

3、BN 思路

- 如果输入数据是白化的 (whitened)，网络会更快的收敛
- 白化目的是降低数据的冗余性和特征的相关性，例如通过线性变换使数据为0均值和单位方差



1



- 并非直接标准化每一层那么简单，如果不考虑归一化的影响，可能会降低梯度下降的影响

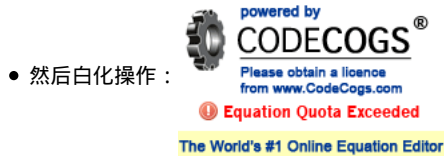
- 标准化与某个样本和所有样本都有关系
- 解决上面的问题，我们希望对于任何参数值，都要满足想要的分布；



- 对于反向传播，需要计算：

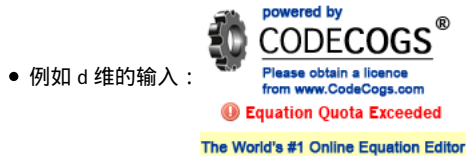


- 这样做的计算代价是非常大的，因为需要计算x的协方差矩阵



- 然后白化操作：

- 上面两种都不行或是不好，进而得到了BN的方法
- 既然白化每一层的输入代价非常大，我们可以进行简化
- 简化1
- 标准化特征的每一个维度而不是去标准化所有的特征，这样就不用求协方差矩阵了



- 例如 d 维的输入：

- 标准化操作：



- 需要注意的是标准化操作可能会降低数据的表达能力,例如我们之前提到的Sigmoid函数：

[//blog.csdn.net/u013082989/article/details/51971121](http://blog.csdn.net/u013082989/article/details/51971121))

18889

scrapy爬虫框架将数据保存Mysql数据库中 (<http://blog.csdn.net/u013082989/article/details/52589791>)

18015

[Python]使用Scrapy爬虫框架简单爬取图片并保存本地 (<http://blog.csdn.net/u013082989/article/details/52507601>)

11047

GitHub合并 (merge) 代码时冲突解决 (<http://blog.csdn.net/u013082989/article/details/51099269>)

8769

递归求数组的最大值，主要说下递归执行的过程 (<http://blog.csdn.net/u013082989/article/details/50784385>)

5607

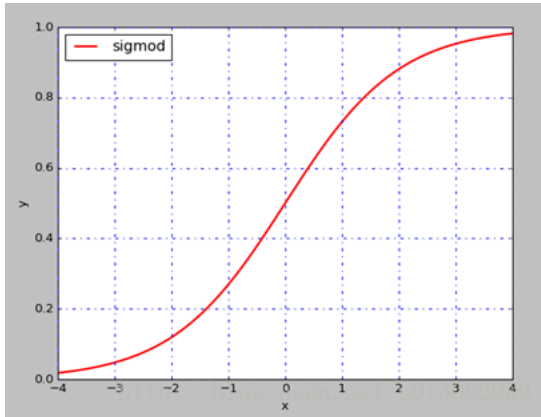


内容举报



返回顶部





- 标准化之后均值为0，方差为1，数据就会落在近似线性的函数区域内，这样激活函数的意义就不明显

- 所以对于每个，对应一对参数：



- 从式子来看就是对标准化的数据进行缩放和平移，不至于使数据落在线性区域内，增加数据的



恢复到原来的值了)

- 但是这里还是使用的全部的数据集，但是如果使用随机梯度下降，可以选取一个batch进行训练

• 简化2

- 第二种简化就是使用 mini-batch 进行 随机梯度下降
- 注意这里使用 mini-batch 也是标准化每一个维度上的特征，而不是所有的特征一起，因为若果 mini-batch 中的数据量小于特征的维度时，会产生奇异协方差矩阵，对应的行列式的值为0，非满秩
- 假设mini-batch 大小为 m 的 B



- 作者给出的批标准化的算法如下：

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

登录

注册

×



内容举报



返回顶部

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1...m}\}$;
 Parameters to be learned: γ, β
Output: $\{y_i = \text{BN}_{\gamma, \beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad // \text{ mini-batch mean}$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\mathcal{B}})^2 \quad // \text{ mini-batch variance}$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \quad // \text{ normalize}$$

$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i) \quad // \text{ scale and shift}$$

Algorithm 1: Batch Normalizing Transform, applied to activation x over a mini-batch: [log.csdn.net/u013082989](http://blog.csdn.net/u013082989)

- 算法中的 ϵ 是一个常量，为了保证数值的稳定性

- 反向传播求梯度：

- 因为：

 powered by
CODECOGS[®]
 Please obtain a licence
 from www.CodeCogs.com
 Equation Quota Exceeded
 The World's #1 Online Equation Editor

- 所以：

 powered by
CODECOGS[®]
 Please obtain a licence
 from www.CodeCogs.com
 Equation Quota Exceeded
 The World's #1 Online Equation Editor

- 因为：

 powered by
CODECOGS[®]
 Please obtain a licence
 from www.CodeCogs.com
 Equation Quota Exceeded
 The World's #1 Online Equation Editor

- 所以：

 powered by
CODECOGS[®]
 Please obtain a licence
 from www.CodeCogs.com
 Equation Quota Exceeded
 The World's #1 Online Equation Editor

- 
 powered by
CODECOGS[®]
 Please obtain a licence
 from www.CodeCogs.com
 Equation Quota Exceeded
 The World's #1 Online Equation Editor



内容举报



返回顶部



- 因为： powered by CODECOGS®
Please obtain a licence from www.CodeCogs.com
① Equation Quota Exceeded
The World's #1 Online Equation Editor
- 和  powered by CODECOGS®
Please obtain a licence from www.CodeCogs.com
① Equation Quota Exceeded
The World's #1 Online Equation Editor

- 所以： powered by CODECOGS®
Please obtain a licence from www.CodeCogs.com
① Equation Quota Exceeded
The World's #1 Online Equation Editor

- 所以： powered by CODECOGS®
Please obtain a licence from www.CodeCogs.com
① Equation Quota Exceeded
The World's #1 Online Equation Editor

-  powered by CODECOGS®
Please obtain a licence from www.CodeCogs.com
① Equation Quota Exceeded
The World's #1 Online Equation Editor

- 对于BN变换是可微分的，随着网络的训练，网络层可以持续学到输入的分布。

4、BN 网络的训练和推断

- 按照BN方法，输入数据 x 会经过变化得到 $BN(x)$ ，然后通过随机梯度下降进行训练，标准化是在 mini-batch 上所以是非常高效的。
- 但是对于推断我们希望输出只取决于输入，而对于输入只有一个实例数据，无法得到 mini-batch 的其他实例，就无法求对应的均值和方差了。
- 可以通过从所有训练实例中获得的统计量来**代替**mini-batch 中 m 个训练实例获得统计量均值和方差
- 我们对每个 mini-batch 做标准化，可以对记住每个 mini-batch 的 B ，然后得到全局统计量

-  powered by CODECOGS®
Please obtain a licence from www.CodeCogs.com
① Equation Quota Exceeded
The World's #1 Online Equation Editor

-  powered by CODECOGS®
Please obtain a licence from www.CodeCogs.com
① Equation Quota Exceeded
The World's #1 Online Equation Editor (这里方差采用的是无偏方差估计)

- 所以推断采用 BN 的方式为：



内容举报



返回顶部





1



- 作者给出的完整算法：

Input: Network N with trainable parameters Θ ;
subset of activations $\{x^{(k)}\}_{k=1}^K$

Output: Batch-normalized network for inference, $N_{\text{BN}}^{\text{inf}}$

- 1: $N_{\text{BN}}^{\text{tr}} \leftarrow N$ // Training BN network
- 2: **for** $k = 1 \dots K$ **do**
- 3: Add transformation $y^{(k)} = \text{BN}_{\gamma^{(k)}, \beta^{(k)}}(x^{(k)})$ to $N_{\text{BN}}^{\text{tr}}$ (Alg. 1)
- 4: Modify each layer in $N_{\text{BN}}^{\text{tr}}$ with input $x^{(k)}$ to take $y^{(k)}$ instead
- 5: **end for**
- 6: Train $N_{\text{BN}}^{\text{tr}}$ to optimize the parameters $\Theta \cup \{\gamma^{(k)}, \beta^{(k)}\}_{k=1}^K$
- 7: $N_{\text{BN}}^{\text{inf}} \leftarrow N_{\text{BN}}^{\text{tr}}$ // Inference BN network with frozen parameters
- 8: **for** $k = 1 \dots K$ **do**
- 9: // For clarity, $x \equiv x^{(k)}, \gamma \equiv \gamma^{(k)}, \mu_{\mathcal{B}} \equiv \mu_{\mathcal{B}}^{(k)}$, etc.
- 10: Process multiple training mini-batches \mathcal{B} , each of size m , and average over them:

$$\mathbb{E}[x] \leftarrow \mathbb{E}_{\mathcal{B}}[\mu_{\mathcal{B}}]$$

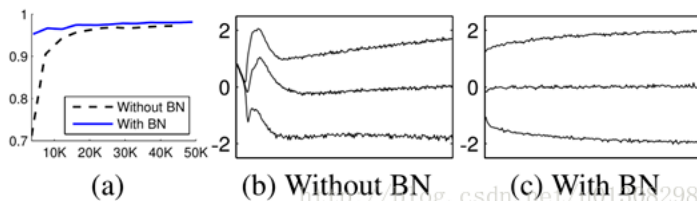
$$\text{Var}[x] \leftarrow \frac{m}{m-1} \mathbb{E}_{\mathcal{B}}[\sigma_{\mathcal{B}}^2]$$
- 11: In $N_{\text{BN}}^{\text{inf}}$, replace the transform $y = \text{BN}_{\gamma, \beta}(x)$ with

$$y = \frac{\gamma}{\sqrt{\text{Var}[x] + \epsilon}} \cdot x + \left(\beta - \frac{\gamma \mathbb{E}[x]}{\sqrt{\text{Var}[x] + \epsilon}} \right)$$
- 12: **end for**

Algorithm 2: Training a Batch-Normalized Network

5、实验

- 最后给出的实验可以看出使用BN的方式训练精准度很高而且很稳定。



版权声明：本文为博主原创文章，未经博主允许不得转载。

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

登录

注册



内容举报




返回顶部




<深度学习优化策略-1>Batch Normalization（BN）

今天给大家带来深度学习的优化策略篇的第一篇Batch Normalization（BN）。BN可以看做对输入样本的一种约束，最大作用是加速收敛，减少模型对dropout，careful weigh...

lqfarmer 2017年05月03日 21:47 1281

(http://blog.csdn.net/lqfarmer/article/details/71155733)

matlab标准化和反标准化——zscore

zb1165048017 2016年03月25日 09:27 5138

先来看自带函数zscore的使用 >> A=[1 2 3;4 5 6]; >> [B,A_mean,A_std]=zscore(A) B = -0.7071 -0.7071 -0.7...

(http://blog.csdn.net/zb1165048017/article/details/50976858)

Python薪酬到底多高？

Python2017年表现异常突出，这次我们拿出10个招聘网站的Python数据，对比后发现....你猜到了吗？



广告

(http://www.baidu.com/cb.php?c=lgF_pyfqHmknjnvPjc0IZ0qnfK9ujYzP1nYPH0k0Aw-5Hc3rHnYnHb0TAq15HfLPWRznjb0T1Y3mVn1PjIbrARsnjmYmhfs0AwY5HDdnHRsn1m3rjn0lgF_5y9YIZ0IQzq-uZR8mLPbUB48ugfElAqspynETZ-YpAq8nWqdlAdxTvqdThP-5yF_UvTkn0KzujY4rHb0mhYqn0KsTWYs0ZNGujYkPHTYn1mk0AqGuJYknWb3rjDY0APGuJYLnWm4n1c0ULI85H00TZbqnW0v0APzm1Ykn16kP0)

数据标准化/归一化normalization


pipisorry 2016年08月19日 09:42 42206

http://blog.csdn.net/pipisorry/article/details/52247379数据的标准化（normalization）和归一化 数据的标准化（normaliza...

(http://blog.csdn.net/pipisorry/article/details/52247379)

<深度学习优化策略-1>Batch Normalization（BN）


今天给大家带来深度学习的优化策略篇的第一篇Batch Normalization（BN）。BN可以看做对输入样本的一种约束，最大作用是加速收敛，减少模型对dropout，careful weigh...

lqfarmer 2017年05月03日 21:47 1281


(http://blog.csdn.net/lqfarmer/article/details/71155733)


【深度学习】论文导读：google的批正则方法（Batch Normalization: Accelerat...

google2015年的论文，首次提出批正则方法，优化深度神经网络的学习摘要在深度网络的训练中，每一层网络的输入都会因为前一层网络参数的变化导致其分布发生改变，这就要求我们必须使用一个很小的学...

mao_xiao_feng 2016年11月13日 17:07 829

(http://blog.csdn.net/mao_xiao_feng/article/details/53150037)



内容举报



返回顶部



基于Theano的深度学习(Deep Learning)框架Keras学习随笔-05-模型


原地址：<http://blog.csdn.net/niuwei22007/article/details/49207187>可以查看更多文章介绍完了优化器和目标函数，那么剩下的就是训练模型了。这...


 dlaicxf 2016年10月18日 11:02 734

 (<http://blog.csdn.net/dlaicxf/article/details/52847051>)

TensorFlow 学习笔记-05]批标准化(Bacth Normalization，BN)


[版权说明] TensorFlow 学习笔记参考： 李嘉璇 著 TensorFlow技术解析与实战黄文坚 唐源 著 TensorFlow实战郑泽宇 顾恩宇 著 TensorFlow实战Googl...


 caicaiatnbu 2017年05月25日 19:22 677

 (<http://blog.csdn.net/caicaiatnbu/article/details/72742293>)

深度学习论文笔记][Weight Initialization] Batch Normalization: Accelerating De...

Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by re...


 Hao_Zhang_Vision 2016年09月20日 13:54 706

 (http://blog.csdn.net/Hao_Zhang_Vision/article/details/52595249)

Neural Networks and Deep Learning-神经网络与深度学习-zh.zip

2017年12月26日 17:42 3.06MB

下载



Deep Learning Book 深度学习 [带标签PDF完整版]


2017年12月09日 08:14 15.91MB


下载



基于深度学习的图像分类Image classification with deep learning常用模型

本文中，我会根据下大家image classification常用的cnn模型，针对cifar10（for 物体识别），mnist（for 字符识别）& ImageNet（for 物体识别）做一个mo...

 GarfieldEr007 2016年03月31日 12:52 5714

 (<http://blog.csdn.net/GarfieldEr007/article/details/51025323>)

MIT Deep Learning book 深度学习 中文+英文版本

2017年12月06日 17:16 28.88MB

下载



深度学习 英文原版 Deep Learning（Ian Goodfellow Yoshua Be...


2017年11月28日 14:38 24.15MB

下载



【deep learning】深度学习用于图片的分类和检测总结

【deep learning】深度学习用于图片的分类和检测总结

 qq_30214939 2017年06月11日 10:10 427

加入CSDN博客，享受更精准的内容推荐，与500万程序员共同成长！

登录

注册

✕

- （最新python版深度学习）Deep Learning with python

2017年12月06日 08:53

5.74MB

下载

ZIP

深度学习方法及应用 英文版 Deep Learning Methods and Appli...

2017年12月05日 12:28

7.6MB

下载

RAR

大牛的《深度学习》笔记，Deep Learning速成教程

雷锋网(搜索“雷锋网”公众号关注)按：本文由Zouxy责编，全面介绍了深度学习的发展历史及其在各个领域的应用，并解释了深度学习的基本思想，深度与浅度学习的区别和深度学习与神经网络之间的关系。深度学...

huangtingting1

2016年08月23日 00:05

5743

(http://blog.csdn.net/huangtingting1/article/details/52281889)

Neural Network and Deep Learning（神经网络和深度学习）中...

2017年11月05日 00:23

4.78MB

下载

PDF

deep learning（深度学习）中文版

2017年11月05日 00:26

26.5MB

下载

ZIP

Multimodal Deep Learning（多模态深度学习）未完待续

摘要： 提出一种在深度网络上的新应用，用深度网络学习多模态。特别的是，我们证明了跨模态特征学习——如果在特征学习过程中多模态出现了，对于一个模态而言，更好的特征可以被学习。此外，我们展示了如...

s2010241013

2016年06月22日 08:53

7464

(http://blog.csdn.net/s2010241013/article/details/51731657)

内容举报

返回顶部

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

登录

注册

第9页 共9页

2018/1/4 上午11:38