

CSDN新首页上线啦，邀请你来立即体验！(http://blog.csdn.net/)

立即体

验



博客 (/blog.csdn.net?ref=toolbar) 学院 (/edu.csdn.net?ref=toolbar)

下载 (/download.csdn.net?ref=toolbar) GitChat (/gitbook.cn/?ref=csdn)

更多 ▾



weixin\_3506... ▾

(//write.blog.csdn.net/postedit/activity?ref=toolbar)\_source=csdnblog



mishidemudong (http://bl...

+ 关注

(http://blog.csdn.net/u010159842)

码云

原创	粉丝	喜欢	未开通
182	169	1	(https://gite
			utm_sourc

## XGBoost解决多分类问题

转载

2016年11月30日 17:58:54

2374

# XGBoost解决多分类问题 写在前面的话

XGBoost官方给的二分类问题的例子

(https://github.com/dmlc/xgboost/tree/master/demo/binary\_classification)是区别蘑菇有无毒，数据集和代码都可以在xgboost中的demo文件夹对应找到，我是用的Anaconda安装的XGBoost，实现起来比较容易。唯一的梗就是在终端中运行所给命令： `../xgboost mushroom.conf` 时会报错，是路径设置的问题，所以我干脆把xgboost文件夹下的xgboost.exe拷到了mushroom.conf配置文件所在文件夹下，这样直接定位到该文件夹下就可以运行：`xgboost mushroom.conf`。二分类数据预处理，也就是data wrangling部分的代码有一定的借鉴意义，值得一看。

多分类问题给的例子 (https://github.com/dmlc/xgboost/tree/master/demo/multiclass\_classification)是根据34个特征识别6种皮肤病，由于终端中运行runexp.sh没有反应，也不报错，所以我干脆把数据集下载到对应的demo文件夹下了,主要的代码如下，原来有部分比较难懂的语句我自己加了一些注释，这样理解起来就会顺畅多了。

```
[python]
01.  #!/usr/bin/python
02.  import numpy as np
03.  import xgboost as xgb
04.
05.  # label need to be 0 to num_class -1
06.  # if col 33 is '?' let it be 1 else 0, col 34 subtract 1
```



```

07. data = np.loadtxt('./dermatology.data', delimiter=',', converters=
08. {33: lambda x:int(x == '?'), 34: lambda x:int(x)-1 } )
09. sz = data.shape
10. train = data[:int(sz[0] * 0.7), :] # take row 1-256 as training set
11. test = data[int(sz[0] * 0.7):, :] # take row 257-366 as testing set
12.
13. train_X = train[:,0:33]
14. train_Y = train[:, 34]
15.
16.
17. test_X = test[:,0:33]
18. test_Y = test[:, 34]
19.
20. xg_train = xgb.DMatrix( train_X, label=train_Y)
21. xg_test = xgb.DMatrix(test_X, label=test_Y)
22. # setup parameters for xgboost
23. param = {}
24. # use softmax multi-class classification
25. param['objective'] = 'multi:softmax'
26. # scale weight of positive examples
27. param['eta'] = 0.1
28. param['max_depth'] = 6
29. param['silent'] = 1
30. param['nthread'] = 4
31. param['num_class'] = 6
32.
33. watchlist = [ (xg_train, 'train'), (xg_test, 'test') ]
34. num_round = 5
35. bst = xgb.train(param, xg_train, num_round, watchlist );
36. # get prediction
37. pred = bst.predict( xg_test );
38.
39. print ('predicting, classification error=%f' % (sum( int(pred[i]) != test_Y[i] for i in range(len(t
40.
41. # do the same thing again, but output probabilities
42. param['objective'] = 'multi:softprob'
43. bst = xgb.train(param, xg_train, num_round, watchlist );
44. # Note: this convention has been changed since xgboost-unity
45. # get prediction, this is in 1D array, need reshape to (ndata, nclass)
46. yprob = bst.predict( xg_test ).reshape( test_Y.shape[0], 6 )
47. ylabel = np.argmax(yprob, axis=1) # return the index of the biggest prob
48.
49. print ('predicting, classification error=%f' % (sum( int(ylabel[i]) != test_Y[i] for i in range(1er

```

结果如下：

```

[python]
01. [0] train-merror:0.011719 test-merror:0.127273
02. [1] train-merror:0.015625 test-merror:0.127273
03. [2] train-merror:0.011719 test-merror:0.109091

```



## 在线课程



腾讯云容器服务架构实现  
介绍 ()  
讲师：董晓杰



容器技术在58同城的实践  
([http://edu.csdn.net/huizhi/course/series\\_detail/73?](http://edu.csdn.net/huizhi/course/series_detail/73?utm_source=blog9)  
[utm\\_source=blog9](http://edu.csdn.net/huizhi/course/series_detail/73?utm_source=blog9))

## 他的热门文章

Python 数据处理扩展包：pandas 模块的 DataFrame 介绍（创建和基本操作）(<http://blog.csdn.net/u010159842/article/details/52759224>)  
15622

ubuntu16.04卸载NVIDIA驱动 (<http://blog.csdn.net/u010159842/article/details/54344583>)  
13541

pandas之dataframe移动复制删除列 (<http://blog.csdn.net/u010159842/article/details/53102281>)  
13474

如何保存Keras模型 (<http://blog.csdn.net/u010159842/article/details/54407745>)  
12661

python删除pandas DataFrame的某一/几列 (<http://blog.csdn.net/u010159842/artic>)

```
04. [3] train-merror:0.007812 test-merror:0.081818
05. [4] train-merror:0.007812 test-merror:0.090909
06. predicting, classification error=0.090909
07. [0] train-merror:0.011719 test-merror:0.127273
08. [1] train-merror:0.015625 test-merror:0.127273
09. [2] train-merror:0.011719 test-merror:0.109091
10. [3] train-merror:0.007812 test-merror:0.081818
11. [4] train-merror:0.007812 test-merror:0.090909
12. predicting, classification error=0.090909
```

不管是直接返回诊断类型，还是返回各类型的概率，然后取概率最大的那个对应的类型的index，结果都是一样的。

## 结语

强烈建议大家使用python notebook来实现代码，当有不明白的代码时看一下执行后的结果能帮助我们很快理解。同时要感叹一下，看大神们的代码感觉好牛X，对我这个XGBoost paper看过两遍还没能完全领略算法精髓的人来说只能拿来主义了，希望后面有机会去读一读算法源码。



发表你的评论

([http://my.csdn.net/weixin\\_35068028](http://my.csdn.net/weixin_35068028))

## 相关文章推荐

### XGBoost：二分类问题 (<http://blog.csdn.net/zc02051126/article/details/46709599>)

二分类问题本文介绍XGBoost的命令行使用方法。Python和R的使用方法见<https://github.com/dmlc/xgboost/blob/master/doc/README.md>。...



zc02051126 (<http://blog.csdn.net/zc02051126>) 2015年07月01日 15:38 16328

### XGBoost：参数解释 (<http://blog.csdn.net/zc02051126/article/details/46711047>)

XGBoost参数在运行XGboost之前，必须设置三种类型成熟：general parameters，booster parameters和task parameters：

e/details/52859035)

12552

## 相关推荐



XGBoost：二分类问题 (<http://blog.csdn.net/zc02051126/article/details/46709599>)

XGBoost：参数解释 (<http://blog.csdn.net/zc02051126/article/details/46711047>)

xgboost 二分类问题实例 (<http://blog.csdn.net/shenxiaoming77/article/details/76037930>)

xgboost使用案例二 (<http://blog.csdn.net/hb707934728/article/details/70739382>)

General para...

 zc02051126 (<http://blog.csdn.net/zc02051126>) 2015年07月01日 17:06  49470



广告



### 【程序员之路】我是前端工程师，怎么了？

今天我30岁了，在此之际，回想我的程序生涯之路，十分感慨，谈谈我作为程序员的选择之路..

([http://www.baidu.com/cb.php?c=lgF\\_pyfqHmknj0dP1f0IZ0qnfK9ujYzP1nYPH0k0Aw-5Hc3rHnYnHb0TAq15HfLPWRznjb0T1dBnv7bPvDkuHTsnj0LPWTd0AwY5HDdnHckrHRYrHc0lgF\\_5y9YIZ0IQzq-uZR8mLPbUB48ugfEIAqspynElvNBnHqdlAdxTvqdThP-5yF\\_UvTkn0KzujYk0AFV5H00TZcqN0KdpyfqHRLPjnvnfKEpyfqHc4rj6kP0KWpyfqP1cvrHnz0AqLUWYs0ZK45HcsP6KWThnqPjDzP1T](http://www.baidu.com/cb.php?c=lgF_pyfqHmknj0dP1f0IZ0qnfK9ujYzP1nYPH0k0Aw-5Hc3rHnYnHb0TAq15HfLPWRznjb0T1dBnv7bPvDkuHTsnj0LPWTd0AwY5HDdnHckrHRYrHc0lgF_5y9YIZ0IQzq-uZR8mLPbUB48ugfEIAqspynElvNBnHqdlAdxTvqdThP-5yF_UvTkn0KzujYk0AFV5H00TZcqN0KdpyfqHRLPjnvnfKEpyfqHc4rj6kP0KWpyfqP1cvrHnz0AqLUWYs0ZK45HcsP6KWThnqPjDzP1T))



### xgboost 二分类问题实例 (<http://blog.csdn.net/shenxiaoming77/article/details/76037930>)

二分类问题 本文介绍XGBoost的命令行使用方法。Python和R的使用方法见<https://github.com/dmlc/xgboost/blob/master/doc/README.md>。...

 shenxiaoming77 (<http://blog.csdn.net/shenxiaoming77>) 2017年07月24日 20:49  591



### xgboost使用案例二 (<http://blog.csdn.net/hb707934728/article/details/70739382>)

# -\*- encoding:utf-8 -\*- #xgboost安装教程 参考 [http://blog.csdn.net/lht\\_okk/article/details/54311333](http://blog.csdn.net/lht_okk/article/details/54311333) #xgbo...

 hb707934728 (<http://blog.csdn.net/hb707934728>) 2017年04月25日 14:49  1708

### XGBoost解决多分类问题 ([http://blog.csdn.net/Leo\\_Xu06/article/details/52424924](http://blog.csdn.net/Leo_Xu06/article/details/52424924))

XGBoost解决多分类问题

 Leo\_Xu06 ([http://blog.csdn.net/Leo\\_Xu06](http://blog.csdn.net/Leo_Xu06)) 2016年09月03日 19:25  6632



广告

### 一学就会的 WordPress 实战课



学习完本课程可以掌握基本的 WordPress 的开发能力，后续可以根据需要开发适合自己的主题、插件，打造最个性的 WordPress 站点。

([http://www.baidu.com/cb.php?c=lgF\\_pyfqHmknjfvP1m0IZ0qnfK9ujYzP1f4Pjnz0Aw-5Hc4nj6vPjm0TAq15Hf4rjn1n1b0T1YYnHRZmvmw9PWf1nA79mHTd0AwY5HDdnHckrHRYrHc0lgF\\_5y9YIZ0IQzqMpgwBUvqoQhP8QvIGIAPCmgfEmvq\\_lyd8Q1N9nHmvnj7hnHPWnjFhPAD1Pyn4uW99ujqdlAdxTv5HDknWw9mhkEusKzujYk0AFV5H00TZcqN0KdpyfqHRLPjnvnfKEpyfqHnsnj0YnsKWpyfqP1cvrHnz0AqLUWYs0ZK45HcsP6KWThnqnHcvP6](http://www.baidu.com/cb.php?c=lgF_pyfqHmknjfvP1m0IZ0qnfK9ujYzP1f4Pjnz0Aw-5Hc4nj6vPjm0TAq15Hf4rjn1n1b0T1YYnHRZmvmw9PWf1nA79mHTd0AwY5HDdnHckrHRYrHc0lgF_5y9YIZ0IQzqMpgwBUvqoQhP8QvIGIAPCmgfEmvq_lyd8Q1N9nHmvnj7hnHPWnjFhPAD1Pyn4uW99ujqdlAdxTv5HDknWw9mhkEusKzujYk0AFV5H00TZcqN0KdpyfqHRLPjnvnfKEpyfqHnsnj0YnsKWpyfqP1cvrHnz0AqLUWYs0ZK45HcsP6KWThnqnHcvP6))

### XGBoost : 二分类问题 ([http://blog.csdn.net/levy\\_cui/article/details/60877008](http://blog.csdn.net/levy_cui/article/details/60877008))





二分类问题 本文介绍XGBoost的命令行使用方法。Python和R的使用方法见<https://github.com/dmlc/xgboost/blob/master/doc/README.md>。...

 levy\_cui ([http://blog.csdn.net/levy\\_cui](http://blog.csdn.net/levy_cui)) 2017年03月08日 17:37  479



## XGBoost-Python完全调参指南-参数解释篇 (<http://blog.csdn.net/wzmsltw/article/details/50...>)

关于XGBoost的参数，发现已经有比较完善的翻译了。故本文转载其内容，并作了一些修改与拓展。原文链接见：<http://blog.csdn.net/zc02051126/article/detail...>

 wzmsltw (<http://blog.csdn.net/wzmsltw>) 2016年03月27日 22:28  38724


## xgboost原理 (<http://blog.csdn.net/a819825294/article/details/51206410>)

文章内容可能会相对比较多，读者可以点击上方目录，直接阅读自己感兴趣的章节。1.序 距离上一次编辑将近10个月，幸得爱可老师（微博）推荐，访问量陡增。最近毕业论文与xgboost相关，于是重新写一下...

 a819825294 (<http://blog.csdn.net/a819825294>) 2016年04月21日 10:15  87236

## XGBoost：二分类问题 ([http://blog.csdn.net/flyinghorse\\_2012/article/details/50533363](http://blog.csdn.net/flyinghorse_2012/article/details/50533363))

本文介绍XGBoost的命令行使用方法。Python和R的使用方法见<https://github.com/dmlc/xgboost/blob/master/doc/README.md>。下面将介...

 flyinghorse\_2012 ([http://blog.csdn.net/flyinghorse\\_2012](http://blog.csdn.net/flyinghorse_2012)) 2016年01月17日 20:43  937

## 从数组中找出一对元素，其和是一个给定的目标数字。假设数组中只存在一个符合要求的数值对...

【解题分析】对于数组中某个下标i，如何判断它是否属于符合条件的两个数字之一？最直观的就是再次扫描数字，判断target - array【i】是否存在于数组中。这样做时间复杂度O(n^2)，效率不高，原因是...

 woshilingdaoren (<http://blog.csdn.net/woshilingdaoren>) 2016年04月12日 09:42  858

## Matlab中的并交子集运算 ([http://blog.csdn.net/pretend\\_b/article/details/5941357](http://blog.csdn.net/pretend_b/article/details/5941357))

以下函数都可以在Matlab的help中找到说明与例子。1、并集：unionc = union(A, B) 返回矢量A与B的并集，A与B必须是数字或字符矢量或者字符元胞数组。结果是排序的。c = un...

 pretend\_b ([http://blog.csdn.net/pretend\\_b](http://blog.csdn.net/pretend_b)) 2010年10月14日 17:44  7580



## XGBoost：多分类问题 (<http://blog.csdn.net/xuxiatian/article/details/62224175>)



0





转自：[http://blog.csdn.net/leq\\_xu06/article/details/52424924](http://blog.csdn.net/leq_xu06/article/details/52424924) 下面用数据 UCI Dermatology dataset演示XGBoos...

 xuxiatian (<http://blog.csdn.net/xuxiatian>) 2017年03月15日 14:37  301



### XGBoost：多分类问题 ([http://blog.csdn.net/levy\\_cui/article/details/60877083](http://blog.csdn.net/levy_cui/article/details/60877083))

下面用数据 UCI Dermatology dataset演示XGBoost的多分类问题 首先要安装好XGBoost的C++版本和相应的Python模块，然后执行如下脚本，如果本地没有训练所需要...

 levy\_cui ([http://blog.csdn.net/levy\\_cui](http://blog.csdn.net/levy_cui)) 2017年03月08日 17:40  306

### SVM解决多分类问题的方法 (<http://blog.csdn.net/qing101hua/article/details/50687816>)

SVM解决多分类问题的方法 SVM算法最初是为二值分类问题设计的，当处理多类问题时，就需要构造合适的多类分类器。目前，构造SVM多类分类器的方法主要有两类：一类是直接法，直接在目标函数上进行修改，...

 qing101hua (<http://blog.csdn.net/qing101hua>) 2016年02月18日 18:02  3207

### SVM如何解决多分类问题 (<http://blog.csdn.net/keepreder/article/details/47339105>)

从 SVM的那几张图可以看出来，SVM是一种典型的两类分类器，即它只回答属于正类还是负类的问题。而现实中要解决的问题，往往是多类的问题（少部分例外，例如垃圾邮件过滤，就只需要确定“是”还是“不是”垃圾...

 keepreder (<http://blog.csdn.net/keepreder>) 2015年08月07日 14:41  1835



0



#### 多分类问题中混淆矩阵(Confusion Matrix)的Matlab画法 ([http://download.csdn.net/download/qq\\_14453054/9608888](http://download.csdn.net/download/qq_14453054/9608888))

[http://download.csdn.net/download/qq\\_14453054/9608888](http://download.csdn.net/download/qq_14453054/9608888) 2015年12月07日 11:36 52KB [下载](#)





#### 一种用于多分类问题的改进支持向量机\_黄景涛 ([http://download.csdn.net/download/qq\\_14453054/9608888](http://download.csdn.net/download/qq_14453054/9608888))

[http://download.csdn.net/download/qq\\_14453054/9608888](http://download.csdn.net/download/qq_14453054/9608888) 2015年04月09日 09:51 66KB [下载](#)

### MATLAB实现多分类问题,使用libsvm,1-vs-rest和1-vs-1两种方法代码 (<http://blog.csdn.net/u014381600>)

最近的项目需要实现多分类,主要的平台是MATLAB+LIBSVM. 如果需求只是实现二分类,那非常简单,可以参见这篇博文多分类存在的主要问题是,LIBSVM主要是为二分类设计的,为了实现多分类就需要...

 u014381600 (<http://blog.csdn.net/u014381600>) 2016年09月23日 14:31  2019



### 模式识别多分类问题求解方法 (<http://download.csdn.net/detail/shark0532/...>)

[/http://download.csdn.net/detail/shark0532/...](http://download.csdn.net/detail/shark0532/...) 2015年11月02日 22:53 515KB [下载](#)

### SVM 的推导、特点、优缺点、多分类问题及应用 (<http://blog.csdn.net/keepreder/article/details/473547>)

关于SVM的推导网上有一大堆的资料可以参考，在这里就不在叙述了，重点讲解SVM的其他问题 SVM有如下主要几个特点：  
(1)非线性映射是SVM方法的理论基础,SVM利用内积核函数代替向高维空间的非线性映射...



keepreder (<http://blog.csdn.net/keepreder>) 2015年07月30日 11:07 [3547](#)