

ML: Neural Network

Peng Li

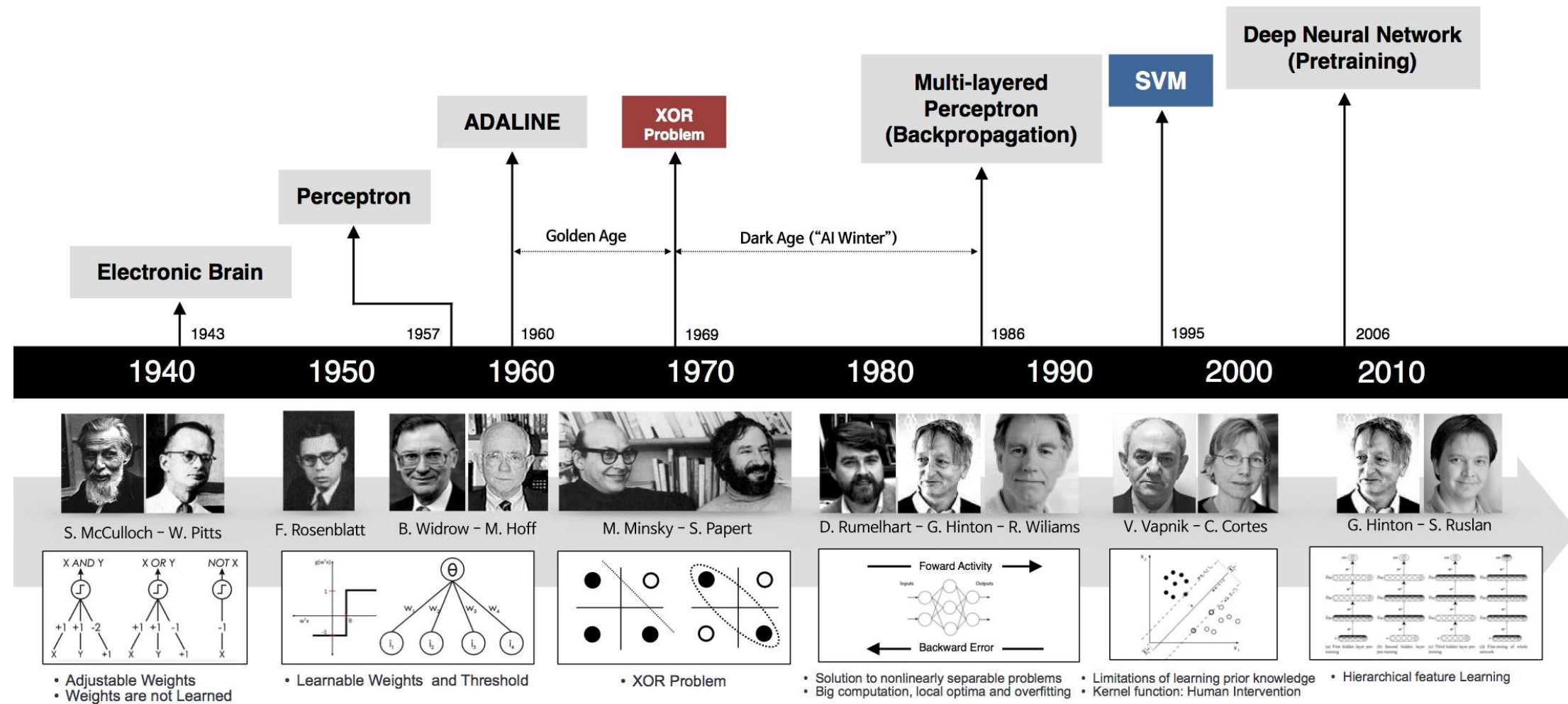
<https://simplelp.github.io/>

2019/05/08

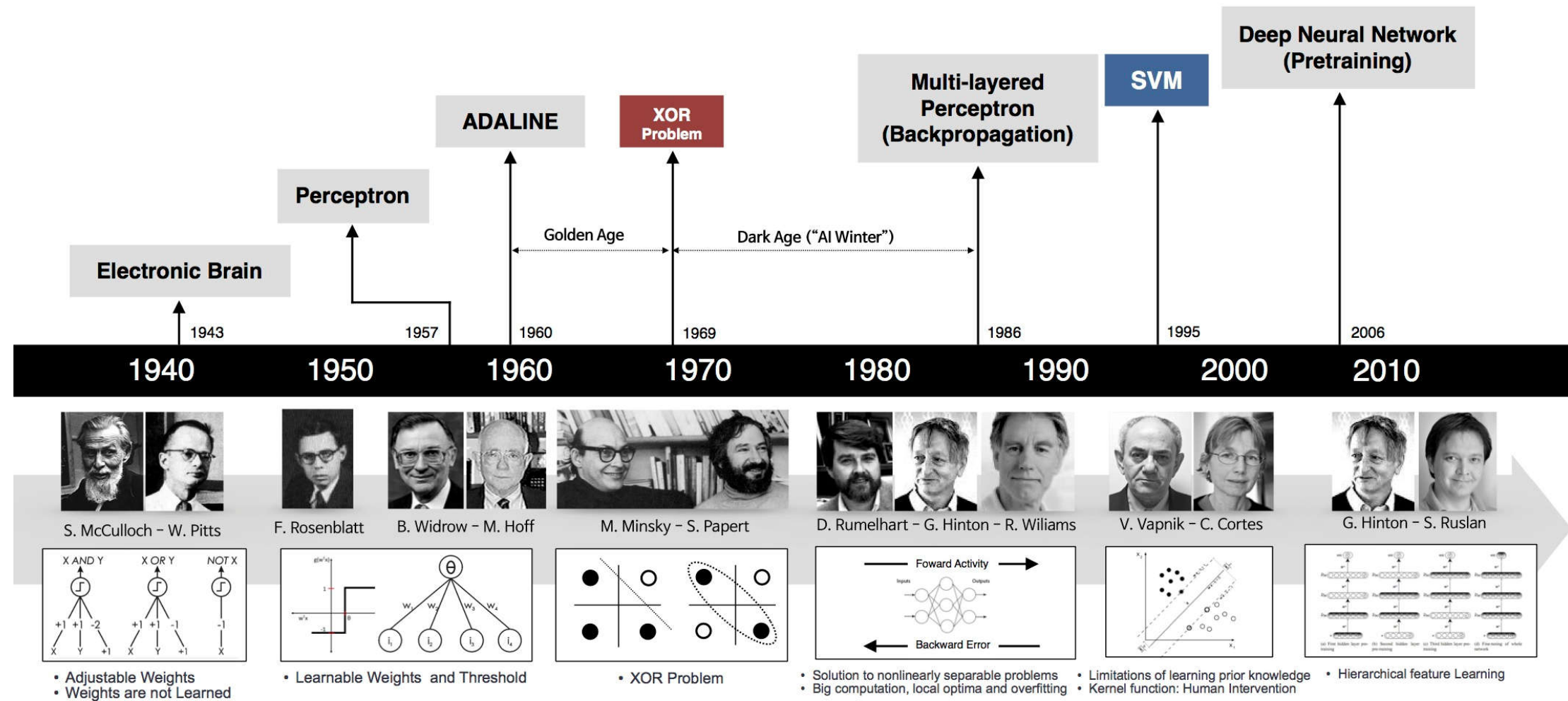
主要参考

- 《机器学习》，周志华著

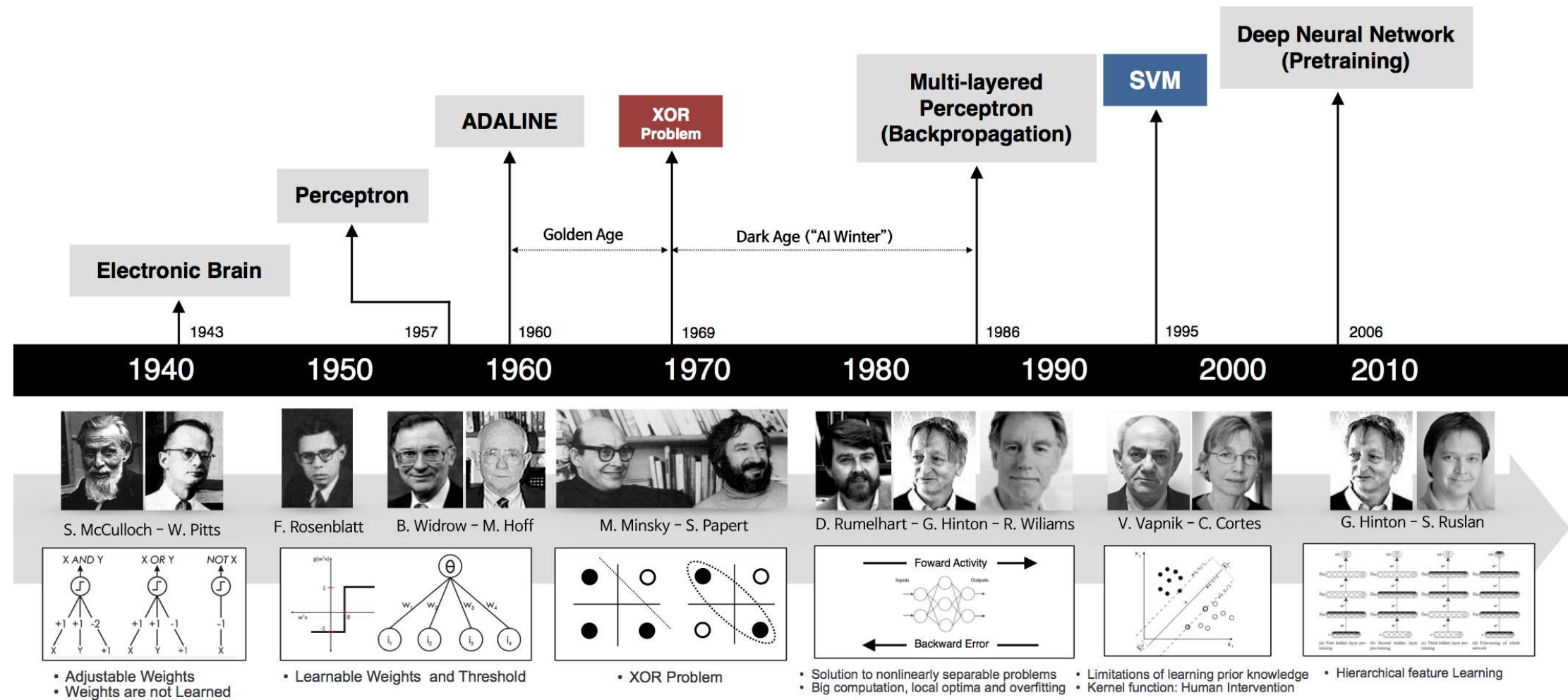
神经网络发展史



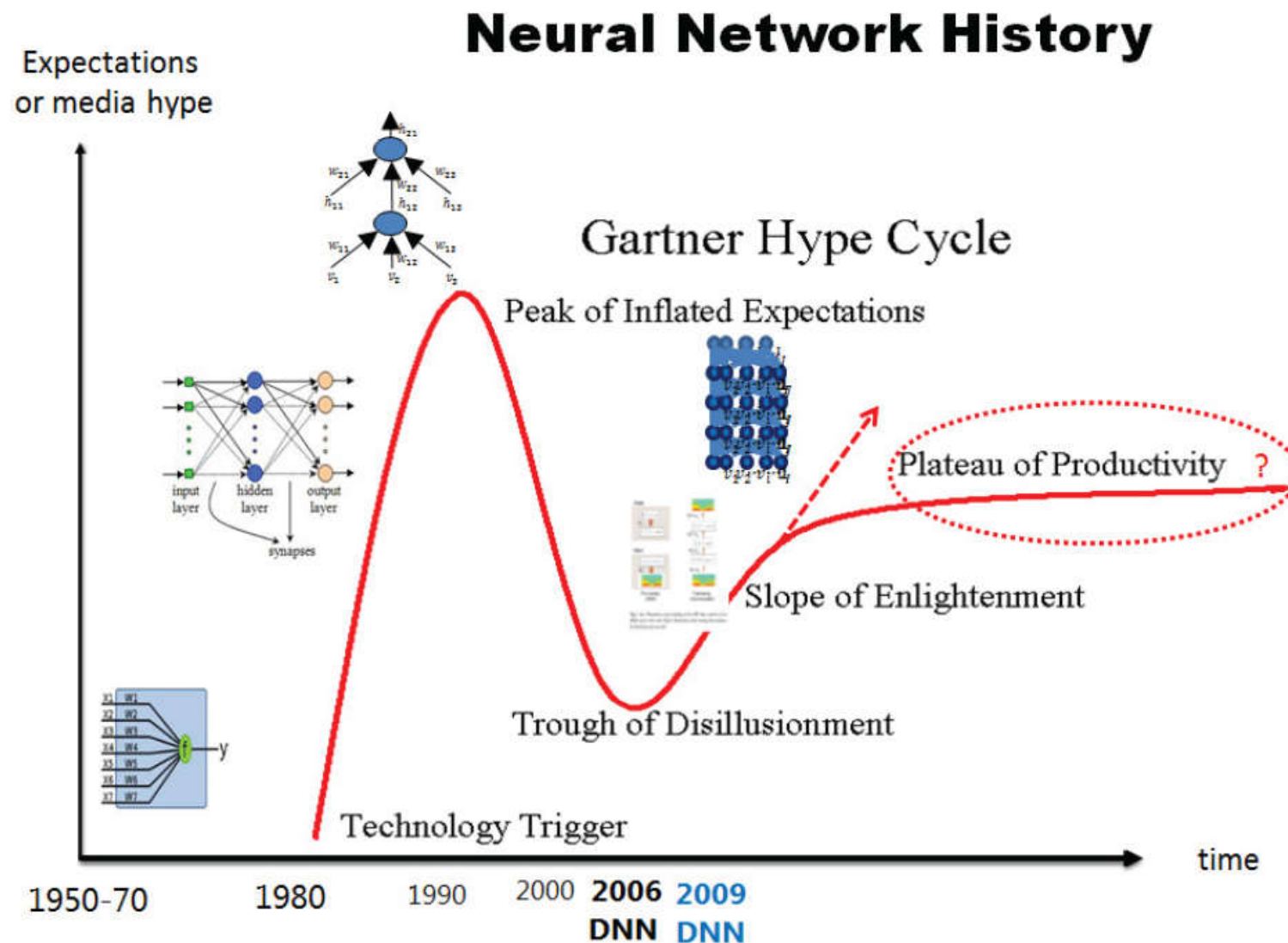
神经网络发展史



神经网络发展史



神经网络发展史



周志华老师寄语

科学的发展总是“螺旋式上升”

三十年河东、三十年河西

坚持才能有结果！

追热门、赶潮流 —— 三思而后行

目录

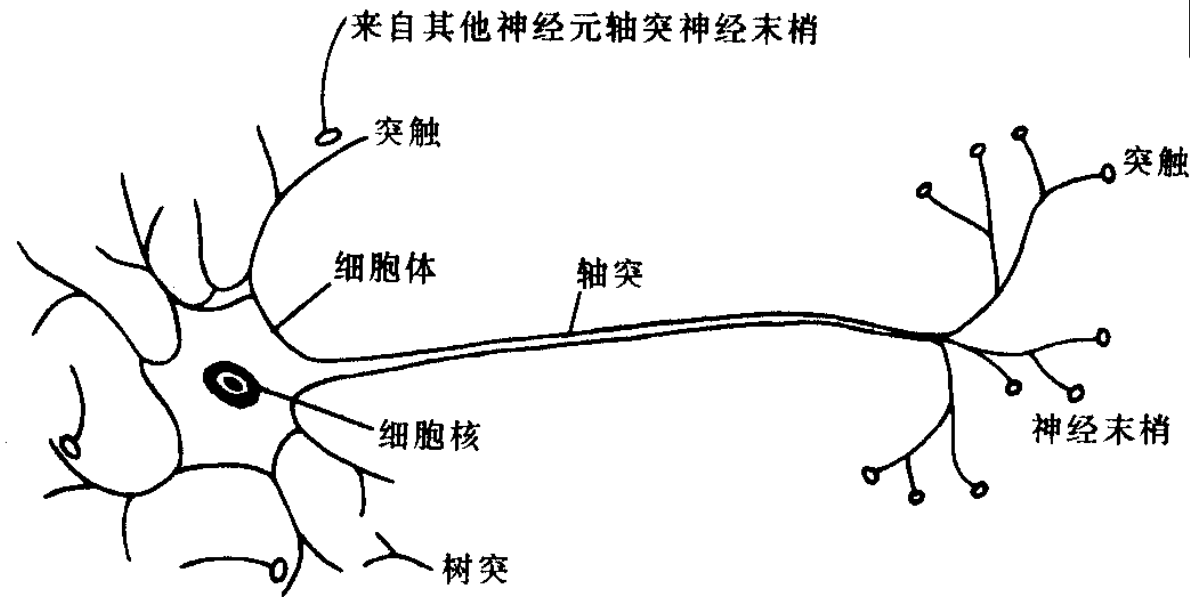
- 神经元模型
- 感知机与多层网络
- 误差反向传播算法
- 全局最小与局部最小
- 其他常见神经网络
- 深度学习

神经元模型

神经网络

- “神经网络是由具有**适应性的简单单元**组成的**广泛并行互联**的网络, 它的组织能够**模拟生物神经系统**对真实世界物体所作出的反应”

[Kohonen, 1988]



神经元模型

- M-P 神经元模型 [McCulloch and Pitts, 1943]

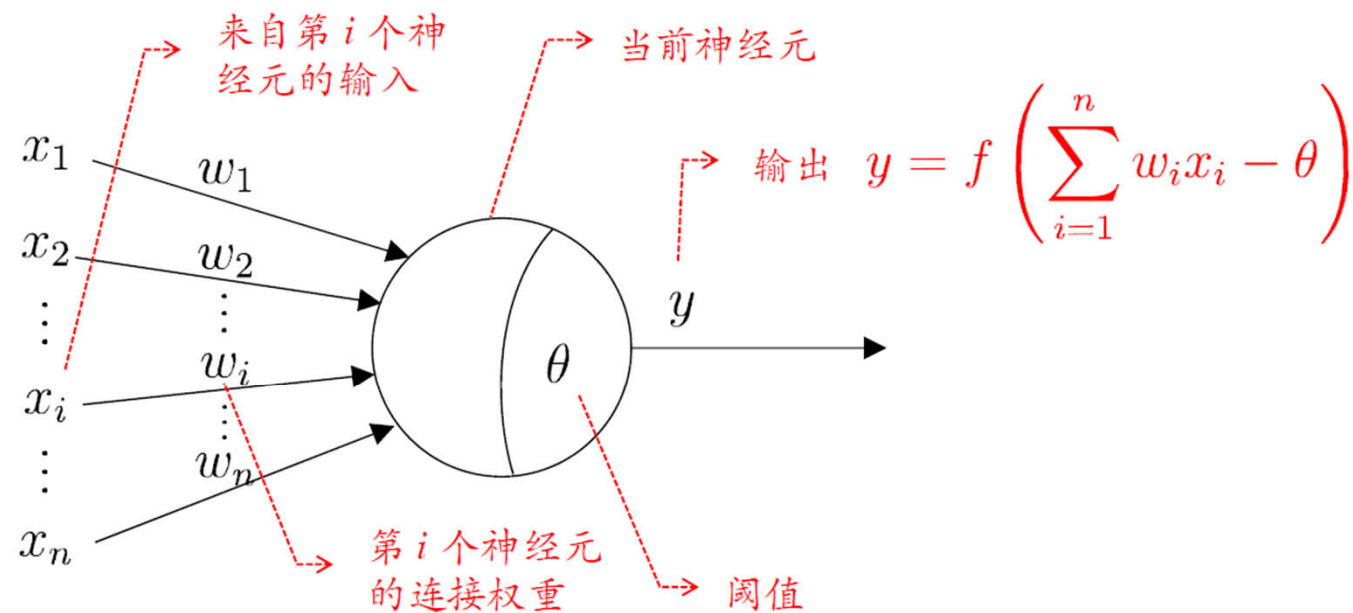
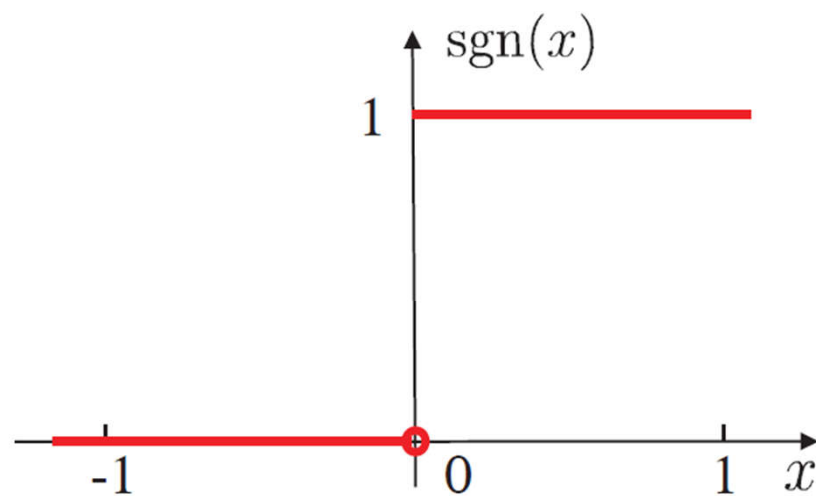


图 5.1 M-P 神经元模型

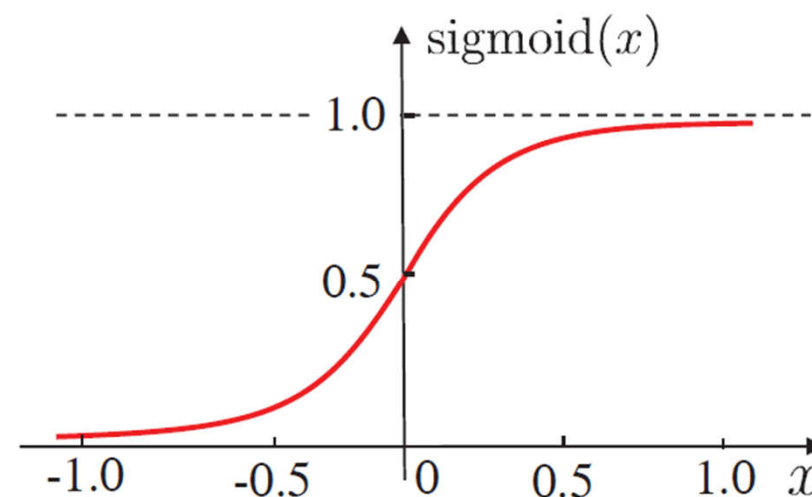
神经元模型

- 激活函数



$$\text{sgn}(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases}$$

(a) 阶跃函数



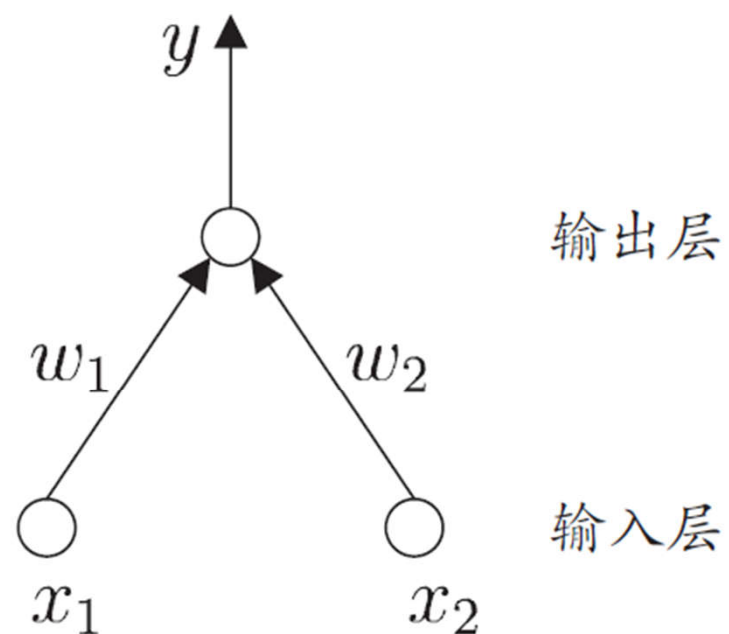
$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$$

(b) Sigmoid 函数

图 5.2 典型的神经元激活函数

感知机与多层网络

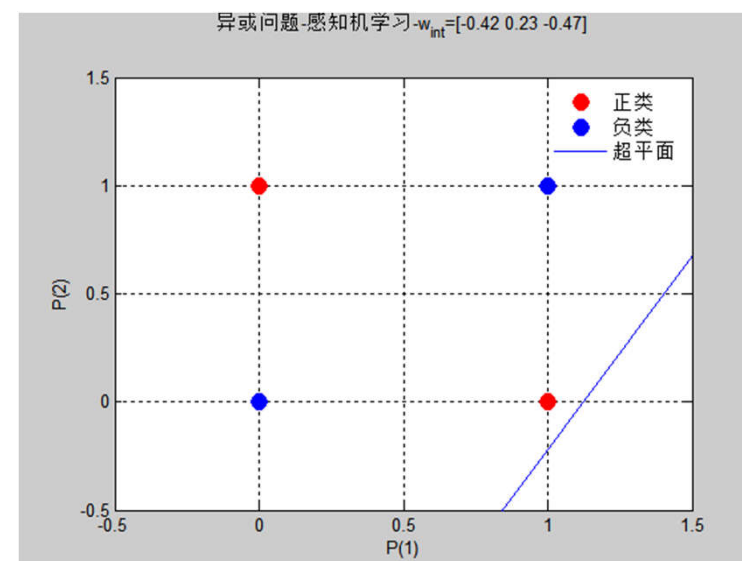
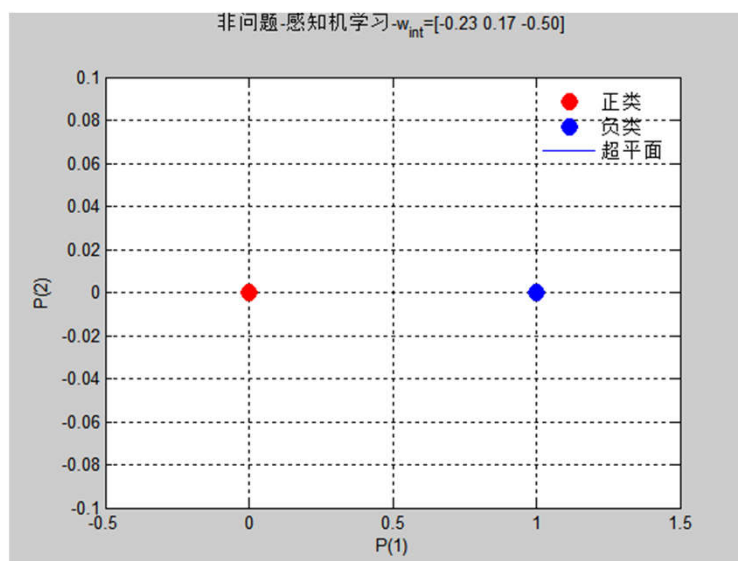
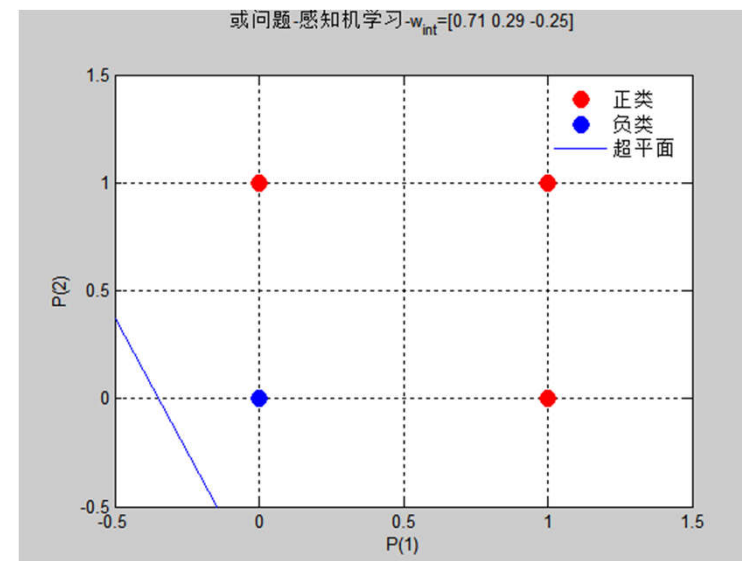
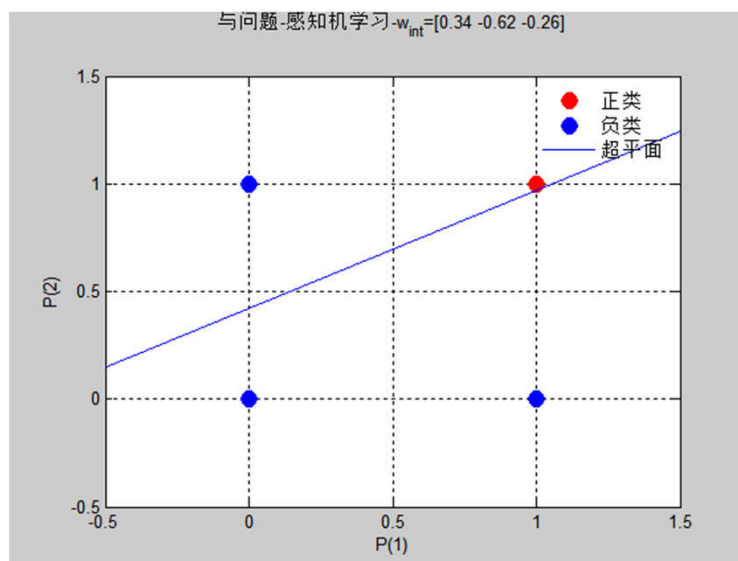
感知机及其学习



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

图 5.3 两个输入神经元的感知机网络结构示意图

感知机学习



多层感知机

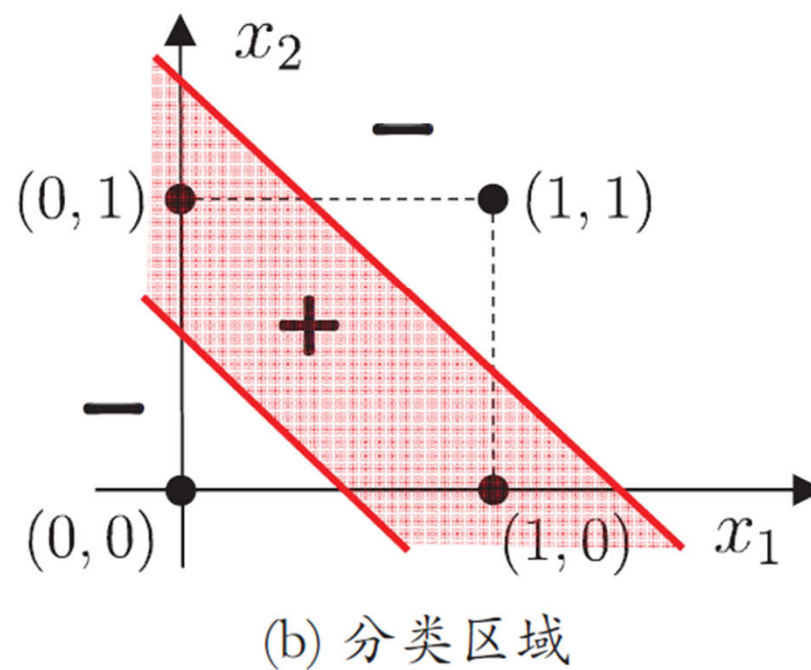
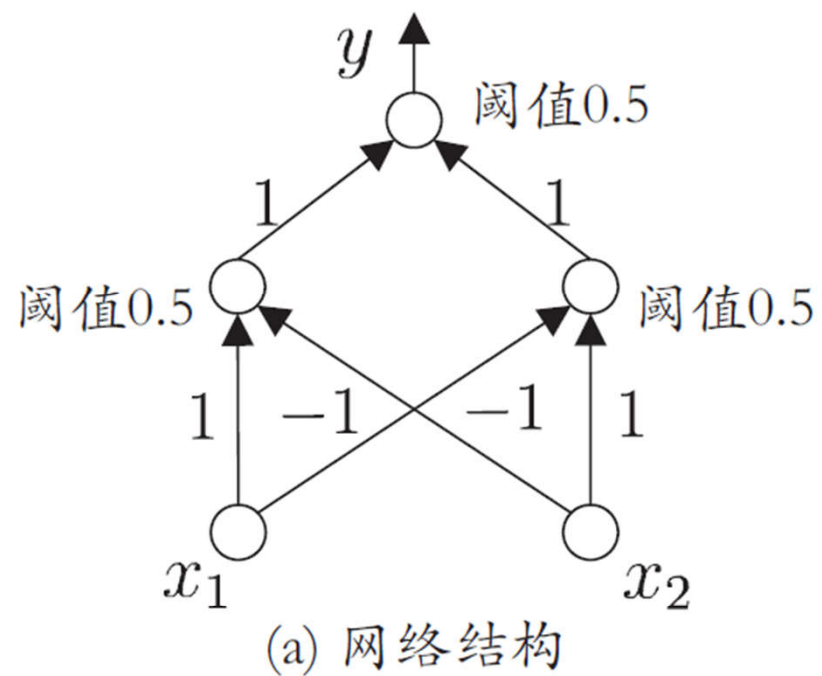
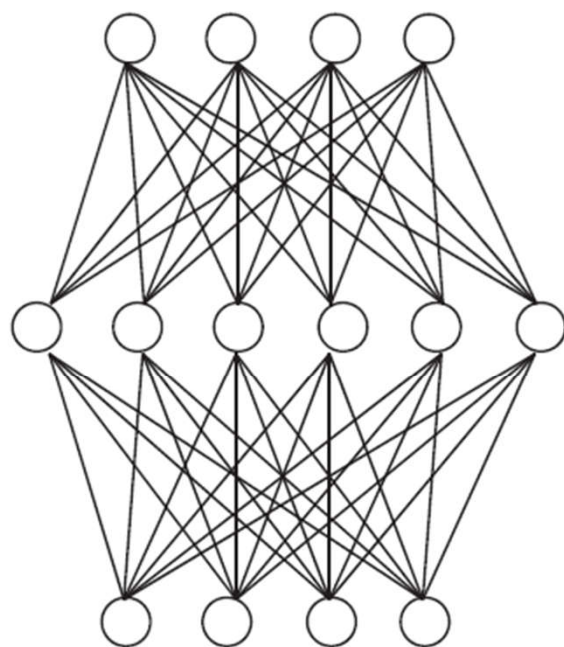


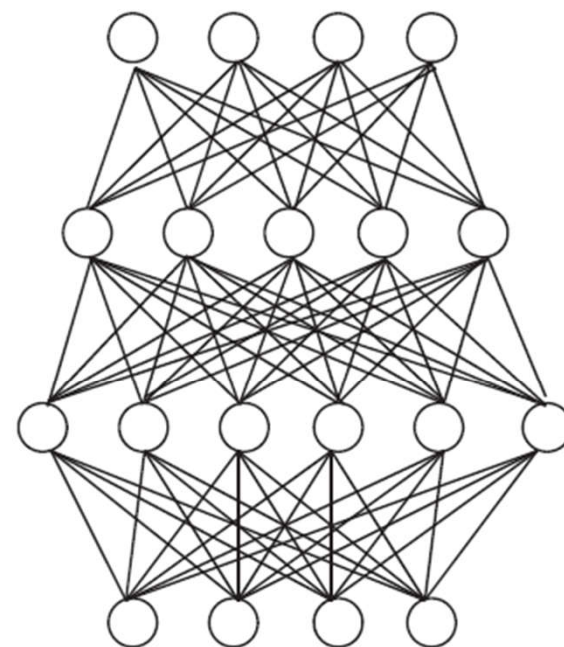
图 5.5 能解决异或问题的两层感知机

多层前馈神经网络

- 每层神经元与下一层神经元全互联, 神经元之间不存在同层连接也不存在跨层连接



(a) 单隐层前馈网络



(b) 双隐层前馈网络

多层前馈网络表示能力

- 只需要一个包含足够多神经元的隐层, 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

[Hornik et al. , 1989]

多层前馈网络局限性

- 神经网络由于强大的表示能力, 经常遭遇**过拟合**. 表现为: 训练误差持续降低, 但测试误差却可能上升
- 如何**设置隐层神经元的个数**仍然是个未决问题. 实际应用中通常使用“试错法”调整

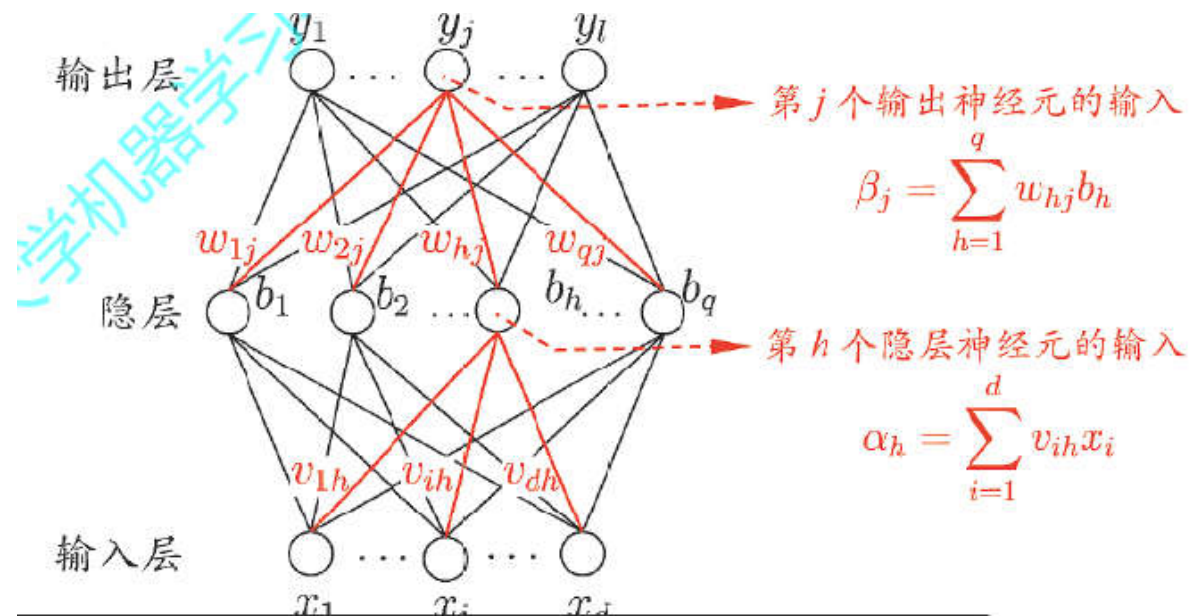
多层前馈网络局限性

- 缓解过拟合
 - 早停
 - 正则化
 - Dropout
 - 更多的数据
 - ...

误差反向传播算法

误差反向传播算法

- 误差反向传播算法（Error BackPropagation, 简称BP）是最成功的训练多层前馈神经网络的学习算法。



<https://google-developers.appspot.com/machine-learning/crash-course/backprop-scroll/>

误差反向传播算法

输入：训练集 $D = \{(\mathbf{x}_k, \mathbf{y}_k)\}_{k=1}^m$;
学习率 η .

过程：

- 1: 在(0, 1)范围内随机初始化网络中所有连接权和阈值
- 2: **repeat**
- 3: **for all** $(\mathbf{x}_k, \mathbf{y}_k) \in D$ **do**
- 4: 根据当前参数和式(5.3) 计算当前样本的输出 $\hat{\mathbf{y}}_k$;
- 5: 根据式(5.10) 计算输出层神经元的梯度项 g_j ;
- 6: 根据式(5.15) 计算隐层神经元的梯度项 e_h ;
- 7: 根据式(5.11)-(5.14) 更新连接权 w_{hj} , v_{ih} 与阈值 θ_j , γ_h
- 8: **end for**
- 9: **until** 达到停止条件

输出：连接权与阈值确定的多层前馈神经网络

$$\Delta w_{hj} = \eta b_j g_h \quad (5.11)$$

$$\Delta \theta_j = -\eta g_j, \quad (5.12)$$

$$\Delta v_{ih} = \eta e_h x_i, \quad (5.13)$$

$$\Delta \gamma_h = -\eta e_h, \quad (5.14)$$

图 5.8 误差逆传播算法

BP 算法实验

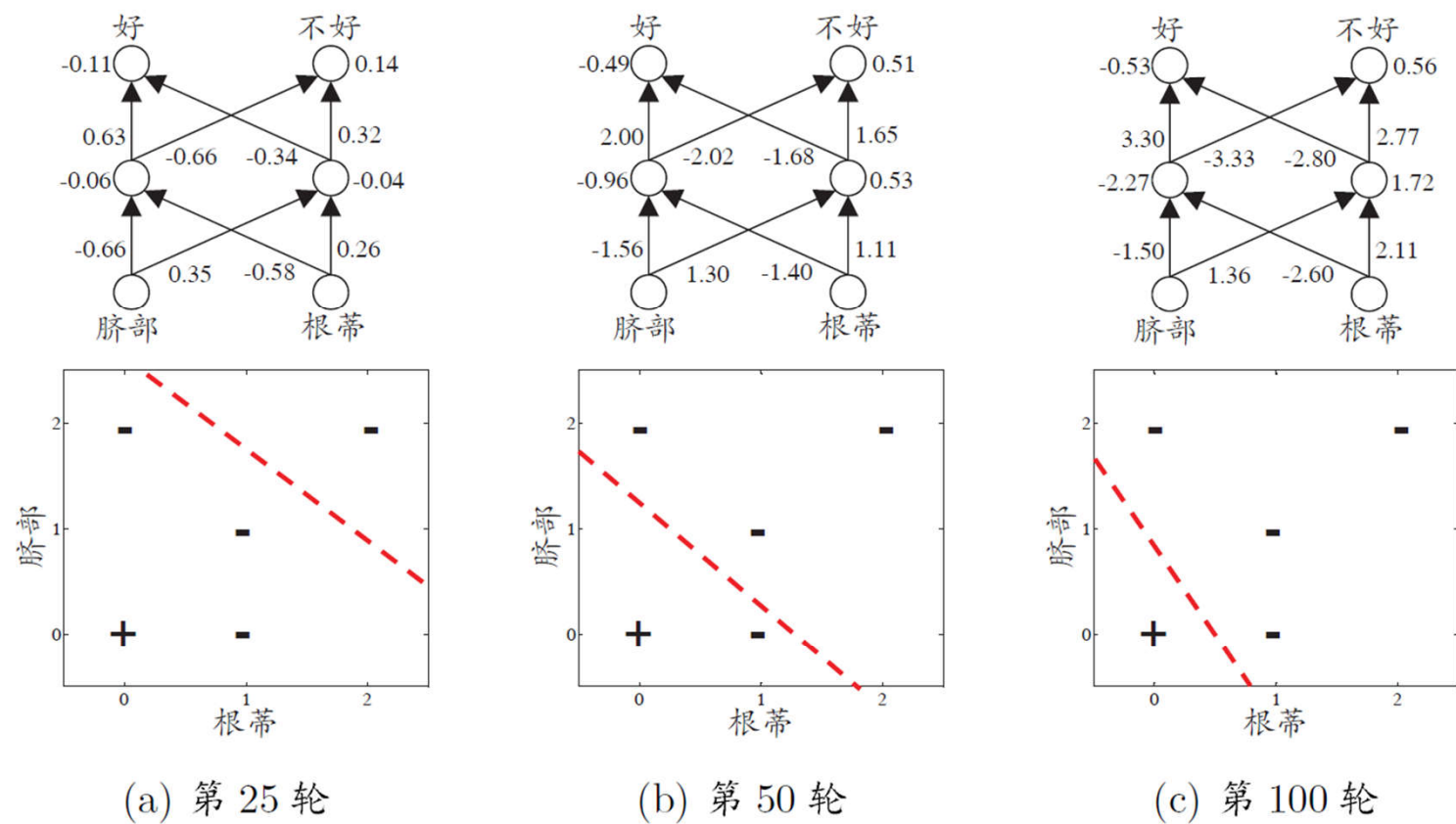


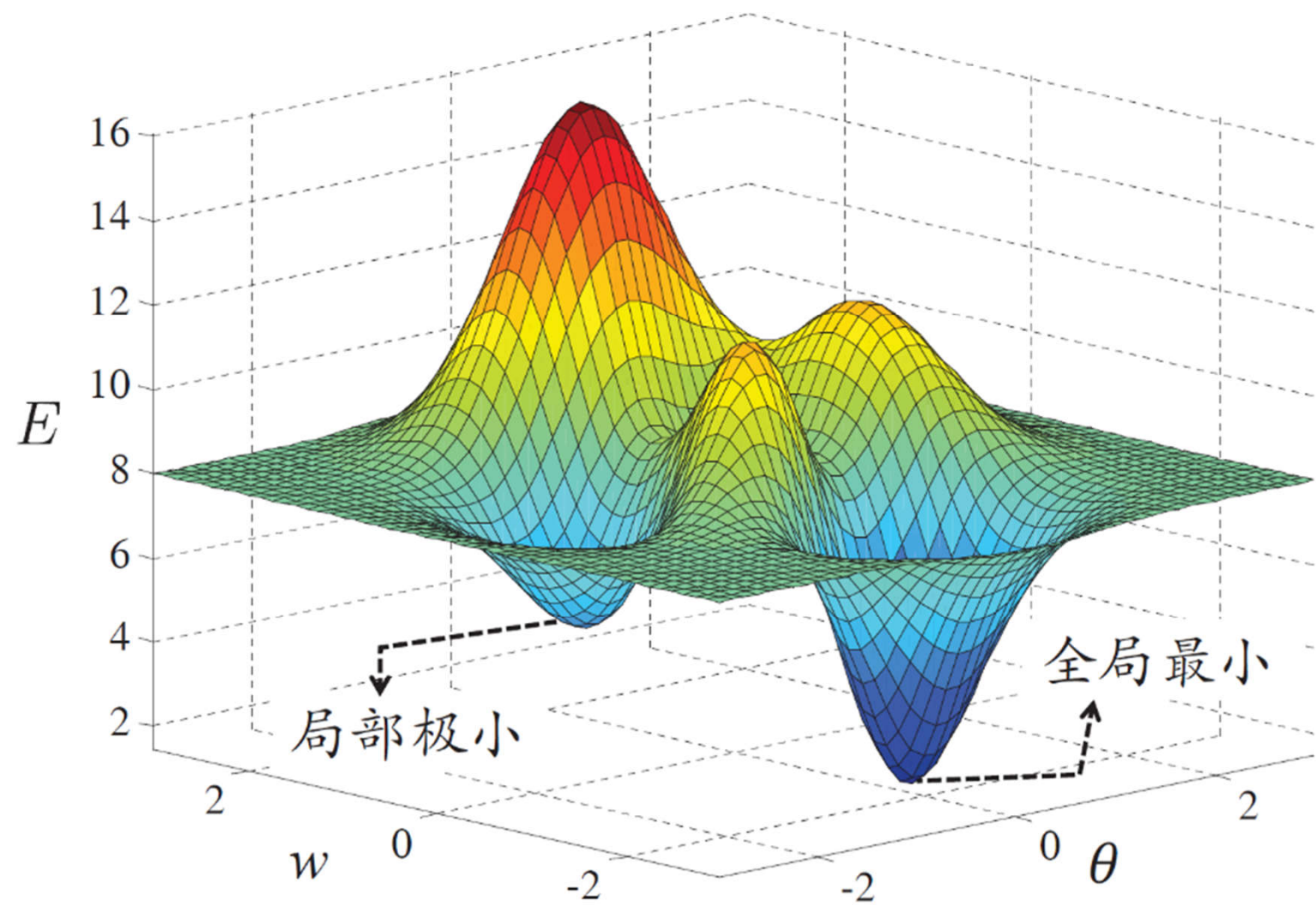
图 5.9 在 2 个属性、5 个样本的西瓜数据上, BP 网络参数更新和分类边界的变化情况

误差反向传播算法

- 梯度下降
- 随机梯度下降
- 小批量梯度下降

全局最小与局部最小

全局最小与局部最小

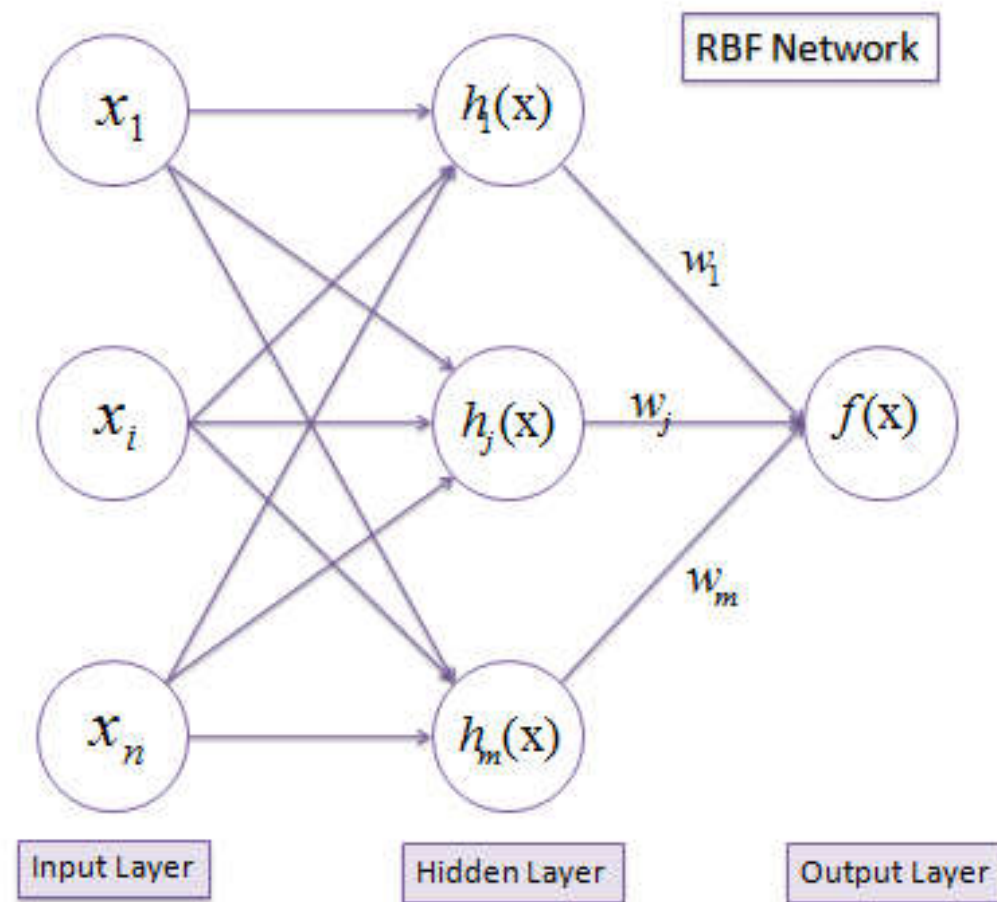


“跳出”局部最小的策略

- 不同的初始参数
- 模拟退火技术 [Aarts and Korst, 1989].
- 随机梯度下降
- 遗传算法 [Goldberg, 1989]

其他常见神经网络

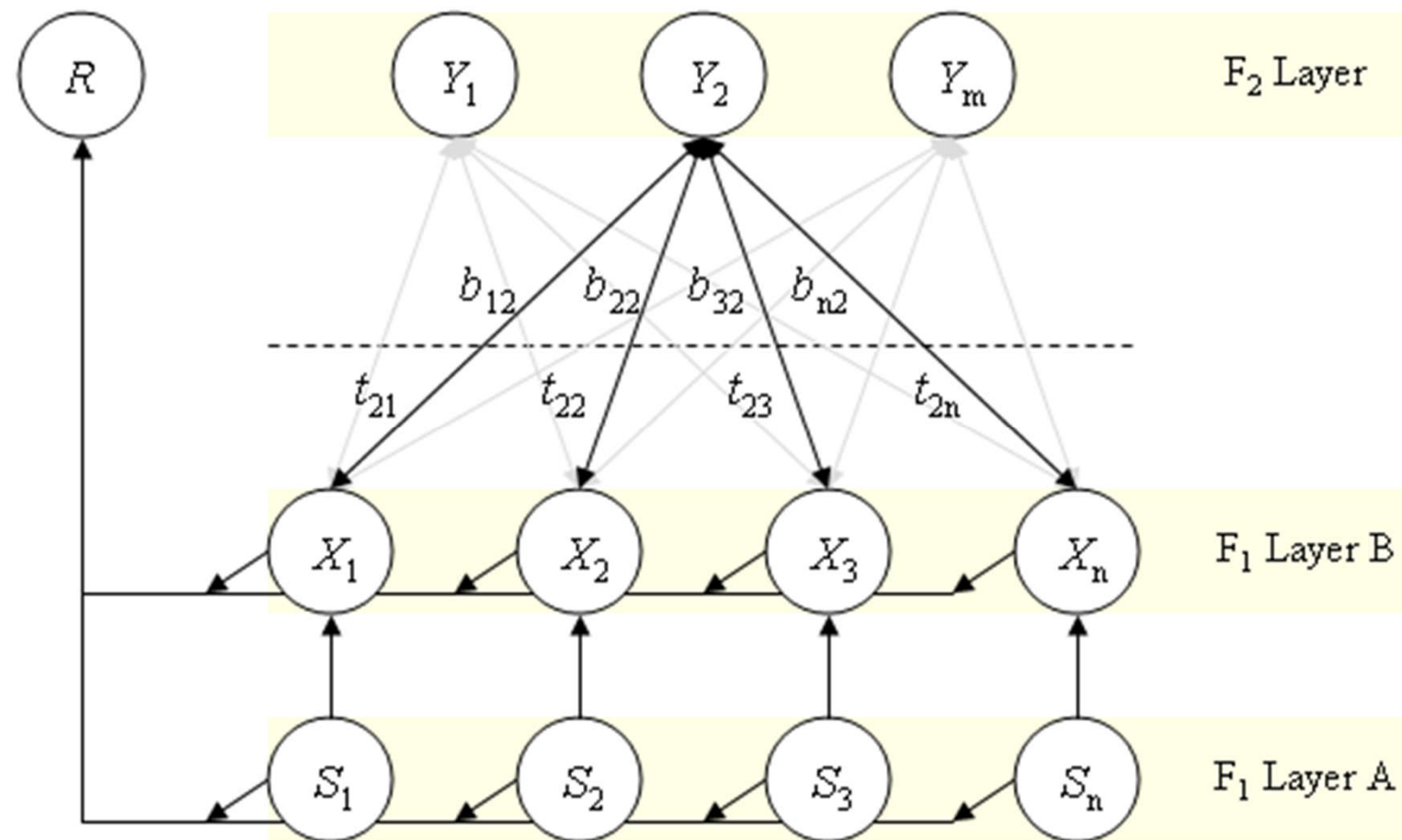
RBF 网络 [Broomhead and Lowe, 1988]



$$f(x) = \sum_{j=1}^m w_j h_j(x)$$

$$h(x) = \exp\left(-\frac{(x-c)^2}{r^2}\right)$$

ART 网络 [Carpenter and Grossberg, 1987]



SOM 网络 [Kohonen, 1982]

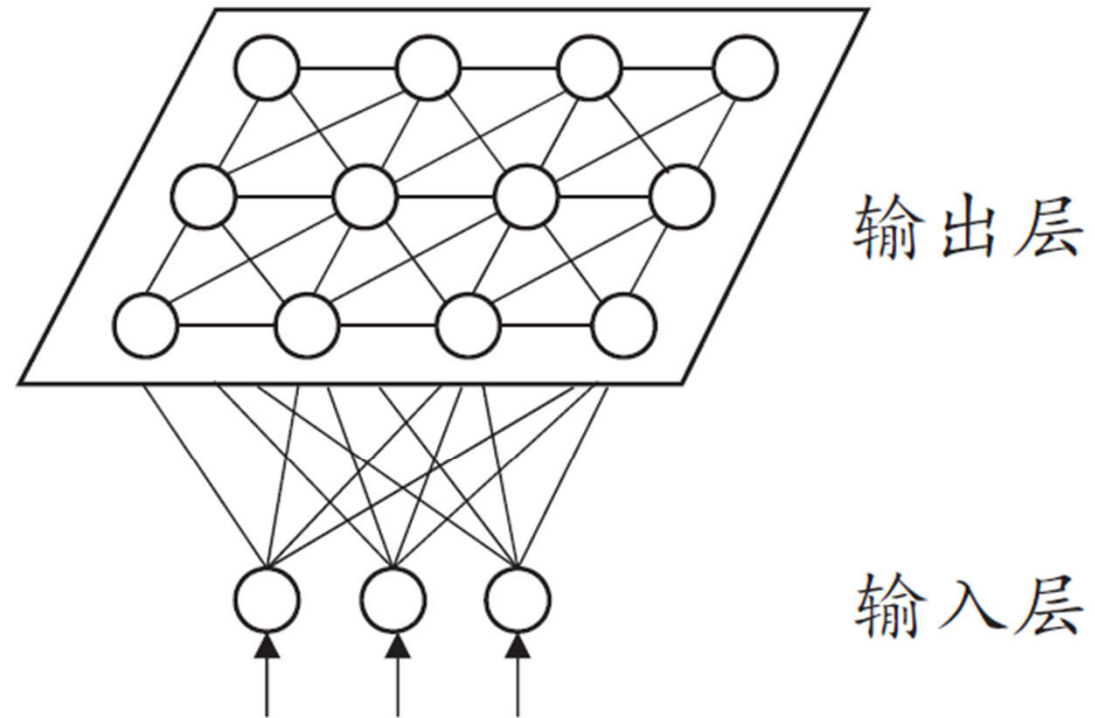
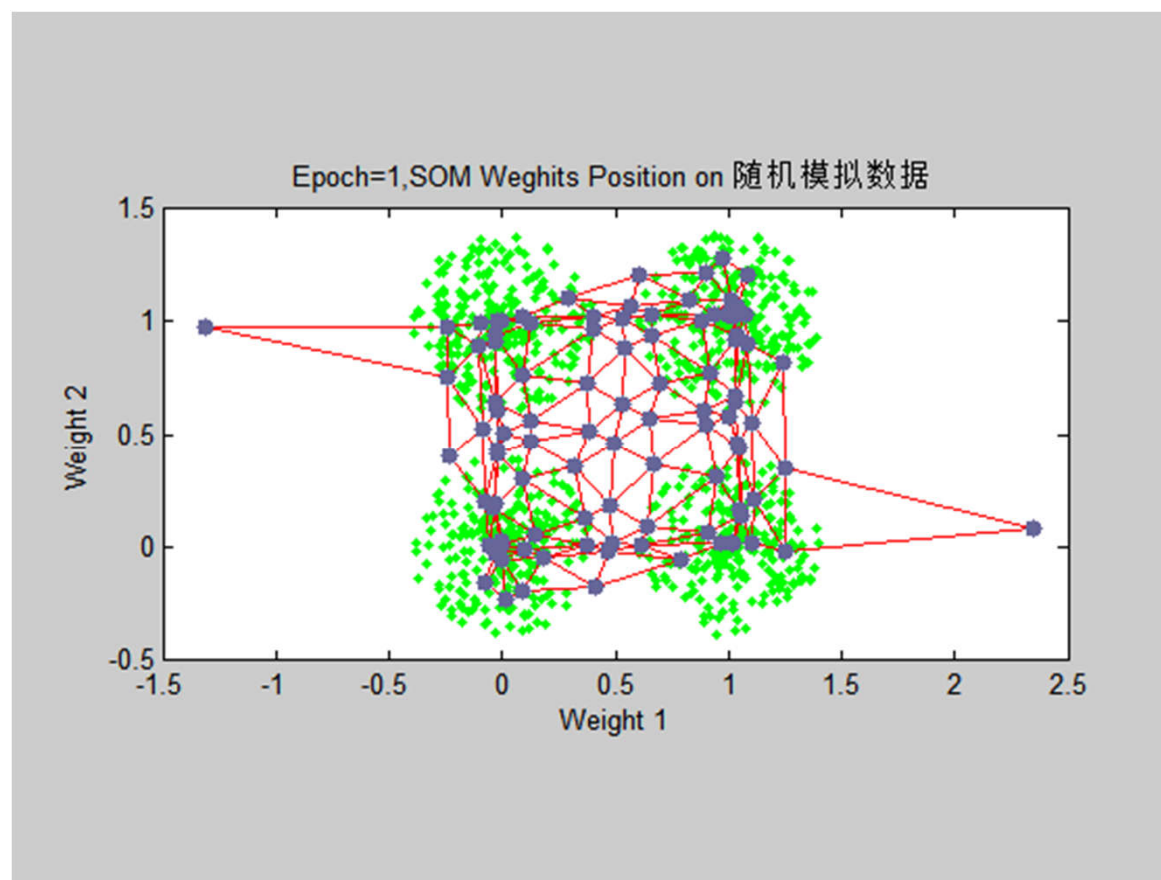


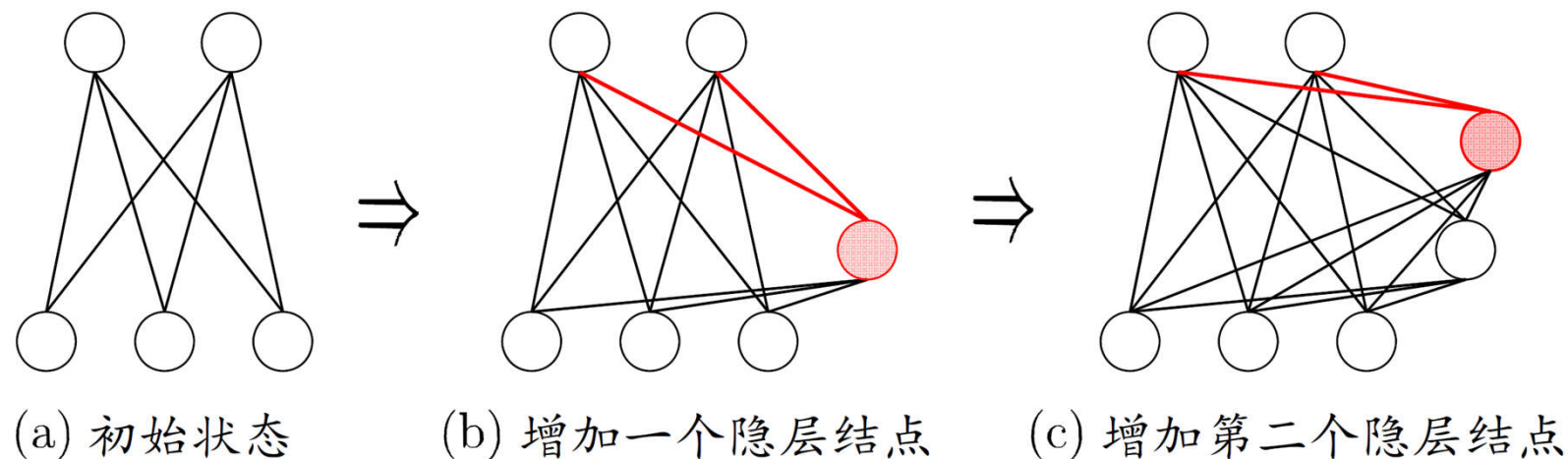
图 5.11 SOM 网络结构

SOM 网络训练



级联相关网络 [Fahlman and Lebiere 1990]

- 级联：建立层次连接的层级结构
- 相关：最大化神经元的输出与网络误差之间的相关性来训练相关参数



Elman 网络 [Elman 1990]

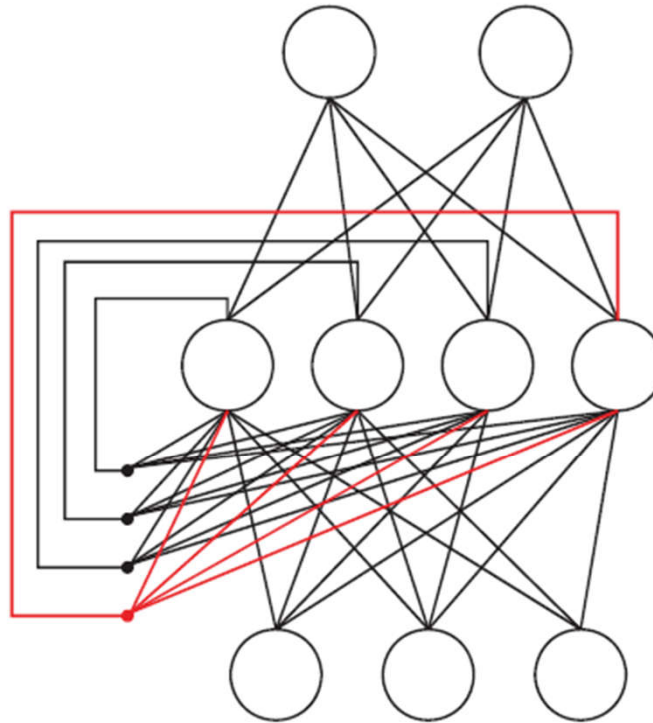
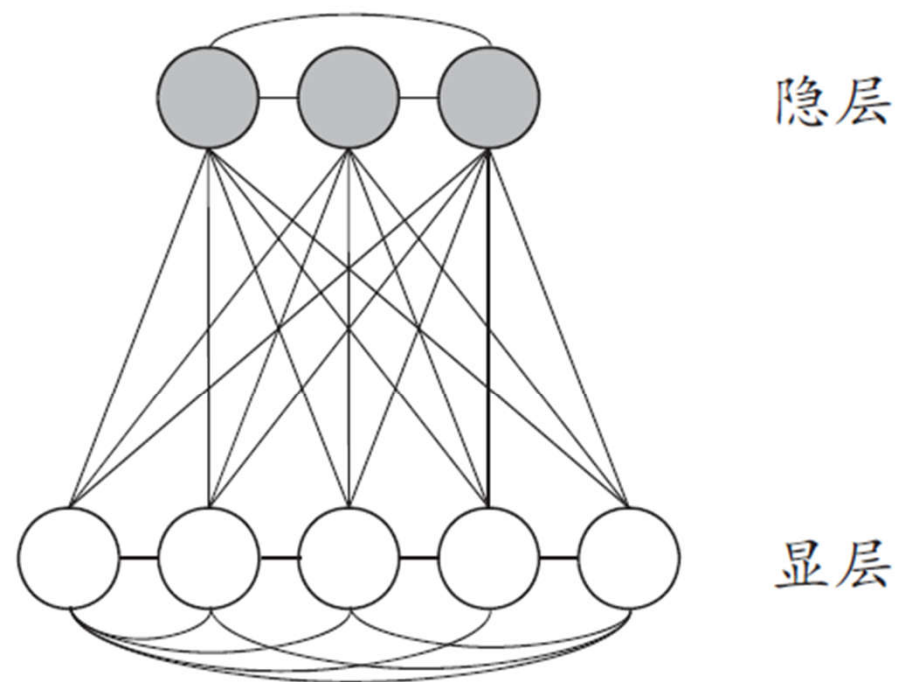


图 5.13 Elman 网络结构

Boltzmann 机



(a) Boltzmann机

受限Boltzmann 机

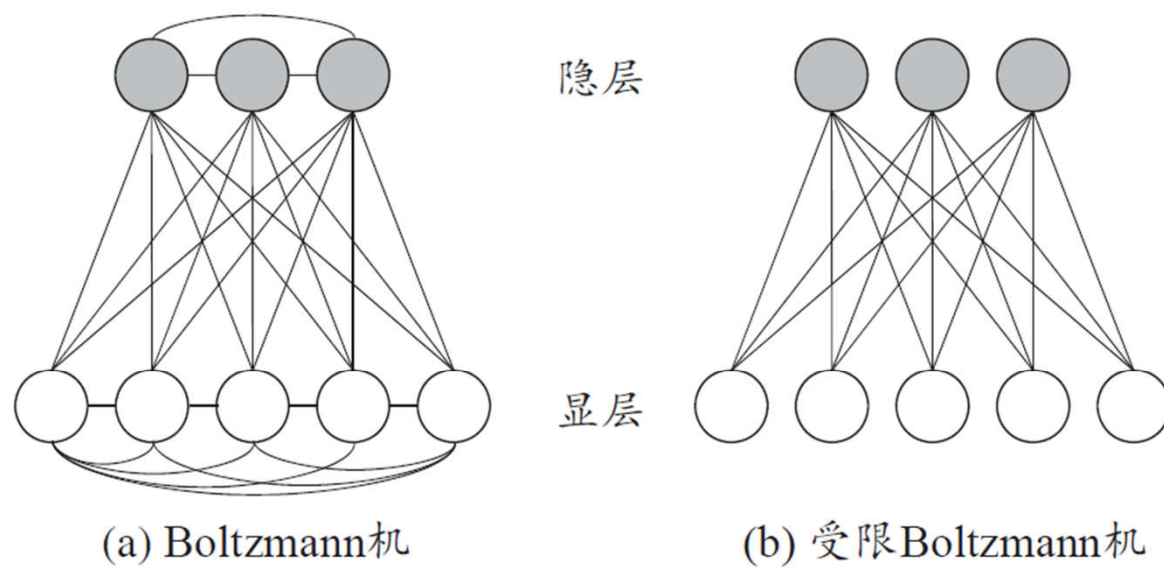
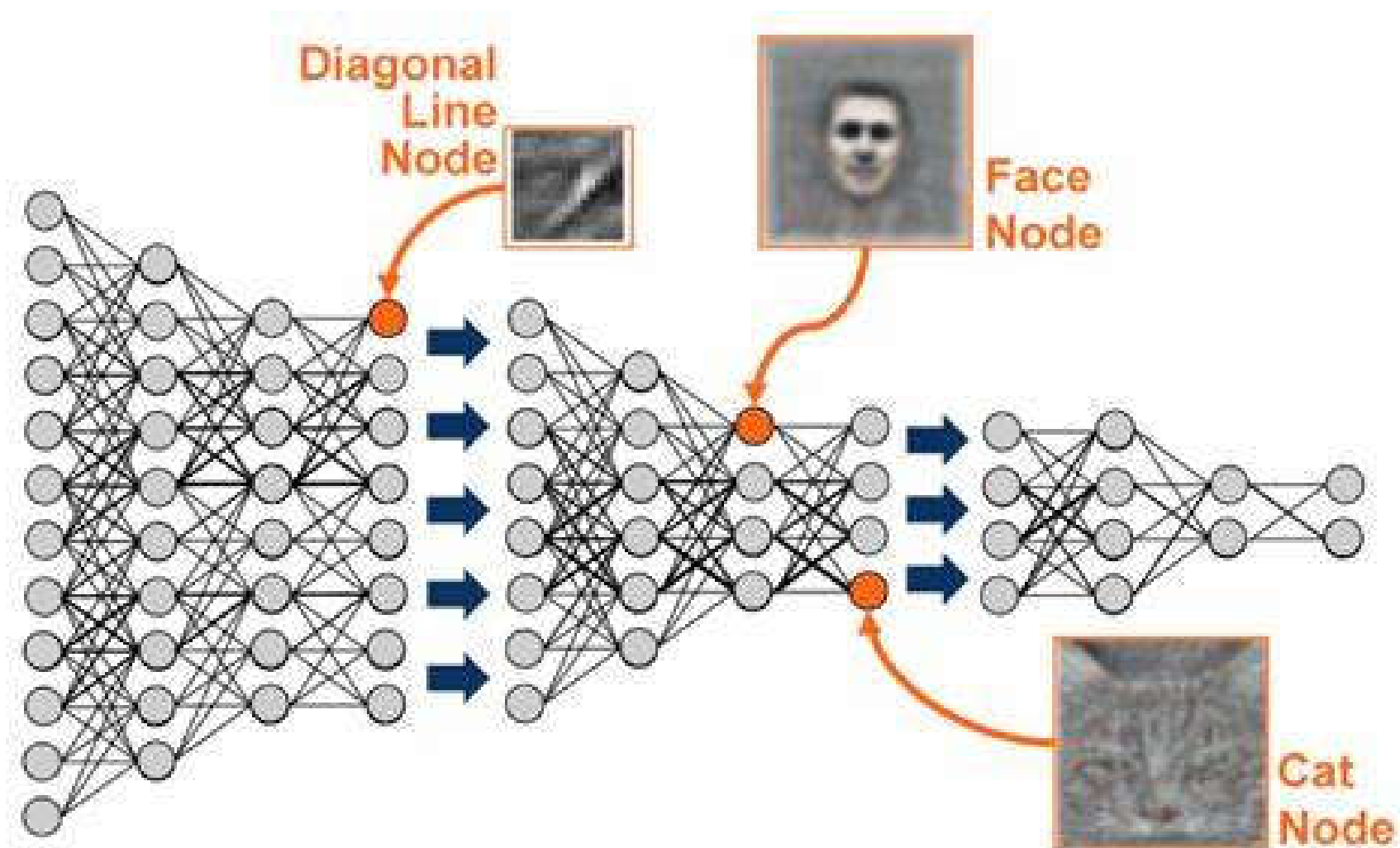


图 5.14 Boltzmann 机与受限 Boltzmann 机

深度学习

深度学习模型



深度学习模型复杂度

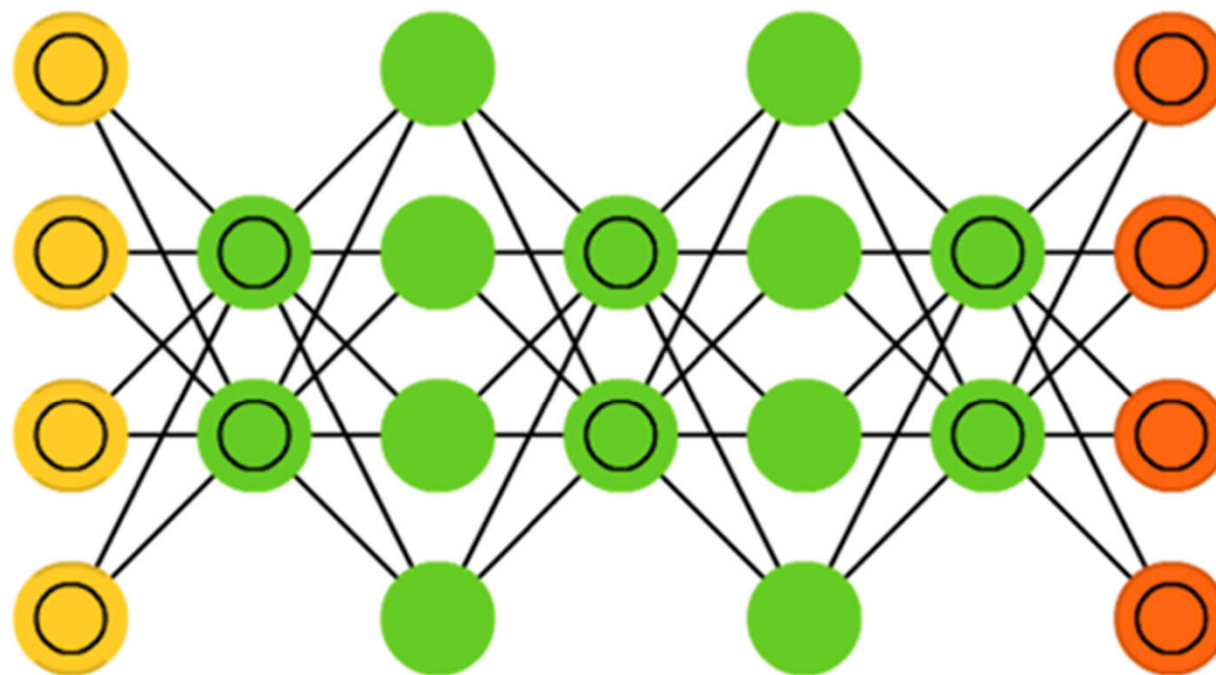
- 增加隐层神经元的数目 (模型宽度)
- 增加隐层数目 (模型深度)

复杂模型训练方法

- **预训练:** 监督逐层训练是多隐层网络训练的有效手段, 每次训练一层隐层结点, 训练时将上一层隐层结点的输出作为输入, 而本层隐结点的输出作为下一层隐结点的输入, 这称为“预训练”.
- **微调:** 在预训练全部完成后, 再对整个网络进行微调训练. 微调一般使用BP算法.

复杂模型训练方法

- 深度信念网络



复杂模型训练方法

- 权值共享

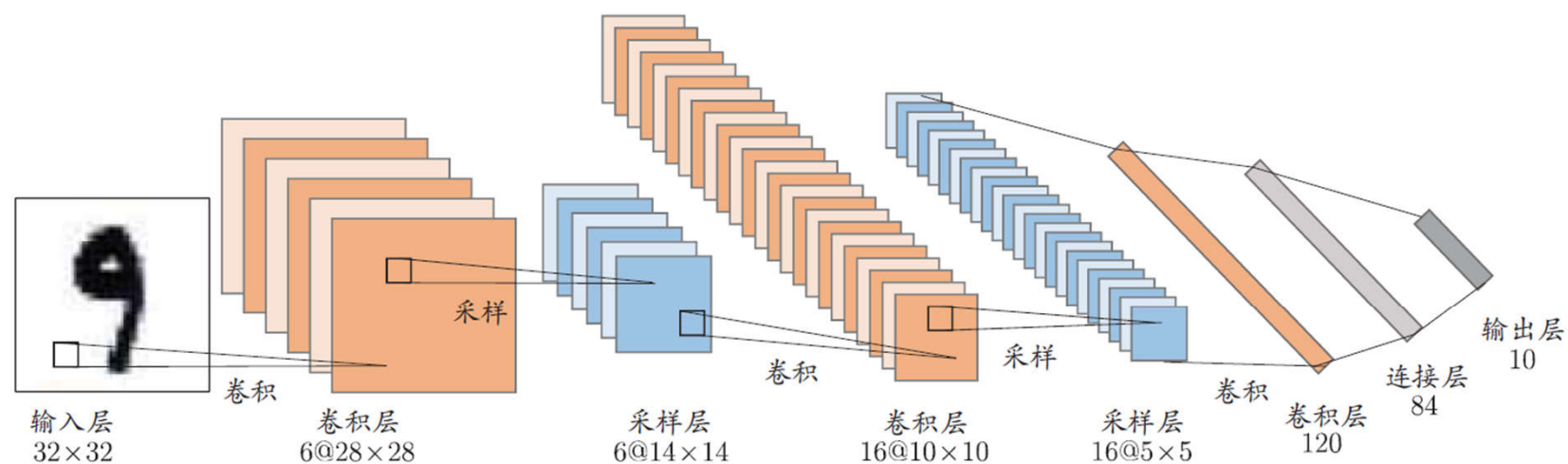


图 5.15 卷积神经网络用于手写数字识别 [LeCun et al., 1998]

深度学习理解

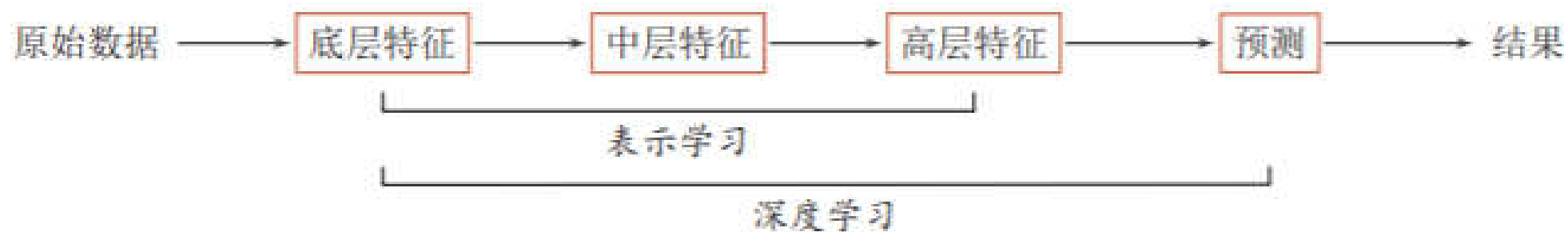


图 1.5 深度学习的数据处理流程

Thanks !