

W必须初始化为一个很小的值，为了使得sigmoid或tanh激活后仍然有一个比较大的梯度防止下降速度太慢。B可以初始化为0