



码龄6年 暂无认证

63 8 7 4 2万+
原创 粉丝 获赞 评论 访问

904 44 4万+ 10万+
积分 收藏 周排名 总排名 等级



TA的主页

私信

关注

搜博文文章



笔记本电脑排名

最新文章

搭建配置DHCP服务

Cursor

Transaction processing

Python enviroment management

Derivatives_in_Matrix

分类专栏



Web

1篇



Algorithm

19篇



Markdown

4篇



Data

17篇



Programming

5篇



Project

7篇



归档

2019

10月

1篇

8月

1篇

7月

16篇

6月

12篇

5月

10篇

4月

14篇

3月

7篇

2月

3篇

热门文章

CNN的优缺点 7804

Install TensorFlow in Anaconda 2796



点赞



评论



分享



收藏

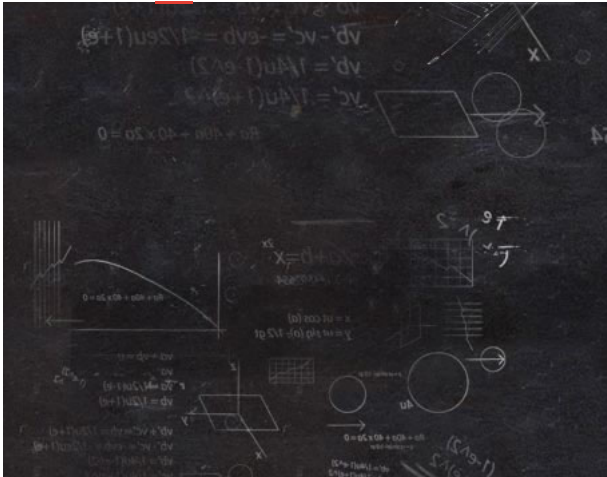


手机看

...



关注



最新评论

Optimizer in Tens...
lrglgy: [reply]gwb281386172[/reply]刚开始看源码都头大，个人感觉看多了就不...
Optimizer in Tens...
gwb281386172: [reply]lrglgy[/reply]嗯嗯，看源码我好头大，看不懂。。老哥有...
Optimizer in Tens...
lrglgy: [reply]gwb281386172[/reply]感谢您的阅读，最近忙着搬砖呢，有时间一...
Optimizer in Tens...
gwb281386172: 我以为你要解读一下1.2呢。。老哥，我满怀期待，这个get_slot...

目录

目录

- 第5章 初学牛顿法
 - 5.1 牛顿法
 - 5.2 拟牛顿法
 - 5.3 残留问题
 - 5.4 参考

Newton's Method

BogeyDa 2019-05-10 18:40:03 279 收藏 版权

分类专栏: Algorithm

目录

第1章 决策树

- 1 基本流程
- 2 流程核心（划分选择）
 - 2.1 信息增益
 - 2.1.1 概念
 - 2.1.2 劣势
 - 2.2 增益率
 - 2.2.1 概念
 - 2.2.2 劣势
 - 2.3 基尼指数
 - 2.3.1 概念
 - 2.3.2 使用
- 3 优化（过拟合 属性过多）
 - 3.1 预剪枝
 - 3.2 后剪枝
 - 3.3 多变量决策树
- 4 特殊值（连续与缺失值）
 - 4.1 连续值

- 5 参考

第2章 数据预处理

- 2.1 归一化
- 2.2 特征二值化
- 2.3 独热编码
- 2.4 缺失值计算
- 2.5 数据变换
- 2.6 样本不均衡
- 2.7 参考

第3章 特征选择

- 3.1 特征选择简介
- 3.2 过滤式 (Filter)
 - 3.2.1 方差过滤法
 - 3.2.2 皮尔森相关系数
 - 3.2.3 互信息和最大信息系数
 - 3.2.4 信息增益
- 3.3 包裹式 (Wrapper)
 - 3.3.1 递归特征消除
 - 3.3.2 Las Vegas Wrapper
- 3.4 嵌入式 (Embedding)
 - 3.5 参考

第4章 梯度下降算法总览

- 4.1 梯度下降算法
 - 4.1.1 批量梯度下降算法 (BGD)
 - 4.1.2 随机梯度下降算法 (SGD)
 - 4.1.3 最小批量梯度下降算法 (MBGD)
- 4.2 梯度下降算法的优化
 - 4.2.1 动量 (Momentum)
 - 4.2.2 Nesterov_momentum
 - 4.2.3 AdaGrad
 - 4.2.4 Adadelta
 - 4.2.5 RMSprop
 - 4.2.6 Adam

- 4.3 参考

第5章 初学牛顿法

- 5.1 牛顿法
- 5.2 拟牛顿法
- 5.3 残留问题
- 5.4 参考

第6章 拟牛顿法详解

- 6.1 DFP算法
- 6.2 BFGS算法
- 6.3 参考

第7章 牛顿法收敛性

[7.2 理论用度](#)[7.3 参考](#)

第8章 拟牛顿法中的正定性

[8.1 问题及解决](#)[8.2 参考](#)

第5章 初学牛顿法

本文是对牛顿法和拟牛顿法的公式进行简单的推导以及介绍，主要解决如何使用牛顿法（即不管怎么来，怎么用），我会在后续博文中继续学习总结它是怎么来的。

本文将分为两个部分，第一部分是牛顿法介绍；第二部分为拟牛顿法的简单介绍。

牛顿法

本文是对牛顿法的公式进行简单的推导以及介绍，主要解决如何使用牛顿法（即不管怎么来，为什么，怎么用），我会在后续博文中继续学习总结它是怎么来的。

牛顿法是在梯度下降算法的基础上引入二阶导数，从而加快优化算法的收敛速度。具体算法思路如下：

有无约束最优化问题：

$$\min_{x \in R^n} f(x) \quad (4.1)$$

假设 $f(x)$ 有二阶连续偏导数， x_k 是第 k 次迭代值，可得 $f(x)$ 的二阶泰勒展开式如下：

$$f(x) = f(x_k) + g_k(x - x_k) + \frac{1}{2}(x - x_k)^T H_k(x - x_k) \quad (4.2)$$

其中 $g_k = g(x_k) = f'(x_k)$ ， H_k 是 $f(x)$ 的海森矩阵（Hesse matrix）。

令上式对 x 求导得：

$$\nabla f(x) = g_k + H_k(x - x_k) \quad (4.3)$$

假设在 x_{k+1} 点满足：

$$\nabla f(x_{k+1}) = 0 \quad (4.4)$$

将式子(3)中的 x 替换为 x_{k+1} 得：

$$g_k + H_k(x_{k+1} - x_k) = 0 \quad (4.5)$$

求解式子(4.5)得：

$$x_{k+1} = x_k - \frac{g_k}{H_k} \quad (4.6)$$

令 $H_k p_k = -g_k$ ，可得：

$$x_{k+1} = x_k + p_k \quad (4.7)$$

将式子(4.6)作为迭代公式的算法就是牛顿算法。

牛顿法

输入：目标函数 $f(x)$ ，梯度 $g(x) = \nabla f(x)$ ，海森矩阵 $H(x)$ ，精度要求 ϵ ；

输出： $f(x)$ 的极小值点 x^* 。

取初始点 x_k ，置 $k = 0$ ；

计算 $g_k = g(x_k)$ ；

若 $\|g_k\| < \epsilon$ ，停止计算，返回近似解 $x^* = x_k$ ；

计算 $H_k = H(x_k)$ ，并求 p_k ：

$$H_k p_k = -g_k$$

置 $x_{k+1} = x_k + p_k$ ；

[点赞](#)[评论](#)[分享](#)[收藏](#)[手机看](#)

...

[关注](#)

优点：1. 牛顿法为二阶收敛，收敛速度快；

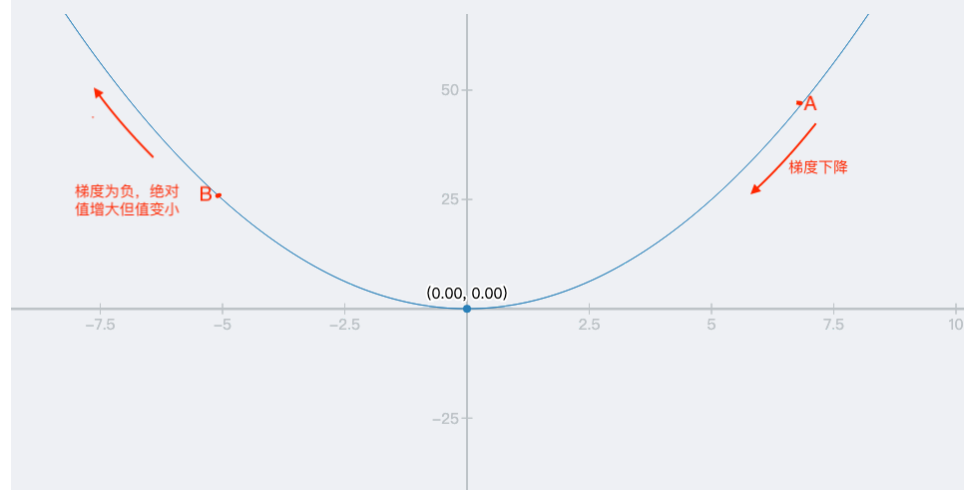
缺点：1. 牛顿法每一步都要求目标函数海森矩阵的逆矩阵，计算复杂度较大；
2. 可能发生被零除错误，因为在零点附近，导数非常小，运算易出现被零除错误；

跳过的坑

为什么求导后的式子 (4.3) 没有 $f(x_k)$ ？

这个问题很简单，我们是对 x 进行求导，而 $f(x_k)$ 中并没有 x_k ，因此求导后 $f(x_k)$ 为零。

我一直陷入一个误区，认为优化就是沿着梯度的反方向移动，以凸函数为例：



初始点为A时，沿梯度反方向（即梯度下降的方向）移动，可以到达最优解，但初始点为B时，沿梯度下降方向移动不是越移越远了吗？

其实我理解是错误的，更加规范的说法是**沿负梯度方向移动**，当位于A点时，梯度为正，负梯度即向梯度减小的方向移动，而B点梯度为负，负梯度是向梯度增大的方向移动。所以不管梯度正还是负，沿梯度负方向移动都是向极小值点移动。

拟牛顿法

由于牛顿法中求海森矩阵及其逆矩阵较为复杂，所以考虑用一个n阶矩阵 G 来代替海森矩阵 H 或其逆矩阵如下所示：

$$\begin{aligned}x_{k+1} &= x_k - G_{k+1}g_k \\ \text{或} \\ -g_k &= G_{k+1}(x_{k+1} - x_k) \\ \text{替代} \\ x_{k+1} &= x_k - H_k^{-1}g_k\end{aligned}$$

什么矩阵 G 是 n 阶矩阵？

我的个人理解是，矩阵可视为一种映射，而这里是将 n 维向量 $(x_{k+1} - x_k)$ 映射到 n 维向量 $(g_{k+1} - g_k)$ ，所以矩阵 G 应为 n 阶矩阵，否则等式无法成立。

为了实现替代，矩阵 G 要满足 H^{-1} 所满足的条件：

满足公式 $x_{k+1} = x_k - G_{k+1}g_k$ ；

G 必须为正定矩阵；

满足条件一可以使用假设法，假设 G_{k+1} 满足等式，并设计 G_{k+1} 的迭代表达方式，然后进行逆推，求出真正表达式。而不同的假设方式对应于拟牛顿法的两种不同算法DFP和BFGS，详情见博客[第7章 拟牛顿法解](#)

满足条件二，可设置 G 的初始值为正定矩阵并在迭代过程中保证其正定性不变。

优点：1. 拟牛顿法为二阶收敛，收敛速度快；

2. 拟牛顿法利用迭代方式求 G_{k+1} ，计算复杂度比牛顿法小；

缺点：暂时还没学习到。



点赞



评论



分享



收藏



手机看



关注

x_{k+1} 是如何假设的?

的解答: [第6章 拟牛顿法详解](#)

为什么要那样假设?

的解答: [第6章 拟牛顿法详解](#)

什么牛顿法收敛速度更快?

的解答: [第7章 牛顿法收敛性](#)

牛顿法如何保证 G_{k+1} 的正定性?

的解答: [第8章 拟牛顿法中的正定性](#)

参考

基百科.黑塞矩阵

outube)CCU MOOCs.3.4 Newton Method 牛顿法

iyong_will.优化算法——牛顿法(Newton Method)

outube)有道学堂公开课.Taylor展式与拟牛顿

优缺点:

胡子皮-.最优方法: 牛顿迭代法和拟牛顿迭代法

法中如何求出 P_k 和 Q_k :

果提.牛顿法与拟牛顿法学习笔记 (三) DFP 算法

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

迭代法(Newton's Method)

無名黑洞 4万+

代法 (简称牛顿法) 由英国著名的数学家牛顿爵士最早提出。但是, 这一方法在牛顿生前并未公开发表。牛顿法的作...

on法 (牛顿法 Newton Method)

春华秋实 1万+

常看到牛顿法怎样怎样, 一直不得要领, 今天下午查了一下维基百科, 写写我的认识, 很多地方是直观理解, 并没有...

优质评论可以帮助作者获得更高权重

抢沙发

评论

构-树(tree)

蜗牛君的奋斗史 2669

关基础概念及其表示方法

ian矩阵和Hessian矩阵

MemRay 2万+

bian转载自: http://jacoxu.com/?p=146在向量分析中, 雅可比矩阵是一阶偏导数以一定方式排列成的矩阵, 其行列...

【方法】牛顿法——Newton Method

ACM_hades的博客 557

顿法主要有两个应用方向: 求方程的根求函数最优化求解二、求方程的根: 假设我们现在要求方程 $f(x)=0$ $f(x)=0$ $f(x)=...$

ent Descent (二)

likyoo的博客 716

Gradient Descent 需要注意什么呢? Tuning your learning rates η 的大小控制不好, 会造成很多问题: 如果 η 太小...

lessen-burg矩阵的特征值

06-09

essen-burg矩阵的特征值,求上Hessen-burg矩阵的特征值

习笔记之多元梯度下降法

Steven的博客 2338

降法作为机器学习中很基础的方法, 是我们必须要掌握的。随机以一组数据作为例子: $x_1x_2x_3...x_n...y$ 72011363...26...

讲明白梯度下降Gradient Descent (以求解多元线性回归参数为例)

化简可得 253

录1.梯度2.梯度下降梯度下降算法在机器学习中出现频率特别高, 是非常常用的优化算法。本文借多元线性回归, 用...

度下降方法对比 (Batch gradient descent、Mini-batch gradient desc...

天泽28的专栏 1万+

度下降方法对比 (Batch gradient descent、Mini-batch gradient descent 和 stochastic gradient descent) &a...

习之Hessian矩阵在牛顿法中的应用

维函数, 每个点在每一个方向上的导数是不同的, 如果使用梯度

点赞

评论

分享

收藏

手机看

...

关注

下降法首先梯度下降法不是一个机器学习算法，而是一种基于搜索的最优化方法。它的作用是最小化一个损失函数。...

ent Descent

liupc的学习笔记 1975

毅视频官网: <http://speech.ee.ntu.edu.tw/~tlkagk/courses.html>

点击此处返回...

算法——牛顿法(Newton Method)

null的专栏 6万+

顿法概述

学习入门系列04, Gradient Descent (梯度下降法)

yofer张耀琦的专栏 2万+

梯度下降法? 学习速率的引入; 如何调整学习速率; Adagrad算法介绍; 用泰勒展开式对梯度下降法进行数学理论支持

ing to learn by gradient descent by gradient descent - PyTorch实践

senius的博客 5328

文, 转载请注明来源 (csdn or blog) 引言Learning to learn, 即学会学习, 是每个人都具备的能力, 具体指的是一...

学习- (二十六) 无约束优化算法——牛顿法、拟牛顿法及无约束算法总结

欢迎你 188

学习学习笔记一、牛顿法 (Newton's method)Newton's method 1.推导在最速下降...

ent Descent Algorithm

qiongyu0422的专栏 613

ent Descent AlgorithmAll variables should be refreshed simultaneously but not one by one.

ows版YOLOv4目标检测实战: 训练自己的数据集

04-26

示环境: Windows10; cuda 10.2; cudnn7.6.5; Python3.7; VisualStudio2019; OpenCV3.4 需要学习ubuntu系统...

数学建模历年题目及优秀论文

04-02

数学建模历年题目及优秀论文, 好资源与大家分享! !

©2020 CSDN 皮肤主题: 技术黑板 设计师: CSDN官方博客 返回首页

关于我们 招聘 广告服务 网站地图 kefu@csdn.net 客服论坛 400-660-0108 QQ客服 (8:30-22:00)

公安备案号 11010502030143 京ICP备19004658号 京网文〔2020〕1039-165号 版权与免责声明 版权申诉 网络110报警服务

中国互联网举报中心 家长监护 版权申诉 北京互联网违法和不良信息举报中心 ©1999-2020 北京创新乐知网络技术有限公司