

原

随机梯度下降

2016年12月06日 17:28:39

为爱存在

阅读数：877

标签：

神经网络

深度学习

随机梯度下降

更多

随机梯度下降(stochastic gradient descent)

简单的介绍一下什么是随机梯度下降，为什么要用随机梯度下降这个方法。

1.背景

在我们进行深度学习的时候，对于神经网络的输出结果，我们需要知道结果对不对，以及每个神经元的阈值和权重对不对以及我调整一下神经网络的输出结果和我们预期的输出结果会更接近还是误差更大。如果更接近，那么我们可以继续调整权重和阈值，让神经网络输出结果等于预期何进行定量的分析呢？不能只依靠感觉，这个时候，我们引入函数

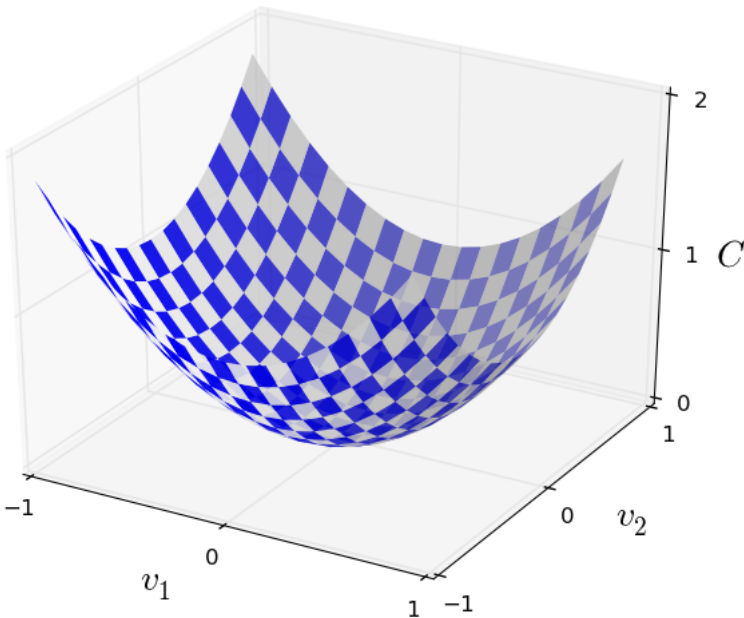
$$C(w,b) \equiv \frac{1}{2n} \sum_x \|y(x) - a\|^2.$$

其中 $y(x)$ 代表我们预期的输出结果，比如说，一个识别手写阿拉伯数字的神经网络，我们网络的输出就是一个10维的向量，比如(1, 0, 0, 0, 0, 0, 0, 0, 0, 0)代表0，(0, 1, 0, 0, 0, 0, 0, 0, 0, 0)代表1，而 a 代表一个神经网络实际输出的结果, w 和 b 就分别代表了权重和阈值。也就是说，我们如果想要让一个神经网络的结果更准确，就需要对 b ，让 C 更小。即求函数 C 的最小值。

也就是说，我们现在是要求一个函数的最小值，让我们仅仅从数学方面思考，如何求解一个函数的最小值。假设有一个函数 $C(v)$,而这个函数有俩个变量，首先想到的就是微积分，假如只有2个变量，我们就要求出函数的二阶偏导，然后再令二阶偏导等于0，算出几个点，再判断这几个点哪个是最大值，哪个是鞍点。可是问题是，我们处理的数据大部分都是多维的，甚至达到百万级别，所以用微积分的话，那就很难算出来了。因此我们就需要梯度下降的思想

2.梯度下降的思想

我们先假设在求一个二元函数的最小值，它的图像如图所示：



<http://blog.csdn.net/baiyu9821179>

假设我们在这个图像曲面上放置一个小球，它受到重力影响，肯定会往下滑落，而最终停止的地方，就是整个函数的最小值，这个就是梯度下降的思想。为了更清楚的说明这个问题,我们把 v_1 在 v_1 方向移动 Δv_1 ，在 v_2 在 v_2 方向移动 Δv_2 ，根据微积分，我们可以得到：

$$\nabla C \equiv \left(\frac{\partial C}{\partial v_1}, \frac{\partial C}{\partial v_2} \right)^T.$$

因此(3)可以改写成

$$\Delta C \approx \nabla C \cdot \Delta v.$$

那么我们如何选择 Δv 令 ∇C 为负数呢？我们可以选择令：

$$\Delta v = -\eta \nabla C$$

原式可变为 $\Delta C \approx -\eta \nabla C \cdot \nabla C = -\eta \|\nabla C\|^2$. 因为 $\|\nabla C\|^2 \geq 0$, 这就保证了 $\Delta C \leq 0$ 。那么 $-\eta$ 怎么求呢？因为 $\Delta v = -\eta \nabla C$ ，因此小，然后求出 $\|\nabla C\|$ 。那么 $-\eta$ 的值就求出来了。

$\frac{\|\Delta v\|}{\|\nabla C\|}$,我们可以事先规

3.随机梯度下降

根据前面的介绍，我们已经可以算出最小点在哪里了(理论上)，但是随机梯度下降还有一些问题，我在这里说两个：（1）容易陷入局部极小值，中，我们只画出了一个全局极小值点，所以梯度下降可以直接找到最小点，但是在实际中，函数会有很多局部极小值，因此梯度下降可能会停止在局部不是全局极小值。（2）计算量太大，注意到式(1)，我们计算所有输入图像的cost function，然后取平均，这样计算量太大了。因此，我们可以通过本里的 ∇C_x 来估计 ∇C 的值。我们随机选取输入 X_1, X_2, \dots, X_m ，把它们看做是一个迷你批(mini-batch). 样本数 m 要足够大以致于我们可以估计式所示：

$$\frac{\sum_{j=1}^m \nabla C_{X_j}}{m} \approx \frac{\sum_x \nabla C_x}{n} = \nabla C,$$

两边互换一下，即可得：

$$\nabla C \approx \frac{1}{m} \sum_{j=1}^m \nabla C_{X_j},$$

应用到神经网络上，可以得到下式：

$$\begin{aligned} w_k &\rightarrow w'_k = w_k - \frac{\eta}{m} \sum_j \frac{\partial C_{X_j}}{\partial w_k} \\ b_l &\rightarrow b'_l = b_l - \frac{\eta}{m} \sum_j \frac{\partial C_{X_j}}{\partial b_l}, \end{aligned}$$

其中 w 和 b 分别代表权值和阈值。当我们遍历完所有迷你批以后，接着在剩下的样本中选取第二个迷你批，直到穷尽所有样本，此时称为完成一次epoch。然后继续上述过程。

参考资料：《Neural Networks and Deep Learning》Michael Nielsen

会这门技术的程序员遭疯抢，年薪20万还是白菜价！

各大公司目前急缺大数据人才，随着技术与运用的成熟，应用集中于互联网、金融、医疗、新能源、通信和房地产等行业。

想对作者说点什么？

我来说两句

学习笔记13：随机梯度下降法（Stochastic gradient descent, SGD）

666

假设我们提供了这样的数据样本（样本值取自于 $y=3 \times x_1 + 4 \times x_2$ ）： $x_1 \ x_2 \ y$
1 4 192 5 265 1 194 2 29
 x_1 和 x_2 是样本值...

来自：kwame211的博客

感知机3 -- 梯度下降与随机梯度下降的对比

2019

声明：1，本篇为个人对《2012.李航.统计学习方法.pdf》的学习总结，不得用作商用。2，由于本人在学...

来自：xueyingxue001的专栏

从0到1学好区块链开发，一年编程经验学完月薪可达40K+

立即申请试学，成为时代颠覆者

随机梯度下降(Stochastic Gradient Descent)浅显理解

77

来自：qq_33720555的博客

批量梯度下降（BGD）、随机梯度下降（SGD）、小批量随机梯度下降（MSGD）实现过程详解

2829

原文地址：http://blog.csdn.NET/xiaoch1222/article/details/52847521 时间：2016.10.18 一、前言 ...

来自：wzy的博客

批梯度下降和随机梯度下降的区别和代码实现

1969

三种梯度下降的方式：批量梯度下降、小批量梯度下降、随机梯度下降

在机器学习领域中，梯度下降的方式有三种，分别是：批量梯度下降法BGD、随机梯度下降法SGD、小批量梯度下...

2万

来自：UESTC_C的博文

白杨女教师发现股票K线规律，声称按此方法炒股爆赚不亏！

知恩网络 · 嫵嫵

批量梯度下降和随机梯度下降

转自：http://www.cnblogs.com/walcott/p/4957098.html 梯度下降与随机梯度下降 梯度下降法先随机给出参数的...

1909

来自：u0100的博文

机器学习笔记（一）：梯度下降算法，随机梯度下降，正规方程

一、符号解释 M 训练样本的数量 x 输入变量，又称特征 y 输出变量，又称目标 (x, y) 训练样本，对应监督学习的输...

1.5万

来自的博文

深度学习中的随机梯度下降(SGD)简介

随机梯度下降(Stochastic Gradient Descent, SGD)是梯度下降算法的一个扩展。 机器学习中反复出现的一个...

676

来自：网络的限的

随机梯度下降法

一、考虑一下线性方程组

5.4万

来自：Forever-守望

文章热词

深度学习框架 选择 图片检测 深度学习 数据增强 深度学习 深度学习模型汇总 深度学习文字匹配

相关热词

随机梯度下降 随机梯度下降图解 异步随机梯度下降 随机梯度下降知乎 随机梯度下降简称

博主推荐



HHTNAN

关注

350篇文章



__gyl__

关注

94篇文章



Steven_ycs

关注

216篇文章

最优化与随机梯度下降

出处：http://blog.csdn.net/han_xiaoyang/article/details/50178505 声明：版权所有，转载请联系作者并注明出处 1...

1656

来自：Focusing on your own Min...

【机器学习笔记04】随机梯度下降

【参考资料】 【1】Spark MLlib 机器学习实践 【2】机器学习之梯度下降 http://www.cnblogs.com/rcfeng/p/39589...

42

来自：Fredric的笔记

深度学习—随机梯度下降（SGD）

2241

来自：Duckie的博客

2018 聪明的白杨人都在靠他赚外快！知道的人已经赚疯了！

骄荣 · 嫵嫵

随机梯度下降和批量梯度下降的区别

最近，看了斯坦福大学讲的梯度下降算法的视频，对其中的批量梯度下降算法(batch gradient descent algorithm , B...

1428

来自：NOT_GUY的博客

深度学习反向传播---随机梯度下降法

一、误差准则函数与随机梯度下降：数学一点将就是，对于给定的一个点集 (X, Y)，找到一条曲线或者曲面，对...

356

来自：weixin_38208741的博客

【sklearn第十一讲】随机梯度下降

机器学习训练营——机器学习爱好者的自由交流空间（qq 群号：696721295） 随机梯度下降(Stochastic Gradient De...

134

来自：wong2016的博客

10、Tensorflow：梯度下降、随机梯度下降和批量梯度下降

这几种方法呢都是在求最优解中经常出现的方法，主要是应用迭代的思想来逼近。在梯度下降算法中，都是围绕以...

1623

来自：Dean

神经网络算法学习---梯度下降和随机梯度下降

申明：本文非笔者原创，原文转载自：http://www.cnblogs.com/549294286/archive/2012/12/13/2817204.html ...

2350

来自：VivienFu的专栏



各种滤波器原理与设计

百度广告

Deep learning系列（十）随机梯度下降

介绍了随机梯度下降法在使用过程中的一些技巧，包括使用动量和在迭代过程中逐步更新学习率。...

来自：du_me的专栏

[deeplearning-001] stotistics gradient descent随机梯度下降算法的最简单例子解释

1.gradient descent梯度下降优化1.1假设要优化一个函数 $f(x)=(x-1)^2$ 。这个函数在 $x=1$ 时有最小值，这...

来自：大数据和机器学习研究



为爱存在

关注

原创84

粉丝246

喜欢25

评论16

等级：博客4

积分：1935

勋章：恒

访问：17万+

排名：2万+



虚拟主机如何用



最新文章

2 . 矩阵消元

1 . 方程组的几何解释

深度学习中比较经典的paper

VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

VIM常用命令

个人分类

机器学习3篇

深度学习7篇

Oracle1篇

凸优化1篇

计算机视觉3篇

展开

归档

2018年7月2篇

2018年6月1篇

2018年2月1篇

2018年1月1篇

2017年12月1篇

展开

热门文章

- numpy的array和python中自带的list之间相互转化

阅读量：78439
- 在Mac安装cmake command line tool

阅读量：16100
- ubuntu14.04安装CUDA8.0

阅读量：8929
- 编译caffe提示错误/usr/bin/ld: cannot find -lhdf5_hl

阅读量：7389
- Numpy中 tile函数的用法

阅读量：6103

最新评论

- 在Mac安装cmake comma...

u010090347：有用
- 在ubuntu编译caffe的ma...

m0_38001337：请问我用了这个脚本之后，就出现这么多问题了，这是什么原因？ [code=plain] MEX-fi...
- OSError: [Errno 1...

koma_csdn：请问您的这个报错解决了吗？ [reply]roguesir[/reply]
- C++重载输入运算符<<...

BeerBread134：请问，为什么要return stream;啊？求解释...
- C++重载输入运算符<<...

BeerBread134：thanks



人脸识别 闸机



联系我们



扫码联系客服



下载CSDN APP

- ✉ kefu@csdn.net

☎ 400-660-0108
- 🗨 QQ客服

🗨 客服论坛

关于我们 招聘 广告服务 网站地图
🔍 百度提供站内搜索 京ICP证09002463号
©2018 CSDN版权所有

经营性网站备案信息 网络110报警服务
北京互联网违法和不良信息举报中心
中国互联网举报中心