



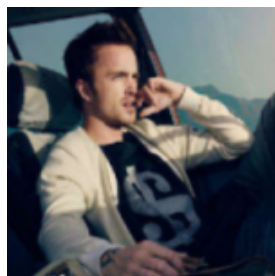
余音、未散的博客

目录视图

摘要视图

RSS 订阅

个人资料



余音、未散

关注

发私信



访问：159798次

积分：2788

等级：BLOG > 5

排名：第14356名

原创：159篇

转载：14篇

图灵赠书——程序员11月书单 【思考】Python这么厉害的原因竟然是！ 感恩节赠书：《深度学习》等另
作译者评选启动！ 每周荐书：京东架构、Linux内核、Python全栈

xgboost使用调参

2016-12-20 15:14

7361人阅读

评论(0)

收藏

分类：

机器学习 (30)

版权声明：本文为博主原创文章，未经博主允许不得转载。

关闭

目录(?)

[+]

github：<https://github.com/dmlc/xgboost>论文参考：<http://www.kaggle.com/blobs/download/forum-messapaper.pdf>

基本思路及优点



望京soho



美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院

译文： 0篇

评论： 38条

github地址

<https://github.com/lytforgood>

文章搜索

文章分类

机器学习 (31)

R (41)

算法 (12)

深度学习 (6)

天池比赛 (6)

机器学习算法MapReduce版 (3)

Hadoop相关 (25)

Spark (16)

linux (22)

Python (8)

爬虫 (6)

深入学习JVM笔记 (3)

Java (12)

神经网络 (11)

面试笔记 (4)

文章存档

<http://blog.csdn.net/q383700092/article/details/60954996>参考<http://dataunion.org/15787.html><http://blog.csdn.net/china1000/article/details/51106856>

在有监督学习中，我们通常会构造一个目标函数和一个预测函数，使用训练样本对目标函数最小化学习到相关的参数，然后用预测函数和训练样本得到的参数来对未知的样本进行分类的标注或者数值的预测。

1. Boosting Tree构造树来拟合残差，而Xgboost引入了二阶导来进行求解，并且引入了节点的数目、参数的L2正则来评估模型的复杂度,构造Xgboost的预测函数与目标函数。

2. 在分裂点选择的时候也以目标函数最小化为目标。

优点：

1. 显示的把树模型复杂度作为正则项加到优化目标中。

2. 公式推导中用到了二阶导数，用了二阶泰勒展开。（GBDT用牛顿法貌似也是二阶

3. 实现了分裂点寻找近似算法。

4. 利用了特征的稀疏性。

5. 数据事先排序并且以block形式存储，有利于并行计算。

6. 基于分布式通信框架rabit，可以运行在MPI和yarn上。（最新已经不基于rabit了）

7. 实现做了面向体系结构的优化，针对cache和内存做了性能1

原理推导及与GBDT区别

<http://blog.csdn.net/q383700092/article/details/60954996>参考<http://dataunion.org/15787.html><https://www.zhihu.com/question/41354392>

参数说明



望京soho



美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院

关闭

2017年09月 (6)
 2017年08月 (1)
 2017年06月 (1)
 2017年03月 (5)
 2017年02月 (4)

展开

阅读排行

xgboost使用调参 (7333)
 朴素贝叶斯之MapReduce版 (7228)
 Xgboost筛选特征重要性 (6484)
 R参考卡片 (5386)
 机器学习面试问题汇总 (5173)
 R可视化绘图三-recharts (5054)
 R-k折交叉验证 (4151)
 阿里音乐流行趋势预测大赛之... (3359)
 R语言-data.table包使用(方便自... (3185)
 机器学习与R之决策树C50算法 (3132)

评论排行

R可视化绘图三-recharts (15)
 机器学习面试编程题汇总 (6)
 去掉CSDN-markdown编辑器的... (4)
 gcForest算法理解 (4)
 机器学习与R之朴素贝叶斯分... (2)

参考http://blog.csdn.net/han_xiaoyang/article/details/52665396

参数

booster：默认 gbtree效果好 (linear booster很少用到)

gbtree：基于树的模型

gbliner：线性模型

silent[默认0]

nthread[默认值为最大可能的线程数]

eta[默认0.3] 学习率 典型值为0.01-0.2

min_child_weight[默认 1] 决定最小叶子节点样本权重和 值较大，避免过拟合 值过高拟合

max_depth[默认6]

gamma[默认0] 指定了节点分裂所需的最小损失函数下降值。这个参数的值越大，算

subsample[默认1] 对于每棵树，随机采样的比例 减小，算法保守，避免过拟合。值设置得太小，它会导致欠拟合 典型值：0.5-1

colsample_bytree[默认1] 每棵随机采样的列数的占比

colsample_bylevel[默认1] 树的每一级的每一次分裂，对列数的采样的占比

lambda[默认1] 权重的L2正则化项

alpha[默认1] 权重的L1正则化项

scale_pos_weight[默认1] 在各类别样本十分不平衡时，参数设
 objective[默认reg:linear] 最小化的损失函数

binary:logistic 二分类的逻辑回归，返回预测的概率(不是类别)
 类器，返回预测的类别(不是概率)。

在这种情况下，你还需要多设一个参数：num_class(类别数目
 数一样，但是返回的是每个数据属于各个类别的概率。

eval_metric[默认值取决于objective参数的取值]

对于回归问题，默认值是rmse，对于分类问题，默认值是erro

关闭



望京soho



美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院

- 对机器学习与数据竞赛的一些... (2)
- R参考卡片 (2)
- GBDT 原理与使用 (1)
- 机器学习与R笔记之线性回归 (1)
- Xgboost筛选特征重要性 (1)

推荐文章

- * 【2017年11月27日】CSDN博客更新周报
- * 【CSDN】邀请您来GitChat赚钱啦！
- * 【GitChat】精选——JavaScript进阶指南
- * 改做人工智能之前，90%的人都没能给自己定位
- * TensorFlow 人脸识别网络与对抗网络搭建
- * Vue 移动端项目生产环境优化
- * 面试必考的计算机网络知识点梳理

最新评论

- gcForest算法理解
qq_40368473 : 为什么在南京大学周志华教授的官方网站下载的代码里面看不到gcforest.py这个文件？
- gcForest算法理解
zhaopanpan_Love_73 : RnadamForest本身就可以用袋外误差来进行验证，进而来代替交叉验证，那篇论文公开的源码里面的...
- gcForest算法理解
余音、未散 : @xiongchengluo1129:图片特征+概率特征
- gcForest算法理解

rmse 均方根误差 mae 平均绝对误差 logloss 负对数似然函数值
error 二分类错误率 merror 多分类错误率 mlogloss 多分类logloss损失函数 auc 曲线下面积
seed(默认0) 随机数的种子 设置它可以复现随机数据的结果

sklearn包，XGBClassifier会改变的函数名

eta -> learning_rate

lambda->reg_lambda

alpha->reg_alpha

常用调整参数：

第一步：确定学习速率和tree_based 参数调优的估目

树的最大深度一般3-10

max_depth = 5

节点分裂所需的最小损失函数下降值0.1到0.2

gamma = 0

采样

subsample= 0.8,

colsample_bytree = 0.8

比较小的值，适用极不平衡的分类问题

min_child_weight = 1

类别十分不平衡

scale_pos_weight = 1

关闭



望京soho



美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院

zhaopanpan_Love_73 :想请教一下, raw input Features 如果为图片, 那么在multi-Grained S C...

去掉CSDN-markdown编辑器的上传图片里...
浮生一城 :有设置项吗? 每次手动删除很麻烦.....

去掉CSDN-markdown编辑器的上传图片里...
ChengKaoAO :感谢博主分享!!!

Xgboost筛选特征重要性

tt_tara :你好, 请问特征重要性的值是怎么计算?

R可视化绘图三-recharts

余音、未散 :@qq_37997902:请看新版本, 新版本已经弃用脚本, 直接用函数就可以, 而且使用方法有大改动

去掉CSDN-markdown编辑器的上传图片里...
Errol_King :赞。

R可视化绘图三-recharts

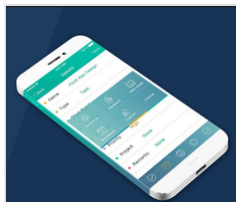
qq_37997902 :您好, 关于echartR.R的脚本



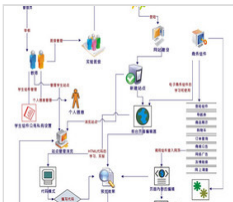
app开发报价单



小学的奥数



app开发公司



电商平台开发



```
1 from xgboost import XGBClassifier
2 xgb1 = XGBClassifier(
3     learning_rate=0.1,
4     n_estimators=1000,
5     max_depth=5,
6     min_child_weight=1,
7     gamma=0,
8     subsample=0.8,
9     colsample_bytree=0.8,
10    objective='binary:logistic',
11    nthread=4,
12    scale_pos_weight=1,
13    seed=27)
```

第二步：max_depth 和 min_weight 参数调优

grid_search参考

http://scikit-learn.org/stable/modules/generated/sklearn.grid_search.GridSearchCV.html

<http://blog.csdn.net/abcjennifer/article/details/23884761>

网格搜索scoring='roc_auc'只支持二分类, 多分类需要修改score

```
1 param_test1 = {
2     'max_depth':range(3,10,2),
3     'min_child_weight':range(1,6,2)
4 }
5 #param_test2 = {
6     'max_depth':[4,5,6],
7     'min_child_weight':[4,5,6]
8 }
```

关闭



望京soho



美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院



```

9  from sklearn import svm, grid_search, datasets
10 from sklearn import grid_search
11 gsearch1 = grid_search.GridSearchCV(
12     estimator = XGBClassifier(
13         learning_rate =0.1,
14         n_estimators=140, max_depth=5,
15         min_child_weight=1,
16         gamma=0,
17         subsample=0.8,
18         colsample_bytree=0.8,
19         objective= 'binary:logistic',
20         nthread=4,
21         scale_pos_weight=1,
22         seed=27),
23     param_grid = param_test1,
24     scoring='roc_auc',
25     n_jobs=4,
26     iid=False,
27     cv=5)
28 gsearch1.fit(train[predictors],train[target])
29 gsearch1.grid_scores_, gsearch1.best_params_,gsearch1.best_score_
30 #网格搜索scoring='roc_auc'只支持二分类，多分类需要修改scoring

```

第三步：gamma参数调优

```

1  param_test3 = {
2      'gamma':[i/10.0 for i in range(0,5)]
3  }
4  gsearch3 = GridSearchCV(
5      estimator = XGBClassifier(

```

关闭



望京soho

美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院



```

6 learning_rate = 0.1,
7 n_estimators = 140,
8 max_depth = 4,
9 min_child_weight = 6,
10 gamma = 0,
11 subsample = 0.8,
12 colsample_bytree = 0.8,
13 objective = 'binary:logistic',
14 nthread = 4,
15 scale_pos_weight = 1,
16 seed = 27,
17 param_grid = param_test3,
18 scoring = 'roc_auc',
19 n_jobs = 4,
20 iid = False,
21 cv = 5)
22 gsearch3.fit(train[predictors], train[target])
23 gsearch3.grid_scores_, gsearch3.best_params_, gsearch3.best_score_

```

第四步：调整subsample 和 colsample

```

1 #取0.6,0.7,0.8,0.9作为起始值
2 param_test4 = {
3     'subsample': [i/10.0 for i in range(6,10)],
4     'colsample_bytree': [i/10.0 for i in range(6,10)]
5 }
6
7 gsearch4 = GridSearchCV(
8     estimator = XGBClassifier(
9         learning_rate = 0.1,

```

关闭



望京soho



美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院



```

10 n_estimators=177,
11 max_depth=3,
12 min_child_weight=4,
13 gamma=0.1,
14 subsample=0.8,
15 colsample_bytree=0.8,
16 objective= 'binary:logistic',
17 nthread=4,
18 scale_pos_weight=1,
19 seed=27),
20 param_grid = param_test4,
21 scoring='roc_auc',
22 n_jobs=4,
23 iid=False,
24 cv=5)
25 gsearch4.fit(train[predictors],train[target])
26 gsearch4.grid_scores_, gsearch4.best_params_, gsearch4.best_score_

```

第五步：正则化参数调优

```

1 param_test6 = {
2   'reg_alpha':[1e-5, 1e-2, 0.1, 1, 100]
3 }
4 gsearch6 = GridSearchCV(
5   estimator = XGBClassifier(
6     learning_rate =0.1,
7     n_estimators=177,
8     max_depth=4,
9     min_child_weight=6,
10    gamma=0.1,

```

[关闭](#)


望京soho



美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院



```

11 subsample=0.8,
12 colsample_bytree=0.8,
13 objective= 'binary:logistic',
14 nthread=4,
15 scale_pos_weight=1,
16 seed=27),
17 param_grid = param_test6,
18 scoring='roc_auc',
19 n_jobs=4,
20 iid=False,
21 cv=5)
22 gsearch6.fit(train[predictors],train[target])
23 gsearch6.grid_scores_, gsearch6.best_params_, gsearch6.best_score_

```

第六步：降低学习速率

```

1 xgb4 = XGBClassifier(
2     learning_rate =0.01,
3     n_estimators=5000,
4     max_depth=4,
5     min_child_weight=6,
6     gamma=0,
7     subsample=0.8,
8     colsample_bytree=0.8,
9     reg_alpha=0.005,
10    objective= 'binary:logistic',
11    nthread=4,
12    scale_pos_weight=1,
13    seed=27)
14 modelfit(xgb4, train, predictors)

```



望京soho

美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院



python示例

```

1 import xgboost as xgb
2 import pandas as pd
3 #获取数据
4 from sklearn import cross_validation
5 from sklearn.datasets import load_iris
6 iris = load_iris()
7 #切分数据集
8 X_train, X_test, y_train, y_test = cross_validation.train_test_split(iris.data, iris.target, test_size=0
9 #设置参数
10 m_class = xgb.XGBClassifier(
11     learning_rate=0.1,
12     n_estimators=1000,
13     max_depth=5,
14     gamma=0,
15     subsample=0.8,
16     colsample_bytree=0.8,
17     objective='binary:logistic',
18     nthread=4,
19     seed=27)
20 #训练
21 m_class.fit(X_train, y_train)
22 test_21 = m_class.predict(X_test)
23 print "Accuracy : %.2f" % metrics.accuracy_score(y_test, test_21)
24 #预测概率
25 #test_2 = m_class.predict_proba(X_test)
26 #查看AUC评价标准
27 from sklearn import metrics
28 print "Accuracy : %.2f" % metrics.accuracy_score(y_test, test_21)

```

关闭



望京soho

美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院



```

29 ##必须二分类才能计算
30 ##print "AUC Score (Train): %f" % metrics.roc_auc_score(y_test, test_2)
31 #查看重要程度
32 feat_imp = pd.Series(m_class.booster().get_fscore()).sort_values(ascending=False)
33 feat_imp.plot(kind='bar', title='Feature Importances')
34 import matplotlib.pyplot as plt
35 plt.show()
36 #回归
37 #m_regress = xgb.XGBRegressor(n_estimators=1000,seed=0)
38 #m_regress.fit(X_train, y_train)
39 #test_1 = m_regress.predict(X_test)

```

整理

xgb原始

```

1 from sklearn.model_selection import train_test_split
2 from sklearn import metrics
3 from sklearn.datasets import make_hastie_10_2
4 import xgboost as xgb
5 #记录程序运行时间
6 import time
7 start_time = time.time()
8 X, y = make_hastie_10_2(random_state=0)
9 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5,
10 #xgb矩阵赋值
11 xgb_train = xgb.DMatrix(X_train, label=y_train)
12 xgb_test = xgb.DMatrix(X_test, label=y_test)

```



望京soho

美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院



```

13 ##参数
14 params={
15 'booster':'gbtree',
16 'silent':1, #设置成1则没有运行信息输出，最好是设置为0.
17 '#nthread':7, # cpu 线程数 默认最大
18 'eta': 0.007, # 如同学习率
19 'min_child_weight':3,
20 # 这个参数默认是 1，是每个叶子里面 h 的和至少是多少，对正负样本不均衡时的 0-1 分类而言
21 #，假设 h 在 0.01 附近，min_child_weight 为 1 意味着叶子节点中最少需要包含 100 个样本。
22 #这个参数非常影响结果，控制叶子节点中二阶导的和的最小值，该参数值越小，越容易 overfit
23 'max_depth':6, # 构建树的深度，越大越容易过拟合
24 'gamma':0.1, # 树的叶子节点上作进一步分区所需的最小损失减少,越大越保守，一般0.1、0.2过
25 'subsample':0.7, # 随机采样训练样本
26 'colsample_bytree':0.7, # 生成树时进行的列采样
27 'lambda':2, # 控制模型复杂度的权重值的L2正则化项参数，参数越大，模型越不容易过拟合。
28 '#alpha':0, # L1 正则项参数
29 '#scale_pos_weight':1, #如果取值大于0的话，在类别样本不平衡的情况下有助于快速收敛。
30 '#objective': 'multi:softmax', #多分类的问题
31 '#num_class':10, # 类别数，多分类与 multisoftmax 并用
32 'seed':1000, #随机种子
33 '#eval_metric': 'auc'
34 }
35 plst = list(params.items())
36 num_rounds = 100 # 迭代次数
37 watchlist = [(xgb_train, 'train'),(xgb_test, 'val')]
38
39 #训练模型并保存
40 # early_stopping_rounds 当设置的迭代次数较大时，early_stoppin
41 model = xgb.train(plst, xgb_train, num_rounds, watchlist, early_st
42 #model.save_model('./model/xgb.model') # 用于存储训练出的模
43 print "best best_ntree_limit", model.best_ntree_limit

```

关闭



望京soho



美国前50名大学

练书法的好处 app开发报价单
 气垫bb使用方法 冰点脱毛价格
 oa系统 app制作 python 周末班
 打印机租赁 美国十大商学院



```

44 y_pred = model.predict(xgb_test, ntree_limit=model.best_ntree_limit)
45 print('error=%f' % (sum(1 for i in range(len(y_pred)) if int(y_pred[i]>0.5)!=y_test[i]) / float(len(y_pred))))
46 #输出运行时长
47 cost_time = time.time()-start_time
48 print "xgboost success!","\n","cost time:",cost_time,"(s)....."
49

```

xgb使用sklearn接口(推荐)

官方

会改变的函数名是：

eta -> learning_rate

lambda -> reg_lambda

alpha -> reg_alpha



```

1 from sklearn.model_selection import train_test_split
2 from sklearn import metrics
3 from sklearn.datasets import make_hastie_10_2
4 from xgboost.sklearn import XGBClassifier
5 X, y = make_hastie_10_2(random_state=0)
6 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5,
7 clf = XGBClassifier(
8 silent=0, #设置成1则没有运行信息输出，最好是设置为0.是否在运
9 #nthread=4, # cpu 线程数 默认最大
10 learning_rate= 0.3, # 如同学习率
11 min_child_weight=1,
12 # 这个参数默认是 1，是每个叶子里面 h 的和至少是多少，对正负
13 #，假设 h 在 0.01 附近，min_child_weight 为 1 意味着叶子节点中
14 #这个参数非常影响结果，控制叶子节点中二阶导的导数的最小值，i

```

关闭



望京soho

美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院



```

15 max_depth=6, # 构建树的深度, 越大越容易过拟合
16 gamma=0, # 树的叶子节点上作进一步分区所需的最小损失减少, 越大越保守, 一般0.1、0.2这样子。
17 subsample=1, # 随机采样训练样本 训练实例的子采样比
18 max_delta_step=0, # 最大增量步长, 我们允许每个树的权重估计。
19 colsample_bytree=1, # 生成树时进行的列采样
20 reg_lambda=1, # 控制模型复杂度的权重值的L2正则化项参数, 参数越大, 模型越不容易过拟合。
21 #reg_alpha=0, # L1 正则项参数
22 #scale_pos_weight=1, #如果取值大于0的话, 在类别样本不平衡的情况下有助于快速收敛。平衡正负权重
23 #objective= 'multi:softmax', #多分类的问题 指定学习任务和相应的学习目标
24 #num_class=10, # 类别数, 多分类与 multisoftmax 并用
25 n_estimators=100, #树的个数
26 seed=1000 #随机种子
27 #eval_metric= 'auc'
28 )
29 clf.fit(X_train,y_train,eval_metric='auc')
30 #设置验证集合 verbose=False不打印过程
31 clf.fit(X_train, y_train, eval_set=[(X_train, y_train), (X_val, y_val)], eval_metric='auc', verbose=False)
32 #获取验证集合结果
33 evals_result = clf.evals_result()
34 y_true, y_pred = y_test, clf.predict(X_test)
35 print("Accuracy : %.4g" % metrics.accuracy_score(y_true, y_pred))
36 #回归
37 #m_regress = xgb.XGBRegressor(n_estimators=1000, seed=0)

```

网格搜索

可以先固定一个参数 最优化后继续调整

第一步：确定学习速率和tree_based 给个常见初始值 根据是否
max_depth, min_child_weight, gamma, subsample, scale_pos_weight

关闭



望京soho



美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院



max_depth=3 起始值在4-6之间都是不错的选择。

min_child_weight比较小的值解决极不平衡的分类问题eg:1

subsample, colsample_bytree = 0.8: 这个是最常见的初始值了

scale_pos_weight = 1: 这个值是因为类别十分不平衡。

第二步：max_depth 和 min_weight 对最终结果有很大的影响

‘max_depth’:range(3,10,2),

‘min_child_weight’:range(1,6,2)

先大范围地粗调参数，然后再小范围地微调。

第三步：gamma参数调优

‘gamma’:[i/10.0 for i in range(0,5)]

第四步：调整subsample 和 colsample_bytree 参数

‘subsample’:[i/100.0 for i in range(75,90,5)],

‘colsample_bytree’:[i/100.0 for i in range(75,90,5)]

第五步：正则化参数调优

‘reg_alpha’:[1e-5, 1e-2, 0.1, 1, 100]

‘reg_lambda’

第六步：降低学习速率

learning_rate = 0.01,

```
1 from sklearn.model_selection import GridSearchCV
2 tuned_parameters= [{'n_estimators':[100,200,500],
3                       'max_depth':[3,5,7], ##range(3,10,2)
4                       'learning_rate':[0.5, 1.0],
5                       'subsample':[0.75,0.8,0.85,0.9]
6                       }]
7 tuned_parameters= [{'n_estimators':[100,200,500,1000]
8                       }]
```

关闭



望京soho

美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院



```

9  clf = GridSearchCV(XGBClassifier(silent=0,nthread=4,learning_rate= 0.5,min_child_weight=1, max_depth=3,gan
10  clf.fit(X_train, y_train)
11  ##clf.grid_scores_, clf.best_params_, clf.best_score_
12  print(clf.best_params_)
13  y_true, y_pred = y_test, clf.predict(X_test)
14  print"Accuracy : %.4g" % metrics.accuracy_score(y_true, y_pred)
15  y_proba=clf.predict_proba(X_test)[:,:1]
16  print "AUC Score (Train): %f" % metrics.roc_auc_score(y_true, y_proba)

```

```

1  from sklearn.model_selection import GridSearchCV
2  parameters= [{'learning_rate':[0.01,0.1,0.3],'n_estimators':[1000,1200,1500,2000,2500]}]
3  clf = GridSearchCV(XGBClassifier(
4      max_depth=3,
5      min_child_weight=1,
6      gamma=0.5,
7      subsample=0.6,
8      colsample_bytree=0.6,
9      objective= 'binary:logistic', #逻辑回归损失函数
10     scale_pos_weight=1,
11     reg_alpha=0,
12     reg_lambda=1,
13     seed=27
14 ),
15     param_grid=parameters,scoring='roc_auc')
16  clf.fit(X_train, y_train)
17  print(clf.best_params_)
18  y_pre= clf.predict(X_test)
19  y_pro= clf.predict_proba(X_test)[:,:1]
20  print "AUC Score : %f" % metrics.roc_auc_score(y_test, y_pro)
    print"Accuracy : %.4g" % metrics.accuracy_score(y_test, y_pre)

```

关闭



望京soho

美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院

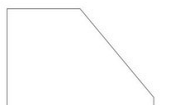


开发一个app多少钱

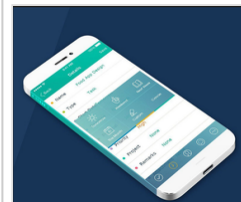


app开发报价单

一道小学四年级的奥数题
在图上画一条直线使下面的图形划为两个三角形



小学的奥数



app开发公司



电商平台开



输出特征重要性

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 feat_imp = pd.Series(clf.booster().get_fscore()).sort_values(ascending=False)
4 feat_imp.plot(kind='bar', title='Feature Importances')
5 plt.ylabel('Feature Importance Score')
6 plt.show()
```

顶

3

踩

0

- 上一篇 GBDT 原理与使用
- 下一篇 并发和并行

相关文章推荐

- 机器学习系列(12)_XGBoost参数调优完全指南（附...)
- 机器学习竞
- 腾讯云容器服务架构实现介绍--董晓杰
- Python可以j
- xgboost入门与实战（实战调参篇）
- Sk-learn及算
- 容器技术在58同城的实践--姚远
- 华为工程师
- XGBoost-Python完全调参指南-参数解释篇
- xgboost入门



望京soho



美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院

关闭



开发一个app多少钱



- Tensorflow项目实战-文本分类
- XGBoost参数调优完全指南（附Python代码）
- MySQL深入浅出
- XGBoost：在Python中使用XGBoost
- kdd2017---踩坑
- Kaggle房价预测进阶版/bagging/boosting/AdaBoost/...



客户管理系统



练习书法



学习书法



图书管理系统



OA系统



怎

查看评论

暂无评论

发表评论

用户名： weixin_35068028

评论内容：



提交

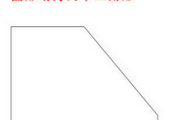
* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场

关闭

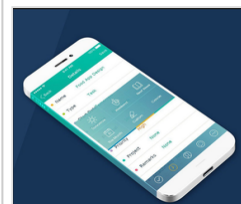


app开发报价单

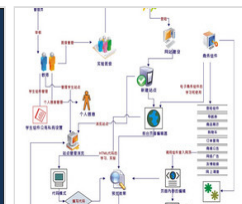
一道小学四年级的奥数题
在图上画一条直线使下面的
图形划为两个三角形



小学的奥数



app开发公司



电商平台开



望京soho



美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院

方式 | 版权声明 | 法律顾问 | 问题报告 | 合作伙伴 | 论坛反馈

webmaster@csdn.net

400-660-0108

| 北京创新乐知信息技术有限公司 版权所有 | 江苏知之为计算机有限公司 | 江苏乐知网络技术有限公司

SDN.NET, All Rights Reserved



开发一个app多少钱



app开发报价单

小学的奥数



app开发公司

电商平台开



关闭



望京soho



美国前50名大学

练书法的好处 app开发报价单

气垫bb使用方法 冰点脱毛价格

oa系统 app制作 python 周末班

打印机租赁 美国十大商学院