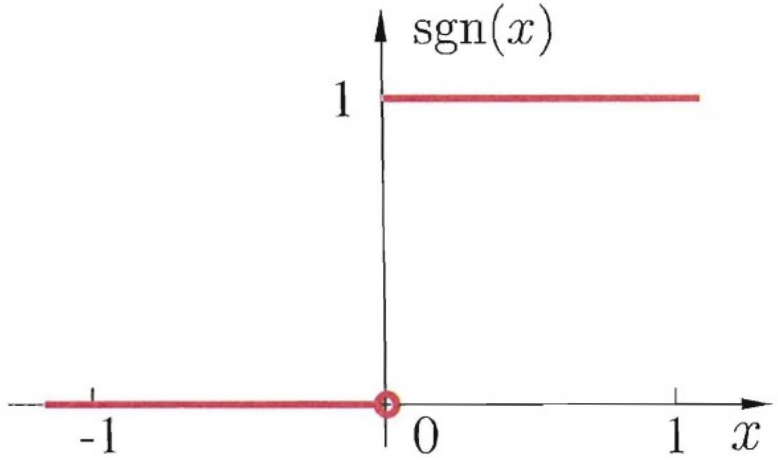
(a) 阶跃函数



感知机与多层网络

感知机

- 逻辑与
- 令 $w_1 = w_2 = 1,\ \theta = 2$
- 则 $y = f(1 \cdot x_1 + 1 \cdot x_2 - 2)$
- 那么，仅在 $x_1 = x_2 = 1$ 时
- $y = 1$



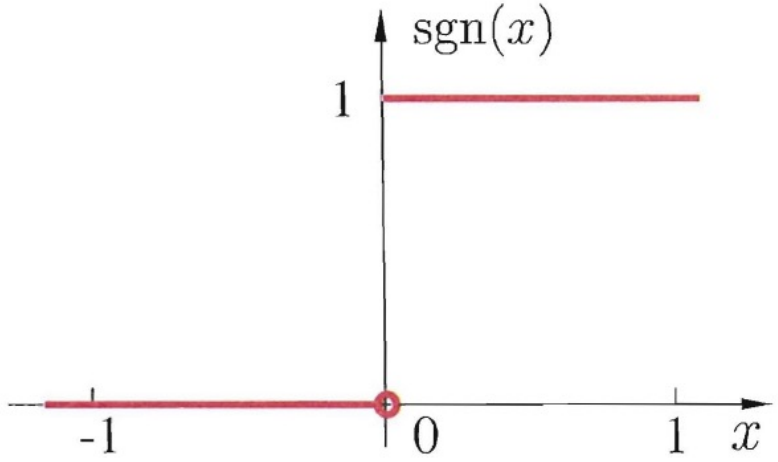
$$\mathrm{sgn}(x) = \begin{cases} 1, & x \geqslant 0; \\ 0, & x < 0. \end{cases}$$

(a) 阶跃函数

感知机与多层网络

感知机

- 逻辑或
- 令 $w_1 = w_2 = 1,\ \theta = 0.5$
- 则 $y = f(1 \cdot x_1 + 1 \cdot x_2 - 0.5)$
- 那么，当 $x_1 = 1$ 或 $x_2 = 1$ 时
- $y = 1$



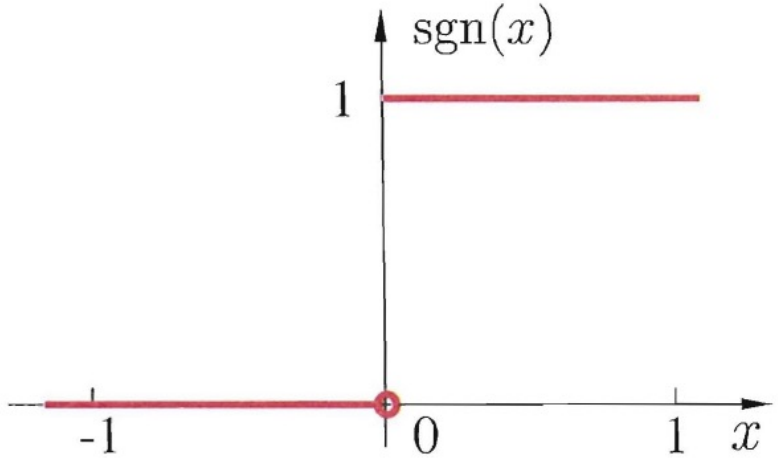
$$\mathrm{sgn}(x) = \begin{cases} 1, & x \geqslant 0; \\ 0, & x < 0. \end{cases}$$

(a) 阶跃函数

感知机与多层网络

感知机

- 逻辑非
- 令 $w_1 = -0.6$，$w_2 = 0$，$\theta = -0.5$
- 则 $y = f(-0.6 \cdot x_1 + 0 \cdot x_2 + 0.5)$
- 那么：
- 当 $x_1 = 1$ 时，y = 0
- 当 $x_1 = 0$ 时，y = 1



$$\mathrm{sgn}(x) = \begin{cases} 1, & x \geqslant 0; \\ 0, & x < 0. \end{cases}$$

(a) 阶跃函数

感知机与多层网络

感知机

- 给定训练数据集
- 神经元的权重 $w_i(i = 1,2,\cdots,n)$ 和阈值 θ 可以通过学习得到
- 假设我们先忽略 θ：“哑结点（Dummy Node）”
- 感知机的学习规则就变得非常简单

感知机与多层网络

感知机

- 对训练样本 (x, y)，若当前感知机的输出为 $\hat{y}$
- 则感知机权重将调整为：

$$w_i \leftarrow w_i + \Delta w_i$$
$$\Delta w_i = \eta(y - \hat{y})x_i$$

- 其中 $\eta \in (0,1)$ 称为学习率（Learning Rate）
- 若感知机对训练样本预测正确，即 $\hat{y} = \mathrm{y}$，则感知机不发生变化
- 否则将根据错误的程度进行权重调整

感知机与多层网络

感知机

- 感知机只有输出层神经元进行激活函数处理
- 其学习能力非常有限
- 事实上，上述与、或、非问题都是线性可分的问题
- 即存在一个线性超平面能将它们分开
- 因此感知机的学习过程一定会收敛（Converge）
- 但是对于非线性可分的问题来说，感知机就很难训练了

感知机与多层网络

感知机

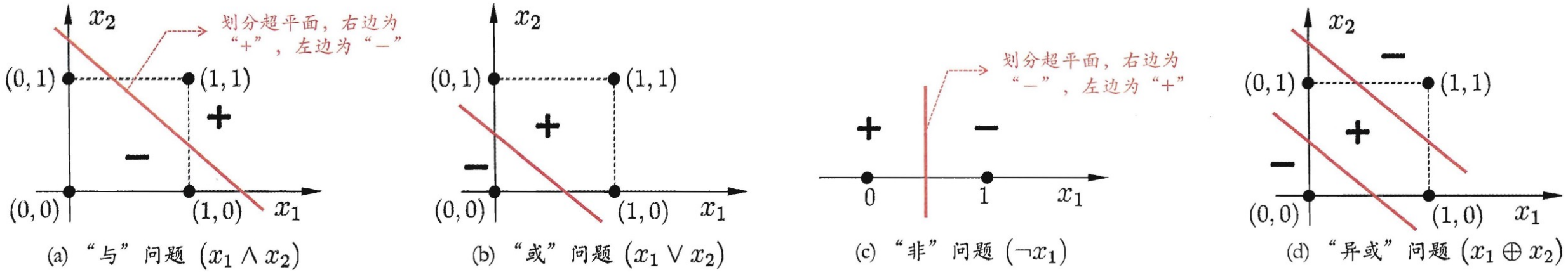
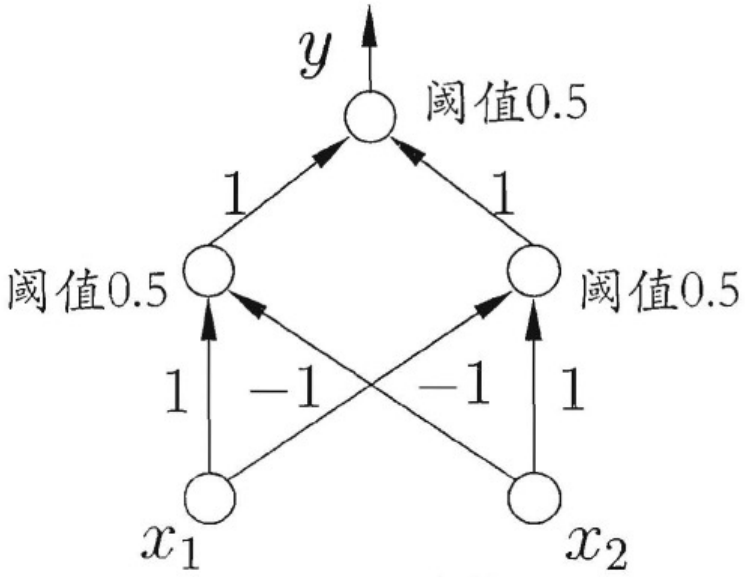


图 5.4 线性可分的“与” “或” “非”问题与非线性可分的“异或”问题

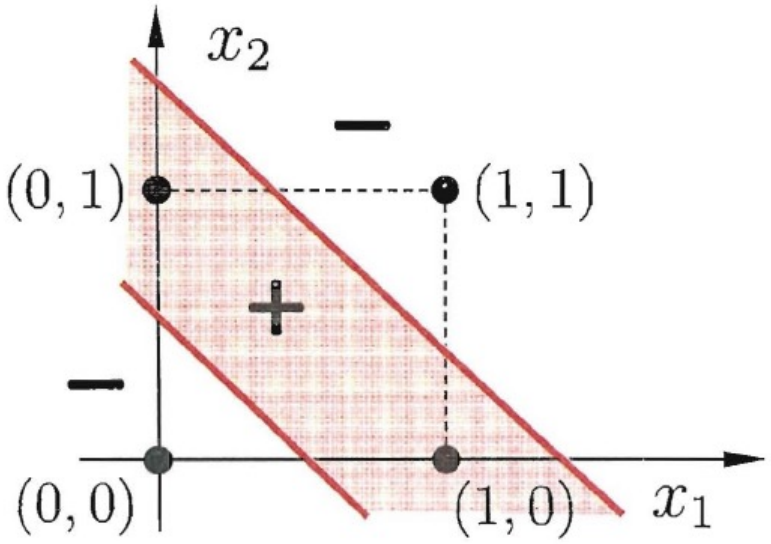
感知机与多层网络

多层网络

- 要解决非线性可分问题，需考虑使用多层神经元
- 例如下图的两层感知机就能解决异或问题



(a) 网络结构

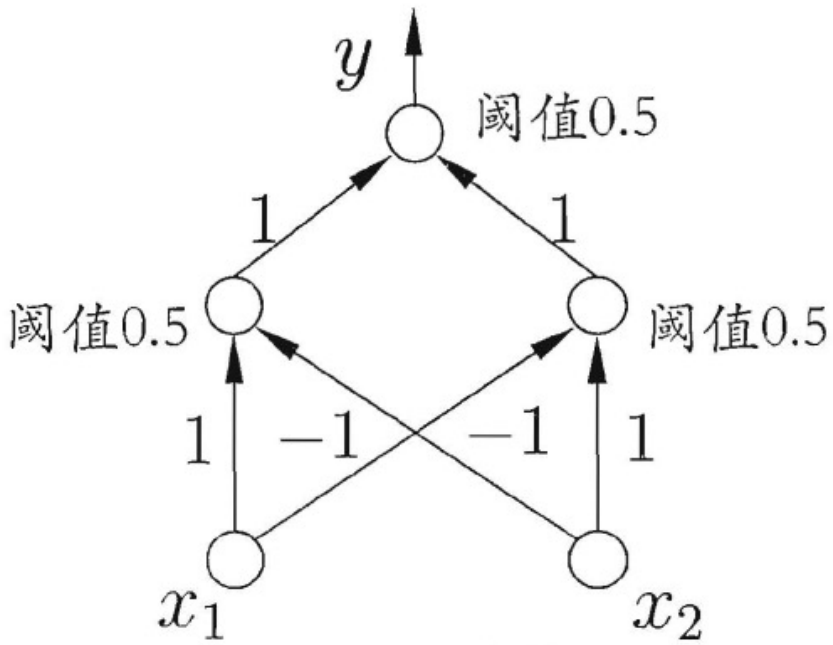


(b) 分类区域

感知机与多层网络

多层网络

- 输出层与输入层之间的神经元被称为：
- 隐含层（Hidden Layer）
- 隐含层和输出层神经元都是拥有激活函数的功能神经元

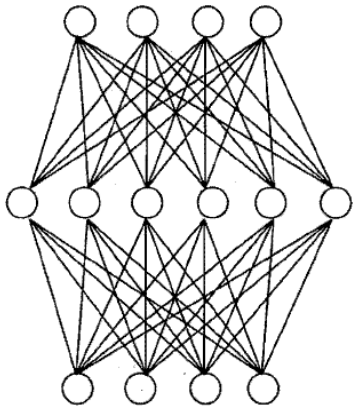


(a) 网络结构

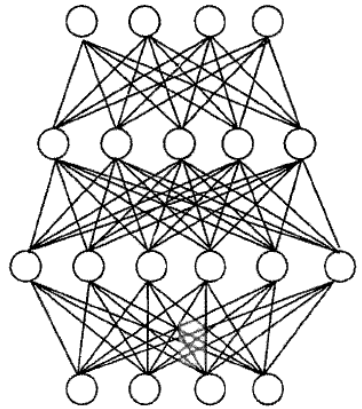
感知机与多层网络

多层网络

- 常见的神经网络是形如右图所示的层级结构
- 每层神经元与下一层神经元全互连
- 神经元之间不存在同层连接
- 也不存在跨层连接
- 这样的神经网络结构通常称为：
- “多层前馈神经网络”
- Multi-layer Feedforward Neural Network



(a) 单隐层前馈网络



(b) 双隐层前馈网络

感知机与多层网络

多层网络

- 多层前馈神经网络有强大的表示能力
- 只需一个包含足够多神经元的隐含层
- 多层前馈神经网络就能以任意精度逼近任意复杂度的连续函数
- 然而，如何设置隐含层神经元的个数仍是个未决问题
- 实际应用中通常靠“试错法”调整

感知机与多层网络

多层网络

- 神经网络的学习过程，就是根据训练数据来调整神经元之间的：
- “连接权重（Connection Weight）”
- 以及每个功能神经元的阈值
- 换言之，神经网络“学”到的东西，蕴涵在连接权重与阈值中

目录

- 神经元模型
- 感知机与多层网络
- 误差逆传播算法
- 全局最小与局部极小

误差逆传播算法

基本概念

- 对训练样本 (x, y)，若当前感知机的输出为 $\hat{y}$
- 则感知机权重将调整为：

$$w_i \leftarrow w_i + \Delta w_i$$
$$\Delta w_i = \eta(y - \hat{y})x_i$$

- 多层神经网络的学习能力比单层感知机强得多
- 训练时，上述公式不足以表达神经元之间的连接权重

误差逆传播算法

基本概念

- 误差逆传播（Backpropagation，BP）
- 对多层人工神经网络进行梯度下降（Gradient Decent）的算法
- 也就是用链式法则以网络每层的权重为变数计算损失函数的梯度
- 以更新权重来最小化损失函数

误差逆传播算法

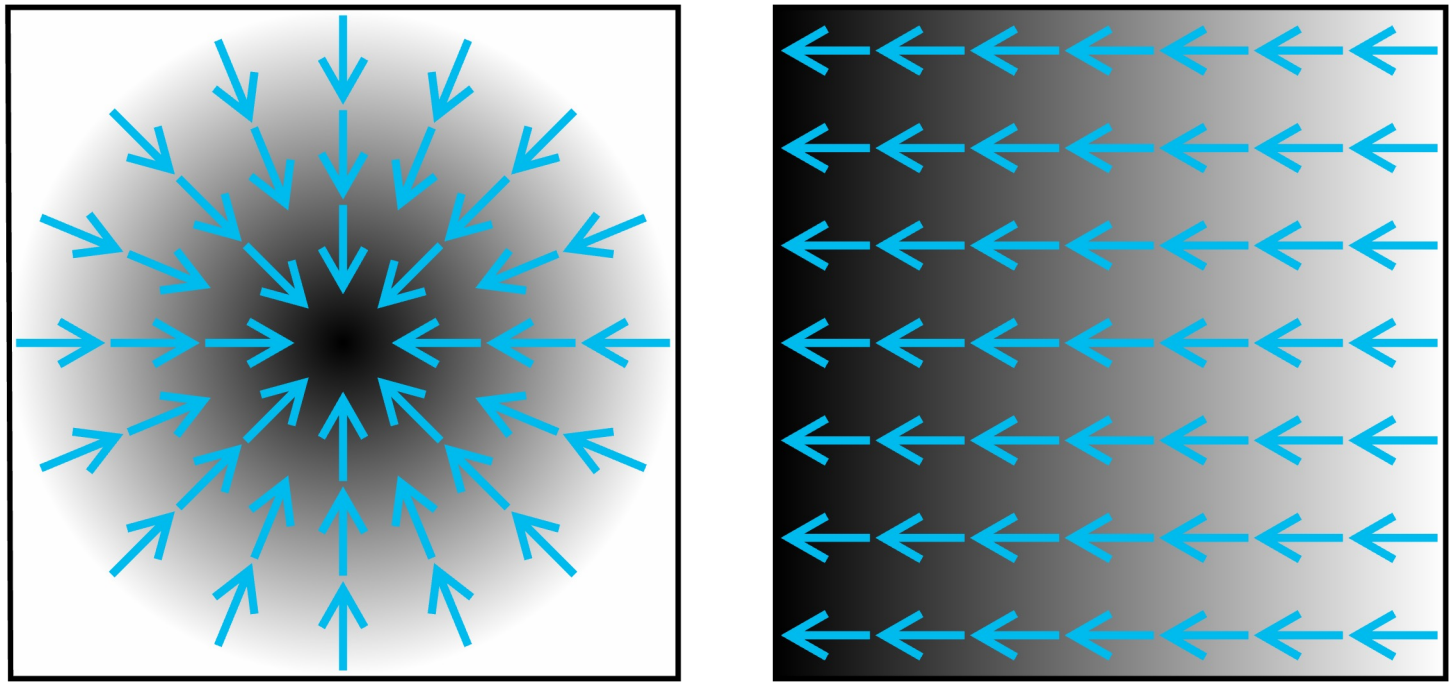
梯度下降法

- 梯度（Gradient）
- 在向量微积分中，梯度是一种关于多元导数的概括
- 一元函数的导数是标量值函数
- 多元函数的梯度是向量值函数
- 多元可微函数 f 在点 P 上的梯度
- 是以 f 在 P 上的偏导数为分量的向量

误差逆传播算法

梯度下降法

- 标量场的值用灰度表示，越暗表示越大的数值
- 而其相应的梯度用蓝色箭头表示



误差逆传播算法

梯度下降法

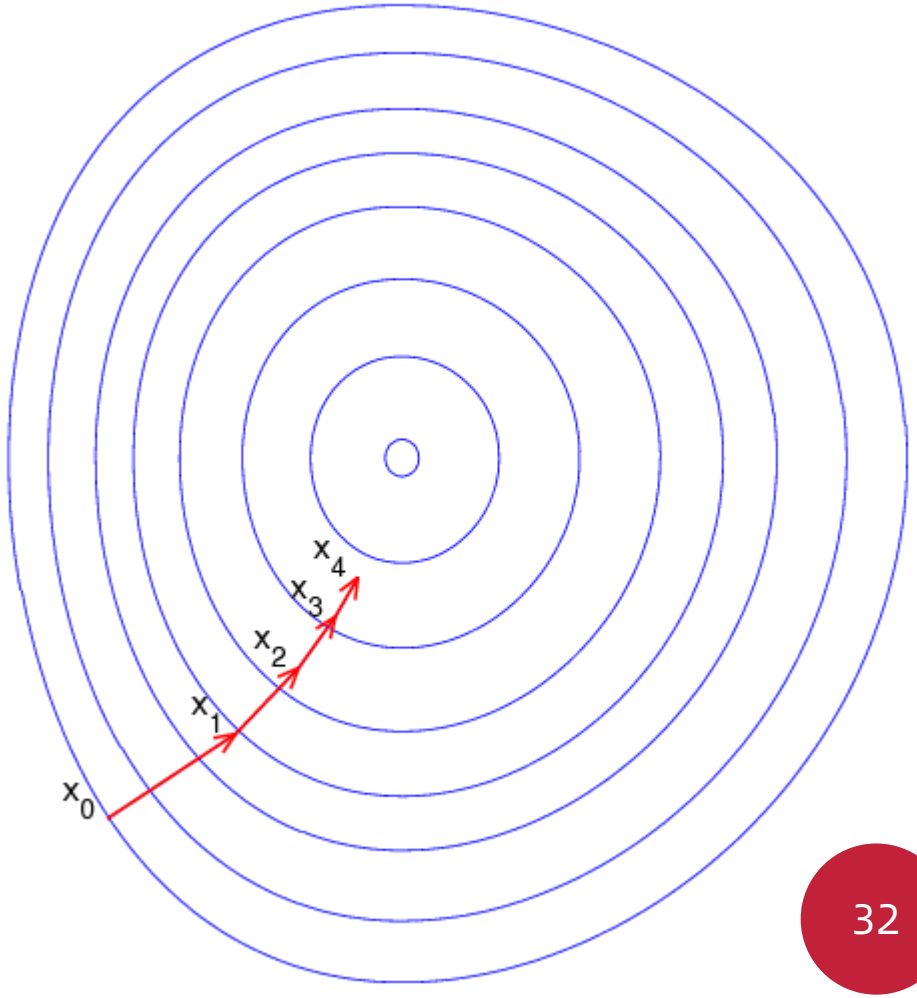
- 假设有一个房间，房间内所有点的温度由一个标量场给出
- 在房间的每一点，该点的梯度将显示变热最快的方向
- 梯度的大小将表示在该方向上变热的速率

- 考虑一座高度在 (x, y) 点是 $H(x, y)$ 的山
- H 这一点的梯度是在该点坡度（或者说斜度）最陡的方向
- 梯度的大小告诉我们坡度到底有多陡

误差逆传播算法

梯度下降法

- 梯度下降（Gradient Decent）
- 一个一阶最优化算法
- 不断的对函数上当前点对应梯度的反方向的规定步长距离点进行迭代搜索
- 直到找到函数的一个局部极小值



误差逆传播算法

误差逆传播算法

- 第一阶段：激励传播
- 每次迭代中的传播环节包含两步：
- 前向传播阶段：
- 将训练样本输入神经网络并获得预测结果
- 反向传播阶段：
- 通过“损失函数（Loss Function）”计算预测结果与真相的误差

误差逆传播算法

误差逆传播算法

- 第二阶段：权重更新
- 对于每个神经元上的权重，按照以下步骤进行更新：
- 将激活函数的值与误差相乘，从而获得权重的梯度
- 将这个梯度乘上一个比例并取反后加到权重上

- 注意：这个比例将会影响训练过程的速度和效果

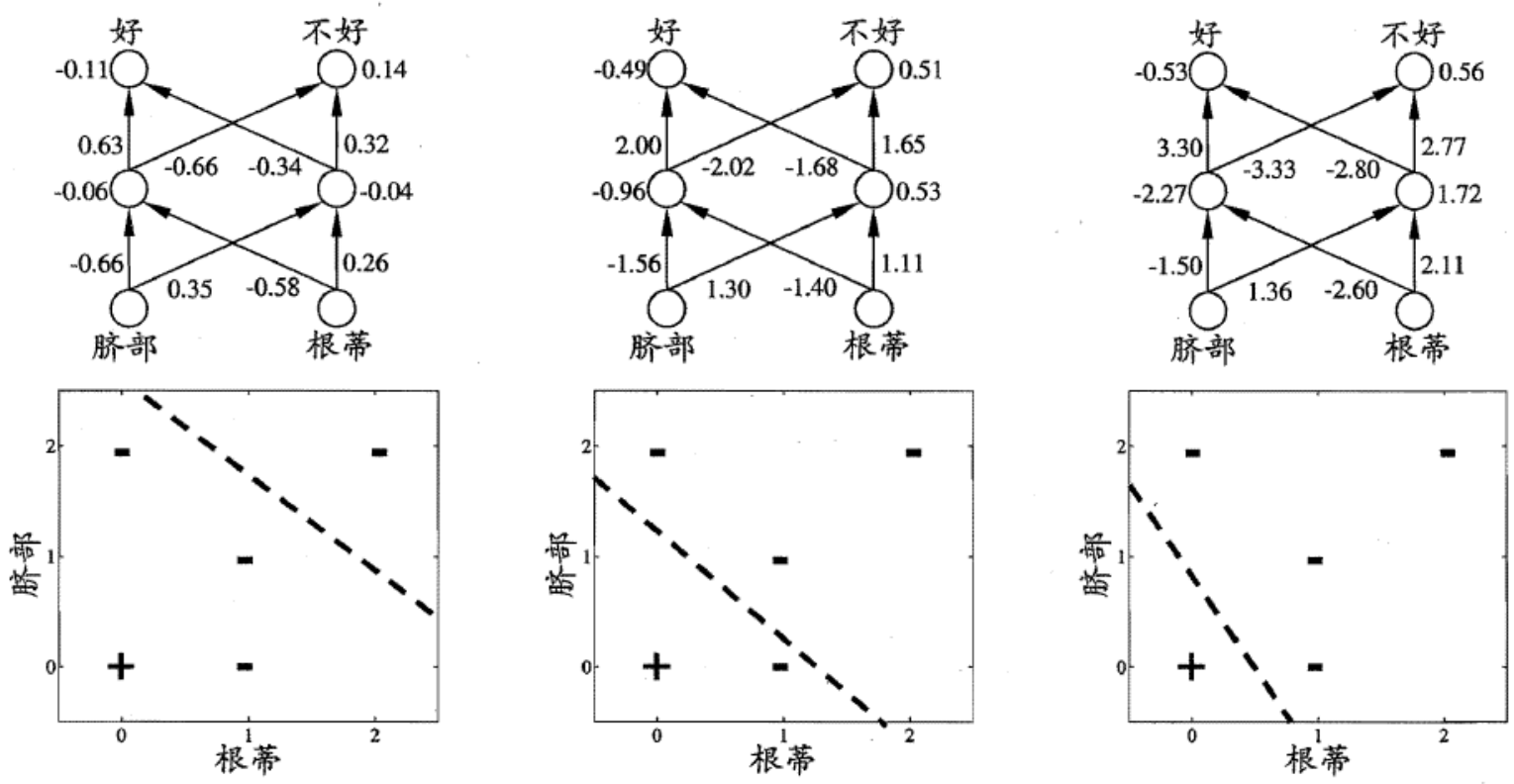
误差逆传播算法

误差逆传播算法

- 第一阶段和第二阶段可以反复循环迭代
- 直到误差达到一定阈值范围之内（“收敛”）
- 则训练完毕，得到模型

误差逆传播算法

误差逆传播算法



(a) 第 25 轮　　(b) 第 50 轮　　(c) 第 100 轮

图 5.9　在2个属性、5个样本的西瓜数据上, BP网络参数更新和分类边界的变化情况

目录

- 神经元模型
- 感知机与多层网络
- 误差逆传播算法
- 全局最小与局部极小

全局最小与局部极小

基本概念

- 若用 E 表示神经网络在训练集上的误差
- 则它显然是关于连接权重 w 和阈值 θ 的函数
- 此时，神经网络的训练过程可看作一个参数寻优过程
- 即在参数空间中，寻找一组最优参数使得 E 最小

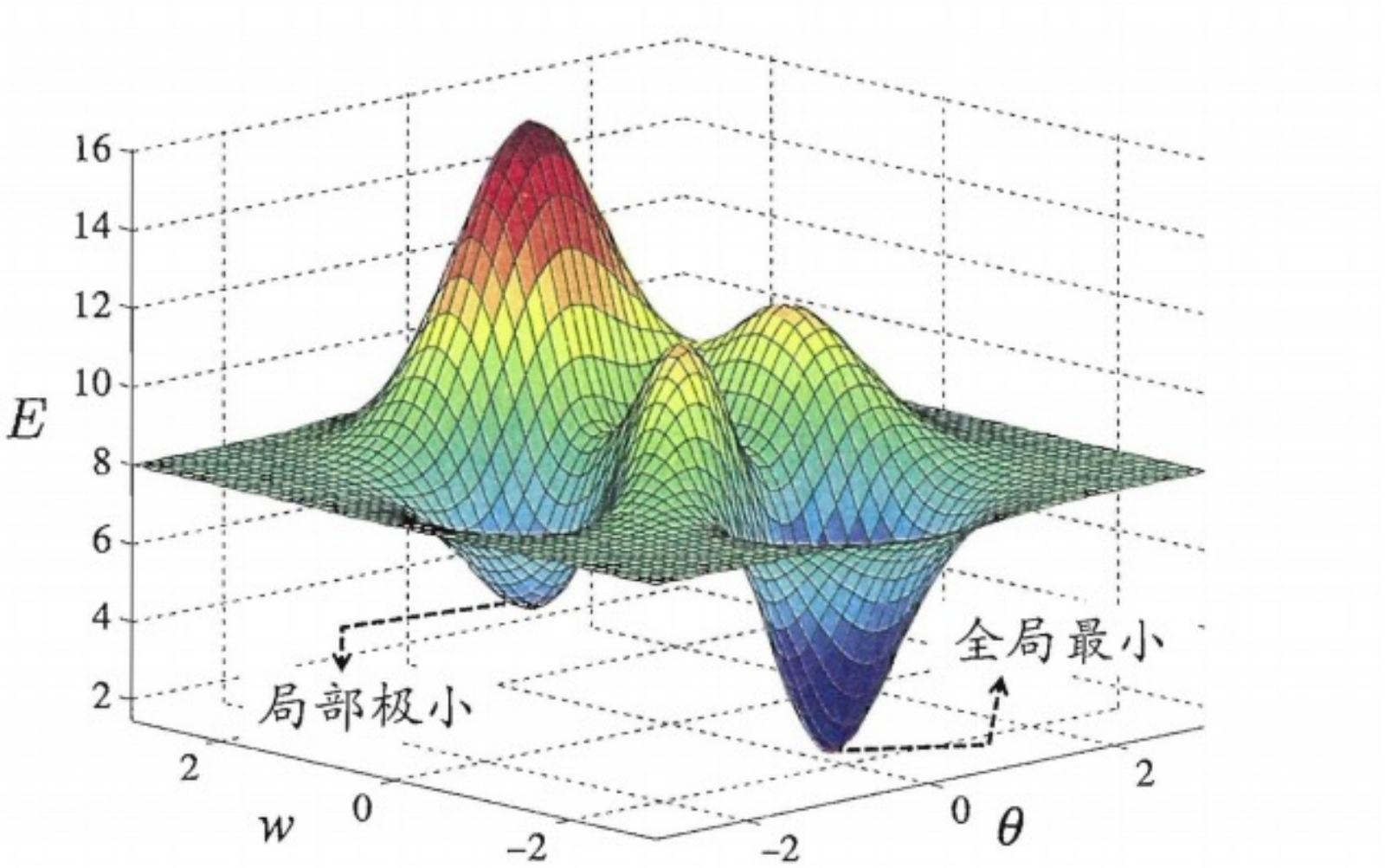
全局最小与局部极小

基本概念

- 我们常会谈到两种误差的“最优”：
- 局部极小（Local Minimum）
- 全局最小（Global Minimum）
- 对于参数空间中的某个点：
- 局部极小解是指其邻域点的误差函数值均不小于该点的误差函数值
- 全局最小解是指所有点的误差函数值均不小于该点的误差函数值

全局最小与局部极小

基本概念



全局最小与局部极小

解决方案

- 神经网络常采用以下策略来试图“跳出”局部极小
- 从而进一步接近全局最小
- 模拟退火（Simulated Annealing，SA）
- 随机梯度下降（Stochastic Gradient Descent，SGD）
- 遗传算法（Genetic Algorithm，GA）

- 注意：上述算法大多是启发式算法，理论上缺乏保障

全局最小与局部极小

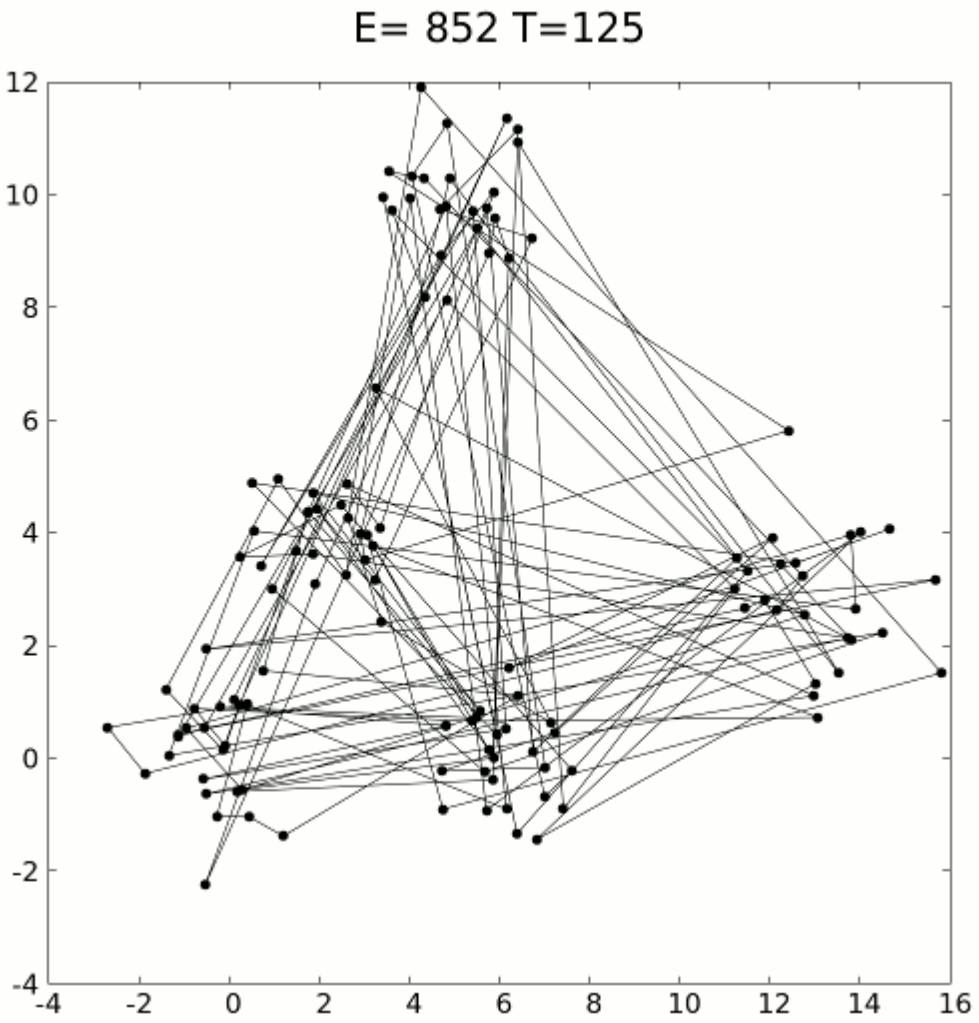
解决方案

- 模拟退火（Simulated Annealing，SA）
- 退火是指冶金学中将材料加热后再经特定速率冷却
- 材料中的原子起初会停留在内能局部极小值的位置
- 加热使能量变大，原子会离开原来位置，而随机在其他位置中移动
- 退火冷却时速度较慢
- 使得原子有较多可能可以找到内能比原先更低的位置

全局最小与局部极小

解决方案

- 模拟退火的原理和金属退火的原理近似
- 将搜寻空间内每一点想像成原子
- 算法先以搜寻空间内一个任意点开始
- 每一步先选择一个“邻居”
- 然后再计算到达“邻居”的概率
- 此处将模拟退火算法应用于求解旅行商问题：求出连接 125 个点的最小路线长度



全局最小与局部极小

解决方案

- 随机梯度下降（Stochastic Gradient Descent，SGD）
- 梯度下降算法的一个随机估计（Stochastic Approximation）
- 在小批量随机数据上计算损失函数的梯度
- 由于加入了随机性，即便陷入局部极小点，也有一定概率跳出

全局最小与局部极小

解决方案

- 遗传算法（Genetic Algorithm，GA）
- 在寻找全局最小的过程中加入随机“突变”
- 由于加入了随机性，即便陷入局部极小点，也有一定概率跳出

总结

- 神经网络是深度学习的重要基础
- 深度学习
- 2024-05-27
- 10:25 - 12:00

总结

- 「西瓜书」周志华《机器学习》，清华大学出版社
- https://item.jd.com/12762673.html

- 「南瓜书」谢文睿、秦州《机器学习公式详解》，人民邮电出版社
- https://github.com/datawhalechina/pumpkin-book/

总结

- Scikit-Learn
- https://scikit-learn.org

- TensorFlow
- https://www.tensorflow.org

TensorFlow

東莞理工學院
DONGGUAN UNIVERSITY OF TECHNOLOGY

Thank You!

丁烨，计算机科学与技术学院
dingye@dgut.edu.cn