

# 如何选择激活函数；关于激活函数，你想知道的都在这

作者: [lai-bluejay](#)

原文链接: <https://ld246.com/article/1555560699441>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

# 激活函数

对于神经网络，网络的每一层计算可以理解为 $f(wx+b)=f(w \cdot x)$ ，每一层可以认为是每个权重矩阵 $W$ 输入特征矩阵 $X$ 。根据矩阵乘法，如果是线性激励，即激活函数为 $f(x) = x$ ，多层网络相当于一层网络比如： $f(W_1 * f(W_2 x)) = W_1 W_2 x = Wx$ 。

因此，需要引入非线性激活函数，对特征进行充分组合。也可以说相当于进行空间变换，使得原来线不可解的问题有非线性的表达，找到非线性的解。

## 好的激活函数的特性

1. 非线性，保证不会退回线性网络。
2. 几乎处处可微：可微性保证了在优化中梯度的可计算性。（有限不可微的点影响不大）
3. 计算简单/计算量小：
4. 非饱和性 (no-saturation): 饱和指的是在某些区间梯度接近于零（即梯度消失），使得参数无法续更新的问题。ReLU在 $x < 0$ 的情况下，梯度为0，也饱和，dying ReLU
5. 单调性：导数符号不变
6. 输出范围有限：有限的输出范围使得网络对于一些比较大的输入也会比较稳定（但可能会导致梯度失，或者限制神经元表达能力）
7. 接近恒等变换 (identity)：即约等于 $x$ 。这样的好处是使得输出的幅值不会随着深度的增加而发生著的增加，从而使网络更为稳定，同时梯度也能够更容易地回传。
8. 参数少：Maxout，尽管本身没有参数，但在同样输出通道数下 $k$ 路Maxout需要的输入通道数是其函数的 $k$ 倍，这意味着神经元数目也需要变为 $k$ 倍；但如果不考虑维持输出通道数的情况下，该激活函数又能将参数个数减少为原来的 $k$ 倍。
9. 归一化 (normalization)：这个是最近才出来的概念，对应的激活函数是SELU[8]，主要思想是使本分布自动归一化到零均值、单位方差的分布，从而稳定训练。在这之前，这种归一化的思想也被用网络结构的设计，比如Batch Normalization。

我们来看看，哪些激活函数能满足这些特性。

## 可视化

从下面的菜单中选择激活函数，以绘制它及其一阶导数。右侧的框中提供了与神经网络相关的一些属。

[可视化详细页面](#)