

peghoty

学习是一种态度!

RSS订阅

个人资料

?

皮果提

关注

原创

104

粉丝

2659

喜欢

213

评论

897

等级：

博客

7

访问：167万+

积分：2万+

排名：518

贵

PC建站费用1万元

PHP

MySQL

APP建站费用1万元

ASP.NET

手机建站费

微站建设费用1万元

SQLSERVER

建站周期

HTML

网站建设

网站做完后无法更改样式

关键字

做网站需要多少钱

归档

2014年11月

1篇

2014年10月

4篇

2014年8月

1篇

2014年7月

6篇

2014年6月

4篇

展开

热门文章

word2vec 中的数学原理详解（一）目录和前言  
阅读量：141844

word2vec 中的数学原理详解（四）基于 Hierarchical Softmax 的模型  
阅读量：83222

word2vec 中的数学原理详解（三）背景知识  
阅读量：59585

word2vec 中的数学原理详解（五）基于 Negative Sampling 的模型  
阅读量：59145

word2vec 中的数学原理详解（二）预备知识  
阅读量：58235

word2vec 中的数学原理详解...

LINTEBI：博主，方便的话可以发一份PDF给我吗？感激不尽！感激不尽！604156331@qq.com

word2vec 中的数学原理详解...

Qinjian666：NB闪闪的博主

word2vec 中的数学原理详解...

leayc：#6.9.3: <https://github.com/RaRe-Technologies/gens...>

word2vec 中的数学原理详解...

p03721：博主可以发一份PDF吗？感谢！875051204@qq.com

受限玻尔兹曼机（RBM）学习笔记（...

jocelynxyq：赞楼主的分享！

## 原 牛顿法与拟牛顿法学习笔记（三）DFP 算法

2014年03月24日 00:52:06

阅读数：22670

机器学习算法中经常碰到非线性优化问题，如 **Sparse Filtering** 算法，其主要工作在于求解一个非线性极小化问题。在具体实现中，大多调用的是成熟的软件包做支撑，其中最常用的一个算法是 **L-BFGS**。为了解这个算法的数学机理，这几天做了一些调研，现把学习过程中理解的一些东西整理出来。

### 目录链接

- (1) 牛顿法
- (2) 拟牛顿条件
- (3) DFP 算法
- (4) BFGS 算法
- (5) L-BFGS 算法

## §2.2 DFP 算法

DFP 算法是以 William C. Davidon、Roger Fletcher、Michael J. D. Powell 三个人的名字的首字母命名的, 它由 Davidon 于 1959 年首先提出, 后经 Fletcher 和 Powell 加以发展和完善, 是最早的拟牛顿法. 该算法的核心是: 通过迭代的方法, 对  $H_{k+1}^{-1}$  做近似. 迭代格式为

$$D_{k+1} = D_k + \Delta D_k, \quad k = 0, 1, 2, \dots \quad (2.22)$$

其中的  $D_0$  通常取为单位矩阵  $I$ . 因此, 关键是每一步的校正矩阵  $\Delta D_k$  如何构造.

注意, 迭代格式 (2.22) 将嵌套在算法 1.2 中, 因此, 我们猜想  $\Delta D_k$  可能与  $s_k$ ,  $y_k$  和  $D_k$  发生关联. 这里, 我们采用“待定法”, 即首先将  $\Delta D_k$  待定成某种形式, 然后结合拟牛顿条件 (2.21) 来进行推导.

那将  $\Delta D_k$  待定成什么形式呢? 这个说起来比较 tricky, 我们将其待定为

$$\Delta D_k = \alpha \mathbf{u} \mathbf{u}^T + \beta \mathbf{v} \mathbf{v}^T, \quad (2.23)$$

其中,  $\alpha, \beta$  为待定系数,  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$  为待定向量. 从形式上看, 这种待定公式至少保证了矩阵  $\Delta D_k$  的对称性 (因为  $\mathbf{u} \mathbf{u}^T$  和  $\mathbf{v} \mathbf{v}^T$  均为对称矩阵).

将 (2.23) 代入 (2.22), 并结合指导条件 (2.21), 可得

$$\mathbf{s}_k = D_k \mathbf{y}_k + \alpha \mathbf{u} \mathbf{u}^T \mathbf{y}_k + \beta \mathbf{v} \mathbf{v}^T \mathbf{y}_k, \quad (2.24)$$

从 (2.24) 似乎也看不出什么啊! 别急, 我们将其改写一下

$$\begin{aligned} \mathbf{s}_k &= D_k \mathbf{y}_k + \mathbf{u}(\alpha \mathbf{u}^T \mathbf{y}_k) + \mathbf{v}(\beta \mathbf{v}^T \mathbf{y}_k) \\ &= D_k \mathbf{y}_k + (\alpha \mathbf{u}^T \mathbf{y}_k) \mathbf{u} + (\beta \mathbf{v}^T \mathbf{y}_k) \mathbf{v} \end{aligned} \quad (2.25)$$

看到了吧? 括号中的  $\alpha \mathbf{u}^T \mathbf{y}_k$  和  $\beta \mathbf{v}^T \mathbf{y}_k$  是两个数, 既然是数, 我们不妨作如下简单赋值

$$\alpha \mathbf{u}^T \mathbf{y}_k = 1, \quad \beta \mathbf{v}^T \mathbf{y}_k = -1, \quad (2.26)$$

即

$$\alpha = \frac{1}{\mathbf{u}^T \mathbf{y}_k}, \quad \beta = -\frac{1}{\mathbf{v}^T \mathbf{y}_k}, \quad (2.27)$$

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

×

其中向量  $\mathbf{u}, \mathbf{v}$  仍有待确定.

那么,  $\mathbf{u}, \mathbf{v}$  如何确定呢? 将 (2.26) 代入 (2.25), 得到

$$\mathbf{u} - \mathbf{v} = \mathbf{s}_k - D_k \mathbf{y}_k, \quad (2.28)$$

要上式成立, 不妨直接取

$$\mathbf{u} = \mathbf{s}_k, \quad \mathbf{v} = D_k \mathbf{y}_k, \quad (2.29)$$

再将 (2.29) 代入 (2.27), 便得

$$\alpha = \frac{1}{\mathbf{s}_k^T \mathbf{y}_k}, \quad \beta = -\frac{1}{(D_k \mathbf{y}_k)^T \mathbf{y}_k} = -\frac{1}{\mathbf{y}_k^T D_k \mathbf{y}_k}, \quad (2.30)$$

其中第二个等式用到了  $D_k$  的对称性.

至此, 我们已经将校正矩阵  $\Delta D_k$  构造出来了, 将 (2.29) 和 (2.30) 代入 (2.23), 便得

$$\Delta D_k = \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{s}_k^T \mathbf{y}_k} - \frac{D_k \mathbf{y}_k \mathbf{y}_k^T D_k}{\mathbf{y}_k^T D_k \mathbf{y}_k}. \quad (2.31)$$

综上, 我们给出 DFP 算法的一个完整算法描述.

#### 算法 2.1 (DFP 算法)

1. 给定初值  $\mathbf{x}_0$  和精度阈值  $\epsilon$ , 并令  $D_0 = I$ ,  $k := 0$ .
2. 确定搜索方向  $\mathbf{d}_k = -D_k \cdot \mathbf{g}_k$ .
3. 利用 (1.13) 得到步长  $\lambda_k$ , 令  $\mathbf{s}_k = \lambda_k \mathbf{d}_k$ ,  $\mathbf{x}_{k+1} := \mathbf{x}_k + \mathbf{s}_k$ .
4. 若  $\|\mathbf{g}_{k+1}\| < \epsilon$ , 则算法结束.
5. 计算  $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$ .
6. 计算  $D_{k+1} = D_k + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{s}_k^T \mathbf{y}_k} - \frac{D_k \mathbf{y}_k \mathbf{y}_k^T D_k}{\mathbf{y}_k^T D_k \mathbf{y}_k}$ .
7. 令  $k := k + 1$ , 转至步 2.

## 参考文献

- [1] <http://www.materialssimulation.com/node/625>
- [2] <http://www.codelast.com/?p=2780>
- [3] [http://www.tydxq.cn/kuai\\_su/youhuasheji/suanfayuanli/3.1.asp](http://www.tydxq.cn/kuai_su/youhuasheji/suanfayuanli/3.1.asp)
- [4] [http://en.wikipedia.org/wiki/BFGS\\_method](http://en.wikipedia.org/wiki/BFGS_method)
- [5] [http://en.wikipedia.org/wiki/Sherman-Morrison\\_formula](http://en.wikipedia.org/wiki/Sherman-Morrison_formula)
- [6] Dai Y H. A perfect example for the BFGS method[J]. Mathematical Programming, 2013, 138(1-2): 501-530.
- [7] Liu D C, Nocedal J. On the limited memory BFGS method for large scale optimization[J]. Mathematical programming, 1989, 45(1-3): 503-528.
- [8] Nocedal J. Updating quasi-Newton matrices with limited storage[J]. Mathematics of computation, 1980, 35(151): 773-782.
- [9] <http://baike.baidu.com/view/6062086.htm>

作者: peghoty

出处: <http://blog.csdn.net/itplus/article/details/21896981>

欢迎转载/分享, 但请务必声明文章出处.

版权声明: 本文为博主原创文章, 未经博主允许不得转载。 <https://blog.csdn.net/peghoty/article/details/21896981>

文章标签: [拟牛顿法](#) [DFP算法](#) [Davidon](#) [Fletcher](#) [Powell](#)

个人分类: 数学天地

### 从小白到AI工程师的学习经验分享


这是转型AI的励志故事, 从非科班到拿下阿里云栖一等奖, 他经历的坑足够你学习100天! 以下为他的正文分享, 你可以清晰地看到他趟过的每一个坑, 希望借他的肩, 让你勇敢前行。

[查看更多>>](#)

12522

[查看更多>>](#)

想对作者说点什么? [我来说一句](#)

 随煜而安 2015-06-27 09:32:07 #2楼

[查看回复\(1\)](#)

博主我想问一下, 怎么保证每一个由迭代公式求出的 $D(k+1)$ 都是正定的? 最开始 $D_0$ 是正定的,  $\Delta D_0$ 是对称的。或者说一个正定矩阵加上一个对称矩阵怎么证明是正定的?

 Mr\_Jia\_Yuan 2015-01-04 09:16:10 #1楼

[查看回复\(3\)](#)

博主写的很好, 求教一下: 1.  $(2.26)$  怎么来的, 为什么可以直接等于 $+1$ ; 2.  $(2.28) \rightarrow (2.29)$  这样有什么依据, 能简单解释一下吗

## 优化算法——拟牛顿法之DFP算法

一、牛顿法 在博文“优化算法——牛顿法(Newton Method)”中介绍了牛顿法的思路，牛顿法具有二阶收敛性，相比较最速下降法，收敛的速度更快。在牛顿法中使用到了函数的二阶导数的信息，对...

 google19890102 2015-05-19 22:11:04 阅读数：6811

## 最优化学习笔记(十八)——拟牛顿法(4)DFP算法

秩2算法可以保证在任意第 $k$ 步迭代下，只要一维搜索是精确的，近似矩阵 $H_k$ 就是正定的。DFP算法 令 $k=0$ ,选择初始点 $x(0)$

 chunyun0716 2017-01-15 11:21:27 阅读数：887

## 苏坡90后小伙在家无聊玩微信，存款惊呆父母

威能科技 · 顶新


## 优化算法——拟牛顿法之L-BFGS算法

一、BFGS算法 在“优化算法——拟牛顿法之BFGS算法”中，我们得到了BFGS算法的校正公式：利用Sherman-Morrison公式可对上式进行变换，得到令，则...

 google19890102 2015-06-06 17:39:57 阅读数：6697


## 拟牛顿法、DFP算法及BFGS算法

拟牛顿法/Quasi-Newton DFP算法/Davidon-Fletcher-Powell BFGS算法/Broyden-Fletcher-Goldfarb-Shanno 转载...

 bestlinjialyin 2014-03-03 20:29:38 阅读数：1813

## 拟牛顿法之DFP算法

拟牛顿法(Quasi-Newton Methods)是求解非线性优化问题最有效的方法之一，于20世纪50年代由美国Argonne国家实验室的物理学家W. C. Davidon所提出来。Davidon设...

 lming\_08 2015-04-26 20:55:32 阅读数：1306

## 最优化 - 拟牛顿法DFP算法

一、牛顿法 在博文“优化算法——牛顿法(Newton Method)”中介绍了牛顿法的思路，牛顿法具有二阶收敛性，相比较最速下降法，收敛的速度更快。在牛顿法中使用到了函数的二阶导数的信息，对...

 ZK\_J1994 2017-06-20 13:30:53 阅读数：565

## 手机新出的赚钱方式，第一批人已经赚嗨翻了！！！！

盛图科技 · 顶新

## 拟牛顿法/Quasi-Newton，DFP算法/Davidon-Fletcher-Powell，及BFGS算法/Broyden-Fletcher-Goldfarb-Shanno

转自：<http://www.codelast.com/?p=2780> 在最优化领域，有几个你绝对不能忽略的关键词：拟牛顿、DFP、BFGS。名字很怪，但是非常著名。下面会依次地说明它们分别是什么...

 u014568921 2015-06-17 14:35:44 阅读数：1100

## 优化算法——拟牛顿法之DFP

转载地址：<http://blog.csdn.net/google19890102/article/details/45848439> 一、牛顿法 在博文“优化算法——牛顿法(Newton ...

 u012341849 2017-04-05 16:31:51 阅读数：299

## 牛顿法与拟牛顿法

牛顿法求函数的根牛顿法的最初提出是用来求解方程的根的。我们假设点 $x^*$ 为函数 $f(x)$ 的根，那么有 $f(x^*)=0$ 。现在我们把函数 $f(x)$ 在点 $x_k$ 处一阶泰勒展...

 batuwuhanpei 2016-07-21 11:49:43 阅读数：8316

## 最优化学习笔记(十九)——拟牛顿法(5)BFGS算法

一、BFGS算法的更新公式 为了推导BFGS算法，需要用到对偶或者互补的概念，前边已经讨论过hessian矩阵逆矩阵的近似矩阵需要满足以下条件： $H_k+1\Delta g(i)=\Delta x(i)0\leq i\leq k$  \b...

 chunyun0716

2017-02-12 10:34:08

阅读数：2428

优化算法——拟牛顿法之BFGS算法

一、BFGS算法简介 BFGS算法是使用较多的一种拟牛顿方法，是由Broyden，Fletcher，Goldfarb，Shanno四个人分别提出的，故称为BFGS校正。 同DFP校...


 google19890102

2015-05-20 11:31:14

阅读数：13885

算法细节系列（3）：梯度下降法，牛顿法，拟牛顿法

算法细节系列（3）：梯度下降法，牛顿法，拟牛顿法迭代算法原型话不多说，直接进入主题。在我看来，不管是梯度下降法还是牛顿法，它们都可以归结为一个式子，即  $x=\phi(x)$   $x=\phi(x)$  也就是...

 u014688145

2016-12-16 09:17:14

阅读数：3941

梯度下降、牛顿法、拟牛顿法

介绍 在向量微积分中，标量场的梯度是一个向量场。标量场中某一点上的梯度指向标量场增长最快的方向，梯度的长度是这个最大的变化率。更严格的说，从欧几里得空间 $R^n$ 到 $R$ 的函数的梯度是在 $R^n$ 某一点最佳的线性...

 a819825294

2016-08-10 18:50:23

阅读数：7144

程序猿不会英语怎么行？英语文档都看不懂！

老司机教你用数学公式读懂天下英文→



拟牛顿法

转自:ACdreamer 今天，我来讲一种在机器学习中常用到的优化算法，叫做BFGS算法。BFGS算法被认为是数值效果最好的拟牛顿法，并且具有全局收敛性和超线性收敛速度。那么接下...

 lishuandao

2016-03-31 16:10:35

阅读数：1917

牛顿法与拟牛顿法学习笔记（五）L-BFGS 算法

机器学习算法中经常碰到非线性优化问题，如 Sparse Filtering 算法，其主要工作在于求解一个非线性极小化问题。在具体实现中，大多调用的是成熟的软件包做支撑，其中最常用的一个算法是 L-BF...

 peghoty

2014-03-24 00:53:27

阅读数：45952

牛顿法与拟牛顿法学习笔记（一）牛顿法

机器学习算法中经常碰到非线性优化问题，如 Sparse Filtering 算法，其主要工作在于求解一个非线性极小化问题。在具体实现中，大多调用的是成熟的软件包做支撑，其中最常用的一个算法是 L-BF...

 peghoty

2014-03-24 00:51:18

阅读数：57082

最优化学习笔记（十四）——共轭梯度法

共轭梯度法不需要预先给定 $Q$ 共轭方向，而是随着迭代的进行不断产生 $Q$ 共轭方向。在每次的迭代中，利用上一个搜索方向和目标函数在当前迭代点的梯度向量 之...

 chunyun0716

2016-12-18 21:32:04

阅读数：5175

无约束最优化方法——牛顿法、拟牛顿法、BFGS、LBFGS

好久没写博客了，今天打开一看csdn终于可以用latex，不用到处去粘贴标签，方便了许多。且先试试效果如何。先讲讲一些优化方法。 最速下降法 牛顿法 拟牛顿法 SF 1 BFGS DFP LBFGS 【...

 lansatiankongxxc

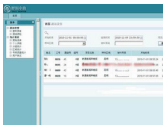
2015-05-20 17:44:32

阅读数：4544

呼叫中心系统

专业的呼叫中心系统

百度广告



梯度-牛顿-拟牛顿优化算法和实现

一、线搜索技术和Armijo准则 二、最速下降法 三、牛顿法 四、阻尼牛顿法 五、牛顿法的修正算法 六、拟牛顿法 6.1、 DFP算法及其Python实现 6.2、BFGS算法及其..

 u012478504

2015-05-20 10:10:11

阅读数：8277

牛顿、拟牛顿法以及其他优化方法的R实现

\quad牛顿法 ( Newton method ) 和拟牛顿法 ( quasi Newton method ) 是求解无约束最优化问题的常用方法，有收敛速度快的优点。1. 牛顿法考虑无约束最优化问题  $n$   $inx\in R...$