




博客 ([//blog.csdn.net/](http://blog.csdn.net/))


学院 ([//edu.csdn.net/](http://edu.csdn.net/)) 下载 (<http://download.csdn.net/>)


GitChat (<http://gitbook.cn/?ref=csdn>) 论坛 (<http://bbs.csdn.net>)


...



3




 写博客

 发Chat

登录 (<https://passport.csdn.net/account/login>) 注册 (<https://passport.csdn.net/account/mobileregister?action=mobileRegister>)

[http://so.csdn.net/so/new/gitchat/activity?](http://so.csdn.net/so/new/gitchat/activity?utm_source=csdnhtml)





Python中的虚拟变量(dummy variables)



原创 2017年07月14日 14:26:39



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

-  虚拟变量 (<http://so.csdn.net/so/search/s.do?q=虚拟变量&t=blog>) /
-  dummy variables (<http://so.csdn.net/so/search/s.do?q=dummy variables&t=blog>) /
-  哑变量 (<http://so.csdn.net/so/search/s.do?q=哑变量&t=blog>) /
-  离散特征编码 (<http://so.csdn.net/so/search/s.do?q=离散特征编码&t=blog>)

3193

虚拟变量(dummy variables)
虚拟变量，也叫哑变量和离散特征编码，可用来表示分类变量、非数量因素可能产生的影响。

① 离散特征的取值之间有大小的意义
例如：尺寸(L、XL、XXL)
离散特征的取值有大小意义的处理函数map
pandas.Series.map(dict)
参数 dict:映射的字典

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

② 离散特征的取值之间没有大小的意义



Erin_data (<http://blog.cs...>)

+ 关注

(<http://blog.csdn.net/III1528238733>)

码云

原创

粉丝

喜欢

未开通

58

86

11

(<https://gite...>)

utm_sourc



全息手机



他的最新文章

更多文章 (<http://blog.csdn.net/III1528238733>)

pandas.get_dummies

例如：颜色(Red,Blue,Green)

处理函数：

get_dummies(data,prefix=None,prefix_sep="_",dummy_na=False,columns=None,drop_first=False)

- ① data 要处理的DataFrame
- ② prefix 列名的前缀，在多个列有相同的离散项时候使用
- ③ prefix_sep 前缀和离散值的分隔符，默认为下划线，默认即可
- ④ dummy_na 是否把NA值，作为一个离散值进行处理，默认为不处理
- ⑤ columns 要处理的列名，如果不指定该列，那么默认处理所有列
- ⑥ drop_first 是否从备选项中删除第一个，建模的时候为避免共线性使用



[python]

```
01. # -*- coding: utf-8 -*-
02. import pandas
03.
04. data = pandas.read_csv(
05.     'D:\\PDA\\4.18\\data.csv',
06.     encoding='utf8'
07. )
08.
09. data['Education Level'].drop_duplicates()
10.
11. """
12. 博士后      Post-Doc
13. 博士         Doctorate
14. 硕士         Master's Degree
15. 学士         Bachelor's Degree
16. 副学士       Associate's Degree
17. 专业院校    Some College
18. 职业学校    Trade School
19. 高中         High School
20. 小学         Grade School
21. """
22. educationLevelDict = {
23.     'Post-Doc': 9,
24.     'Doctorate': 8,
25.     'Master's Degree': 7,
```

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

中国消费金融的未来，已经到来 - 消费金融真经 (http://blog.csdn.net/III1528238733/article/details/78690265)

评分卡上线后如何进行评分卡的监测 (http://blog.csdn.net/III1528238733/article/details/78356803)

数据分析中非常实用的自编函数和代码模块整理 (http://blog.csdn.net/III1528238733/article/details/76779651)

懒癌必备-dplyr和data.table让你的数据分析事半功倍 (http://blog.csdn.net/III1528238733/article/details/76762734)

信用标准评分卡模型开发及实现 (http://blog.csdn.net/III1528238733/article/details/76602006)

文章分类

信用风险评级模型的开发 (http://blog.csdn.net/III1528238733/article/details/76602006) 10篇

非常干的干货分享！ (http://blog.csdn.net/III1528238733/article/details/76602006) 3篇

R数据挖掘篇 (http://blog.csdn.net/III1528238733/article/details/76602006) 16篇

R数据分析篇 (http://blog.csdn.net/III1528238733/article/details/76602006) 15篇

Python数据分析篇 (http://blog.csdn.net/III1528238733/article/details/76602006) 14篇

展开 ▼

文章存档 登录

注册



```

26.         'Bachelor\'s Degree': 6,
27.         'Associate\'s Degree': 5,
28.         'Some College': 4,
29.         'Trade School': 3,
30.         'High School': 2,
31.         'Grade School': 1
32.     }
33.
34. data['Education Level Map'] = data[
35.     'Education Level'
36. ].map(
37.     educationLevelDict
38. )
39.
40. data['Gender'].drop_duplicates()
41.
42. dummies = pandas.get_dummies(
43.     data,
44.     columns=['Gender'],
45.     prefix=['Gender'],
46.     prefix_sep="_",
47.     dummy_na=False,
48.     drop_first=False
49. )
50.
51. dummies['Gender'] = data['Gender']

```

版权声明：本文为博主原创文章，未经博主允许不得转载。

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

2017年12月 (<http://blog.csdn.net/111528238733/article/details/76602006>) 1篇

2017年10月 (<http://blog.csdn.net/111528238733/article/details/76601897>) 11篇

2017年8月 (<http://blog.csdn.net/111528238733/article/details/75114360>) 45篇

2017年7月 (<http://blog.csdn.net/111528238733/article/details/75107729>) 3157

他的热门文章

信用标准评分卡模型开发及实现 (<http://blog.csdn.net/111528238733/article/details/76602006>)

8730

评分卡模型开发-基于逻辑回归的标准评分卡实现 (<http://blog.csdn.net/111528238733/article/details/76601897>)

7011

Python中的相关分析correlation analysis (<http://blog.csdn.net/111528238733/article/details/75114360>)

3921

Python中的虚拟变量(dummy variables) (<http://blog.csdn.net/111528238733/article/details/75107729>)

3157

评分法模型开发-WOE值计算 (<http://blog.csdn.net/111528238733/article/details/76600598>)

2336

登录

注册





【Python】使用 multiprocessing.dummy 执行多线程任务

-*- coding: utf-8 -*- from multiprocessing.dummy import Pool as Thr ns2250225 2015年09月26日 22:11 8445
 3
 eadPool import time import url...
 (http://blog.csdn.net/ns2250225/article/details/48755741)

Pandas 处理 dummy variable



weiwei9363 2017年10月16日 22:20 461

Pandas 处理 dummy variable Dummy Variable Dummy Variable处理步骤 Step 1 读入数据 Step 2 处理数据 Step 3 与原数据进行拼接 S...

(http://blog.csdn.net/weiwei9363/article/details/78255210)

线性回归分析中的哑变量 - x_i_y_u_e的专栏 - CSDN博客

基础不扎实,高级方法用起来真觉得底气不足,今天看到哑变量在回归分析中的应用,...Python(20) JavaScript/CSS(4) Java(77)
 Java阅读整理(1) Java/J2EE(...

2018-1-10

(http://blog.csdn.net/x_i_y_u_e/article/details/39675055?>)

生成哑变量的几种方法 - CSDN博客

Python中的虚拟变量(dummy variables) 虚拟变量(dummy variables) 虚拟变量,也叫哑变量和离散特征编码,可用来表示分类变量、非数量因素可能产生的影响。① 离散特征...
 加入CSDN,享受更精准的内容推荐,与500万程序员共同成长!



程序员要学多久



联系我们

网站客服 (http://wpa.qq.com/msgrd?v=3&uin=2431299880&site=qq&menu

webmaster@csdn.net
 (mailto:webmaster@csdn.net)

微博客服
 (http://e.weibo.com/csdnsupport/profil
 400-660-0108

关于 招聘 广告服务 ali

©2018 CSDN 京ICP证09002463号

(http://www.miibeian.gov.cn/)



经营性网站备案信息

(http://www.hd315.gov.cn/beian/view.asp?

bianhao=010202001032100010) 网络110

报警服务 (http://www.cyberpolice.cn/)

登录

注册



2018-1-17

(https://passport.csdn.net/a

(http://blog.csdn.net/yugao1986/article/details/6259887??>)

程序员不会英语怎么行？



👍
老司机教你一个数学公式秒懂天下英语
3

哑变量或虚拟变量介绍(dummyvar)



qy20115549 2016年10月13日 08:38 3140

虚拟变量介绍虚拟变量是指回归分析中使用的一个数值变量来代表你的研究中的样本的子组。在研究设计中，一个虚拟变量经常用来区分不同的治疗组。在最简单的情况下，我们会使用一个0,1的虚拟变量，其中一个人是给...

(http://blog.csdn.net/qy20115549/article/details/52803647)

哑变量或虚拟变量介绍(dummyvar) - qy20115549的博客 - CSDN博客

Python中的虚拟变量(dummy variables) 虚拟变量(dummy variables) 虚拟变量,也叫哑变量和离散特征编码,可用来表示分类变量、非数量因素可能产生的影响。① 离散特征...

2018-1-18

(http://blog.csdn.net/qy20115549/article/details/52803647??>)

Python中的虚拟变量(dummy variables) - CSDN博客

虚拟变量,也叫哑变量和离散特征编码,可用来表示分类变量、非数量因素可能产生的影响...Python语言中,Pandas中的get_dummmy()函数是将拥有不同值的变量转换为0/1数值...

2018-1-19

(http://blog.csdn.net/lll1528238733/article/details/75107729??>)

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

登录

注册



python-数据清洗



julyorangesea 2017年03月06日 20:50 726

数据因为存在数据错误，数据缺失，离群值的存在，所以需要数据

据进行清洗 1.数据错误： 错误类型 – 脏数据或错误数据 • 比如, Age = -2003 – 数据不正确 • ‘...

(<http://blog.csdn.net/julyorangesea/article/details/60595260>)

Python数据分析笔记



shangwei1991 2016年01月06日 14:39 6830

第一章 NumPy基础：数组和矢量计算 1.Numerical Python是高性能科学计算和数据分析的基础包。它提供了一个具有矢量算术运算和复杂广播能力的快速且节省空间的多维数组ndarray；用于...

(<http://blog.csdn.net/shangwei1991/article/details/50468773>)



pandas哑变量 - CSDN博客

利用python-pandas和sklearn进行天池移动推荐离线赛的全过程 python,pandas,sklearn LY_ysys629 2017年06月30日 22:23
2116 R语言中生成虚拟变量/哑变量 转载自...

2018-1-21

(http://blog.csdn.net/szj_huhu/article/details/79120149?>)

线性回归分析中的哑变量 - CSDN博客

哑变量(Dummy Variable),也叫虚拟变量,引入哑变量的目的是,将不能够定量处理的...R 14篇 文本分析 1篇 金融 7篇 Python 1篇 机器学习 30篇 展开 文章...

2018-1-19

(<http://blog.csdn.net/mousever/article/details/50500205?>>)

python--因子变量转换为哑变量



hch_power 2016年03月14日 17:24 327

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

登录


注册



在数据样本中有一类特征属性彼此是平行的关系，不能简单的以数值或字符赋予其含义。例如 职业：教师 工人 商人..... 面对这种情况可以通过构建哑变量来解决。例如含有三个因子的特征可以将其转化为三列每...

(http://blog.csdn.net/hch_power/article/details/50888232)

生成哑变量的几种方法

 yugao1986 2011年03月18日 20:06 2242

1.logistic回归中可以生成哑变量 proc logistic data=yuyu; class a_status / param=ref; run; 2.IF-THE...

(<http://blog.csdn.net/yugao1986/article/details/6259887>)

利用python进行数据分析——之数据结构pandas(一) - AI..._CSDN博客

这一系列的文章主要是自己读《利用python进行数据分析》所做的总结,详细内容可以...将生成的哑变量和其他数据合并: 3. 7、groupby方法 DataFrame的groupby方法和SQL...

2018-1-17

(<http://blog.csdn.net/u014380165/article/details/68486673?>>)


python-数据清洗 - CSDN博客

•每个有缺失值的变量生成一个指示哑变量,参与后续的建模,原始变量不使用。3...了解清洗后,接下来,就来学习一下Python的数据清洗吧!现在有一份心脏病患者的...

2018-1-18

(<http://blog.csdn.net/julyorangesea/article/details/60595260?>>)

线性回归分析中的哑变量

 x_i_y_u_e 2014年09月29日 20:28 1567

最近偶尔在重温统计学，发现自己工作后用了各种高级的统计分析方法，各种统计模型，却忽视了统计学中一些最基础的知识，而这些知识是所有这些高级方法的基础，基础不扎实，高级方法用起来真觉得底气不足，今...

加入CSDNlog享受更精准的内容推荐,与500万程序员共同成长!

登录

注册



Python实现逻辑回归



u013790563

2016年03月19日 21:13

1724

Python实现逻辑回归

[\(http://blog.csdn.net/u013790563/article/details/50923147\)](http://blog.csdn.net/u013790563/article/details/50923147)

Python数据分析笔记 - CSDN博客

1.Numerical Python是高性能科学计算和数据分析的基础包。它提供了一个具有矢量算术...(8)计算指标/哑变量。一种常用于统计建模或机器学习的转换方式是:将分类变量(...)

2018-1-18

[\(http://blog.csdn.net/shangwei1991/article/details/50468773?>\)](http://blog.csdn.net/shangwei1991/article/details/50468773?>)

python_openpyxl_操作excel - CSDN博客

Openpyxl is a Python library for reading and writing Excel 2010 xlsx/xlsm/...python_pandas_dataframe_行列选择_切片操作
486 R语言生成虚拟变量/哑变量 376...

2018-1-19

[\(http://blog.csdn.net/fanlei_lianjia/article/details/78225857?>\)](http://blog.csdn.net/fanlei_lianjia/article/details/78225857?>)

Python语言——Pandas包中的get_dummy()函数用法

Python语言中，Pandas中的get_dummy()函数是将拥有不同值的变量 LCJhust2012 2018年01月04日 20:05 23
转换为0/1数值。 举例说明：一群样本的年龄分别为19，32,56,94岁，
19岁用1表示，32岁用2表示，56岁用...

[\(http://blog.csdn.net/LCJhust2012/article/details/78974793\)](http://blog.csdn.net/LCJhust2012/article/details/78974793)

加入Dummy variable (变量虚拟/哑变量) 程序员共同成长 u011467621 2015年09月26日 21:21 3754

登录

注册



总是可以从身边同学口中听到说到dummy化，dummy lr, 今天自己看了dummy variable的wiki，这里记下笔记总结：dummy variable是用0或1表示某个类别是否出现 为什...

(https://passport.csdn.net/a

(http://blog.csdn.net/u011467621/article/details/48754679)

【数据挖掘】dummy variable 虚拟变量



Linh47 2018年01月08日 11:19 45

一般地，在虚拟变量的设置中：基础类型、肯定类型取值为1；比较类型，否定类型取值为0。适用于无序的离散数字

变量。例如：若用数字1-12表示1-12月，那么就潜在表示了12月...

(http://blog.csdn.net/Linh47/article/details/79000837)



使用python玩跳一跳超详细使用教程



LittleBeautiful 2018年01月02日 21:02 53958

在上一篇文章里介绍了使用Python玩微信跳一跳的详细使用教程，不过依旧还是有很多小伙伴有各种各样的问题。为了让各位小伙伴都能使用黑科技，于是再做一个超详细教程。从Python的安装开始，手把手教你一...

(http://blog.csdn.net/LittleBeautiful/article/details/78955792)

将自己的python程序打包成.exe/.app(秀同学一脸呐)

Windows 10 x64macOS Sierra 10.12.4Python 2.7 准备好装哔~了 MrLevo520 2016年07月06日 16:24 62621
么，来吧，做个真正意义上的绿色小软件 Win下发布应用起因 今天
实验室同学看到我的乞丐版...

(http://blog.csdn.net/MrLevo520/article/details/51840217)

python核心编程第三版中文版PDF

2017年09月28日 20:43 24.83MB 下载



Python - if语句



lynn_yan 2010年04月08日 22:38 161639

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

条件语句1) if语句例子：# 执行if语句内的程序a = input("a:") b = input("b:")if(a > b):print a, "> ", bNote：如果是在py...

登录

注册



(http://blog.csdn.net/lynn_yan/article/details/5464911)

(<https://passport.csdn.net/a>)

在Ubuntu下配置舒服的Python开发环境



kingppy 2013年10月26日 00:10 112930

Ubuntu 提供了一个良好的 Python 开发环境，但如果想使我们的开发效率最大化，还需要进行很多定制化的安装和配置。下面的是我们团队开发人员推荐的一个安装和配置步骤，基于 Ubuntu 12.0...

(<http://blog.csdn.net/kingppy/article/details/13080919>)

3

python代码过长的换行



codechelle 2017年03月10日 16:20 43311

python代码换行就是每行后面加个\ 举个栗子：time = "2017" print "one" + "," + "two" + "," + "three" + "\", " + time ...

(<http://blog.csdn.net/codechelle/article/details/61199212>)

【MySQL】经典数据库SQL语句编写练习题——SQL语句扫盲

【MySQL】数据库原理复习——SQL语言



WolfOfSiberian 2014年09月17日 19:23 13533

(<http://blog.csdn.net/WolfOfSiberian/article/details/39346781>)

Python的unittest自动化测试框架



zzmfish 2015年01月14日 21:22 3233

项目代码越来越复杂，并且对质量要求越来越高，就开始需要自动化测试。测试框架对比 python有很多自动化测试框架可以选择：python自带的unittest第三方的Robot ...

(<http://blog.csdn.net/zzmfish/article/details/42716645>)

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

登录

注册

