

学习是一种态度!





等级: 博客 7 访问: 167万+ 积分: 2万+ 排名: 518



归档

1篇 2014年11月 2014年10月 4篇 1篇 2014年8月 6篇 2014年7月 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 Hi

erarchical 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 N

egative Sampling 的模型

阅读量:59145

word2vec 中的数学原理详解 (二)预备知

阅读量:58235

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...

 $leayc: \verb§\#6.9.3: https://github.com/RaRe-Technolo$

gies/gens...

word2vec 中的数学原理详解...

p03721:博主可以发一份PDF吗?感谢!875051

204@qq.com

受限玻尔兹曼机(RBM)学习笔记(...

jocelynxyq: 赞楼主的分享!

廖 牛顿法与拟牛顿法学习笔记(一)牛顿法

2014年03月24日 00:51:18 阅读数: 57081

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

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 提出. Davidon 设计的这种算法在当时看来是非线性优化领域最具创造性的发明之一.不久 R. Fletcher 和 M. J. D. Powell 证实了这种新的算法远比其他方法快速和可靠,使得非线性优化这门学科在一夜之间突飞猛进. 在之后的 20 年里, 拟牛顿方法得到了蓬勃发展, 出现了大量的变形公式以及数以百计的相关论文 ([9]).

DFP 方法、BFGS 方法以及 L-BFGS 算法都是重要的拟牛顿法,本文将对这些方法进行简要介绍. 当然,在介绍拟牛顿法之前,我们先看看什么是牛顿法,以及拟牛顿法和牛顿法之间有什么关系. 为此,考虑如下**无约束的极小化问题**

$$\min_{\mathbf{x}} f(\mathbf{x}), \tag{0.1}$$

其中 $\mathbf{x} = (x_1, x_2, \cdots, x_N)^T \in \mathbb{R}^N$. 由于本文不准备对收敛性理论进行讨论, 因此不妨对**目标 函数** $f: \mathbb{R}^N \to \mathbb{R}$ 作一个较苛刻的假设, 这里我们假定 f 为**凸函数**, 且**两阶连续可微**. 此外, 记极小问题 (0.1) 的解为 x^* .

§1 牛顿法

§1.1 原始牛顿法

为简单起见, 首先考虑 N=1 的简单情形 (此时目标函数 $f(\mathbf{x})$ 变为 f(x)). 牛顿法的基本思想是: **在现有极小点估计值的附近对** f(x) 做二阶泰勒展开, 进而找到极小点的下一个估计值. 设 x_k 为当前的极小点估计值, 则

$$\varphi(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2}f''(x_k)(x - x_k)^2$$
(1.2)

表示 f(x) 在 x_k 附近的二阶泰勒展开式 (略去了关于 $x-x_k$ 的高阶项). 由于求的是最值, 由极值必要条件可知, $\varphi(x)$ 应满足

$$\varphi'(x) = 0, (1.3)$$

即

$$f'(x_k) + f''(x_k)(x - x_k) = 0, (1.4)$$

从而求得

$$x = x_k - \frac{f'(x_k)}{f''(x_k)},\tag{1.5}$$

于是, 若给定初始值 x_0 , 则可以构造如下的迭代格式

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}, \quad k = 0, 1, \cdots$$
 (1.6)

产生序列 $\{x_k\}$ 来逼近 f(x) 的极小点. 在一定条件下, $\{x_k\}$ 可以收敛到 f(x) 的极小点.

对于 N>1 的情形, 二阶泰勒展开式 (1.2) 可以做推广, 此时

$$\varphi(\mathbf{x}) = f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k) \cdot (\mathbf{x} - \mathbf{x}_k) + \frac{1}{2} \cdot (\mathbf{x} - \mathbf{x}_k)^T \cdot \nabla^2 f(\mathbf{x}_k) \cdot (\mathbf{x} - \mathbf{x}_k), \tag{1.7}$$

其中 ∇f 为 f 的**梯度向量**, $\nabla^2 f$ 为 f 的**海森矩阵** (Hessian matrix), 其定义分别为

注意, ∇f 和 ∇ 中的元素均为关于 \mathbf{x} 的函数, 以下分别将其简记为 \mathbf{g} 和 H. 特别地, 若 f 的混合偏导数 换次序 (即对 \forall i, j, 成立 $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$), 则海森矩阵 H 为**对称矩阵**. 而 $\nabla f(\mathbf{x}_k)$ 和 ∇^2 则表示将 \mathbf{x} 取为 \mathbf{x}_k 后得到的实值向量和矩阵, 以下分别将其简记为 \mathbf{g}_k 和 H_k (这里字 \mathfrak{S} 表示 gradient, H 表示 **H**essian).

同样地, 由于是求极小点, 极值必要条件要求它为 $\varphi(\mathbf{x})$ 的**驻点**, 即

$$\nabla \varphi(\mathbf{x}) = 0, \tag{1.9}$$

亦即 (通过在 (1.7) 两边作用一个梯度算子)

$$\mathbf{g}_k + H_k \cdot (\mathbf{x} - \mathbf{x}_k) = 0. \tag{1.10}$$

进一步, 若矩阵 H_k 非奇异, 则可解得

$$\mathbf{x} = \mathbf{x}_k - H_k^{-1} \cdot \mathbf{g}_k,\tag{1.11}$$

于是, 若给定初始值 x_0 , 则同样可以构造出迭代格式

$$\mathbf{x}_{k+1} = \mathbf{x}_k - H_k^{-1} \cdot \mathbf{g}_k, \quad k = 0, 1, \cdots$$
 (1.12)

这就是**原始的牛顿迭代法**,其迭代格式中的搜索方向 $\mathbf{d}_k = -H_k^{-1} \cdot \mathbf{g}_k$ 称为**牛顿方向**. 下面给 电一个**牛顿注**的宗教管注描述

加入CSDN,享受更精准的内容推荐,与500万程序员共同成

长!

- 1. 给定初 $_{\mathbb{S}_{\mathbb{R}}}$ 精 $_{\mathbb{H}}$ ϵ , 并 $_{\mathbb{S}}$ k:=0.
- 2. 计算 \mathbf{g}_k 和 H_k .
- 3. 若 $\|\mathbf{g}_k\| < \epsilon$, 则停止迭代; 否则确定搜索方向 $\mathbf{d}_k = -H_k^{-1} \cdot \mathbf{g}_k$.
- 4. 计算新的迭代点 $\mathbf{x}_{k+1} := \mathbf{x}_k + \mathbf{d}_k$.
- 5. $\Diamond k := k+1$, 转至步 2.

当目标函数是二次函数时,由于二次泰勒展开函数与原目标函数不是近似而是完全相同的二次式,海森矩阵退化成一个常数矩阵,从任一初始点出发,利用 (1.12) 只需一步迭代即可达到 $f(\mathbf{x})$ 的极小点 x^* ,因此牛顿法是一种具有**二次收敛性**的算法.对于非二次函数,若函数的二次性态较强,或迭代点已进入极小点的邻域,则其收敛速度也是很快的,这是牛顿法的主要优点.

但原始牛顿法由于迭代公式中没有步长因子, 而是定步长迭代, 对于非二次型目标函数, 有时会使函数值上升, 即出现 $f(\mathbf{x}_{k+1}) > f(\mathbf{x}_k)$ 的情况, 这表明原始牛顿法**不能保证函数值稳定地下降**, 在严重的情况下甚至可能造成迭代点列 $\{\mathbf{x}_k\}$ 的发散而导致计算失败.

§1.2 阻尼牛顿法

为消除 §1.1 中提到的关于原始牛顿法的弊病, 人们提出了"**阻尼牛顿法**", 阻尼牛顿法 每次迭代的方向仍采用 \mathbf{d}_k , 但每次迭代需沿此方向作一维搜索 (line search), 寻求最优的步长因子 λ_k , 即

$$\lambda_k = \arg\min_{\lambda \in \mathbb{R}} f(\mathbf{x}_k + \lambda \mathbf{d}_k). \tag{1.13}$$

下面给出一个阻尼牛顿法的完整算法描述.

算法 1.2 (阻尼牛顿法)

- 1. 给定初值 x_0 和精度阀值 ϵ , 并令 k := 0.
- 2. 计算 \mathbf{g}_k 和 H_k .
- 3. 若 $\|\mathbf{g}_k\| < \epsilon$, 则停止迭代; 否则确定搜索方向 $\mathbf{d}_k = -H_k^{-1} \cdot \mathbf{g}_k$.
- 4. 利用 (1.13) 得到步长 λ_k , 并令 $\mathbf{x}_{k+1} := \mathbf{x}_k + \lambda_k \mathbf{d}_k$.
- 5. 令 k := k + 1, 转至步 2.

注 1.1 算法 1.2 的步 3 中涉及到 H_k^{-1} 的计算, 实际应用中, 通常并不直接对 H_k 进行求逆, 而是将其转化为求解线性代数方程组 $H_k \mathbf{d}_k = -\mathbf{g}_k$, 此时可根据系数矩阵 H_k 的性态来选择适合的迭代法, 如预条件共轭梯度法 (PCG)、代数多重网格法 (AMG) 等.

注 1.2 有些文献资料里将算法 1.2 步 3 中的搜索方向写成 $\mathbf{d}_k = H_k^{-1} \cdot \mathbf{g}_k$, 这是没问题的,原因是在步 4 中有一个求步长的过程,将搜索方向写成 $\mathbf{d}_k = H_k^{-1} \cdot \mathbf{g}_k$, 无非是求得的最佳步长和原来求得的最佳步长相差一个符号罢了.

至此已完成了牛顿法的算法介绍,接下来对其做个小结([3]):

牛顿法是梯度(下降)法的进一步发展,梯度法利用目标函数的一阶偏导数信息、以负梯度方向作为搜索方向,只考虑目标函数在迭代点的局部性质;而牛顿法不仅使用目标函数的一阶偏导数,还进一步利用了目标函数的二阶偏导数,这样就考虑了梯度变化的趋势,因而能更全面地确定合适的搜索方向以加快收敛,它具二阶收敛速速.但牛顿法主要存在以下两个缺点:

- 对目标函数有较严格的要求. 函数必须具有连续的一、二阶偏导数, 海森矩阵必须正定.
- 计算相当复杂. 除需计算梯度而外, 还需计算二阶偏导数矩阵和它的逆矩阵. 计算量、存储量均很大, 且均以维数 N 的平方比增加, 当 N 很大时这个问题更加突出.

参考文献

- [1] http://www.materialssimulation.com/node/625
- [2] http://www.codelast.com/?p=2780
- [3] http://www.tydxyq.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/21896453

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/21896453

个人分类: 数学天地

为什么说大数据工程师会在2018年越过越滋润?

大数据生态系统在2017年终于实现了火力全开!这次我们总结了下,给大家完整的展现下大数据工程师的详情,希望这篇分享可以帮到所有程序员!

查看更多>>

13596

查看更多>>

想对作者说点什么?

我来说一句

qq_41933057 2018-04-27 11:29:03 #14楼

为什么我的拟牛顿法会出现 梯度差为0的情况啊 但是出来的结果明显是没有优化好的 求大神帮我分析一下

gyksfwit877074 2017-11-04 11:54:15 #13楼

尊重他人知识产权,转载请标明 http://blog.csdn.net/itplus/article/details/21896453

ssdut_209 2017-10-21 15:30:15 #12楼

写得真好

查看 16 条热评

牛顿法与拟牛顿法学习笔记(三)DFP 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 00:52:06 阅读数: 22669

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

机器学习——优化算法:牛顿法-伪代码描述算法 - CSDN博客

至此,整个牛顿法就能已经在自己心中有个蓝图。 本次主要是用伪代码来描述整个算法过程,对于一个码农来说,感觉很有意思。 伪代码: ... 2018-4-11

拟牛顿法——伪代码描述 - CSDN博客

3、由牛顿法中,Xk的递推关系式可得: 近似策略有: 1、逼近Hession矩阵--DFP 2、逼近Hession矩阵的逆矩阵--BFGS DFP伪代码描述: ;(PDM表示为正定矩阵) ...

2018-3-18

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

再谈 牛顿法/Newton's Method In Optimization

转自:http://www.codelast.com/?p=8052 牛顿法是最优化领域的经典算法,它在寻优的过程中,使用了目标函数的二阶导数信息,具体说来就是:用迭代点的梯度和二阶导数对目...

🏶 u014568921 2015-06-17 14:28:47 阅读数:1078

梯度下降法,牛顿法,高斯-牛顿迭代法,附代码实现 - CSDN博客

例如,改进的随机梯度下降算法,伪代码如下: /***...从本质上去看,牛顿法是二阶收敛,梯度下降是一阶收敛...

2018-5-3

数值计算方法程序代码

数值计算方法:最小二乘法的曲线拟合、四阶龙格库塔法求解常微分方程、牛顿法...数值计算方法 数值积分(伪代码 c/c++ python) 数值计算方法程序 代码 立即下载...

2018-5-4

牛顿法解机器学习中的Logistic回归

这仍然是近期系列文章中的一篇。在这一个系列中,我打算把机器学习中的Logistic回归从原理到应用详细串起来。本文将结合之前介绍的Logistic回归的原理,解释牛顿法在优化问题中的应用,特别是在求解...

参 baimafujinji 2016-04-18 19:26:09 阅读数: 7484

【斯坦福---机器学习】复习笔记之牛顿方法

本讲大纲:1.牛顿方法(Newton's method) 2.指数族(Exponential family) 3.广义线性模型(Generalized linear models)1.牛顿方法假设...

Andrewseu 2015-07-06 16:42:25 阅读数:9499

数值最优化 - CSDN博客

其中精确线性搜索有公式法和黄金分割法,非精确线性搜索主要有Armijo型线性搜索和Wolfe-Powell型线性搜索,都是比较简单的一元函数知识,因为步长是一元的。由此,数值...

2018-5-3

牛顿迭代求根算法的分析与实现 论文 完整版

摘要:牛顿迭代法是《数值分析》这门课程中一个重要的计算方法和思想。这次的课程设计是通过在学习中所学习到的牛顿迭代的方法的思想计算方程:求方程 x3+x2-3x-3...

2018-5-7

Andrew Ng机器学习课程之学习笔记---牛顿方法

牛顿方法 本次课程大纲: 1、 牛顿方法:对Logistic模型进行拟合 2、 指数分布族 3、 广义线性模型(GLM):联系Logistic回归和最小二乘模型 复习: Logi...

() wjk7186912 2017-06-03 11:32:40 阅读数:404

新出的微投资赚钱方式,2018聪明的苏坡人都在闷声发大财!

仕永投资管理·顶新

Logistic回归 - CSDN博客

牛顿法 局部加权回归(Locally weighted linear regression) Pavel and barbecue...找到最佳回归系数,也就是拟合出Logistic回归模型的最佳参数,梯度上升法的伪代码如...

2018-5-3

梯度下降、牛顿法、拟牛顿法 - CSDN博客

(2)随机梯度下降伪代码: 每个回归参数初始化为1 重复R次 计算每个样本的梯度...一般认为牛顿法可以利用到曲线本身的信息,比梯度下降法更容易收敛(迭代更少次数)...

2018-5-5

【机器学习-斯坦福】学习笔记4 - 牛顿方法

网"主题有奖征文 【机器学习-斯坦福】学习笔记4 - 牛顿方法 分类: 机器学习2013-10-10 12:45 74人阅读 评论(0) 收藏 举报 机器学习牛...

● pi9nc 2013-10-10 14:53:27 阅读数:8859

最优化——拟牛顿方法matlab程序

 $\% \ BFGS \ function \ [x, output] = bfgs(fun, \ dfun, \ x0, \ varargin) \ \% \ Step \ 1: initialization \ epsi = 1.0e-...$

● beatbean 2012-12-31 15:57:04 阅读数: 3101

牛顿迭代法求根 - CSDN博客

牛顿迭代法求根的原理: 设r是的根,选取作为r的初始近似值,过点做曲线的...背包问题---01背包(原理,伪代码,编程实现) 3986 eclipse 项目转as项目

2018-3-29

近200篇机器学习&深度学习资料分享(含各种文档,视频,源..._CSDN博客

介绍:加州伯克利大学博士 Aria Haghighi 写了一篇超赞的数值优化博文,从牛顿法讲到拟牛顿法,再讲到 BFGS 以及L-BFGS, 图文并茂,还有伪代码。强烈推荐。《简明深度…

2018-5-5

Matlab拟牛顿法以及实例

2012年12月26日 756B 下载



Logistic回归与牛顿法 (附Matlab实现)

回归,是一种连续模型,受噪声的影响较大,一般都是用来做预测的,但也有除外,比如本文要讲的Logistic回归就是用来做分类的。 Logistic Regre ss Logistic一般用于二分...

● u013571752 2014-03-12 20:06:49 阅读数:3823

理解牛顿法

牛顿法与梯度下降法相比,收敛速度更快,在搜索空间中进行二阶收敛,即以椭圆曲面去逼近最优解,也可以将牛顿法看作二次曲面下的梯度下降法。牛顿法对于凸二次最优问题,迭代一次即可得到最优解。首先给出无约束最优...

C shijing_0214 2016-04-11 21:17:48 阅读数:1869

牛顿法 Newton Method

上一次我们讨论了具有 Q-线性收敛性的普通的 gradient descent 方法,今天我们要介绍一种收敛速度更快的算法:Newton Method(或者叫 Newton's Method)。 ...

🚇 GarfieldEr007 2016-01-08 19:46:52 阅读数: 1162

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

C#程序加壳,虚拟机外壳,强度堪比VMP

集自动代码移植、混淆、外壳加密于一身,无需编程就能达到极高的保护强度



【机器学习详解】解无约束优化问题:梯度下降、牛顿法、拟牛顿法

无约束优化问题是机器学习中最普遍、最简单的优化问题。 x*=minx f(x),x∈Rnx^*=min_{x}\ f(x),x\in R^n1.梯度下降梯度下降是最简单的迭代优化 算法,每一次迭代需求…

🔞 luoshixian099 2016-07-06 20:58:22 阅读数:6100

牛顿迭代法 && 高斯牛顿法

相关知识 多元函数的泰勒展开: 原文:http://blog.sina.com.cn/s/blog_6819cb9b0100mkkx.html 雅克比矩阵和海森矩阵: 原文:http://jaco...

🧐 lgfun 2016-11-21 17:32:09 阅读数:6822

最优化学习笔记(一)——牛顿法(一维搜索方法)

一、一维搜索方法讨论目标函数为一元单值函数f:R \rightarrow Rf: \mathbb{R}\to \mathbb{R}\phi的最优化问题的迭代求解方法。二、局部极小点的条件n元实值函数ff的一阶导数DfDf为: ...

● chunyun0716 2016-04-17 12:02:46 阅读数: 2289

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

优化算法——牛顿法(Newton Method)

一、牛顿法概述

● google19890102 2014-11-13 22:00:59 阅读数:25935

免费云主机试用一年

云主机免费推荐吗

百度广告



(斯坦福机器学习课程笔记)牛顿法算法学习

代码均使用PYTHON3.x 牛顿法算法复杂些,作以下笔记,免忘。 下图(来自http://www.myexception.cn/cloud/1987100.html,侵删) 为了找到 f(...

cson qq_32231743 2016-12-20 19:18:05 阅读数: 2048

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 $x*x^*$ 为函数f(x)f(x)的根,那么有 $f(x*)=0f(x^*)=0$ 。现在我们把函数 f(x)f(x)在点 xkx_k 上处一阶泰勒展…

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