

深度学习网络大杀器之Dropout——深入解析Dropout

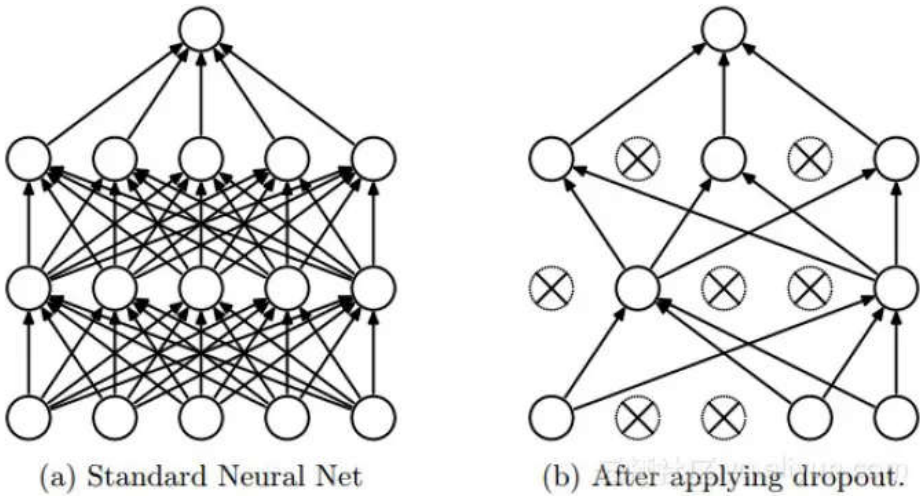
 阿里云社区 关注

 0.352 2018.01.31 13:58:36 字数 1,512 阅读 4,402

摘要：本文详细介绍了深度学习中dropout技巧的思想，分析了Dropout以及Inverted Dropout两个版本，另外将单个神经元与伯努利随机变量相联系让人耳目一新。

过拟合是深度神经网络（DNN）中的一个常见问题：模型只学会在训练集上分类，这些年提出的许多过拟合问题的解决方案；其中dropout具有简单性并取得良好的结果：

Dropout



上图为Dropout的可视化表示，左边是应用Dropout之前的网络，右边是应用了Dropout的同一个网络。

Dropout的思想是训练整体DNN，并平均整个集合的结果，而不是训练单个DNN。DNNs是以概率 p 舍弃部分神经元，其它神经元以概率 $q=1-p$ 被保留，舍去的神经元的输出都被设置为零

引述作者：

在标准神经网络中，每个参数的导数告诉其应该如何改变，以致损失函数最后被减少。因此神经元元可以通过这种方式修正其他单元的错误。但这可能导致复杂的协调，反过来导致过拟合，因为这些协调没有推广到未知数据。Dropout通过使其他隐藏单元存在不可靠性来防止共拟合。

简而言之：Dropout在实践中能很好工作是因为其在训练阶段阻止神经元的共适应。

Dropout如何工作

Dropout以概率 p 舍弃神经元并让其它神经元以概率 $q=1-p$ 保留。每个神经元被关闭的概率是相同的。这意味着：

假设：

$h(x)=xW+b$ ， d 维的输入 x 在 d 维输出空间上的线性投影；

$a(h)$ 是激活函数

写下你的评论...

 评论0  赞13 ...

推荐阅读

蔡康永：“你不必在朋友圈假装生活”
阅读 1,200

华为，好狠一公司，应届生你给201万
阅读 57,402

我哥的葬礼上，我当着138位亲朋的面，打了我妈三个巴掌
阅读 17,093

郭德纲：于谦家里真有钱
阅读 3,597

有了docsify神器，从此爱上看文档
阅读 13,837

其中 $D=(X_1, \dots, X_{dh})$ 是 dh 维的伯努利变量 X_i ，伯努利随机变量具有以下概率质量分布：

$$f(k; p) = \begin{cases} p & \text{if } k = 1 \\ 1 - p & \text{if } k = 0 \end{cases}$$

其中 k 是可能的输出。

将Dropout应用在第 i 个神经元上：

$$O_i = X_i a\left(\sum_{k=1}^{d_i} w_k x_k + b\right) = \begin{cases} a\left(\sum_{k=1}^{d_i} w_k x_k + b\right) & \text{if } X_i = 1 \\ 0 & \text{if } X_i = 0 \end{cases}$$

其中 $P(X_i=0)=p$

由于在训练阶段神经元保持 q 概率，在测试阶段必须仿真出在训练阶段使用的网络集的行为。

为此，作者建议通过系数 q 来缩放激活函数：

训练阶段：

$$O_i = X_i a\left(\sum_{k=1}^{d_i} w_k x_k + b\right)$$

测试阶段：

$$O_i = qa\left(\sum_{k=1}^{d_i} w_k x_k + b\right)$$

Inverted Dropout

与dropout稍微不同。该方法在训练阶段期间对激活值进行缩放，而测试阶段保持不变。

倒数Dropout的比例因子为

，因此：

训练阶段：

$$O_i = \frac{1}{q} X_i a\left(\sum_{k=1}^{d_i} w_k x_k + b\right)$$

测试阶段：

$$O_i = a\left(\sum_{k=1}^{d_i} w_k x_k + b\right)$$

Inverted Dropout是Dropout在各种深度学习框架实践中实现的，因为它有助于一次性定义模型，并只需更改参数（保持/舍弃概率）就可以在同一模型上运行训练和测试过程。

一组神经元的Dropout

n 个神经元的第 h 层在每个训练步骤中可以被看作是 n 个伯努利实验的集合，每个成功的概率等于 p 。

因此舍弃部分神经元后 h 层的输出等于：

$$Y = \sum_{i=1}^{d_h} X_i$$

推荐阅读

蔡康永：“你不必在朋友圈假装生活”
阅读 1,200

华为，好狠一公司，应届生你给201万
阅读 57,402

我哥的葬礼上，我当着138位亲朋的面，打了我妈三个巴掌
阅读 17,093

郭德纲：于谦家里真有钱
阅读 3,597

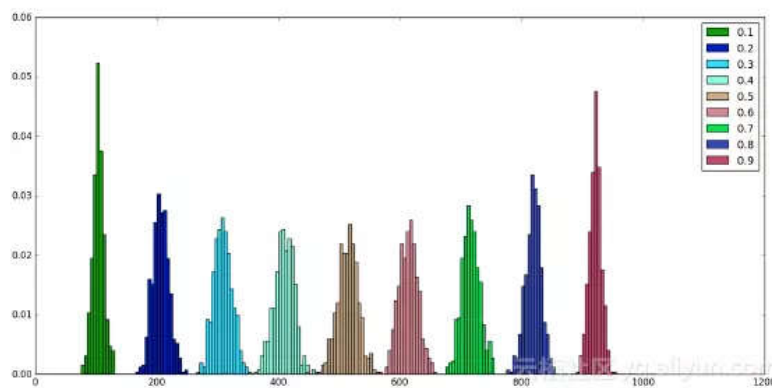
有了docsify神器，从此爱上看文档
阅读 13,837

$$Y \sim Bi(d_h, p)$$

n 次尝试中有 k 次成功的概率由概率质量分布给出：

$$f(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

当使用dropout时，定义了一个固定的舍去概率 p ，对于选定的层，成比例数量的神经元被舍弃。



从上图可以看出，无论 p 值是多少，舍去的平均神经元数量均衡为 np ：

$$E[Bi(n, p)] = np$$

此外可以注意到，围绕在 $p = 0.5$ 值附近的分布是对称。

Dropout与其它正则化

Dropout通常使用L2归一化以及其他参数约束技术。正则化有助于保持较小的模型参数值。

L2归一化是损失的附加项，其中 λ 是一种超参数、 $F(W; x)$ 是模型以及 ϵ 是真值 y 与和预测值 \hat{y} 之间的误差函数。

$$\mathcal{L}(y, \hat{y}) = \mathcal{E}(y, F(W; x)) + \frac{\lambda}{2} W^2$$

通过梯度下降进行反向传播，减少了更新数量。

$$w \leftarrow w - \eta \left(\frac{\partial F(W; x)}{\partial w} + \lambda w \right)$$

Inverted Dropout和其他正则化

由于Dropout不会阻止参数增长和彼此压制，应用L2正则化可以起到作用。

明确缩放因子后，上述等式变为：

$$w \leftarrow w - \eta \left(\frac{1}{q} \frac{\partial F(W; x)}{\partial w} + \lambda w \right)$$

推荐阅读

蔡永康：“你不必在朋友圈假装生活”
阅读 1,200

华为，好狠一公司，应届生你给201万
阅读 57,402

我哥的葬礼上，我当着138位亲朋的面，打了我妈三个巴掌
阅读 17,093

郭德纲：于谦家里真有钱
阅读 3,597

有了docsify神器，从此爱上看文档
阅读 13,837

将 q 称为推动因素，因为其能增强学习速率，将 $r(q)$ 称为有效的学习速率。

有效学习速率相对于所选的学习速率而言更高：基于此约束参数值的归一化可以帮助简化学习速率选择过程。

总结

- 1 Dropout存在两个版本：直接（不常用）和反转
- 2 单个神经元上的dropout可以使用伯努利随机变量建模
- 3 可以使用二项式随机变量来对一组神经元上的舍弃进行建模
- 4 即使舍弃神经元恰巧为 np 的概率是低的，但平均上 np 个神经元被舍弃。
- 5 Inverted Dropout提高学习率
- 6 Inverted Dropout应该与限制参数值的其他归一化技术一起使用，以便简化学习速率选择过程
- 7 Dropout有助于防止深层神经网络中的过度拟合

作者介绍：Paolo Galeone，计算机工程师以及深度学习研究者，专注于计算机视觉问题的研究。

13人点赞 >

日记本

...

"小礼物走一走，来简书关注我"

赞赏支持

还没有人赞赏，支持一下

阿里云云栖社区 阿里云官方内容社区！博客、直播、问答、技术团队、资料下载...
总资产3,124 (约309.53元) 共写了759.0W字 获得65,917个赞

关注

写下你的评论...

全部评论 0

只看作者

按时间倒序 按时间正序

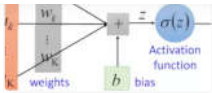
被以下专题收入，发现更多相似内容

首页投稿（暂停...
 我是程序员
 @IT·互联网
 程序员
 互联网科技

@产品
 深度学习 神经...
展开更多

也提到过，机器学习的本质就是寻找最...

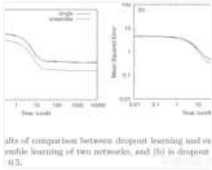
我偏笑_NSNirvana



2017-6-25 今日收集

时间序列 金融 Deeptrader: Deep Learning in Finance (Auto-encod...

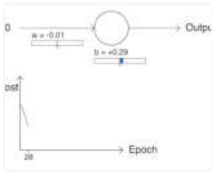
hzyido



神经网络：改进神经网络的学习方法（上）

改进神经网络的学习方法（上） 当一个高尔夫球员刚开始学习打高尔夫时，他们通常会在挥杆的练习上花费大多数时间。慢慢地...

nightwish夜愿



BAT机器学习面试1000题系列（二）

101.深度学习（CNN RNN Attention）解决大规模文本分类问题。用深度学习（CNN RNN Att...

huangqjduter



生活中的仪式感

1 昨天妈妈生日，我忘记电话给她，直到今天下午，才猛然想起来。我很懵，手机日历明明设置了提醒，在关键时刻却没提醒...

涂涂小排屋



推荐阅读

蔡康永：“你不必在朋友圈假装生活”
阅读 1,200

华为，好狠一公司，应届生你给201万
阅读 57,402

我哥的葬礼上，我当着138位亲朋的面，打了我妈三个巴掌
阅读 17,093

郭德纲：于谦家里真有钱
阅读 3,597

有了docsify神器，从此爱上看文档
阅读 13,837

