

神经网络权值为什么不能初始化为零

写在前面：该篇文章的内容以及相关代码（代码在最后），都是我亲自手敲出来的，相关结论分析也是花了挺长时间做出来的，如需转载该文章，请务必先联系我，在后台留言即可。

在深度学习中，神经网络的权重初始化方式非常重要，其对模型的收敛速度和性能有着较大的影响。一个好的权值初始值有以下优点：

- 梯度下降的收敛速度较快
- 深度神经中的网络模型不易陷入梯度消失或梯度爆炸问题

该系列共两篇文章，我们主要讨论以下两个话题：

1. 为什么在线性回归和逻辑回归中可以采用0初始化，而在神经网络中不能采用（实际上不光是0初始化，将权值初始化为任意相同值，都很有可能使模型失效）；
2. 常用的三种权值初始化方法：随机初始化、Xavier initialization、He initialization

在这一篇文章中，我们主要谈论第一个话题

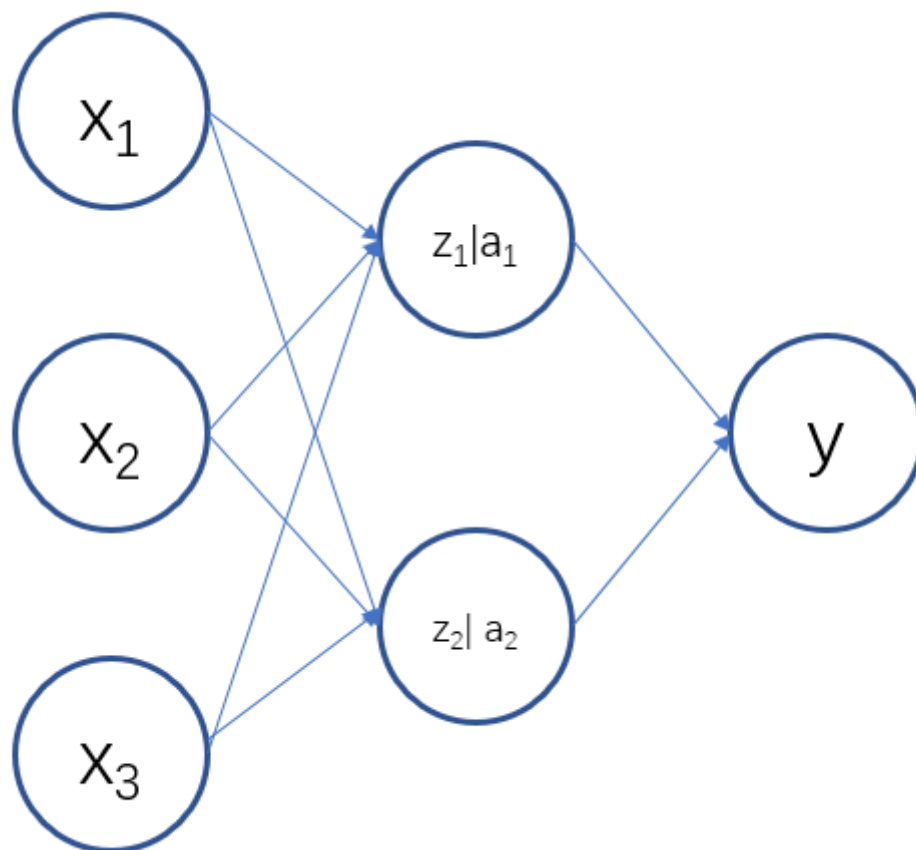
0 初始化

在线性回归和逻辑回归中，我们通常把权值 w 和偏差项 b 初始化为0，并且我们的模型也能取得较好的效果。在线性回归和逻辑回归中，我们采用下面的代码将权值初始化为0（tensorflow框架下）：

12	<pre>w = tf.Variable([[[0], [0], [0]]], dtype=tf.float32, name='weights') b = tf.Variable([0], dtype=tf.float32, name='bias')</pre>

但是，当在神经网络中的权值全部都使用 0 初始化时，模型无法正常工作了。

原因是：在神经网络中因为存在隐含层。我们假设模型的输入为 $[x_1, x_2, x_3]$ ，隐含层数为1，隐含层单元数为2，输出为 y ，模型如下图所示：



则通过正向传播计算之后，可得：

$$z_1 = w_{10} * x_0 + w_{11} * x_1 + w_{12} * x_2 + w_{13} * x_3$$

$$z_2 = w_{20} * x_0 + w_{21} * x_1 + w_{22} * x_2 + w_{23} * x_3$$

在所有的权值 w 和偏差值 b （可以看做是 w_{10} ）初始化为 0 的情况下，即计算之后的 z_1 和 z_2 都等于 0

那么由于 $a_1 = g(z_1)$ 、 $a_2 = g(z_2)$ ，经过激活函数之后得到的 a_1 和 a_2 也肯定是相同的数了，即 $a_1 = a_2 = g(z_1)$

则输出层： $y = g(w_{20} * a_0 + w_{21} * a_1 + w_{22} * a_2)$ 也是固定值了。

重点：在反向传播过程中，我们使用梯度下降的方式来降低损失函数，但在更新权值的过程中，代价函数对不同权值参数的偏导数相同，即 Δw 相同，因此在反向传播更新参数时：

$$w_{21} = 0 + \Delta w$$

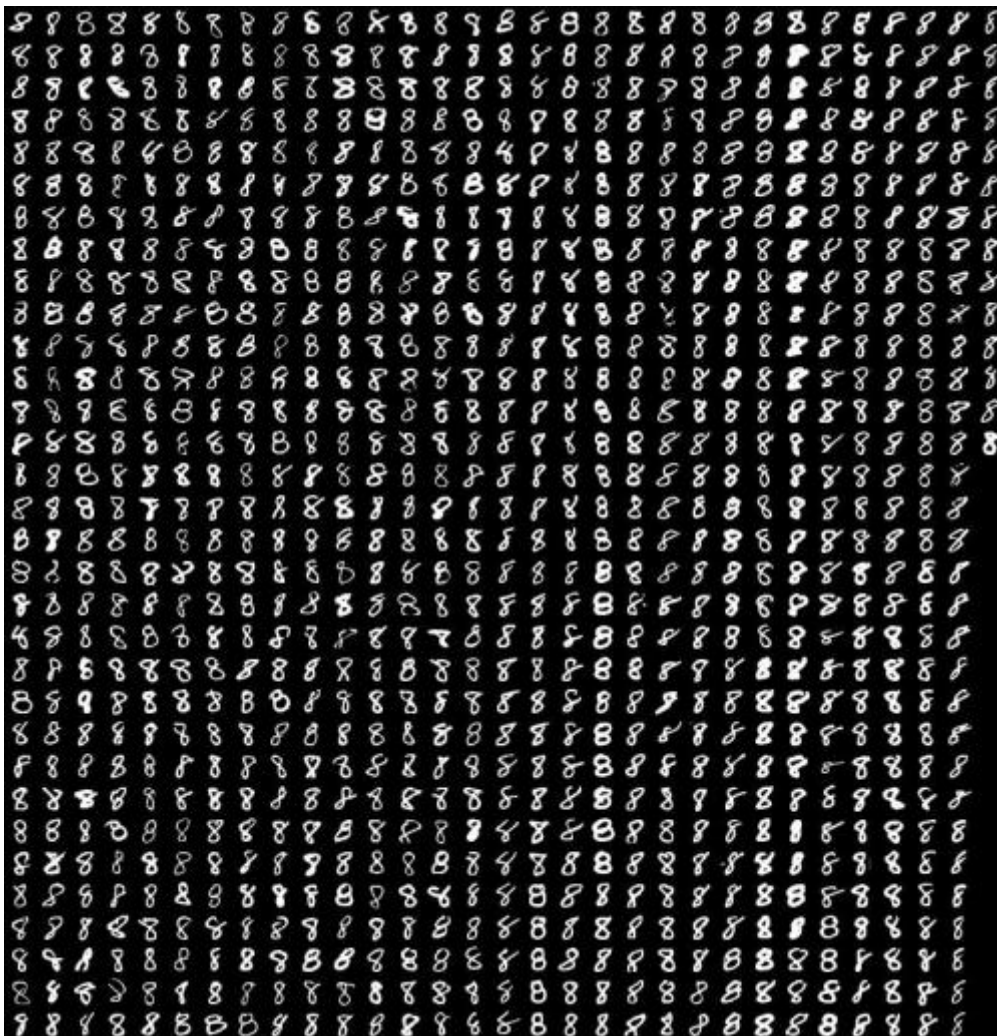
$$w_{22} = 0 + \Delta w$$

实际上使得更新之后的不同节点参数相同，同理可以得到其他更新之后的参数也都是相同的，不管进行多少轮的正向传播和反向传播，得到的参数都一样！因此，神经网络就失去了其特征学习的能力。

在神经网络中使用0 初始化的效果

我们来看一下使用 0 初始化会出现什么样的情况：

我们使用 MNIST 手写数字数据集进行测试：手写数据集是图像处理和机器学习研究最多的数据集之一，在深度学习的发展中起到了重要的作用。



我们看一下使用权值 0 初始化的神经网络训练并测试该数据集的结果：

```
Extracting MNIST_DATA\train-images-idx3-ubyte.gz
Extracting MNIST_DATA\train-labels-idx1-ubyte.gz
Extracting MNIST_DATA\t10k-images-idx3-ubyte.gz
Extracting MNIST_DATA\t10k-labels-idx1-ubyte.gz
Epoch = 0 loss = 2.3016761328957305
Epoch = 10 loss = 2.3013003058867
Epoch = 20 loss = 2.3013187430121675
Epoch = 30 loss = 2.3013124903765636
Epoch = 40 loss = 2.301304516358809
Epoch = 50 loss = 2.3013011997396275
Epoch = 60 loss = 2.301330511353229
Epoch = 70 loss = 2.30131157614968
Epoch = 80 loss = 2.301319170865146
Epoch = 90 loss = 2.301305441422897
Accuracy:11.35%
```

- 在100次的迭代中，每一次迭代，损失值都没有变化
- 模型检测的准确率为11.35%，几乎完全没有检测出来

总结一下：在神经网络中，如果将权值初始化为 0，或者其他统一的常量，会导致后面的激活单元具有相同的值，所有的单元相同意味着它们都在计算同一特征，网络变得跟只有一个隐含层节点一样，这使得神经网络失去了学习不同特征的能力！

