

激活函数

<https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0>

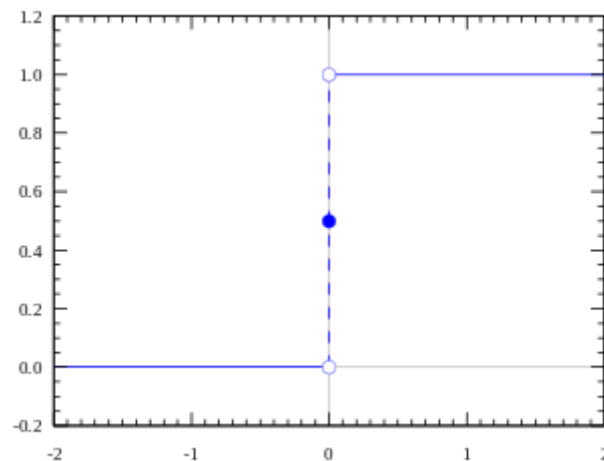
1. 激活函数

一个神经元，输入值带有权重的累加计算，信息是否继续向下传导？

$$Y = \sum (weight * input) + bias$$

神经元的输入可以从负无穷到正无穷，神经元是否传导信息呢（激活函数决定）。

2. 阶跃函数



Its output is 1 (activated) when value > 0 (threshold) and outputs a 0 (not activated) otherwise.

神经元大于阈值时输出1，否则输出0，起到了激活的目的。

缺点

- “yes” or “no” 阶跃函数只能处理二分类问题；
- 多分类问题需要 % 模拟量

3. 线性函数 $y=cx$

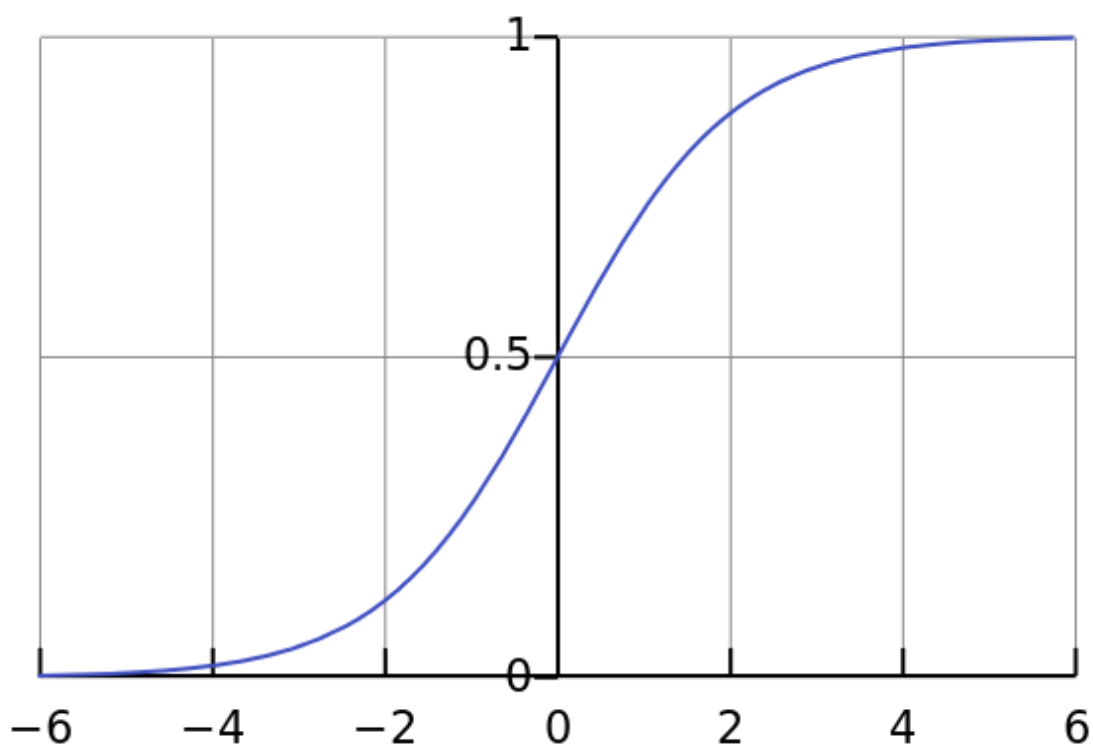
线性激活函数的输出值，正比于神经元的输入值。

缺点

- 线性函数的梯度是个常数，对loss error没有变化；
- 神经网络变成线性网络；

4. Sigmoid激活函数

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



神经网络具有了非线性；

非二进制的激活输出，连续变化，可以处理多分类问题；

梯度变化是平滑的；

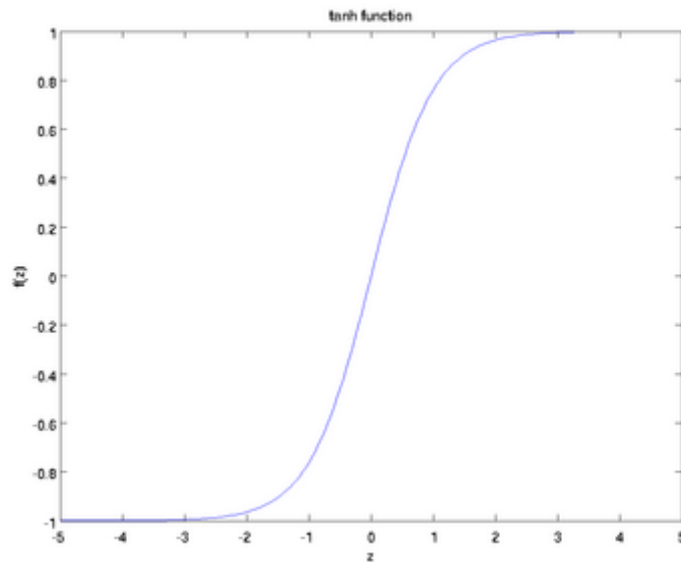
[-2,2] 曲线变化明显，有将函数值变成0、1的趋势，是分类问题理想的激活函数；

值域[0,1]，将神经网络的数值压缩；

缺点：

曲线两端变化不明显，梯度消失；

5. Tanh Function激活函数



This looks very similar to sigmoid. In fact, it is a scaled sigmoid function!

$$f(x) = \tanh(x) = \frac{2}{1+e^{-2x}} - 1$$

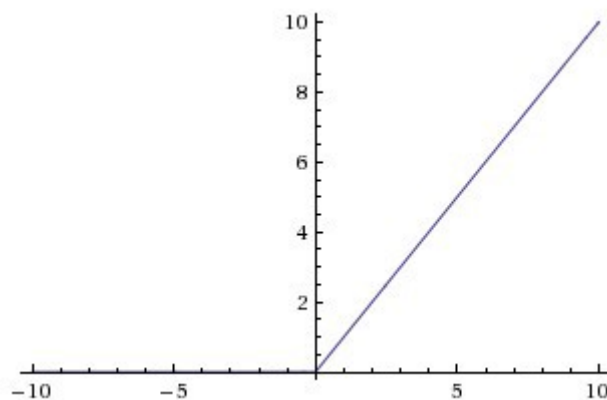
$$\tanh(x) = 2 \operatorname{sigmoid}(2x) - 1$$

具有sigmoid函数的所有特征，并且梯度更大，也有梯度消失的现象。

6. ReLu

ReLU function, $A(x) = \max(0, x)$

It gives an output x if x is positive and 0 otherwise.



relu具有非线性；

relu是稀疏的激活函数，神经网络参数变少了；

更少的计算资源，深度网络使用；

缺点

dying ReLu problem 负数的时候没有梯度，权重不更新；

7. 近似函数的选择

根据神经网络的近似函数的特性进行选择：

- 深度网络--ReLU，稀疏的激活函数，计算量少；
- 分类问题--sigmoid，将数据压缩在零一两端；