

BatchNormalization

笔记本： 深度学习

创建时间： 2018/8/10 8:25

更新时间： 2018/8/10 8:51

作者： beyourselfwb@163.com

URL: <https://www.zhihu.com/question/38102762>

优质回答一：

好了，现在才是重头戏 - - 为什么要用BN？BN work的原因是什么？

说到底，BN的提出还是为了克服深度神经网络难以训练的弊病。

首先来说说 “Internal Covariate Shift” 。文章的title除了BN这样一个关键词，还有一个便是 “ICS” 。大家都知道在统计机器学习中的一个经典假设是 “源空间 (source domain) 和目标空间 (target domain) 的数据分布 (distribution) 是一致的” 。如果不一致，那么就出现了新的机器学习问题，如，transfer learning/domain adaptation等。

大家细想便会发现，的确，对于神经网络的各层输出，由于它们经过了层内操作作用，其分布显然与各层对应的输入信号分布不同，而且差异会随着网络深度增大而增大，可是它们所能 “指示” 的样本标记 (label) 仍然是不变的，这便符合了covariate shift的定义。由于是对层间信号的分析，也即是 “internal” 的来由。

那BN到底是什么原理呢？说到底还是**为了防止 “梯度弥散”**。关于梯度弥散，大家都知道一个简单的栗子：

$0.9^{30} \approx 0.04$ 。在BN中，是通过将activation规范为均值和方差一致的手段使得原本会减小的activation的scale变大。

作者：魏秀参

链接：

<https://www.zhihu.com/question/38102762/answer/85238569>

来源：知乎

著作权归作者所有。商业转载请联系作者获得授权，非商业转载请注明出处。

优质回答二：

不请自来，**BN本质上解决的是反向传播过程中的梯度问题。**

详细点说，反向传播时经过该层的梯度是要乘以该层的参数的，即前向有：

$$h_l = w_l^T h_{l-1}$$

那么反向传播时便有：

$$\frac{\partial l}{\partial h_{l-1}} = \frac{\partial l}{\partial h_l} \cdot \frac{\partial h_l}{\partial h_{l-1}} = \frac{\partial l}{\partial h_l} w_l$$

那么考虑从l层传到k层的情况，有：

$$\frac{\partial l}{\partial h_k} = \frac{\partial l}{\partial h_l} \prod_{i=k+1}^l w_i$$

上面这个 $\prod_{i=k+1}^l w_i$ 便是问题所在。因为网络层很深，如果 w_i 大多小于1，那么传到这里的时候梯度会变得很小比如 0.9^{100} ；而如果 w_i 又大多大于1，那么传到这里的时候又会有梯度

爆炸问题 比如 1.1^{100} 。BN所做的就是解决这个问题，因为BN作用抹去了 w 的scale影响。

具体有：

$$h_l = BN(w_l h_{l-1}) = BN(\alpha w_l h_{l-1})$$

那么反向求导时便有了：

$$\frac{\partial h_l}{\partial h_{l-1}} = \frac{\partial BN w_l h_{l-1}}{\partial h_{l-1}} = \frac{\partial BN \alpha w_l h_{l-1}}{\partial h_{l-1}}$$

可以看到此时反向传播乘以的数不再和 w 的尺度相关，也就是说尽管我们在更新过程中改变了 w 的值，但是反向传播的梯度却不受影响。更进一步：

$$\frac{\partial h_l}{\partial w_l} = \frac{\partial BN w_l h_{l-1}}{\partial w_l} = \frac{1}{\alpha} \cdot \frac{\partial BN \alpha w_l h_{l-1}}{\partial w_l}$$

即尺度较大的 w 将获得一个较小的梯度，在同等的学习速率下其获得的更新更少，这样使得整体 w 的更新更加稳健起来。

总结起来就是BN解决了反向传播过程中的梯度问题（梯度消失和爆炸），同时使得不同scale的 w 整体更新步调更一致。

更详细的解释可以看我写的一篇BN的文章[Batch Normalization详解](https://www.zhihu.com/question/38102762/answer/164790133)

作者：Jiang

链接：

<https://www.zhihu.com/question/38102762/answer/164790133>

来源：知乎

著作权归作者所有。商业转载请联系作者获得授权，非商业
转载请注明出处。