



$$\begin{aligned}\frac{\partial \ell}{\partial \hat{x}_i} &= \frac{\partial \ell}{\partial y_i} \cdot \gamma \\ \frac{\partial \ell}{\partial \sigma_B^2} &= \sum_{i=1}^m \frac{\partial \ell}{\partial \hat{x}_i} \cdot (x_i - \mu_B) \cdot \frac{-1}{2} (\sigma_B^2 + \epsilon)^{-3/2} \\ \frac{\partial \ell}{\partial \mu_B} &= \sum_{i=1}^m \frac{\partial \ell}{\partial \hat{x}_i} \cdot \frac{-1}{\sqrt{\sigma_B^2 + \epsilon}} \\ \frac{\partial \ell}{\partial x_i} &= \frac{\partial \ell}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} + \frac{\partial \ell}{\partial \sigma_B^2} \cdot \frac{2(x_i - \mu_B)}{m} + \frac{\partial \ell}{\partial \mu_B} \cdot \frac{1}{m} \\ \frac{\partial \ell}{\partial \gamma} &= \sum_{i=1}^m \frac{\partial \ell}{\partial y_i} \cdot \hat{x}_i \\ \frac{\partial \ell}{\partial \beta} &= \sum_{i=1}^m \frac{\partial \ell}{\partial y_i}\end{aligned}$$

3. Where to use BN?

BN可以应用于网络中任意的activation set。文中还特别指出在CNN中，BN应作用在非线性映射前，即对  $\mathbf{x} = \mathbf{W}\mathbf{u} + \mathbf{b}$  做规范化。另外对CNN的“权值共享”策略，BN还有其对应的做法（详见文中3.2节）。

4. Why BN?

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。其实BN背后的insight非常简单，只是在文章中被Google复杂化了。

首先来说“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。而covariate shift就是分布不一致假设之下的一个分支问题，它是指源空间和目标空间的条件概率是一致的，但是其边缘概率不同，即：对所有  $\mathbf{x} \in \mathcal{X}$ ,  $P_s(\mathbf{Y}|\mathbf{X} = \mathbf{x}) = P_t(\mathbf{Y}|\mathbf{X} = \mathbf{x})$ , 但是  $P_s(\mathbf{X}) \neq P_t(\mathbf{X})$ . 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”的由来。那么好，为什么前面我说Google将其复杂化了。其实如果严格按照解决covariate shift的路子来做的话，大概就是上“importance weight”（ref (https://link.zhihu.com/?target=http%3A//120.52.72.36/www.jmlr.org/c3pr90ntcsf0/papers/volume8/sugiyama07a/sugiyama07a.pdf)）之类的机器学习方法。可是这里Google仅仅说“通过mini-batch来规范化某些层/所有层的输入，从而可以固定每层输入信号的均值与方差”就可以解决问题。如果covariate shift可以用这么简单的方法解决，那前人对它的研究也真是白做了。此外，试想，均值方差一致的分布就是同样的分布吗？当然不是。显然，ICS只是这个问题的“包装纸”嘛，仅仅是一种high-level demonstration。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。可以说是一种更有效的local response normalization方法（见4.2.1节）。

5. When to use BN?

OK，说完BN的优势，自然可以知道什么时候用BN比较好。例如，在神经网络训练时遇到收敛速度很慢，或梯度爆炸等无法训练的状况时可以尝试BN来解决。另外，在一般使用情况下也可以加入BN来加快训练速度，提高模型精度。

诚然，在DL中还有许多除BN之外的“小trick”。别看是“小trick”，实则是“大杀器”，正所谓“*The devil is in the details*”。希望了解其它DL trick（特别是CNN）的各位请移步我之前总结的：Must Know Tips/Tricks in Deep Neural Networks（https://link.zhihu.com/?target=http%3A//lamda.nju.edu.cn/weixs/project/CNNTricks/CNNTricks.html）

另外，论文中说的Batch Normalization的作用：

- (1) 可以使用更高的学习率。如果每层的scale不一致，实际上每层需要的学习率是不一样的，同一层不同维度的scale往往也需要不同大小的学习率，通常需要使用最小的那个学习率才能保证损失函数有效下降，Batch Normalization将每层、每维的scale保持一致，那么我们就可以直接使用较高的学习率进行优化。
- (2) 移除或使用较低的dropout。dropout是常用的防止overfitting的方法，而导致overfit的位置往往在数据边界处，如果初始化权重就已经落在数据内部，overfit现象就可以得到一定的缓解。论文中最后的模型分别使用10%、5%和0%的dropout训练模型，与之前的40%-50%相比，可以大大提高训练速度。
- (3) 降低L2权重衰减系数。还是一样的问题，边界处的局部最优往往有几维的权重（斜率）较大，使用L2衰减可以缓解这一问题，现在用了Batch Normalization，就可以把这个值降低了，论文中降低为原来的5倍。



是AI就躲个飞机—纯Python实现人工智能  
(<http://blog.csdn.net/u014365862/article/details/54380422>)  
网站客服 微博客服  
(<http://wpa.qq.com/msgrd?v=3&uin=243129980&site=qq&msg=98>)  
(<http://e.weibo.com/csdnsupport/p/22404>)  
✉ webmaster@csdn.net  
C语言之学生管理系统代码（完整）([mailto:webmaster@csdn.net](http://mailto:webmaster@csdn.net))  
(<http://blog.csdn.net/u014365862/article/details/51791677>)  
☎ 19147  
tf15: 中文语音识别 (<http://blog.csdn.net/u014365862/article/details/53869701>) (<http://www.mii.beian.gov.cn/>)  
☎ 13395 关于  
(<http://www.csdn.net/recruit.html>)  
(<http://blog.csdn.net/u014365862/article/details/53869816>)  
(<http://www.csdn.net/company/recruit.html>)  
☎ 10852 广告服务  
(<http://www.csdn.net/company/marketin>)  
☁ 阿里云  
Copyright © 1999–2018  
CSDN.NET, All Rights Reserved

- (4) 取消Local Response Normalization层。 由于使用了一种Normalization，再使用LRN就显得没那么必要了。而且LRN实际上也没那么work。
- (5) 减少图像扭曲的使用。 由于现在训练epoch数降低，所以要对输入数据少做一些扭曲，让神经网络多看看真实的数据。



版权声明：本文为博主原创文章，未经博主允许不得转载。有问题可以加微信：lp9628(注明CSDN)。



发表你的评论

(http://my.csdn.net/wujuxKkoolerter)

为什么会出现Batch Normalization层

NNNNNNNNNNNNNY 2017年04月21日 16:36 2256

训练模型时的收敛速度问题众所周知，模型训练需要使用高性能的GPU，还要花费大量的训练时间。除了数据量大及模型复杂等硬性因素外，数据分布的不断变化使得我们必须使用较小的学习率、较好的权重初值和不容易饱和...

(http://blog.csdn.net/NNNNNNNNNNNNNY/article/details/70331796)

[深度学习] Batch Normalization算法介绍

lhanchao 2017年04月21日 11:34 2744

很早就打算写这篇博客了，最近遇到的问题比较多，所以拖了又拖，今天问题似乎解决了，等着程序运行的时候再来回顾一下Batch Normalization算法。 Batch Normalization是2...

(http://blog.csdn.net/lhanchao/article/details/70308092)

做了这件事后，这位程序员月薪从15k变身30k！

闲下来的时候，程序员为了个人成长，和培养核心价值，该怎么规划和学习呢...



(http://www.baidu.com/cb.php?c=lgF\_pyfqHmknjTzrb0IZ0qnfK9ujYzP1nsrjD10Aw-5Hc3rHnYnHb0TAq15HfLPWRznjb0T1dbn1-WPvF9PycYnWbdmWmL0AwY5HDdnHRzPjc1P1R0lgF\_5y9YIZ0IQzq-uZR8mLPbUB48ugfEpZNGXy-jULNzTvRETvNzpyN1gww-IA7GUatLPjqdAdxTvqdThP-5yF\_UvTkn0KzujYknfKBUHYs0ZKz5H00ly-b5HDdP1f1PWD0Uv-b5HDzrH63nHf0mv-b5HTzPWb1n6KElv3qn0KsXHYZnm0mLFW5Hczrjmd)

Batch Normalization 神经网络加速算法

u012192662 2016年07月29日 15:01 2963

转自： http://www.cnblogs.com/neopenx/p/5211969.html 从Bayesian角度浅析Batch Normalization 前置阅...

(http://blog.csdn.net/u012192662/article/details/52065021)

【机器学习trick】Batch-Normalization的理解和研究

Batch-Normalization概述15年谷歌发表了一篇文章，名字是《Batch Normalization: Accelerating Deep Network Training by Red...

u013106893 2017年04月06日 23:46 214

(http://blog.csdn.net/u013106893/article/details/69488773)

Batch Normalization简介

2017年11月24日 17:20 3.36MB

下载



联系我们

- 网站客服 微博客服 (http://wpa.qq.com/msgrd?v=3&uin=2431299880&site=qq&me (http://e.weibo.com/csdnsupport/p
- webmaster@csdn.net (mailto:webmaster@csdn.net)
- 400-660-0108

京ICP证09002463号 (http://www.miibeian.gov.cn/) 关于 (http://www.csdn.net/company/about.ht 招聘 (http://www.csdn.net/company/recruit.h 广告服务 (http://www.csdn.net/company/marketin 阿里云 Copyright © 1999-2018 CSDN.NET, All Rights Reserved

免费试听课程，迈出百万年薪第一步



【机器学习】： 免费试听课程，迈出百万年薪第一步

Batch Normalization

2017年12月17日 23:49 311KB 下载



Batch Normalization论文翻译——中英文对照

Batch Normalization论文翻译——中英文对照  
(http://blog.csdn.net/Quincuntial/article/details/78124582)

深度学习Batch Normalization理论

2017年11月09日 15:36 169KB 下载



Batch Normalization: Accelerating Deep Network Training by ...

2016年11月20日 18:53 273KB 下载



caffe+报错 | 深度学习参数调优杂记+caffe训练时的问题+dropout/batch Normaliz...

一、深度学习中常用的调节参数 本节为笔者上课笔记（CDA深度学习实战课程第一期） 1、学习率步长的选择：你走的距离长短，越短当然不会错过，但是耗时间。步长的选择比较麻烦。步长越小，越容易得到局部最...

sinat\_26917383 2017年01月08日 11:27 6985  
(http://blog.csdn.net/sinat\_26917383/article/details/54232791)

[深度学习] Batch Normalization算法介绍

很早就打算写这篇博客了，最近遇到的问题比较多，所以拖了又拖，今天问题似乎解决了，等着程序运行的时候再来回顾一下Batch Normalization算法。Batch Normalization是2...  
(http://blog.csdn.net/lhanchao/article/details/70308092)

caffe+报错 | 深度学习参数调优杂记+caffe训练时的问题+dropout/batch Normaliz...

一、深度学习中常用的调节参数 本节为笔者上课笔记（CDA深度学习实战课程第一期） 1、学习率 步长的选择：你走的距离长短，越短当然不会错过，但是耗时间。步长的选择比较麻烦。步...

shizenglin 2017年04月27日 23:17 1303  
(http://blog.csdn.net/shizenglin/article/details/70880765)

深度学习中 Batch Normalization为什么效果好？

作者：魏秀参 链接：http://www.zhihu.com/question/38102762/answer/85238569 来源：知乎 著作权归作者所有，转载请联系作者获得授权。 这...

zhyj3038 2016年09月21日 09:00 539  
(http://blog.csdn.net/zhyj3038/article/details/52604150)

深度学习（二十九）Batch Normalization 学习笔记

Batch Normalization 学习笔记 原文地址：http://blog.csdn.net/hjimce/article/details/50866313 作者：hjimce 一、背...

haoji007 2016年10月17日 10:25 252



联系我们


网站客服 微博客服  
(http://wpa.qq.com/msgrd?v=3&uin=2431299880&site=qq&me  
(http://e.weibo.com/csdnsupport/p  
webmaster@csdn.net  
(mailto:webmaster@csdn.net)  
400-660-0108

京ICP证09002463号  
(http://www.miibeian.gov.cn/)  
关于  
(http://www.csdn.net/company/about.ht  
招聘  
(http://www.csdn.net/company/recruit.h  
广告服务  
(http://www.csdn.net/company/marketin  
阿里云  
Copyright © 1999-2018  
CSDN.NET, All Rights Reserved



(<http://blog.csdn.net/haoji007/article/details/52788634>)

### Batch Normalization (BN)

 qq\_34420188 2017年04月06日 15:41 148


作者：魏秀参 链接：<https://www.zhihu.com/question/38102762/answer/85238569> 来源：知乎 1. What is BN? 顾...  
([http://blog.csdn.net/qq\\_34420188/article/details/69390255](http://blog.csdn.net/qq_34420188/article/details/69390255))

### Batch Normalization

 u010238520 2017年06月22日 21:01 144


Batch Normalization 在训练深度神经网络的时候，由于前面网络层参数的变化，会使得当前层网络的输入分布发生改变，这种情况需要我们很小心的对网络参数的参数进行初始化，以及使用很小的学...  
(<http://blog.csdn.net/u010238520/article/details/73611072>)

### 数据处理－Batch Normalization


 Ddreaming 2016年11月20日 12:12 689

今年过年之前，MSRA和Google相继在ImagenNet图像识别数据集上报告他们的效果超越了人类水平，下面将分两期介绍两者的算法细节。 这次先讲Google的这篇《Batch Normali...  
(<http://blog.csdn.net/Ddreaming/article/details/53239933>)


### 笔记：batch normalization：accelerating deep network training by reducing int...

概述：机器学习系统中输入分布不均称为covariate shift。对于复杂网络系统，某层输入分布不稳定称为internal covariate s hift.本文提出了batch normaliza...  
 a1154761720 2016年03月06日 11:37 1785  
(<http://blog.csdn.net/a1154761720/article/details/50812607>)

### 《Batch Normalization Accelerating Deep Network Training by Reducing Intern...

本文转自：<http://www.aichengxu.com/view/1422042> 今年过年之前，MSRA和Google相继在ImagenNet图像识别数据集上报告他们的效果超越了人类水平，下面...  
 xiaoyanghijk 2016年08月03日 12:01 155  
(<http://blog.csdn.net/xiaoyanghijk/article/details/52102302>)

### 深度学习Deep Learning（05）：Batch Normalization(BN)批标准化

github地址：[https://github.com/lawlite19/DeepLearning\\_Python](https://github.com/lawlite19/DeepLearning_Python) 四、Batch Normalization（BN）批标准化1、说明 参  
考论文：htt...  
 u013082989 2017年01月09日 21:37 2734  
(<http://blog.csdn.net/u013082989/article/details/54293279>)



#### 联系我们

 网站客服  微博客服  
(<http://wpa.qq.com/msgrd?v=3&uin=2431299880&site=qq&me>)  
(<http://e.weibo.com/csdnsupport/p>)  
 webmaster@csdn.net  
(mailto:webmaster@csdn.net)  
 400-660-0108

京ICP证09002463号  
(<http://www.miibeian.gov.cn/>)  
关于  
(<http://www.csdn.net/company/about.htm>)  
招聘  
(<http://www.csdn.net/company/recruit.htm>)  
广告服务  
(<http://www.csdn.net/company/marketin>)  
 阿里云  
Copyright © 1999-2018  
CSDN.NET, All Rights Reserved