



wepon的专栏

- 目录视图
- 摘要视图
- RSS 订阅

个人资料

wepon\_

关注

发私信

访问： 692832次

积分： 5187

等级： 

BLOG

 6

排名： 第5110名

原创： 72篇

转载： 3篇

译文： 0篇

评论： 370条

About

个人网站：<http://wepon.me/>

github：<https://github.com/wepe>

知乎：<https://www.zhihu.com/people/wepo-n-huang>

很久没上CSDN，有问题欢迎邮件交流：masterwepon@163.com

文章搜索

文章分类

Machine Learning (31)

Kaggle (2)

scikit-learn使用手册 (1)

python (16)

数据库 (1)

openCV (3)

Leetcode (28)

ACM题解析 (1)

品味经典书籍 (1)

杂谈 (0)

文章存档

2016年02月 (1)

程序员，为什么写不好一份简历？[征文 | 你会为 AI 转型么？](#)[福利 | 免费参加 2017 OpenStack Days China](#)

【Python】 csv模块的使用

标签：python csv

2014-12-13 22:39

12356人阅读

评论(4)

收藏

举报

分类：

python ( 15 )

版权声明：本文为博主原创文章，未经博主允许不得转载。

目录(?)

[+]

Python csv模块的使用

1、csv简介

CSV (Comma Separated Values)，即逗号分隔值（也称字符分隔值，因为分隔符可以不是逗号），是一种常用的文本格式，用以存储表格数据，包括数字或者字符。很多程序在处理数据时都会碰到csv这种格式的文件，它的使用是比较广泛的（Kaggle上一些题目提供的数据就是csv格式），csv虽然使用广泛，却没有通用的标准，所以在处理csv格式时常常会碰到麻烦，幸好Python内置了csv模块。下面简单介绍csv模块中最常用的一些函数。

更多内容请参考：<https://docs.python.org/2/library/csv.html#module-csv>

2、csv模块中的函数

- reader(csvfile, dialect='excel', \*\*fmtparams)

参数说明：

csvfile，必须是支持迭代(Iterator)的对象，可以是文件(file)对象或者列表(list)对象，如果是文件对象，打开时需要加"b"标志参数。

dialect，编码风格，默认为excel的风格，也就是用逗号(,)分隔，dialect方式也支持自定义，通过调用register\_dialect方法来注册，下文会提到。

fmtparam，格式化参数，用来覆盖之前dialect对象指定的编码风格。

```
[python]
01. import csv
02. with open('test.csv','rb') as myFile:
03.     lines=csv.reader(myFile)
04.     for line in lines:
05.         print line
```

'test.csv'是文件名，'rb'中的r表示“读”模式，因为是文件对象，所以加'b'。open()返回了一个文件对象myFile，reader(myFile)只传入了第一个参数，另外两个参数采用缺省值，即以excel风格读入。reader()返回一个

2015年09月

(1)

2015年08月

(1)

2015年05月

(3)

2015年04月

(3)

展开

阅读排行

大数据竞赛平台——Kaggle 入...

(79657)

交叉熵代价函数

(54503)

DeepLearning tutorial ( 4 ) C...

(38530)

DeepLearning tutorial ( 6 ) 易...

(38494)

DeepLearning tutorial ( 5 ) C...

(36594)

正则化方法：L1和L2 regulariz...

(24591)

【机器学习算法实现】主成分...

(21406)

DeepLearning tutorial ( 1 ) S...

(20412)

scikit-learn中PCA的使用方法

(17973)

机器学习算法中如何选取超参...

(16929)

评论排行

DeepLearning tutorial ( 6 ) 易...

(76)

DeepLearning tutorial ( 5 ) C...

(59)

大数据竞赛平台——Kaggle 入...

(34)

DeepLearning tutorial ( 7 ) 深...

(33)

DeepLearning tutorial ( 4 ) C...

(21)

正则化方法：L1和L2 regulariz...

(15)

Kaggle入门——使用scikit-lear...

(14)

交叉熵代价函数

(10)

【机器学习算法实现】logistic...

(8)

DeepLearning tutorial ( 3 ) M...

(7)

推荐文章

\* CSDN日报20170714——《从创业到再就业，浅述对程序员职业生涯的看法》

\* Android 逆向 | 锁屏密码算法解析以及破解方案

\* 从单一WAR到多活，记述一个创业公司的架构演变

\* 我的AI转型之路与AI之我见（非985211的奋斗路程与视角）

\* AI大行其道，你准备好了吗？—谨送给徘徊于转行AI的程序员

\* AI转型中的思考和洞见

最新评论

DeepLearning tutorial ( 5 ) CNN卷积神...  
草谷乐 : 博主,你好,请问from theano.tensor.signal.pool import down...

OpenCV人脸检测（C++代码）  
jiachen0212 : 赞！

DeepLearning tutorial ( 6 ) 易用的深度学...  
dittoya : ValueError: &#39;The specified si ze contains a dim...

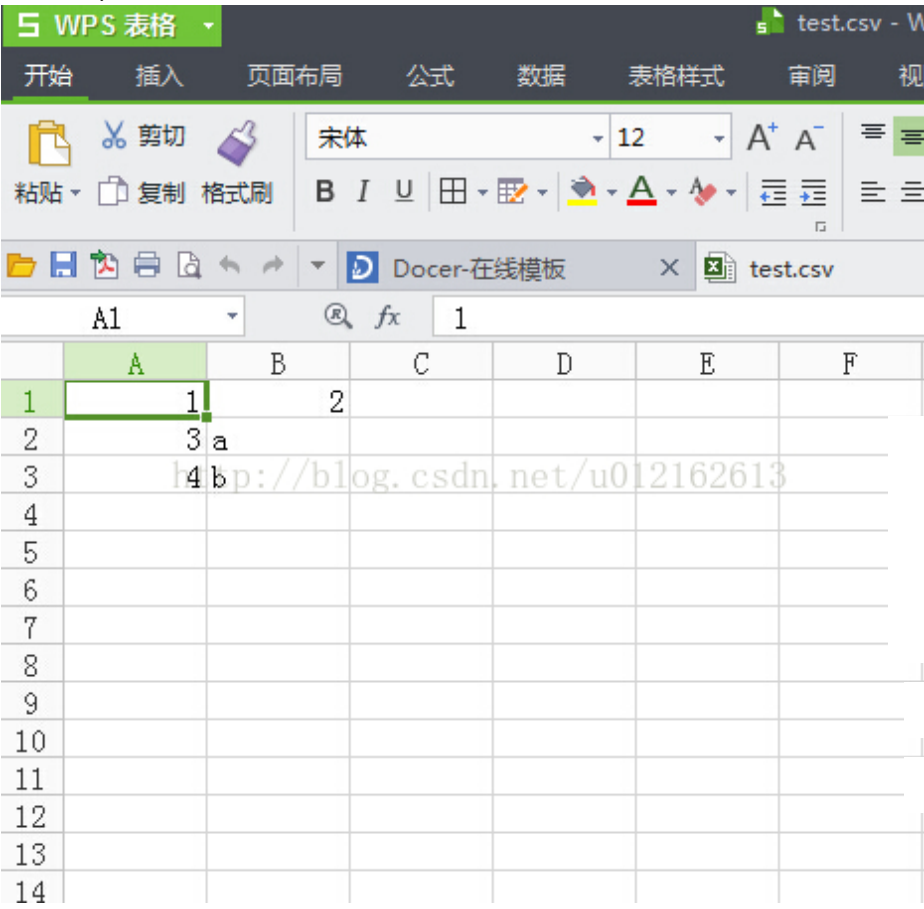
DeepLearning tutorial ( 6 ) 易用的深度学...  
dittoya : ValueError: 'The specified size co ntains a dimensi...

DeepLearning tutorial ( 1 ) Softmax回归...  
gg112324d : @kobecsb:sigmoid二分类soft max多分类

朴素贝叶斯理论推导与三种常见模型  
南宫娜娜 : 看了一下您的代码 def\_calculat

reader对象lines,lines是一个list，当调用它的方法lines.next()时，会返回一个string。上面程序的效果是将csv文件中的文本按行打印，每一行的元素都是以逗号分隔符','分隔得来。

在我的test.csv文件中，存储的数据如图：



程序输出：

```
[1, '2']
[3, 'a']
[4, 'b']
```

补充：reader对象还提供一些方法：line\_num、dialect、next()

- writer(csvfile, dialect='excel', \*\*fmtparams)

参数的意义同上，这里不赘述，直接上例程：

```
[python]
01. with open('t.csv', 'wb') as myFile:
02.     myWriter=csv.writer(myFile)
03.     myWriter.writerow([7, 'g'])
04.     myWriter.writerow([8, 'h'])
05.     myList=[[1,2,3],[4,5,6]]
06.     myWriter.writerows(myList)
```

'w'表示写模式。首先open()函数打开当前路径下的名字为't.csv'的文件，如果不存在这个文件，则创建它，返回myFile文件对象。csv.writer(myFile)返回writer对象myWriter。writerow()方法是一行一行写入，writerows方法是一次写入多行。注意：如果文件't.csv'事先存在，调用writer函数会覆盖文件中的文本。再执行writerow/writerows方法。

补充：除了writerow、writerows，writer对象还提供了其他一些方法：writeheader、dialect

- register\_dialect(name, [dialect, ]\*\*fmtparams)

e\_feature\_prob(self,feature...

朴素贝叶斯理论推导与三种常见模型

南宫娜娜 : 你好，这里2.1.1 举例中计算条件概率的时候，介绍2.1的描述中：“Nyk是类别为yk的样本个数，...

【机器学习算法实现】主成分分析(PCA)—...

Bruin0 : @funny\_QZQ:你好，我想问一下，pca降维之后的数据还能知道代表的什么意思吗？

【机器学习算法实现】主成分分析(PCA)—...

Bruin0 : pca降维之后，降维的数据代表原来数据的什么含义

DeepLearning tutorial ( 6 ) 易用的深度学...

iHunter001 : 按照博主的程序试着跑了一下，出现了错误TypeError: Layer can receive a t...

这个函数是用来自定义dialect的。

参数说明：

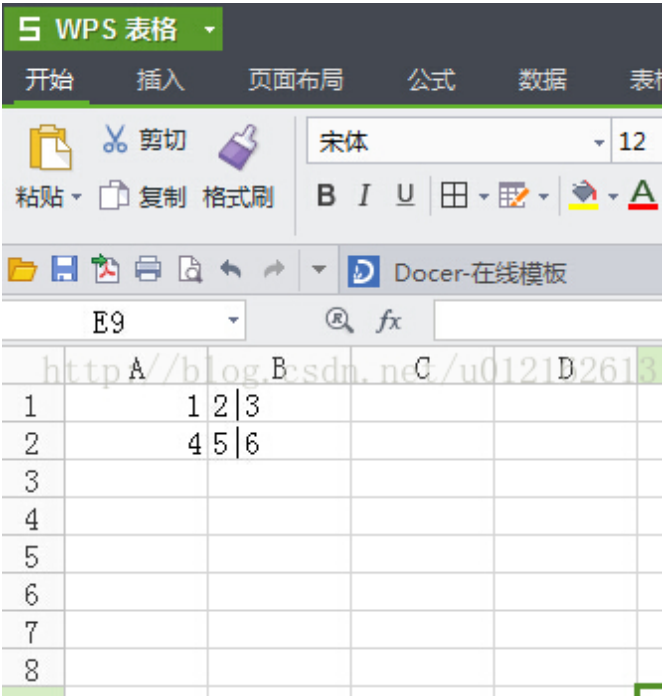
name,你所自定义的dialect的名字，比如默认的是'excel'，你可以定义成'mydialect'

[dialect,]\*\*fmtparams，dialect格式参数，有delimiter（分隔符，默认的就是逗号）、quotechar、quoting等等，可以参考[Dialects and Formatting Parameters](#)

```
[python]
01. csv.register_dialect('mydialect',delimiter='|', quoting=csv.QUOTE_ALL)
```

上面一行程序自定义了一个命名为mydialect的dialect，参数只设置和quoting这两个，其他的仍然采用默认值，其中以'|'为分隔符。接下来我们就可以像使用'excel'一样来'mydialect'了。我们来看看效果：

在我test.csv中存储如下数据：



以'mydialect'风格打印：

```
[python]
01. with open('test.csv','rb') as myFile:
02.     lines=csv.reader(myFile,'mydialect')
03.     print lines.line_num
04.     for line in lines:
05.         print line
```

输出：

['1,2', '3']
['4,5', '6']

可以看到，现在是以'|'为分隔符，1和2合成了一个字符串（因为1和2之间的分隔符是逗号，而mydialect风格的分隔符是'|'），3单独一个字符串。对于writer()函数，同样可以传入mydialect作为参数，这里不赘述。

关闭

• unregister\_dialect(name)

这个函数用于注销自定义的dialect

此外，csv模块还提供get\_dialect(name)、list\_dialects()、field\_size\_limit([new\_limit])等函数，这些都比较简单，可以自己试试。比如list\_dialects()函数会列出当前csv模块里所有的dialect：

```
[python]
```

```
01. | print csv.list_dialects()
```

输出：

```
['excel-tab', 'excel', 'mydialect']
```

'mydialect'是自定义的，'excel-tab', 'excel'都是自带的dialect，其中'excel-tab'跟'excel'差不多，只不过它以tab为分隔符。

csv模块还定义了一些类：DictReader、DictWriter、Dialect等，DictReader和DictWriter类似于reader和writer。  
一些常量：QUOTE\_ALL、QUOTE\_MINIMAL、.QUOTE\_NONNUMERIC等，这些常量可以作为Dialects and Formatting Parameters的值。

先写到这，其他的以后用到再更新。

顶

7

踩

0

- 上一篇      【机器学习算法实现】logistic回归\_基于Python和Numpy函数库
- 下一篇      大数据竞赛平台——Kaggle 入门

相关文章推荐

- Kaggle竞赛入门教程之Kaggle简介（新手向）
- 【机器学习算法实现】kNN算法\_手写识别——基...
- 【机器学习算法实现】logistic回归\_基于Python和...
- 大数据竞赛平台——Kaggle 入门
- python zeros()使用(from numpy import \*)

- 用bind方法保持this上下文
- 线性回归标准方程及其概率解释
- Kaggle 首战拿银总结 | 入门指导 (长文、干货)
- about Kaggle
- kaggle之路（1）



猜你在找

- |                             |                             |
|-----------------------------|-----------------------------|
| C语言及程序设计（讲师：贺利坚）            | Python爬虫工程师培养课程全套（讲师:韦玮）    |
| Python全栈开发入门与实战课（讲师：李杰）     | 2017软考网络规划设计师视频套餐（讲师：任铎）    |
| 2017软考软件设计师视频套餐（讲师：任铎）      | 2017软考-信息系统项目管理师视频套餐（讲师：任铎） |
| 软考(高级)项目经理实战营（讲师：张传波）       | 微信公众平台开发套餐（讲师：刘运强）          |
| 深度学习原理+实战+算法+主流框架套餐（讲师：唐宇迪） | 2017系统集成项目管理工程师通关套餐（讲师：徐朋）  |

查看评论

关闭

- Q-沐风听雨-

3楼 2016-12-28 21:49发表
- 第一个代码，使用rb的方式读取会报错；使用r和rt的方式是正确的，所以，作者上面的内容请核实！
- vipeng

Re: 2017-04-05 11:05发表
- 回复nature\_XD：对的，上述地方不能用rb



2楼 2016-08-11 22:26发表

On Mac, you should type:  
with open('test.csv','rU') as myFile  
.....  
  
instead of 'rb', you should type 'rU', otherwise you will get an error like this:  
"new-line character seen in unquoted field - do you need to open the file in universal-newline mode?"



lipeng1993  
good

1楼 2014-12-17 19:22发表

发表评论

用户名： haijunz

评论内容：

提交

\* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场

