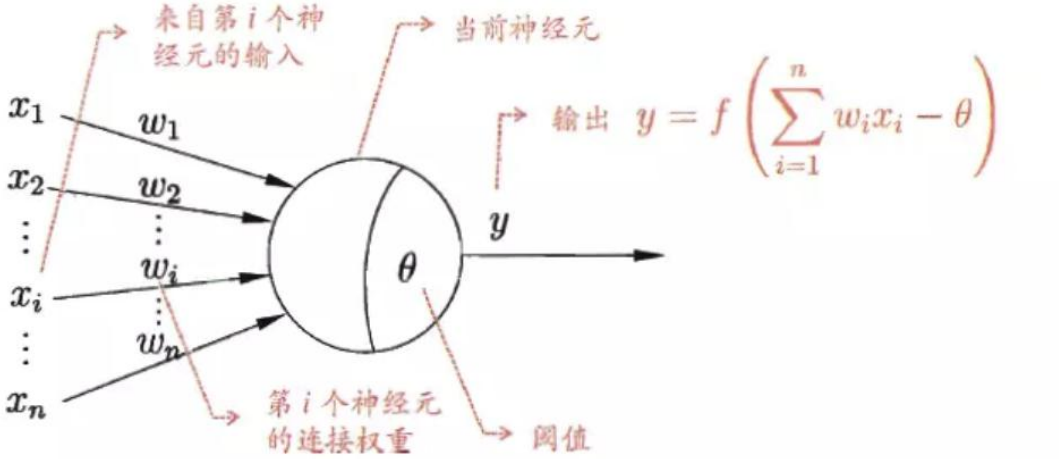
神经网络

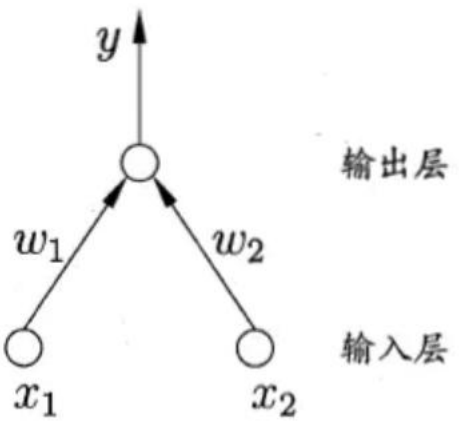
一、神经元模型

神经网络最基本的成分是神经元模型（neuron）。在生物神经网络中，当神经元内的电位超过某个阈值，则会被激活，兴奋起来向其他神经元发送化学物质。
最早将这一神经元模型抽象起来的是 1943 年提出的 M-P 神经元，该神经元将接受 n 个来自其他神经元的信号，并将其带权组合并与阈值比较，并通过激活函数处理。



二、感知机模型与多层神经网络

感知机：两层神经元组成。
输入层：接受外界信号，传递给输出层
输出层：M-P 神经元，亦称“阈值逻辑单元”
可以证明：如果两类模型是线性可分的，那感知机的学习过程一定会收敛

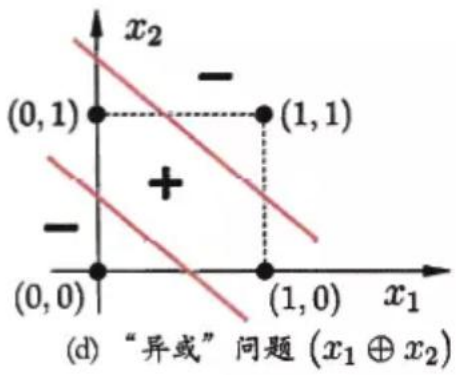


$$w_i \leftarrow w_i + \Delta w_i \ ,$$

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

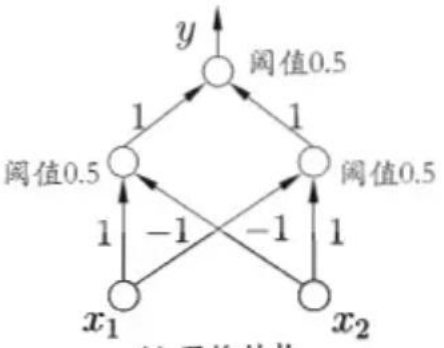
感知机的训练过程：

但是，对于线性不可分的问题，例如：异或问题，感知机模型是无法解决的。

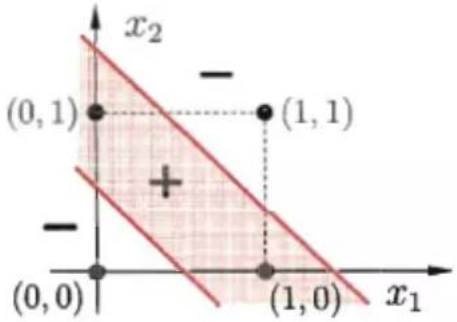


(d) “异或”问题 $(x_1 \oplus x_2)$

解决这种线性不可分问题，我们应该使用多层神经网络。如：

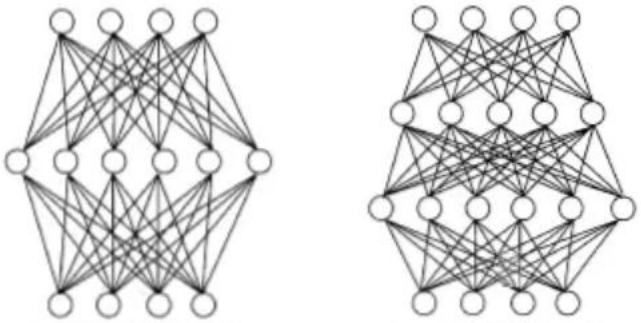


(a) 网络结构



(b) 分类区域

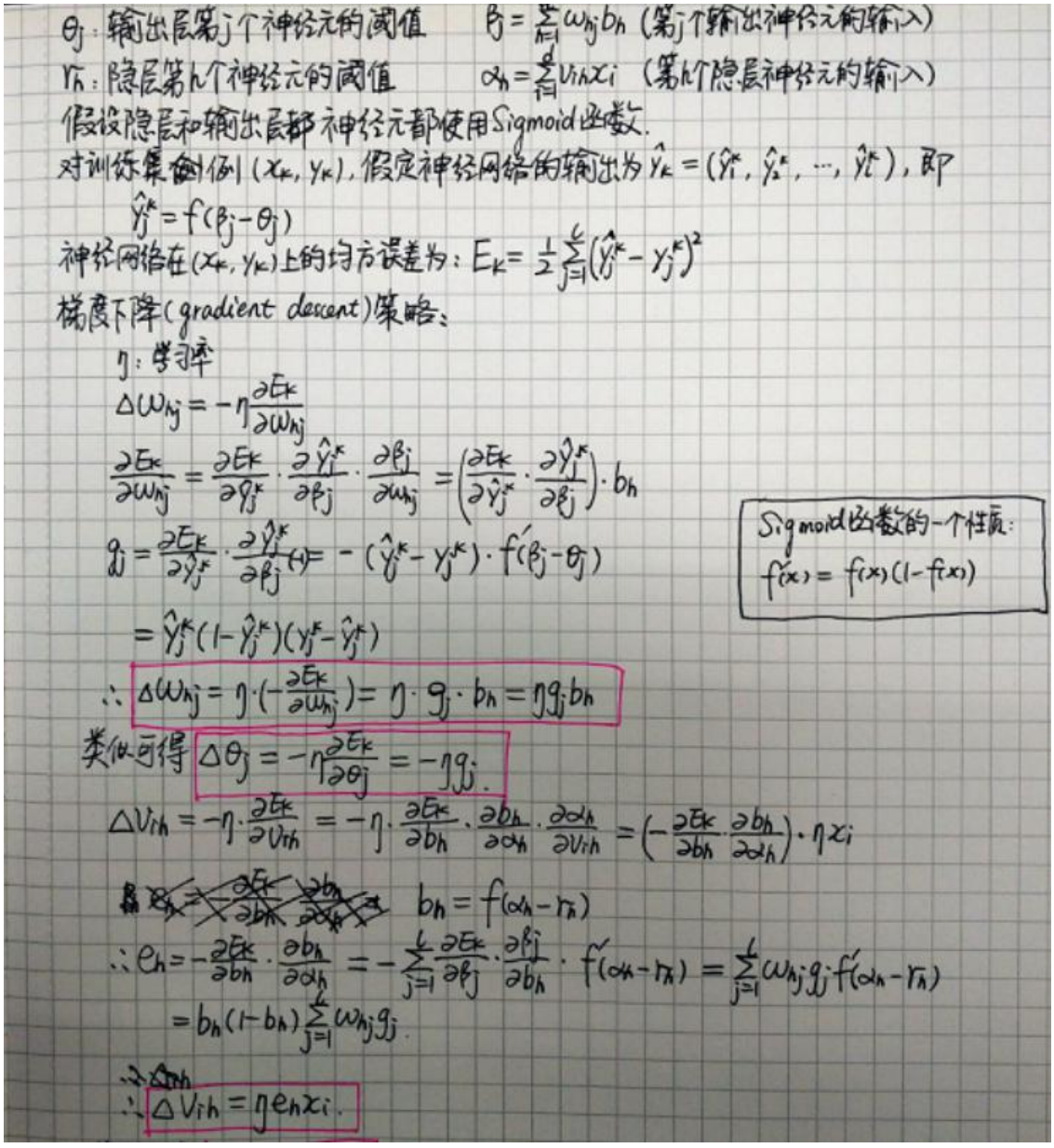
一般的多层神经网络结构如下：


(a) 单隐层前馈网络　　(b) 双隐层前馈网络

三、神经网络训练算法：BP 算法

神经网络的权重和偏置可以通过不断的训练进行优化，优化方法一般采用误差的反向传播，然后进行梯度下降法。

具体推导过程如下：



BP 算法流程如下：

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

过程:

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