

CSDN

首页 博客 学院 下载 GitChat TinyMind ...

搜博主文章

写博客 发Chat 登录 注册

peghoty

学习是一种态度!

RSS订阅

个人资料

皮果提

关注

原创	粉丝	喜欢	评论
104	2659	213	897

等级： 博客 7

访问：167万+

积分：2万+

排名：518

重庆旅游攻略

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

最新评论

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/gensim>...

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（四）BFGS 算法

2014年03月24日 00:53:04

阅读数：44638

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

§2.3 BFGS 算法

BFGS 算法是以其发明者 **Broyden**, **Fletcher**, **Goldfarb** 和 **Shanno** 四个人的名字的首字母命名的。与 DFP 算法相比，BFGS 算法性能更佳。目前它已成为求解无约束非线性优化问题最常用的方法之一。BFGS 算法已有较完善的局部收敛理论，对其全局收敛性的研究也取得了重要成果。

BFGS 算法中核心公式的推导过程和 DFP 完全类似，只是互换了其中 s_k 和 y_k 的位置。为方便自己以后查阅，我打算将上一节的推导过程再重写一遍，已经领会该过程的读者则不妨跳过以下推导，直接看结论。

需要注意的是, BFGS 算法是直接逼近海森矩阵, 即 $B_k \approx H_k$. 仍采用迭代方法, 设迭代格式为

$$B_{k+1} = B_k + \Delta B_k, \quad k = 0, 1, 2, \dots \quad (2.32)$$

其中的 B_0 也常取为单位矩阵 I . 因此, 关键是每一步的校正矩阵 ΔB_k 如何构造. 同样, 将其待定为

$$\Delta B_k = \alpha \mathbf{u} \mathbf{u}^T + \beta \mathbf{v} \mathbf{v}^T. \quad (2.33)$$

将 (2.33) 代入 (2.32), 并结合指导条件 (2.20), 可得

$$\mathbf{y}_k = B_k \mathbf{s}_k + (\alpha \mathbf{u}^T \mathbf{s}_k) \mathbf{u} + (\beta \mathbf{v}^T \mathbf{s}_k) \mathbf{v}, \quad (2.34)$$

通过令 $\alpha \mathbf{u}^T \mathbf{s}_k = 1$, $\beta \mathbf{v}^T \mathbf{s}_k = -1$, 以及

$$\mathbf{u} = \mathbf{y}_k, \quad \mathbf{v} = B_k \mathbf{s}_k, \quad (2.35)$$

可算得

$$\alpha = \frac{1}{\mathbf{y}_k^T \mathbf{s}_k}, \quad \beta = -\frac{1}{\mathbf{s}_k^T B_k \mathbf{s}_k}, \quad (2.36)$$

综上, 便得到了如下校正矩阵 ΔB_k 的公式

$$\Delta B_k = \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{B_k \mathbf{s}_k \mathbf{s}_k^T B_k}{\mathbf{s}_k^T B_k \mathbf{s}_k}. \quad (2.37)$$

好了, 现在把矩阵 ΔB_k 和 ΔD_k 拿出来对比一下, 是不是除了将 D 换成 B 外, 其它只是将 s_k 和 y_k 互调了一下位置呢?

最后, 我们给出 BFGS 算法的一个完整算法描述.

算法 2.2 (BFGS 算法 (I))

1. 给定初值 \mathbf{x}_0 和精度阈值 ϵ , 并令 $B_0 = I$, $k := 0$.
2. 确定搜索方向 $\mathbf{d}_k = -B_k^{-1} \cdot \mathbf{g}_k$.
3. 利用 (1.13) 得到步长 λ_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 $B_{k+1} = B_k + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{B_k \mathbf{s}_k \mathbf{s}_k^T B_k}{\mathbf{s}_k^T B_k \mathbf{s}_k}$.
7. 令 $k := k + 1$, 转至步 2.

算法 2.2 中的步 2 通常是通过求解线性代数方程组 $B_k \mathbf{d}_k = -\mathbf{g}_k$ 来进行. 然而, 更一般的做法是, 通过对步 6 中的递推关系应用 **Sherman-Morrison** 公式, 直接给出 B_{k+1}^{-1} 与 B_k^{-1} 之间的关系式 ([4, 6])

$$B_{k+1}^{-1} = \left(I - \frac{\mathbf{s}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \right) B_k^{-1} \left(I - \frac{\mathbf{y}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \right) + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}, \quad (2.38)$$

或者进一步展开写成 ([4, 6])

$$B_{k+1}^{-1} = B_k^{-1} + \left(\frac{1}{\mathbf{s}_k^T \mathbf{y}_k} + \frac{\mathbf{y}_k^T B_k^{-1} \mathbf{y}_k}{(\mathbf{s}_k^T \mathbf{y}_k)^2} \right) \mathbf{s}_k \mathbf{s}_k^T - \frac{1}{\mathbf{s}_k^T \mathbf{y}_k} (B_k^{-1} \mathbf{y}_k \mathbf{s}_k^T + \mathbf{s}_k \mathbf{y}_k^T B_k^{-1}), \quad (2.39)$$

其中的蓝色部分表示实数.

注 2.1 关于 *Sherman-Morrison* 公式 ([5])

设 $A \in \mathbb{R}^n$ 为非奇异方阵, $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, 若 $1 + \mathbf{v}^T A^{-1} \mathbf{u} \neq 0$, 则有

$$(A + \mathbf{u} \mathbf{v}^T)^{-1} = A^{-1} - \frac{A^{-1} \mathbf{u} \mathbf{v}^T A^{-1}}{1 + \mathbf{v}^T A^{-1} \mathbf{u}}. \quad (2.40)$$

利用 (2.38), 我们很容易将算法 2.2 改写成算法 2.3. 注意, 为了避免出现矩阵求逆符号, 我们统一将 B_i 换用 D_i (这样做仅仅只是符号上看起来舒服起见). 这样, 整个算法中不再需要求解线性方程组, 由矩阵 - 向量运算就可以完成了.

算法 2.3 BFGS 算法 (II)

1. 给定 \mathbf{x}_0 和精度阈值 ϵ , 并令 $D_0 = I$, $k := 0$.
2. 确定方向 $\mathbf{d}_k = -D_k \cdot \mathbf{g}_k$.
3. 利用 3) 得到步长 λ_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 $D_{k+1} = \left(I - \frac{\mathbf{s}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \right) D_k \left(I - \frac{\mathbf{y}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \right) + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}$.
7. 令 $k := k + 1$, 转至步 2.

至此, 关于 DFP 算法和 BFGS 算法的介绍就完成了, 回过头来, 我们对比一下算法 2.1 和算法 2.3, 容易发现, 这两个算法的唯一不同仅在于 D_{k+1} 的迭代公式不同罢了.

最后, 再补充谈谈一维搜索 (line search) 的问题, 在之前几个算法描述中, 为简单起见, 均采用 (1.13) 来计算步长 λ_k , 其实这是一种**精确搜索**. 实际应用中, 还有像 Wolfe 型搜索、Armijo 搜索以及满足 Goldstein 条件的**非精确搜索**. 这里我们以 Wolfe 搜索为例, 简单做个介绍.

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

是指 λ_k 满足如下 Wolfe 条件 ([7])

$$\begin{aligned} f(\mathbf{x}_k + \lambda_k \mathbf{d}_k) &\leq f(\mathbf{x}_k) + \beta \lambda_k \mathbf{d}_k^T \mathbf{g}_k; \\ \mathbf{d}_k^T \mathbf{g}(\mathbf{x}_k + \lambda_k \mathbf{d}_k) &\geq \beta \mathbf{d}_k^T \mathbf{g}_k. \end{aligned} \quad (2.41)$$

带非精确搜索的拟牛顿法的研究是从 1976 年 Powell 的工作开始的，他证明了带 Wolfe 搜索的 BFGS 算法的全局收敛性和超线性收敛性。

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>
- [4] http://en.wikipedia.org/wiki/BFGS_method
- [5] 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/21897443>

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/21897443>

文章标签： [拟牛顿法](#) [BFGS](#) [精确搜索](#) [Wolfe条件](#)

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](#)

- 

gyksfwt877074

2017-11-02 17:49:49

#8楼

查看回复(1)

感觉好棒，这些是从书里整理的吗？想看看完整的书
- 

Schnee_y

2017-09-10 15:27:10

#7楼

查看回复(1)

亲，这个是什么书里面的。方便发我一份电子版的吗？万分感谢！
- 

你大哥不会起名字

2017-03-09 11:00:54

#6楼

查看回复(1)

应该是 $B_k \approx (H_k)^{-1}$ 吧

查看 12 条热评

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

MATLAB编写的BFGS算法

2010年09月08日 1008B

下载



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

BFGS优化算法的理解以及LBFGS源码求解最优化问题

关于最优化求解，吴军有篇blog讲的很不错，证明和解释都很清楚，具体可以参考http://www.cnblogs.com/joneswood/archive/2012/03/11/2390529.ht...
👤 pizibing880909 2014-03-13 20:09:45 阅读数：6631

多变量函数优化的L-BFGS算法matlab程序

2014年08月16日 29KB

下载



LBFGS算法使用

一、首先把两个作者 JORGE NOCEDAL 给的文件看明白sdrive.f和lbfgs.f sdrive.f - is a simple driver lbfgs.f - conta...
👤 settingsun1225 2011-01-15 20:03:00 阅读数：19528

bfgs算法-matlab源程序

2010年09月15日 10KB

下载



苏坡股民福利，这款价值6800元的炒股软件终于免费了！

益盟操盘手 · 顶新

BFGS算法

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。那么接下来将会详细讲解。 Contents ...
👤 ACdreamers 2015-03-27 01:55:41 阅读数：32400

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

最优化学笔记(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

闲扯数学规划问题(3)-等式约束和不等式约束

数学规划问题， $\min_{x_1, x_2} f(x_1, x_2) = c(1)$ $\begin{array}{l} \min & f(x_1, x_2) \\ \text{s.t.} & g(x_1, x_2) = \dots \end{array}$

quicmous 2016-11-29 17:07:38 阅读数：1156

L-BFGS算法（有限内存BFGS算法）

kaka19880812 2015-08-02 19:26:49 阅读数：426

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-...

yanhx1204 2018-02-13 14:55:33 阅读数：34

从牛顿法到L-BFGS的算法演变

前言(本文主要学习自该博主的文章：<http://blog.csdn.net/itplus>，以下是本人的笔记，主要记录了结论部分，省略了推导的部分。对具体推导过程有兴趣的同学请访问原博主的博客~) ...

qq547276542 2017-09-10 09:12:22 阅读数：2807

5分钟完成加壳，防止代码反编译

Virbox Protector加壳工具，堪比VMP安全强度的加壳工具



机器学习优化算法—L-BFGS

L-BFGS算法由于其高效的性能而被广泛运用在实际工程中，本文首先介绍L-BFGS算法和其它算法的比较，然后详细介绍该算法的主要思想以及每一步迭代时近似矩阵的更新细节。...

chenzhijay 2014-11-27 09:53:21 阅读数：5066

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

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 \hat{H}_{k+1} ，则...

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

大规模优化算法 - LBFGS算法

L-BFGS算法比较适合在大规模的数值计算中，具备牛顿法收敛速度快的特点，但不需要牛顿法那样存储Hesse矩阵，因此节省了大量的空间以及计算资源。本文主要通过对于无约束最优化问题的一些常用算法总结，一...

qq_23052951 2017-02-16 22:10:47 阅读数：2425

L-BFGS算法

前面的文章中，讲述过BFGS算法，它是用来求解最优化问题的，在这个算法中，相对于普通的牛顿迭代法有很大的改进。链接：<http://blog.csdn.net/acdreamers/article/de...>

 ACdreamers 2015-03-29 18:12:23 阅读数：18510

呼叫中心系统

热门呼叫中心系统大全

百度广告



L-BFGS

关于优化算法的求解，书上已经介绍了很多的方法，比如有梯度下降法，坐标下降法，牛顿法和拟牛顿法。梯度下降法是基于目标函数梯度的，算法的收敛速度是线性的，并且当问题是病态时或者问题规模较大时，收敛速度尤其...

 seasermymy 2015-11-05 16:35:33 阅读数：1231

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 SR1 BFGS DFP LBFGS 【...

 lansatiankongxxc 2015-05-20 17:44:32 阅读数：4544