

# 激活函数

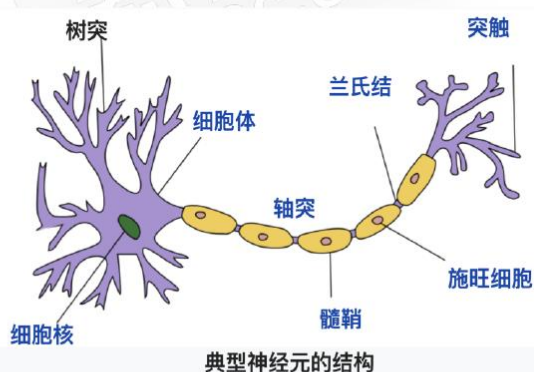
## 目录

- ◆ 什么是激活
- ◆ 常见激活函数

## 什么是激活

## 什么是激活

- ◆ 将输入映射为特定分布的输出，完成非线性变换



树突接收信息，触发区整合电位，产生神经冲动，末端的突触向下一个神经元传递

1、人脑的细胞接受刺激从而产生活动，需要一定的阈值，没有达到阈值，几乎没用。

2、不同的刺激产生的输出也是不同的，达到一定值后就饱和了，再加大也没用。

# 激活函数的简单例子

## ◆ 几种常见的输入输出激活关系

‘打卡’ 激活函数

$\begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$  点



‘加班’ 激活函数

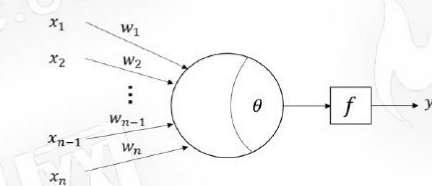
$\begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases}$



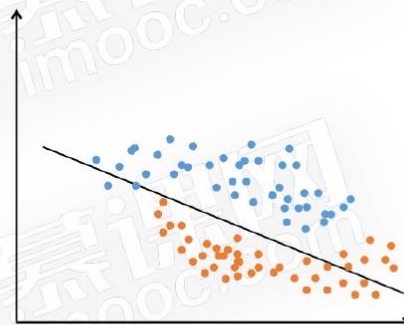
# 网络的表达能力

## ◆ 如果没有非线性激活函数：增加网络层数模型仍然是线性的

$(\quad)$



$(\quad)$   $(\quad)$   $(\quad)$



## 常见激活函数

### S型激活函数

◆ Logistic函数（常被代指Sigmoid函数）与Tanh函数

( ) ——, 0到1的平滑变换

$$\frac{1}{1 + e^{-x}}$$

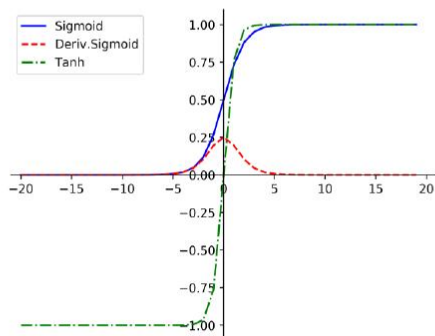
( )

( ) ——

, -1到1的平滑变换

# S型激活函数

## ◆ S型激活函数特点



Sigmoid函数优缺点:

优点: 输出 $0 \sim 1$ , 映射平滑适合预测概率

缺点: 不过零点, 没有负值激活, 影响梯度下降效率; 饱和区**梯度消失问题!**

Tanh函数优缺点:

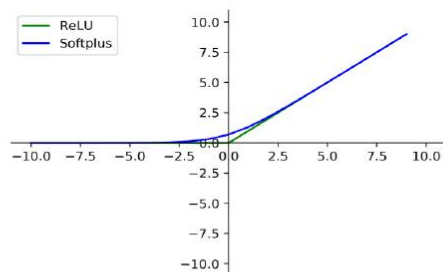
优点: 映射到 $(-1,1)$ 之间, 过零点, 值域比sigmoid更大

缺点: 饱和区**梯度消失问题!**

# ReLU激活函数

## ◆ Rectified Linear Unit, 修正线性单元, 正区间恒等变换, 负区间为0

( )



优点: 计算简单, 导数恒定; 拥有稀疏性, 符合人脑的神经元活跃特性

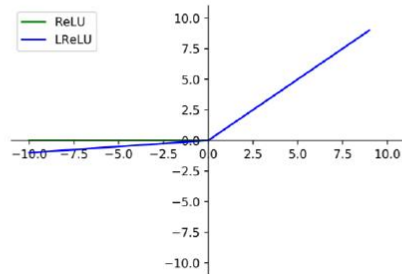
缺点: 非零中心化, 没有负激活值, 影响梯度下降效率; 如果一次不恰当的参数更新后, 所有数据都不能使某个神经元激活, 则其对应的参数梯度为0, 以后也无法激活, 陷入 '死亡' (Dead ReLU)

( ) ( ) ( ) , 的平滑版本

# ReLU激活函数的改进

◆ 统一形式： $\begin{cases} x & x \geq 0 \\ \alpha x & x < 0 \end{cases}$ ：解决ReLU函数"负区间的零激活问题"

$\alpha$ 取固定值则为Leaky ReLU（简称LReLU），为可以学习的参数则为Parametric ReLU（简称PreLU），为随机值则为Random ReLU（简称RReLU）。



LReLU函数：有ReLU的所有优点，并且不会有Dead ReLU问题

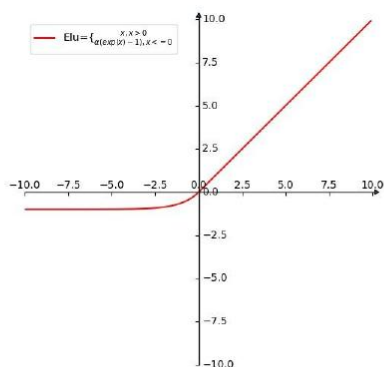
PreLU：网络的浅层尤其是第一层卷积中，学习到的  $\alpha$  会比较大，而到了深层就比较小，更大的  $\alpha$  可以提取到更加稠密的特征。这可以理解网络的浅层学习到了类似于Gabor的浅层特征，随着网络深度的增加，特征变得更加的稀疏。

RReLU：当作一个正则项，用于增强网络的泛化能力。

# ReLU激活函数的改进

◆ Exponential Linear Unit, 简称ELU

$\begin{cases} x & x \geq 0 \\ \alpha(\exp(x) - 1) & x < 0 \end{cases}$ ，将LReLU函数 $x < 0$ 的那一端的函数取非线性函数。

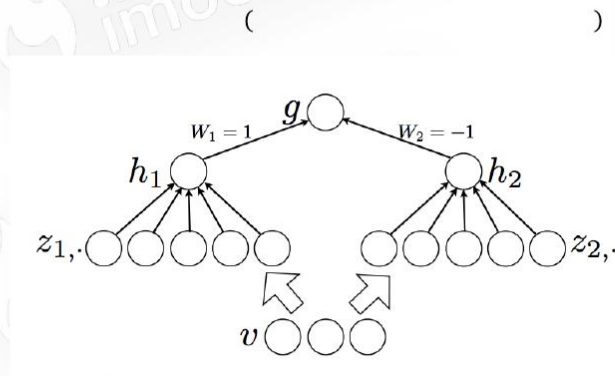


优点：被证实有较高的噪声鲁棒性，能够使得神经元的平均激活均值趋近为0。

缺点：由于需要计算指数，计算量较大。

# Maxout激活函数

◆ 从多个输入中取最大值，求导非常简单，只在最大值的一路有梯度

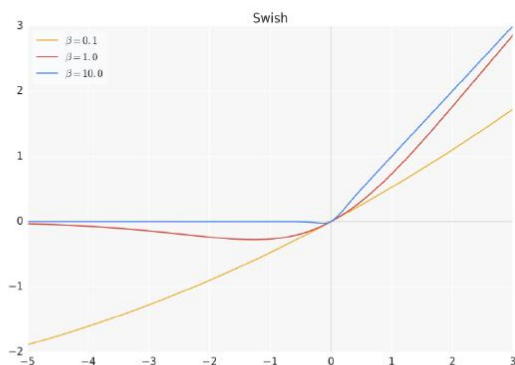


优点：整体学习输入到输出的非线性映射关系，拟合能力非常强

缺点：计算量增加，增加了k-1个神经元

## 自动学习的激活函数代表：Swish函数

◆  $\text{swish}(x) = x \cdot \text{sigmoid}(\beta x)$



- **Unary functions:**  $x, -x, |x|, x^2, x^3, \sqrt{x}, \beta x, x + \beta, \log(|x| + \epsilon), \exp(x) \sin(x), \cos(x), \sinh(x), \cosh(x), \tanh(x), \sinh^{-1}(x), \tan^{-1}(x), \text{sinc}(x), \max(x, 0), \min(x, 0), \sigma(x), \log(1 + \exp(x)), \exp(-x^2), \text{erf}(x), \beta$
- **Binary functions:**  $x_1 + x_2, x_1 \cdot x_2, x_1 - x_2, \frac{x_1}{x_2 + \epsilon}, \max(x_1, x_2), \min(x_1, x_2), \sigma(x_1) \cdot x_2, \exp(-\beta(x_1 - x_2)^2), \exp(-\beta|x_1 - x_2|), \beta x_1 + (1 - \beta)x_2$

介于线性函数与ReLU函数之间，无上界有下界、平滑非单调，从数据中学习参数 $\beta$ ，可以获得任务相关的激活机制



**下次预告：参数初始化**