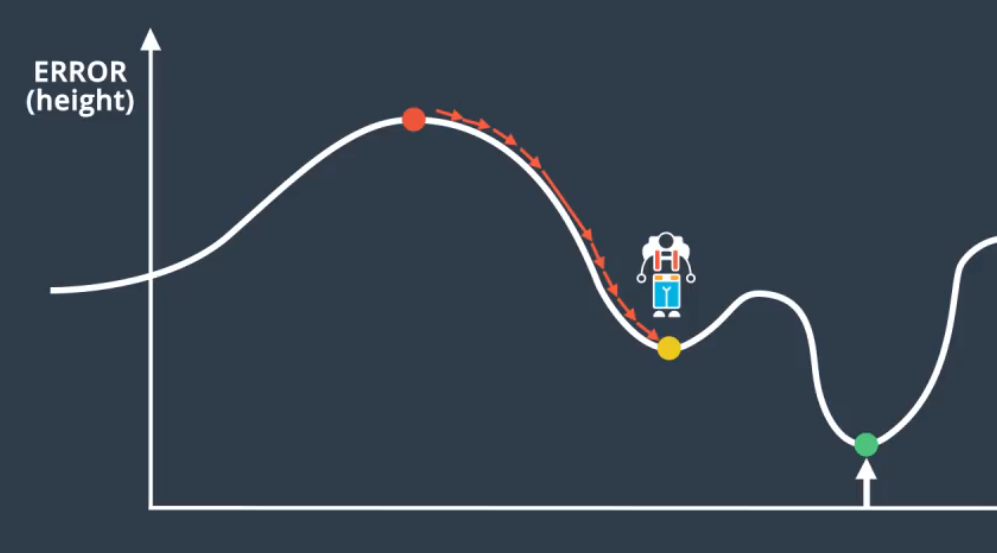
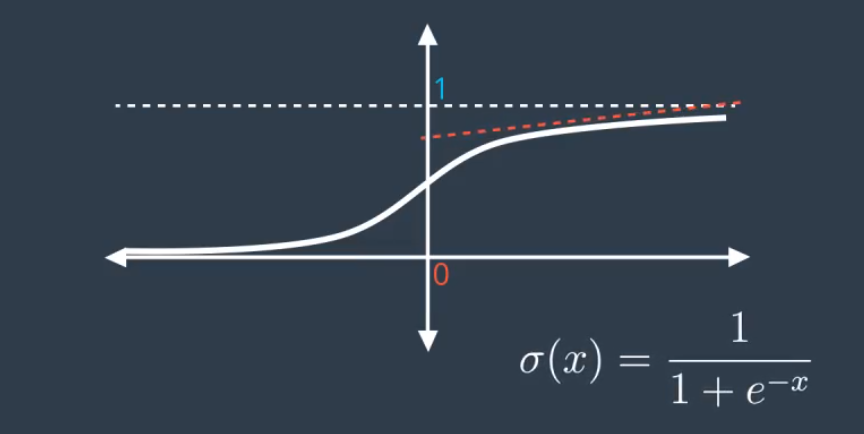
梯度下降中的局部最低点：前提是局部最低点不满足要求

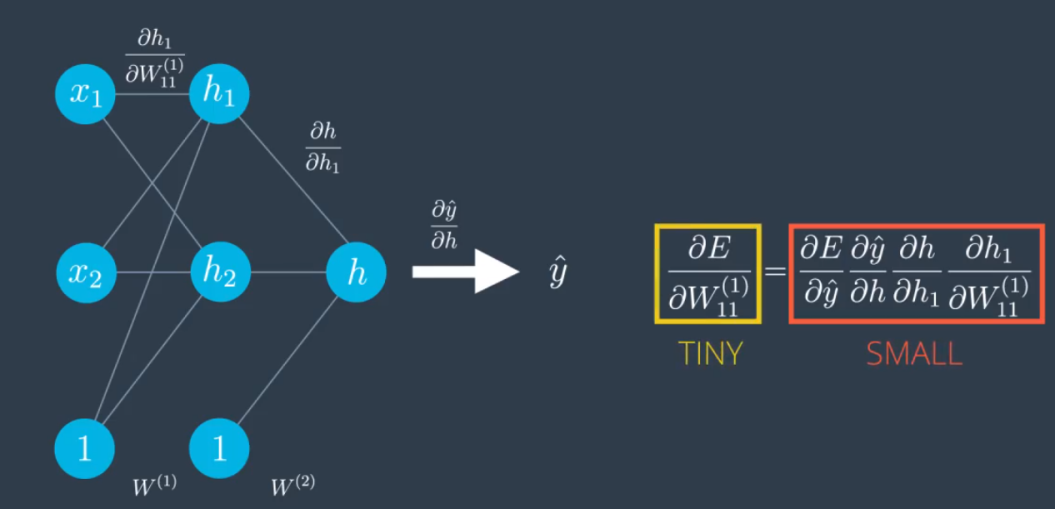


激活函数的缺陷：

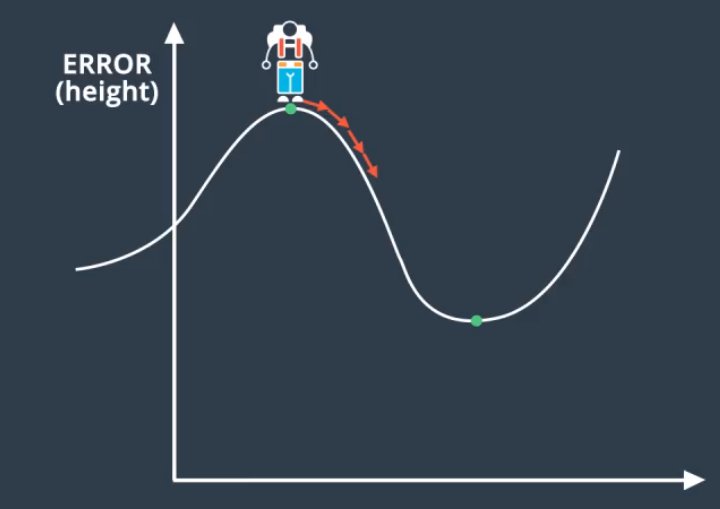


可以看出，其左端或右端的导数非常小，接近于0，这样对梯度下降是不利的。

在多层神经网络中，根据之前推导的公式：



会非常小（梯度消失），这样对权重的修改就非常下，导致下面的情况：



梯度消失：基于导数的链式法则，在多层神经网络中，每一层的导数都小于1，导数会越来越小，梯度后向传播到浅层神经网络时，无法引起参数的扰动，称为梯度消失。

梯度爆炸：每一层（或者有多层）的导数大于1，导致浅层网络的参数变化比深层的还要大，称为梯度爆炸。

从根本上讲无论是梯度消失还是梯度爆炸，其背后的原因是前层网络的梯度是后层网络的乘积，所以神经网络不稳定。唯一可能的情况是以上连续乘积刚好平衡在1左右，但这种几率很小

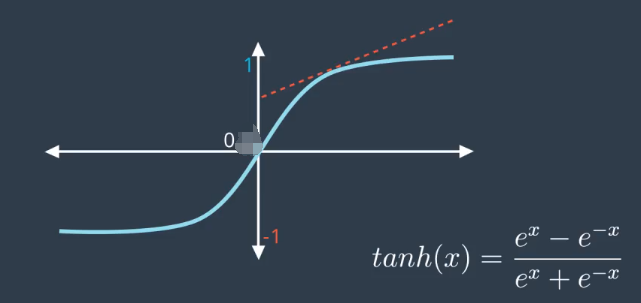
解决的办法一般是更换其他常用的激活函数。

<https://blog.csdn.net/qq_25737169/article/details/78847691>

其他常用的激活函数：

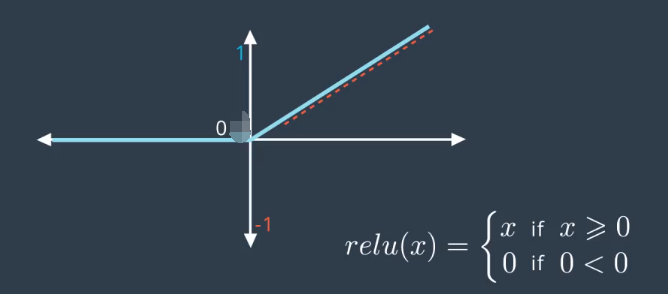
1. 双曲正切函数：tanh(x)



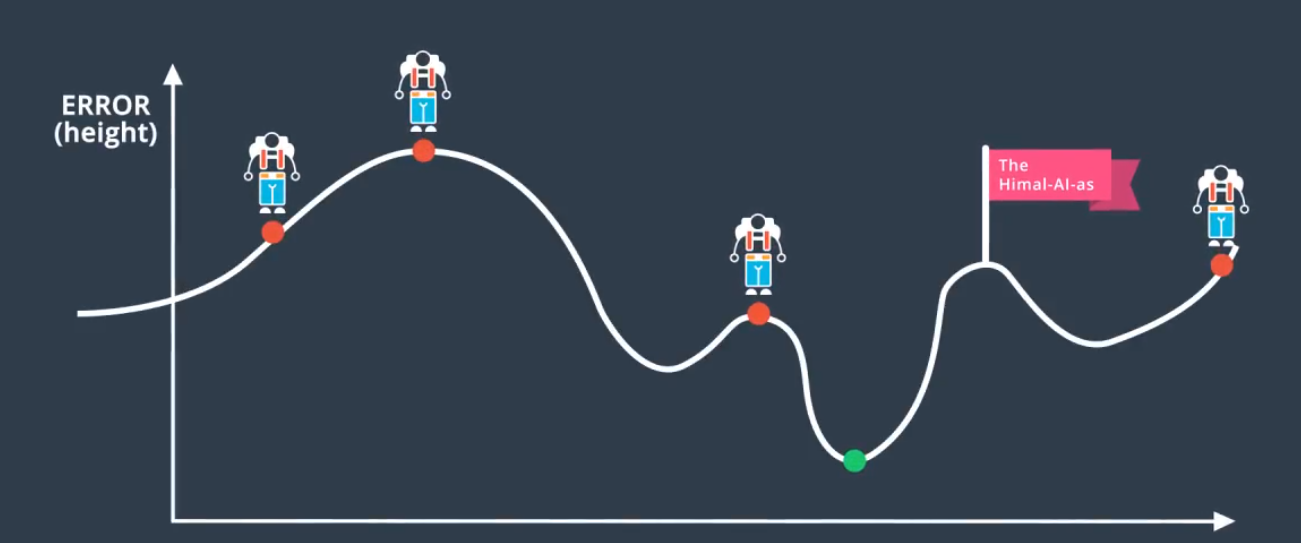


1. 修正线性单元：ReLU





解决局部最小点的方法1：随机重新开始，在不同的地点重新开始计算梯度下降。



方法2：动量

