

神经网络入门之bp算法，梯度下降

原创 2017年07月23日 14:21:20 标签：神经网络

2682

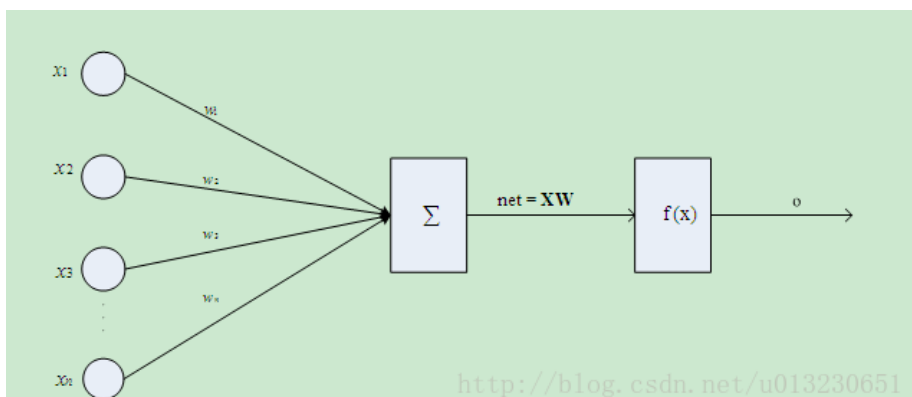
序言

本人作为一个想进行NLP研究的新手，看了很多网络上很好的神经网络的入门代码和数学原理。但是个人数学比较烂，很多东西想了很久才想明白，又害怕忘掉。为此写下这篇大白话入门篇作为自己学习的一个记录，也想为跟我同样想入门的同学们一个参考。希望有问题多多交流。

备注：很多内容都是本人自己想当然的结果，有错误的话，望大神们多多指教。

废话都说完了本文将从一个最简单一个BP网络开始讲起。

bp网络的bp（back propagation）中文就是反向传播的意思，为什么反向传播呢。是为了将配合梯度下降法进行迭代求出好的结果。这个会稍后讲解。



加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

x为输入，w为权重，这个f(x)被称为激活函数（activation function）。如sigmoid，tanh等。他们的特点有一个就是可以容易的求出他们的导数（很关键）。激活函数的意义可以看这篇神经网络激励函数的作用是什么？有没有形象的解释？

可以看到输出的 $o = f(w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4) = f(\sum w_i x_i)$ 。

第一次运算的结果很明显就是上边的o。但是此时问题出来了，运算出来的o和实际的结果肯定是有误差的，该如何利用这个误差优化这个运算呢？也就是得到好的w呢？

梯度下降法

此时就出现了反向传播这个过程。而配合反向传播的就是梯度下降法了。

现在很多同学可能会很晕，当时我第一次看的时候也很晕。

为了容易理解梯度下降法，建议去看Ng的斯坦福网课第二节，非常清楚。比我写的清楚多了。

梯度下降法是求（局部）最好的w。

设误差函数为：

$$J = \frac{1}{2} (y - o)^2 = \frac{1}{2} (y - f(\sum w_i x_i))^2$$

y为实际结果，o为预测结果。

设激活函数f(x)为sigmoid函数，此时就可以很方便的求出其导数了（其他激活函数也是一样）

$$f'(x) = f(x)(1 - f(x))$$

所以我们要求的就是J最小的时候wi的值。a是变化的速率。下式就可以比作从山顶走到山底的过程，而a表示行走的步长或者是速率。



我需要的孤独

原创 1 粉丝 0

等级：博客 1 访问

积分：40 排名：



达内的



文章分类

神经网络

登录

博主热门文章

神经网络入门之bp算法，
2657



网站开发的流程



联系我们



请扫描二

webr

400-6

QQ客

$$W_i = W_i - a \frac{\partial J}{\partial W_i}$$

$$\begin{aligned} \frac{\partial J}{\partial W_i} &= (y - o) * f'(\sum W_i X_i) * (\sum W_i X_i)' \\ &= (y - o) * f(\sum W_i X_i)(1 - f(\sum W_i X_i)) * X_i \end{aligned}$$

此时可以发现每一项都是可以求出的，则经过多次运算，可以求出好的 W_i
一般我们把前两项作为

$$\text{delta} = (y - o) * f(\sum W_i X_i)(1 - f(\sum W_i X_i))$$

此时我们可以发现

$$\frac{\partial J}{\partial W_i} = \text{后一层的delta} * \text{前一层的} X_i。$$

BP的原因

上边介绍完了梯度下降，现在再说反向传播理由。其实很简单了。它用的就是链式法则。我们第一步是前向传播，进行一系列运算得到了预测结果o。为了使用梯度下降法，我们需要得到，上边需要的delta，也就是说 J 这个误差函数。因为实际结果我们知道，而激活函数的导数我们也知道怎么运算。所以我们得到预测结果o时，delta就可以求出来。而delta属于输出层的运算，再乘以输入层的Xi就能得到 ∂W_i ，进一步更新 W_i 。
很明显可以看出整个一轮的运算是：

前向传播：

输入层—w—》输出层（sigmoid）—》预测结果

后向传播：

误差—》输出层(sigmoid)求导—》输入层—》更新 W_i

换句话说，BP算法就是提供了给梯度下降法所需要的所有值。

由链式法则可知，如果网络层数为3层以上时也可以得到每层的delta。

用python代码来说：

```
layer_n_delta = layer_n+1_delta.dot((W_n_n+1).T)
(W_n_n+1) += (Xn).T.dot(layer_n+1_delta)
```

上边有很多符号有点问题，但是我觉得阅读应该没有什么障碍。因为第一次用这个markdown编辑器，很多东西不好弄。有问题请和我讨论。谢谢大家。

[关于](#) [招聘](#) [广告服务](#)

©1999–2018 CSDN版权所有
京ICP证09002463号

[经营性网站备案信息](#)

[网络110报警服务](#)

[中国互联网举报中心](#)

[北京互联网违法和不良信息举报中心](#)



目前您尚未登录，请 [登录](#) 或 [注册](#) 后参与评论



Limk1m 2017-10-19 16:31 #1楼

[回复](#)

写的很好哇，我也是NLP新手，不过好像有个地方有小错误，你对wi求导那边少了个负号

神经网络的梯度下降法



liyiafeia 2016年07月27日 10:49 2276

常用的神经网络如BP神经网络、RBF神经网络等都有三层神经元，即输入层

、隐藏层和输出层。我们知道通过不断的修改神经元之间的权值和偏置使网络的输出能够拟合所有的训练输入，为了量化这个目标，引入一个代价函...

神经网络学习笔记（三） 梯度下降法



cyhbrilliant

2016年09月21日 22:52

📖 5215

梯度下降法在上一张，我们学习过了LMS算法，就是利用了著名的梯度下

降法，但是LMS算法只是一种特殊的实现，是均方差这个特定函数的梯度下降，这次我们来看一下梯度下降对普通函数求极值的一些应用。我们来试一...

为什么互联网公司都在做小程序开发？

微信请来了硅谷大佬一起开发了门小程序课，了解一下

广告



BP神经网络梯度下降算法



prom1201

2016年06月23日 18:21

📖 8379

菜鸟初学人智相关问题，智商低，艰苦学习中，转文只为保存，其中加上了

一些个人注释，便于更简单的理解~新手也可以看，共勉。 转自博客园@编程De： <http://www.cnblogs.com/...>

神经网络与深度学习笔记——神经网络与梯度下降



AndyTeen

2017年10月08日 11:38

📖 441

利用神经网络进行手写体识别 两种重要的人工神经网络：感知机，sigmoid

神经元。神经网络标准学习算法：sgd（随机梯度下降） Perceptrons(感知机) 二进制...

学Python后到底能干什么？

python是啥

百度广告



神经网络梯度下降优化算法及初始化方法小结



bea_tree

2017年10月09日 18:53

📖 599

An overview of gradient descent optimization algorithms and Weight initi

alization methods. 神经网络...

从梯度下降到拟牛顿法：详解训练神经网络的五大学习算法

在神经网络中，系统的学习过程一般是由训练算法所主导。而现如今有许多不同的学习算法，它们每一个都有不同的特征和表现。因此本文力图描述清楚五大学习算法的基本概念及优缺点，给读者们阐明最优化在神经网络中的应...



roslei

2017年03月12日 21:22

📖 1750

BP算法怎样使用梯度下降方法调整权重



tcbdz

2014年05月18日 11:17

📖 3082

最近看论文时，无意间看到B

BP算法与公式推导



Lu597203933

2015年06月20日 21:58

📖 26826

BP(backpropagationalgorithm)：后向传导算法，顾名思义就是从神

经网络的输出(顶层)到输入(底层)进行求解。那么求解什么呢，求解的就是神经网络中的参数的导数，即参数梯度方向，从...

BP算法浅谈（Error Back-propagation）



pennyliang

2011年08月17日 16:32

📖 51763

最近在打基础，大致都和向量有关，从比较基础的人工智能常用算法开始

，以下是对BP算法研究的一个小节。 本文只是自我思路的整理，其中举了个例子，已经对一些难懂的地方做了解释，有兴趣恰好学到人...

BP与梯度下降的关系



P081513083

2017年11月10日 19:16

📖 400

梯度下降是一种优化算法，通过参数的迭代更新的方式来求最优或次优参数

。 $W = W' + \text{梯度}(W)$ BP（误差反向传播算法）是在神经网络领域中，网络权重的梯度的求法。梯度（W）...

腾讯爸爸竟然和外国企业打造小程序开发？

席位有限，立即行动！抢先掌握稀缺技术，成为抢手人才



神经网络中的梯度下降算法

zhanglu_wind2018年01月03日 10:16168

一、概述
梯度下降算法（Gradient Descent Optimization）是神经网络模型训练最常用的优化算法。对于深度学习模型，基本都是采用梯度下降算法来进行优化训练的。梯度下降算法背后的...

神经网络之梯度下降法和反向传播BP

B083701082017年04月18日 22:28690

梯度下降法和反向传播网上资料非常多，记录点自己理解的 1.梯度下降法
是为了使损失函数求最小，而梯度方向是函数增长最快的方向，前面加个负号就变成函数减少最快的方向：...

BP神经网络

ACdreamers2015年03月26日 22:28117540

今天来讲BP神经网络，神经网络在机器学习中应用比较广泛，比如函数逼近，模式识别，分类，数据压缩，数据挖掘等领域。接下来介绍BP神经网络的原理及实现。 Contents 1. BP神经网络...

机器学习中常见问题_几种梯度下降法

u0104027862016年04月19日 14:0910961

随机梯度下降 批梯度下降 minibatch

BP神经网络原理推导（学习笔记）

xujie1262017年01月23日 10:371044

BP神经网络算法部分原理推导（梯度下降法） 20160715xujie 以下图所示网络为例，来源于网上某位大神的BP网络C代码，根据C代码源码进行理解。 上...

学Python后到底能干什么？

python是啥

百度广告

BP神经网络模型及梯度下降法

dcrmg2017年06月26日 23:444894

BP（Back Propagation）网络是1985年由Rumelhart和McClland为首的科学家小组提出，是一种按误差逆传播算法训练的多层前馈网络，是目前应用最广泛的神经网络模型之一。 B...

深入浅出|深度学习算法之BP神经网络 详细公式推导

BP(back propagation)神经网络一种按照误差逆向传播算法训练的多层前馈神经网络，是目前应用最广泛的神经网络。...

weifenglin19972017年08月26日 22:594537

BP算法精彩讲解集合

u0142212662015年09月21日 10:481784

脉络清晰的BP神经网络讲解：http://www.cnblogs.com/wengzilin/archive/2013/04/24/3041019.html BP神经网络模型与学习算法：http://...

BP算法

xueyingxue0012016年10月21日 15:252206

BP算法小问答 Q：BP算法是什么？ A：对于机器学习，基本上任何算法最后都会通过SGD(随机梯度下降)来修正权重，而对于深度学习，其修正权重的方法就是BP算法。 Q：它怎么修正权重？ A：神经...

从神经网络到BP算法（纯理论推导）

FernChen2016年07月08日 15:2211968

本文会从最基本的神经网络结构开始，一步步推导，最终得到一个神经网络利用BP算法进行训练的完整过程，以及中间会用到的公式的推导。 ...