# 神经网络激活函数汇总

## # 激活函数的性质

1. 连续并可导(允许少数点上不可导)的非线性函数

   可导的激活函数可以直接利用数值优化的方法来学习网络参数.

2. 激活函数及其导函数要尽可能的简单

   有利于提高网络计算效率.(如果单个神经元就很复杂,那整体会太复杂)

3. 激活函数的导函数的值域要在一个合适的区间

   不能太大也不能太小,否则会影响训练的效率和稳定性.

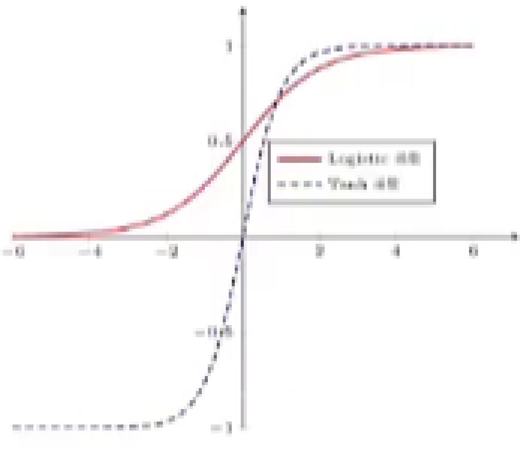
4. ~~单调递增~~

   以前一般需要这样,但是现在不需要了.

## # 常用激活函数

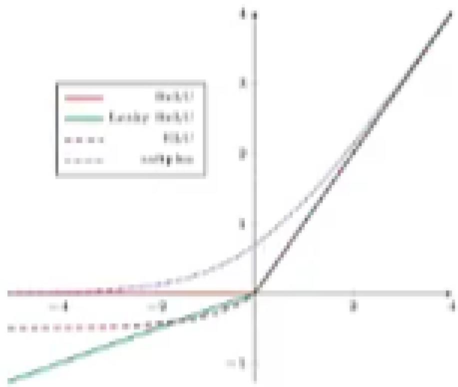
1. S型函数:

   这一类函数叫sigmoid函数,其形状为S型.



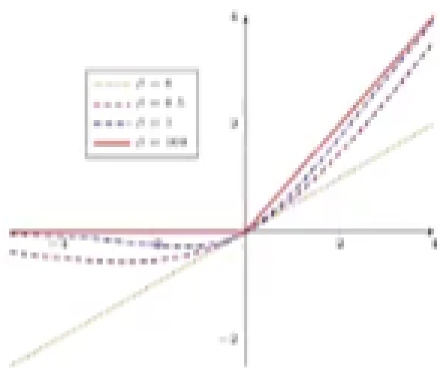
2. 斜坡函数:

   左边一个水平的或者接近水平的形状,右边的斜率基本上是为1的函数.



3. 复合函数:

   既带有斜坡函数的性质,又带有S性函数的性质.

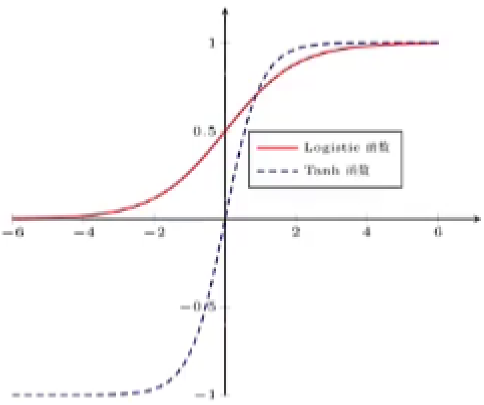


# 常见的S型激活函数

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

$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

如图所示




此外,有$tanh(x) = 2\sigma(2x) - 1$成立.

## 性质

1. 饱和函数 :两端的梯度都是趋近于0.
2. Tanh函数是0中心化的,logistics函数的输出恒大于0.

# # 常见的斜坡函数

1. ReLU函数

   $ReLU = \max(0, x)$

## 性质

- 计算上更加高效
- 生物学合理性:

单侧抑制,宽兴奋边界

- 在一定程度上缓解梯度消失问题

## 问题

死亡relu问题:relu函数的左边,当$x < 0$时,输出则为0,当模型初始化不好时,反向梯度为0,则更新不了.

**2.** LeakyReLU函数

$$
\begin{aligned}
LeakyReLU &= \begin{cases} x\ , if\ x > 0 \\ \gamma x, if x \le 0 \end{cases} \\
&= \max(0, x) + \gamma \min(0, x)
\end{aligned}
$$

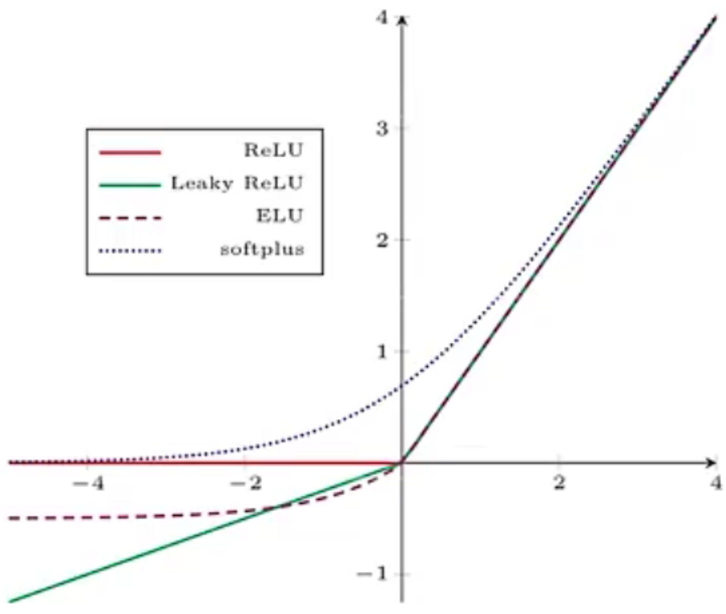
但是这个也是非零中心化函数,也会有问题

**3.** ELU函数

$$
\begin{aligned}
ELU(x) &= \begin{cases} x & if\ x > 0 \\ \gamma(e^x - 1), & if\ x \le 0 \end{cases} \\
&= \max(0, x) + \min(0, \gamma(e^x - 1))
\end{aligned}
$$

**4.** softplus函数

$$softplus(x) = log(1 + e^x)$$



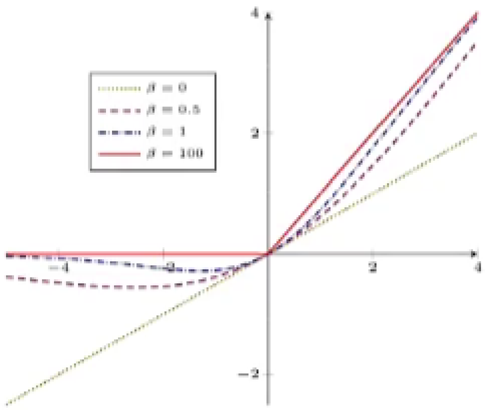

图形如上图所示.

# 常见的复合函数

1. Swish函数:一种自门控(self-Gated)激活函数

   $Swish(x) = x\sigma(\beta x)$

   用$\sigma(\beta x)$控制输出,其中$\beta$控制输出量.如图所示(不同$\beta$值对应的函数曲线)




2. 高斯误差线性单元(Gaussian Error Linear unit, GELU)

   和Swish函数比较类似: $GELU(x) = xP(X \leq x)$

- 其中$P(X \leq x)$是高斯分布$N(0, \sigma^2)$的累计分布函数
- $\mu, \sigma$为超参数,一般设$\mu = 0, \sigma = 1$

  由于高斯分布的累计分布函数为S型函数,因此GELU可以用Tanh函数或logistics函数来近似

  - $GELU(x) \approx 0.5x(1 + tanh(\sqrt{\frac{2}{\pi}}(x + 0.044715x^3)))$
  - $GELU(x) \approx x\sigma(1.702x)$

# 小结

| 激活函数 | 函数 | 导数 |
| --- | --- | --- |
| logistics函数 | $f(x)=\frac{1}{1+e^{-x}}$ | $f'(x)=f(x)(1-f(x))$ |
| Tanh函数 | $f(x)=\frac{e^x-e^{-x}}{e^x+e^{-x}}$ | $f'(x)=1-f(x)^2$ |
| ReLU函数 | $f(x)=\max(0,x)$ | $f'(x)=I(x>0)$ |
| ELU函数 | $f(x)=\max(0,x)+min(0,\gamma(e^x-1))$ | $f'(x)=I(x>0)+I(x\leq 0)\cdot\gamma e^x$ |
| Softplus函数 | $f(x)=log(1+e^x)$ | $f'(x)=\frac{1}{1+e^-x}$ |