

多层感知器神经网络

人工神经网络(artificial neural networks, ANN) ,
简称 神经网络(neural networks, NN) ,
多层感知器(multi-layer perceptron,MLP)模型

神经元与感知器

一个神经元(neuron)就是一个神经细胞,它是神经系统的基本组成单位。根据目前的认识,一个典型的神经元由以下几部分组成(如图5-6所示): (1)细胞体(cell body),是神经细胞的主体,内有细胞核和细胞质,除了实现细胞生存的各种基本功能外,这里是神经细胞进行信息加工的主要场所;(2)树突(dendrites),是细胞体外围的大量微小分支,是细胞的“触角”,一个神经元的树突可达 10^3 数量级,多数长度很短,主要担负着从外界(其他细胞或体液环境)接收信息的功能;(3)轴突(axon),是细胞的输出装置,负责把信号传递给另外的神经细胞,通常每个神经元有一个轴突,有的轴突会很长,比如人体四肢的某些神经细胞的轴突可以长达1m以上;(4)突触(synapse),是一个神经元的轴突与另一个神经元的树突相“连接”的部位,这种连接并不是物理上的直接接触,而是二者的细胞膜充分靠近,通过之间的微小缝隙传递带电离子。神经系统中的信号是电化学信号,是靠带电离子在细胞膜内外的浓度差来形成和维持的,这种信号可以以脉冲的形式沿着轴突传播,并经由突触把电荷传递给下一个神经元。突触的不同状态可以影响信号传递的效率,可以称之为突触的连接强度,同时,信号的传递效率也可以受到细胞所在的体液环境中相关离子浓度等的影响。一个神经系统就是由大量神经元组成的,人的神经系统中各种神经元的总数可达 $10^2 \sim 10^4$.神经元之间通过突触连接,构成了复杂的神经网络系统。

仿生学魅力

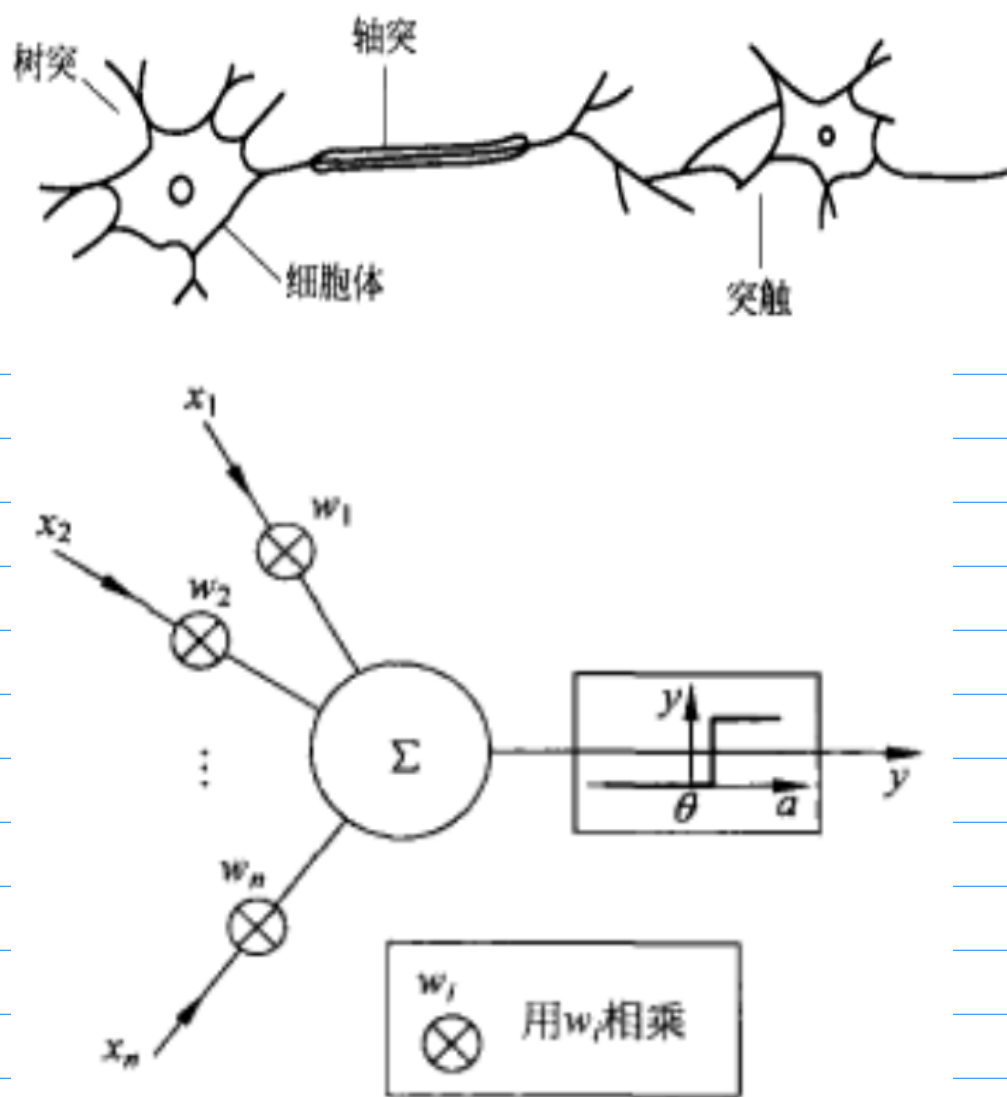


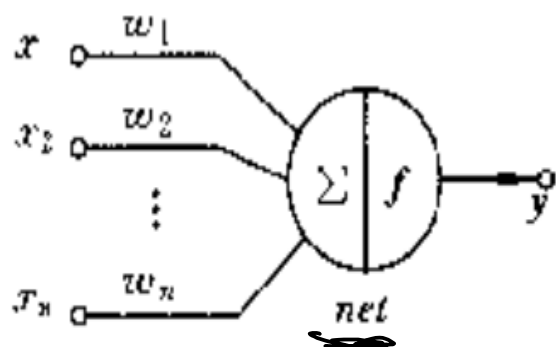
图 5-7 阈值逻辑单元：McCulloch-Pitts 神经元模型

如何来表示neuron功能？

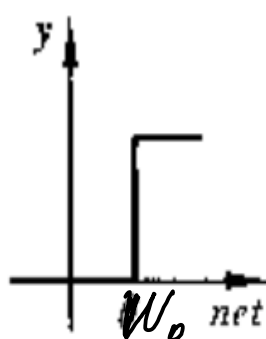
这个模型可以用下面的公式表示

$$y = \theta \left(\sum_{i=1}^n w_i x_i + w_0 \right) \quad (5-14)$$

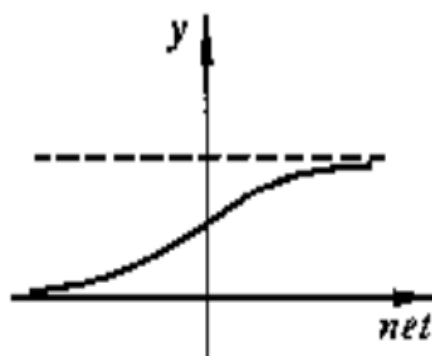
也称作阈值逻辑单元(threshold logic unit, TLU)。其中, $\theta(\cdot)$ 为单位阶跃函数(当自变量为正时函数取值为 1, 否则取值为 0)。在某些情况下, 也可以用符号函数 $\text{sgn}(\cdot)$ 替代 $\theta(\cdot)$, 这时输出 y 的取值就是 1 或 -1。在这个神经元模型中, x 称作神经元的输入, w 称作神经元的权值, $\theta(\cdot)$ 或 $\text{sgn}(\cdot)$ 函数称作神经元的传递函数, y 称作神经元的输出。



(a) 人工神经元模型



(b) 阈值函数



(c) Sigmoid函数

$$net = \sum_{i=1}^n W_i x_i$$

$$y = f(net)$$

若 f 为阈值函数

$$y = \text{sgn} \left(\sum_{i=1}^n W_i x_i - W_0 \right)$$

或写为

$$y = f(W^T X)$$

$$W = [w_0, w_1, \dots, w_n]^T$$

$$X = [1, x_1, \dots, x_n]^T$$

阈值函数不可微, 这会带来一系列问题, 如何解决?

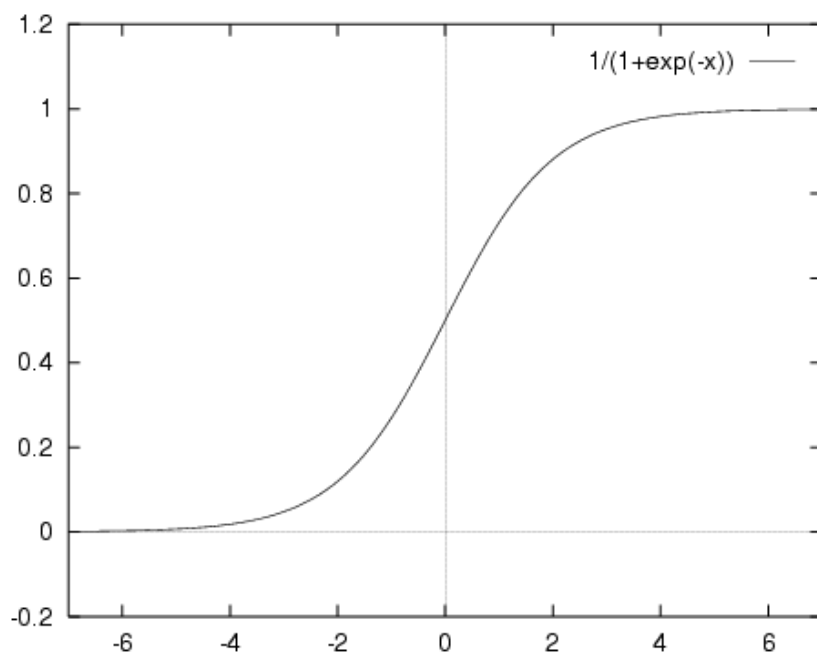
$$f(x) = \frac{1}{1 + e^{-x}}, \text{ sigmoid}$$

sigmoid activation有什么好处?

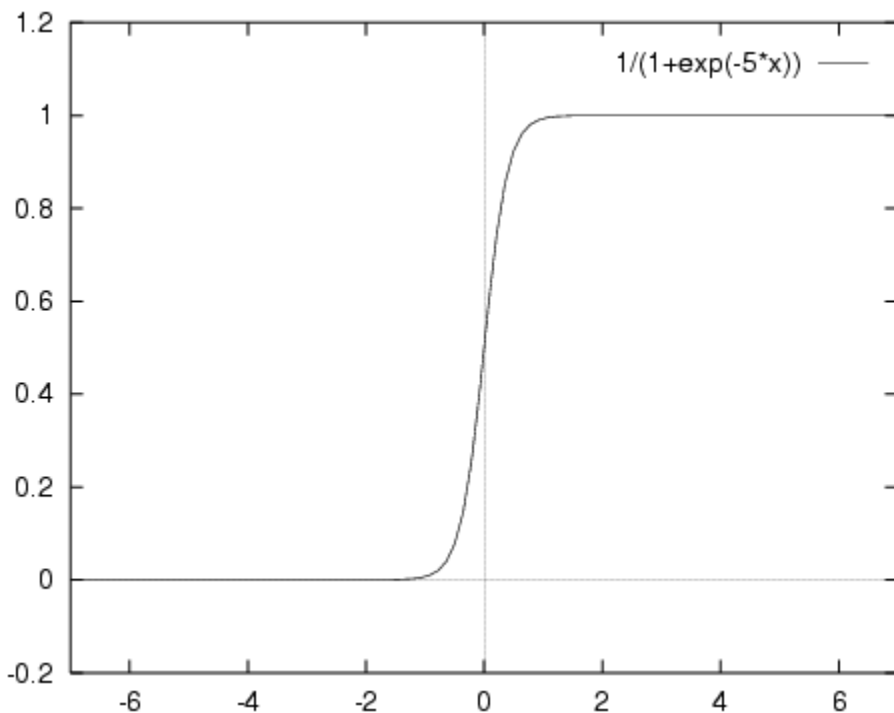
选择Sigmoid函数作为输出函数是由于它具有以下有益的特性: ①非线性, 单调性。②无限次可微。③当权值很大时可近似阈值函数。④当权值很小时可近似线性函数。

淘宝店铺-酷流科技 掌柜: 我是雷锋的朋友

该函数具有如下的特性：当x趋近于负无穷时，y趋近于0；当x趋近于正无穷时，y趋近于1；当x=1/2时，y=0.

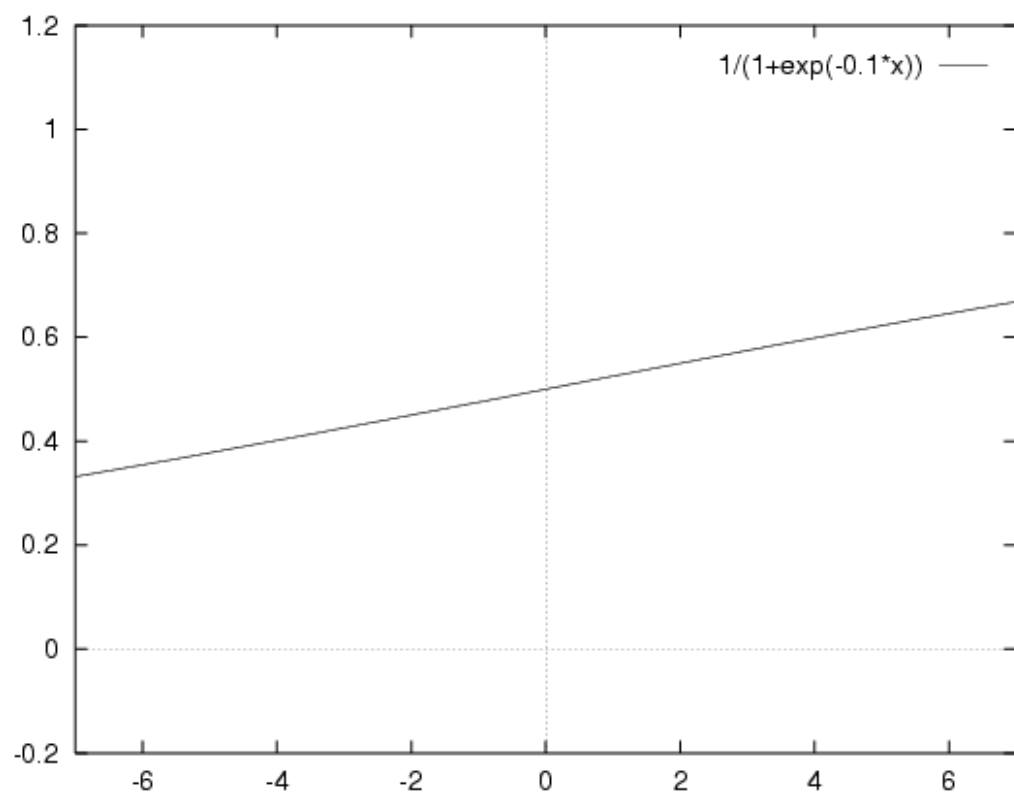


阈值化或者阶梯化：
增加该连接的权值，也就是增加总输入，可以使sigmoid函数越来越趋近于阈值函数，或者叫阶梯函数。
将总输入变为原来的5倍，则sigmoid函数变为如下的形式：



线性化：

只要 w 为非零值，即使 w 非常小， wx 最终都会趋近于 $y=0$ 和 $y=1$ 。下图为 $w=0.1$ ， $-7 < x < 7$ 时的sigmoid函数，看上去 y 是取值范围在0.35到0.65 之间的一条直线。



淘宝店铺-酷流科技 掌柜：我是雷锋的朋友

神经元的学习算法

几乎所有神经网络学习算法都可以看作Hebb学习规则的变形。Hebb学习规则的基本思想是：

如果神经元 μ_j 接收来自另一神经元 μ_i 的输出，则当这两个神经元同时兴奋时 $\mu_i \rightarrow \mu_j$ 权值 W_{ij} 就得到加强。

具体到前述的神经元模型，可以将Hebb规则表现为如下的算法：

$$\Delta W_i = \eta y x_i$$

ΔW_i 是对第 i 个权值的修正量
 η 学习率。

x_i 是 μ_i 的输出 + 兴奋/抑制
 y 是 μ_j 的输出 - 兴奋/抑制

若 x_i 与 y 都兴奋则 ΔW_i 加强。