



## 第4章 经典人工神经网络



## 五、BP算法的程序实现步骤

- (1)初始化初始化  $V$ 、 $W$ 、 $E_{min}$ 、计数器（样本数，总循环次数）  $q=1, p=1, E=0$
  - (2)输入训练样本对  $X \leftarrow X^p$ 、 $d \leftarrow d^p$ ，计算各层输出；
  - (3)计算网络输出误差；
  - (4)计算各层误差信号；
  - (5)调整各层权值；
  - (6)检查是否对所有样本完成一次训练；
  - (7)检查网络总误差是否达到精度要求，方法不唯一，出口可选最大值、平均值
- 另一种方法是在所有样本输入之后，计算网络的总误差：
- $$E = \frac{1}{2P} \sum_{p=1}^P \sum_{k=1}^L (d_k^p - o_k^p)^2$$
- 用 $E$ 计算各层误差信号→调整权值



## 六、多层感知器主要能力

### ① 非线性映射能力

$$\begin{cases} y_1 = f_1(x_1, x_2, \dots, x_n) \\ y_2 = f_2(x_1, x_2, \dots, x_n) \\ \vdots \\ y_m = f_m(x_1, x_2, \dots, x_n) \end{cases}$$

$$X \xrightarrow[BP]{f} Y$$

$$X^1 \quad Y^1$$

$$X^2 \quad Y^2$$

$$\vdots \quad \vdots$$

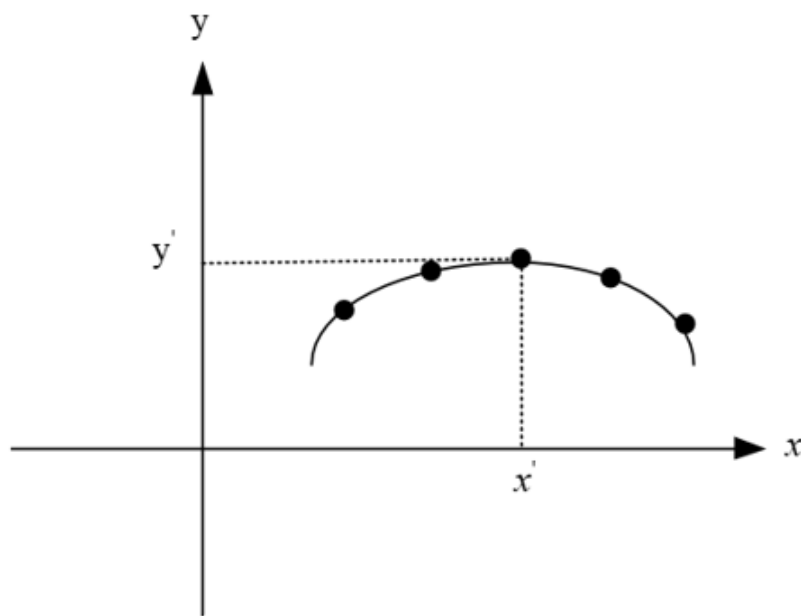
$$X^P \quad Y^P$$

$$X \rightarrow Y$$



## 六、多层感知器主要能力

### ② 泛化能力



### ③ 容错能力

输入样本带有误差或错误，对输出影响很小。

### ④ 分类能力

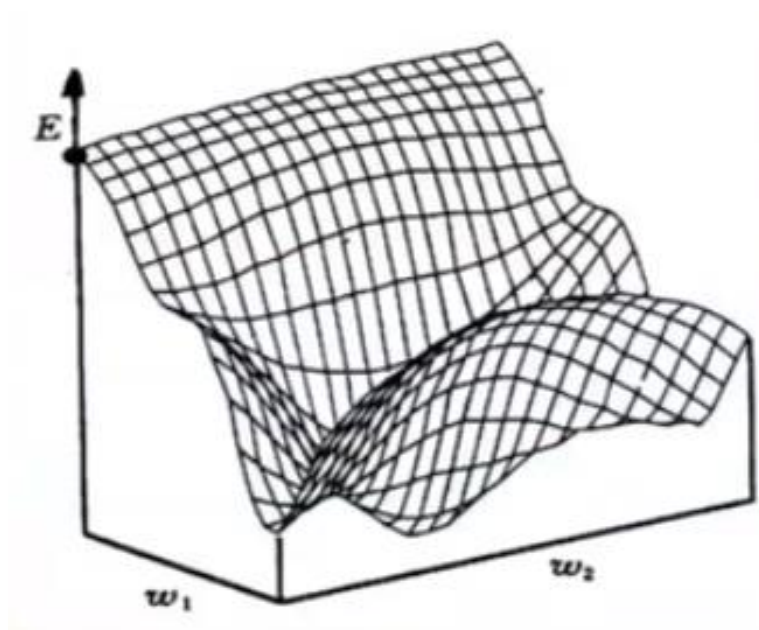
## 七、误差曲面与BP算法的局限

$$E = F(w_{jk}, v_{ij})$$

误差函数的可调整参数的个数  $n_w$  等于各层权值数加上阈值数，即：

$$n_w = m \times (n+1) + l \times (m+1)$$

误差  $E$  是  $n_w+1$  维空间中一个形状极为复杂的曲面，该曲面上的每个点的“高度”对应于一个误差值，每个点的坐标向量对应着  $n_w$  个权值，因此称这样的空间为误差的权空间。





## 七、误差曲面与BP算法的局限

误差曲面的分布有两个特点：

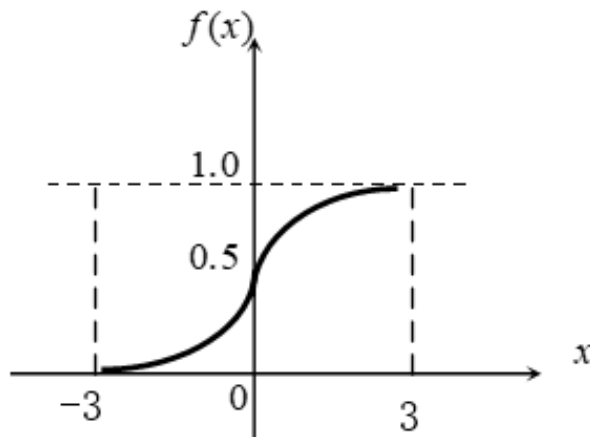
特点之一：存在平坦区域

$$\frac{\partial E}{\partial w_{ik}} = -\delta_k^o y_j \rightarrow 0$$



$$\delta_k^o = (d_k - o_k) o_k (1 - o_k) \rightarrow 0$$

$$\left| \sum_{j=0}^m w_{jk} y_j \right| > 3$$



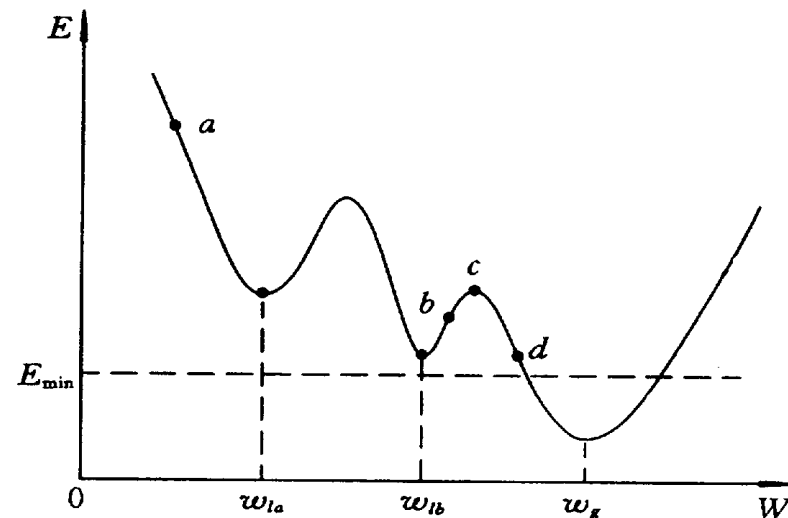


## 七、误差曲面与BP算法的局限

### 特点之二：存在多个极小点

多数极小点都是局部极小，即使是全局极小往往也不是唯一的，但其特点都是误差梯度为零。

误差曲面的平坦区域会使训练次数大大增加，从而影响了收敛速度；而误差曲面的多极小点会使训练陷入局部极小，从而使训练无法收敛于给定误差。



极小值点误差梯度为 0，无法收敛到最小值点。

需总改变初值  
重新初始化



## 八、标准BP算法的改进

标准的BP算法在应用中暴露出不少内在的缺陷：

- (1) 易形成局部极小而得不到全局最优；
- (2) 训练次数多使得学习效率低，收敛速度慢；
- (3) 隐节点的选取缺乏理论指导；
- (4) 训练时学习新样本有遗忘旧样本的趋势。

针对上述问题，国内外已提出不少有效的改进算法，下面仅介绍其中3种较常用的方法。





## 八、标准BP算法的改进

### 1 增加动量项

$$\Delta \mathbf{W}(t) = \eta \delta \mathbf{X} + \alpha \Delta \mathbf{W}(t-1)$$

$\alpha$  为动量系数，一般有  $\alpha \in (0, 1)$

### 2 自适应调节学习率

设一初始学习率，若经过一批次权值调整后使总误差  $\uparrow$ ，则本次调整无效，且  $\eta = \beta \eta$  ( $\beta < 1$ )；

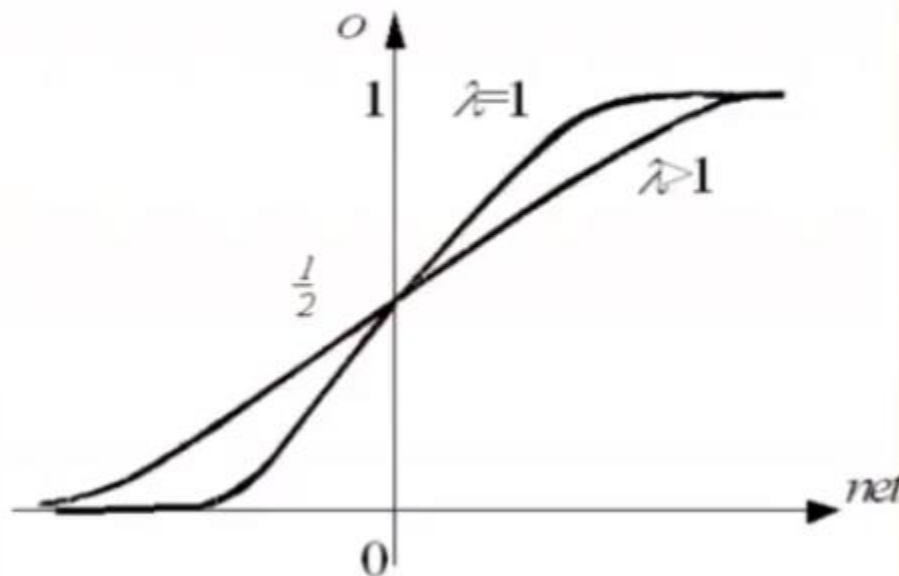
若经过一批次权值调整后使总误差  $\downarrow$ ，则本次调整有效，且  $\eta = \theta \eta$  ( $\theta > 1$ )。

## 八、标准BP算法的改进

### 3 引入陡度因子

实现这一思路的具体作法是，在原转移函数中引入一个陡度因子  $\lambda$

$$o = \frac{1}{1 + e^{-net / \lambda}}$$





# 九、BP网络设计

## 一、训练样本集的准备

### 1. 输入输出量的选择

**输出量**代表系统要实现的功能目标，例如系统的性能指标，分类问题的类别归属，或非线性函数的函数值等等。

**输入量**必须选择那些对输出影响大且能够检测或提取的变量，此外还要求各输入变量之间互不相关或相关性很小，这是输入量选择的两条基本原则。

♠ 从输入、输出量的性质来看，可分为两类：一类是数值变量，一类是语言变量。

♠ **数值变量**的值是数值确定的连续量或离散量。

♠ **语言变量**是用自然语言表示的概念，其“语言值”是用自然语言表示的事物的各种属性。当选用语言变量作为网络的输入或输出变量时，需将其语言值转换为离散的数值量。

## 九、BP网络设计

### 2. 输入量的提取与表示

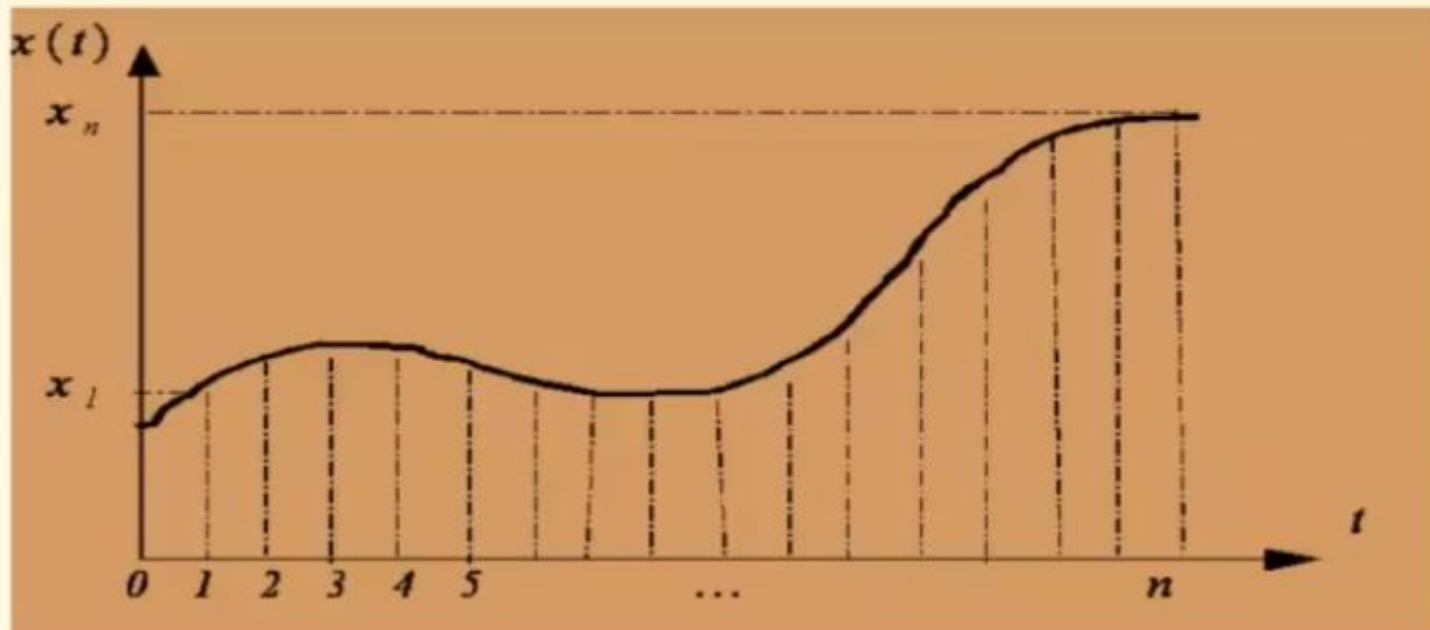
#### (1) 文字符号输入



$$X^C=(111100111)^T \quad X^I=(111100111)^T \quad X^T=(111100111)^T$$

## 九、BP网络设计

### (2) 曲线输入



$$X^p = (x_1^p, x_2^p, \dots, x_i^p, \dots, x_n^p)^T \quad p=1, 2, \dots, P$$

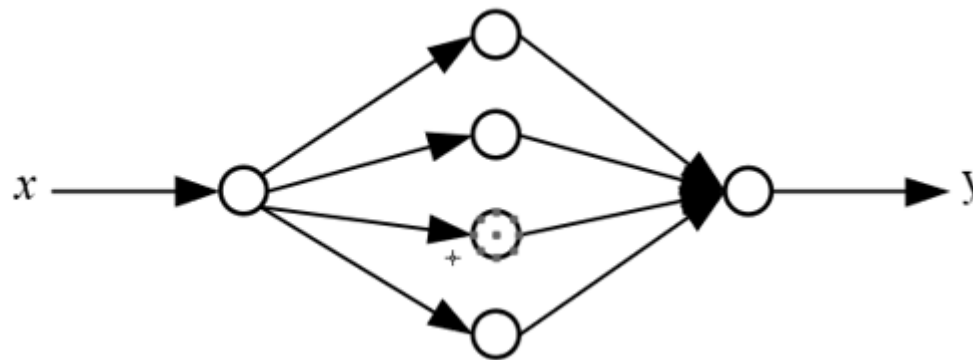


## 九、BP网络设计

### (3) 函数自变量输入

一般有几个输入量就设几个分量，1个输入分量对应1个输入层节点。

$$y=x^2$$





## 九、BP网络设计

### (4) 图像输入

在这类应用中，一般先根据识别的具体目的从图像中提取一些有用的特征参数，再根据这些参数对输入的贡献进行筛选，这种特征提取属于图像处理的范畴。

图像输入：首先进行特征提取，不能只用像素值，不然维数太高。





## 九、BP网络设计

### 3. 输出量的表示

#### (1)“ $n$ 中取1”表示法

“ $n$ 中取1”是令输出向量的分量数等于类别数，输入样本被判为哪一类，对应的输出分量取1，其余  $n-1$  个分量全取0。例如，用**0001**、**0010**、**0100**和**1000**可分别表示优、良、中、差4个类别。

#### (2)“ $n-1$ ”表示法

如果用 $n-1$ 个全为0的输出向量表示某个类别，则可以节省一个输出节点。例如，用**000**、**001**、**010**和**100**也可表示优、良、中、差4个类别。

#### (3)数值表示法

对于渐进式的分类，可以将语言值转化为二值之间的数值表示。数值的选择要注意保持由小到大的渐进关系，并要根据实际意义拉开距离。





# 九、BP网络设计

## 二、输入输出数据的归一化

归一化也称为或标准化，是指通过变换处理将网络的输入、输出数据限制在 $[0, 1]$ 或 $[-1, 1]$ 区间内。

进行归一化的主要原因：

①网络的各个输入数据常常具有不同的物理意义和不同的量纲，归一化给各输入分量以同等重要的地位；

②BP网的神经元均采用Sigmoid转移函数，变换后可防止因净输入的绝对值过大而使神经元输出饱和，继而使权值调整进入误差曲面的平坦区；

③Sigmoid转移函数的输出在 $0 \sim 1$ 或 $-1 \sim 1$ 之间。教师信号如不进行归一化处理，势必使数值大的输出分量绝对误差大，数值小的输出分量绝对误差小。



## 九、BP网络设计

将输入输出数据变换为[0, 1]区间的值常用以下变换式

$$\bar{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

其中， $x_i$ 代表输入或输出数据， $x_{\min}$ 代表数据变化的最小值， $x_{\max}$ 代表数据的最大值。

将输入输出数据变换为[-1, 1]区间的值常用以下变换式

$$x_{\text{mid}} = \frac{x_{\max} + x_{\min}}{2} \quad \bar{x}_i = \frac{x_i - x_{\text{mid}}}{\frac{1}{2}(x_{\max} - x_{\min})}$$

其中， $x_{\text{mid}}$ 代表数据变化范围的中间值。



## 十、网络训练与测试

训练样本集  $\longleftrightarrow$  测试样本集

60%

40%

70%

30%

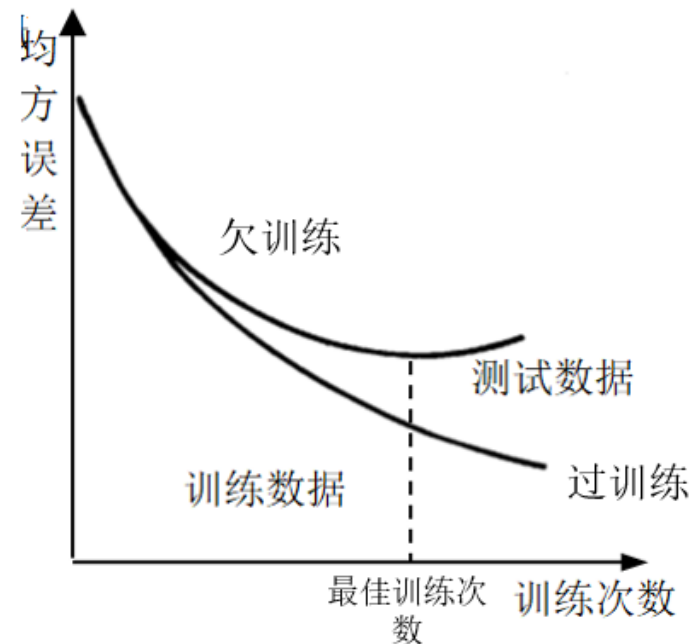
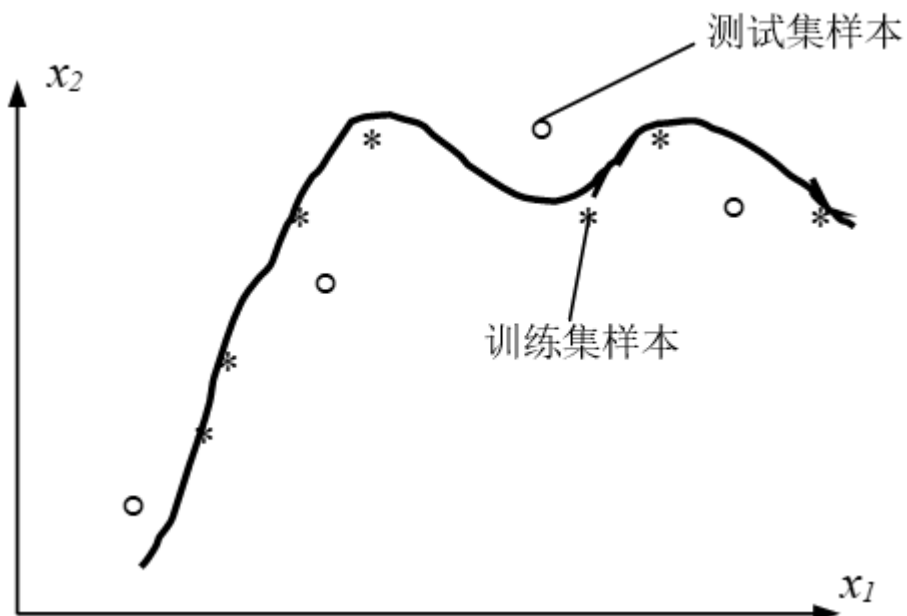
训练出口  $E < \underline{E_{min}}$

思考  $\underline{E_{min}}$  越小约好吗？

核心是泛化能力强



## 十、网络训练与测试



训练精度高并不代表泛化能力强，需要协调。