

[CSDN首页 \(http://www.csdn.net?ref=toolbar\)](http://www.csdn.net?ref=toolbar)[学院 \(http://edu.csdn.net?ref=toolbar\)](http://edu.csdn.net?ref=toolbar)[下载 \(http://download.csdn.net?ref=toolbar\)](http://download.csdn.net?ref=toolbar)

更多 ▾

[下载 CSDN APP \(http://www.csdn.net/app?ref=toolbar\)](http://www.csdn.net/app?ref=toolbar)[写博客 \(http://write.blog.csdn.net/postedit?ref=toolbar\)](http://write.blog.csdn.net/postedit?ref=toolbar)[登录 \(https://passport.csdn.net/account/login?ref=toolbar\)](http://blog.csdn.net/) | [注册 \(http://passport.csdn.net/account/mobile/register?ref=toolbar&action=mobileRegister\)](http://passport.csdn.net/account/mobile/register?ref=toolbar&action=mobileRegister)[首页 \(http://blog.csdn.net/\)](http://blog.csdn.net/)[移动开发 \(http://blog.csdn.net/mobile/newarticle.html\)](http://blog.csdn.net/mobile/newarticle.html)

全部 ▾

CSDN (<http://www.csdn.net>)喜欢 **xgboost快速入门**

原创 2015年11月12日 20:57:27

收藏

9184

3

2



评论

xgboost快速入门



xgboost是gbdt算法的实现，可以做回归，分类，和排序。支持各种语言调用，支持单机和分布式。非常适合于大规模数据集。

项目主页

<https://github.com/dmlc/xgboost> (<https://github.com/dmlc/xgboost>)安装https://github.com/dmlc/xgboost/blob/master/doc/python/python_intro.md(https://github.com/dmlc/xgboost/blob/master/doc/python/python_intro.md)

我选择了python调用xgboost的方式。

1 从项目主页下载源码，解压。

2 在解压后的目录下执行make命令安装。

3 在子文件夹python-package目录下，执行pythonsetup.py install。

当然，你的电脑可能会缺失一些依赖库需要安装。比如在步骤二需要你安装g++，在步骤三需要你安全python的一些数学库。

分类算法实践

**lingerlanlan** (<http://blog.csdn.net/lingerlanlan>)

+ 关注

(<http://blog.csdn.net/linger2012liu>)

码云

原创

粉丝

喜欢

未开通

161

642

0

(<https://github.com/lingerlanlan>)**他的最新文章**更多文章 (<http://blog.csdn.net/linger2012liu>)

Tensorflow入门：数据结构和编程思想
(lingerlanlan/article/details/61616906)


C++实现获取DOTA玩家名称（反汇编
查找指针地址和跨进程读取war3内存）
(lingerlanlan/article/details/53933133)

mahout的itembased推荐算法改造
(lingerlanlan/article/details/50673495)


xgboost快速入门
(lingerlanlan/article/details/49804551)

<https://github.com/dmlc/xgboost/tree/master/demo/guide-python>
 (https://github.com/dmlc/xgboost/tree/master/demo/guide-python)
 这个页面有很多demo都值得研究一下。


下面是一个二分类的问题的具体做法。

 首先，输入数据仍然支持libsvm的格式，这也是比较喜欢的一个格式。
 每一行都是

喜欢 label index1:value1 index2:value2.....

 的格式。
 不过xgboost对label的有个要求，就是要从0开始。

收藏 比如2分类，label只能是0,1。

 3分类，label只能是0,1,2。

评论



分享

```
[python]
1.  #! /usr/bin/python
2.  import numpy as np
3.  import xgboost as xgb
4.
5.
6.  dtrain = xgb.DMatrix('train.txt')
7.  dtest = xgb.DMatrix('test.txt')
8.
9.
10. # specify parameters via map, definition are same as c++ version
11. param = {'max_depth':22, 'eta':0.1, 'silent':0, 'objective':'binary:logistic', 'min_child_weight
12.
13. # specify validations set to watch performance
14. watchlist = [(dtest, 'eval'), (dtrain, 'train')]
15. num_round = 33
16. bst = xgb.train(param, dtrain, num_round, watchlist)
17.
18. # this is prediction
19. preds = bst.predict(dtest)
20. labels = dtest.get_label()
21. print ('error=%f' % ( sum(1 for i in range(len(preds)) if int(preds[i]>0.5)!=labels[i]) /float(
```



Himall打造小天猫，小京东
 你与马云还差一个
 多用户商城系统
 构造生态级多用户商城更专注，更细致，更有品质
 立即申请免费体验

网上商城系统



编辑推荐

最热专栏

如何使用Keras进行分布式/多GPU运算...

深度学习最全优化方法总结比较（SGD...

tensorflow学习笔记（三十一）：构建多..

Focal Loss (/u014380165/article/detail...

xgboost相比传统gbdt有何不同？xgboo...

在线课程

```
22.
23. print ('correct=%f' % ( sum(1 for i in range(len(preds)) if int(preds[i]>0.5)==labels[i]) /float(len(preds))))
```

本文作者:linger

本文链接：<http://blog.csdn.net/lingerlanlan/article/details/49804551>

(<http://blog.csdn.net/lingerlanlan/article/details/49804551>)



喜欢



收藏



评论

版权声明：本文为博主原创文章，未经博主允许不得转载。

▲ 举报

标签：xgboost (<http://so.csdn.net/so/search/s.do?q=xgboost&t=blog>) /

gbdt (<http://so.csdn.net/so/search/s.do?q=gbdt&t=blog>)



分享



xin9910 (/xin9910) 2017-05-04 09:26

2楼

(/xin9910) 这个代码在哪里下载的呢？

回复



wangjiguoqu__ (/wangjiguoqu__) 2016-03-02 12:20

1楼

(/wangjiguoqu__) 请问这个版本是安装在windows系统上的还是mac的os系统上的呢？



用户画像系统应用与技术

解析

(<http://blog.csdn.net/huiyi1990/article/details/51111111>)
Course/series_detail/68?
utm_source=blog9)



实时流计算平台Blink在阿

里集团的应用实践

(<http://blog.csdn.net/huiyi1990/article/details/51111111>)
Course/detail/531?
utm_source=blog9)

回复 1条回复 ▾

相关文章推荐



如何使用Keras进行分布式/多GPU运算？ (/u010159842/article/details/59104029)

喜欢

如何使用Keras进行分布式/多GPU运算？ Keras在使用TensorFlow作为后端的时候可以进行分布式/多GPU的运算，Keras对



多GPU和分布式的支持是通过TF完成的。 with ...



收藏 u010159842 (<http://blog.csdn.net/u010159842>) 2017-03-01 14:11 3537



评论

深度学习最全优化方法总结比较（SGD，Adagrad，Adadelata，Adam，Adamax，Nadam）



(/u010159842/article/details/54574282)

分享

前言（标题不能再中二了）本文仅对一些常见的优化方法进行直观介绍和简单的比较，各种优化方法的详细内容及公式只好去认真啃论文了，在此我就不赘述了。SGD 此处的SGD指mini-batch gr...



u010159842 (<http://blog.csdn.net/u010159842>) 2017-01-16 15:09 244





学刚毕业..

我是如何成为一名python大咖的？ (http://edu.csdn.net/topic/python1?utm_source=blog10)

人生苦短，都说必须python，那么我分享下我是如何从小白成为Python资深开发者的吧。2014年我大

tensorflow学习笔记（三十一）：构建多GPU代码 (/u012436149/article/details/54092928)

构建多GPU代码结构 先构建单GPU代码 写个函数multi_gpu_model(num_gpus)来生成多GPU代码，并将对象保存在collection中 feed data run 如何构建单GP...


 u012436149 (<http://blog.csdn.net/u012436149>) 2017-01-05 14:35  5246

Focal Loss (/u014380165/article/details/77019084)



论文：Focal Loss for Dense Object Detection 论文链接：<https://arxiv.org/abs/1708.02002> RGB和Kaiming大神的新作。我喜欢们...



u014380165 (<http://blog.csdn.net/u014380165>) 2017-08-10 08:21  2835

收藏





xgboost相比传统gbdt有何不同？xgboost为什么快？xgboost如何支持并行？

评论(/xbmatrix/article/details/65450395)





链接：<https://www.zhihu.com/question/41354392/answer/98658997> 传统GBDT以CART作为基分类器，xgboost还支持线性分类器，...

 xbmatrix (<http://blog.csdn.net/xbmatrix>) 2017-03-23 22:08  1344

XGBoost细谈 (/u012771351/article/details/52997120)

注：本文中所有公式和思路来自于邹博先生的《机器学习升级版》，我只是为了加深记忆和理解写的本文。XGBoost是陈天奇先生提出的一种提升方法，较GBDT不同的是，XGBoost采用了二阶导的信息，...

 u012771351 (<http://blog.csdn.net/u012771351>) 2016-11-01 16:22  1049

机器学习、数据挖掘方面好书、网站推荐 (/a819825294/article/details/51158786)

书籍入门1.集体编程智慧 地址 2.机器学习实战 地址 3.统计学习方法 地址 4.机器学习 周志华 地址 5.机器学习 卡内基梅隆大学 地址进阶1.Pattern Recognitio...



a819825294 (<http://blog.csdn.net/a819825294>) 2016-04-15 09:13 1759

一个有趣的排序算法—睡眠排序 (/qq_25827845/article/details/54800064)

今天看到了这么一个排序方法，是感觉有点好玩，它的基本思想是：主要是根据CPU的调度算法实现的，对一组数据进行排序，不能存在负数值，这个数是多大，那么就在线程里睡眠它的10倍再加10，不是睡眠和它的数值...



喜欢



qq_25827845 (http://blog.csdn.net/qq_25827845) 2017-01-31 20:28 1257



收藏

天池 IJCAI17 《口碑商家客流量预测》回顾 (/lyr369852147/article/details/62072109)

2017天池口碑商家客流量预测队伍名：卡文尼尔 第一赛季排名：191 第二赛季排名：168 成绩：0.0824这次比赛是抱着学习做特征工程的心态来做的，所以整体思路是模型加少许规则。一路做下来发...



评论



分享

lyr369852147 (<http://blog.csdn.net/lyr369852147>) 2017-03-14 22:45 1544

R+python | XGBoost极端梯度上升以及forecastxgb (预测) +xgboost (回归) 双案例解读 (/sinat_26917383/article/details/52623754)

一、XGBoost来历 xgboost的全称是eXtreme Gradient Boosting。正如其名，它是Gradient Boosting Machine的一个c++实现，作者为正在华盛顿大...



sinat_26917383 (http://blog.csdn.net/sinat_26917383) 2016-09-22 19:17 8147

python快速入门 (控制结构) (/felaim/article/details/71193384)

控制结构python里的缩进非常重要，这点LZ也是深受其害，之前也讲过这个问题，space和TAB的混用就会带来报错，然后很尴尬的是肉眼基本是辨别不出来的，还有无意中切换至中文输入法下，打了空格，虽...



Felaim (<http://blog.csdn.net/Felaim>) 2017-05-05 11:07 302

Flume快速入门（二）：设计从简 (/manzhizhen/article/details/52606737)

上一篇文章简单介绍了下Flume的背景，接下来本文说说Flume NG的内部设计。注意：本文针对的是Flume1.6.0版本。

上一篇：<http://manzhizhen.iteye.co...>



manzhizhen (<http://blog.csdn.net/manzhizhen>) 2016-09-21 13:06 198



收藏



德州仪器半导体技术（上海）有限公司 通用DSP 技术应用工程师 崔晶 德州仪器（TI）的第一颗达芬奇（DaVinci）芯片（处理器）DM6446已经问世快三年了。继DM644x之后，TI又陆续推...

评论



LG1259156776 (<http://blog.csdn.net/LG1259156776>) 2016-02-29 13:13 968

分享

OpenStack快速入门 (/tomstrong_369/article/details/45914355)



第一部分：OpenStack及其构成简介 一、云计算 云计算是一种计算模型，它将诸如运算能力、存储、网络 and 软件等资源抽象成为服务，以便让用户通过互联网远程享用，付费的形...



Tomstrong_369 (http://blog.csdn.net/Tomstrong_369) 2015-05-22 10:45 965

Flex快速入门系列之二：FlashBuilder创建项目以及远程JAVA对象访问（包括spring）篇 (/zhaoshansanxia/article/details/7017001)



前提：一个blazeds-turnkey，是adobe官方提供的访问远程java对象的工具，相关基本概念可参考：<http://baike.baidu.com/view/1921230.html>，官方文...

 zhaoshansanxia (<http://blog.csdn.net/zhaoshansanxia>) 2011-11-27 17:04  1486

GitHub快速入门简单教程以及常用Git 命令 (/u012146107/article/details/12101069)

如果没有使用过Git，大概这样理解一下：开发者（我们）从服务器克隆了一份数据库到本地，开发过程中，我们在分支中修改代码，然后先向本地提交数据，最后再向服务器push数据 一 注册github账号 ht...



喜欢  u012146107 (<http://blog.csdn.net/u012146107>) 2013-09-27 17:29  1873





MySQL快速入门 (/jx232515/article/details/52760952)



mysql 快速入门 本文转载于：<http://www.cnblogs.com/kunhu/p/3726676.html> 一.入门基础 1.连接MYSQL：格式：mysql -h 主机地址 ...



评论  jx232515 (<http://blog.csdn.net/jx232515>) 2016-10-08 22:04  142



分享



快速入门--使用GoEasy推送在Java后台推送实时消息 (/fengziiti/article/details/50260155)

每个程序猿（媛）都有过很多使用第三方插件或第三方服务的经历，我觉得最让人头疼的事是不能快速将它提供的服务快速运用到自己的代码中，一直沉浸在别人的口水文档中。最近在看关于实时消息推送方面的技术文章，然后...

 FENGZIITI (<http://blog.csdn.net/FENGZIITI>) 2015-12-11 10:09  4154

用星际快速入门PHP面向对象编程(很有意思的教程 赞赞赞) (/yun__yang/article/details/53941604)

用星际快速入门PHP面向对象编程(很有意思的教程 赞赞赞)

 yun__yang (http://blog.csdn.net/yun__yang) 2017-01-01 18:56  2744

HTML5快速入门（一）—— HTML简介 (/yeshaojian/article/details/51615541)

前言：1.HTML5的发展非常迅速，可以说已经是前端开发人员的标配，在电商类型的APP中更是运用广泛，这个系列的文章是本人自己整理，尽量将开发中不常用到的剔除，将经常使用的拿出来，使需要的朋友能够真正...



yeshaojian (<http://blog.csdn.net/yeshaojian>) 2016-06-08 21:50 9463



喜欢



收藏



评论



分享