

CSDN

博客 (/blog.csdn.net/)

学院 (/edu.csdn.net/)

下载 (/download.csdn.net/)

GitChat (http://gitbook.cn/?ref=csdn)

论坛 (http://bbs.csdn.net)

...

写博客

发Chat

登录 (https://passport.csdn.net/account/login)

注册 (https://passport.csdn.net/account/mobileRegister?action=mobileRegister)

Python/scikit-learn机器学习库(特征选取)

原创

2015年11月09日 21:16:15

标签: 机器学习 (http://so.csdn.net/so/search/s.do?q=机器学习&t=blog)

9050

去除方差小的特征

设置一个方差阈值，没有达到这个方差阈值的特征都会被丢弃。
VarianceThreshold，算法输入只要求特征(X),不需要输入结果(Y)。

```
from sklearn.feature_selection import VarianceThreshold
X=[[feature1,feature2,...],...]
sel=VarianceThreshold(threshold=xx)
print(sel.fit_transform(X))
```

单变量特征选取

单变量特征提取的原理是分别计算每个特征的某个统计指标，根据该指标来选取特征。
SelectKBest、SelectPercentile，前者选择排名前k个的特征，后者选择排名在前k%的特征。选择的统计指标需要指定，对于regression问题，使用f_regression指标;对于classification问题，可以使用chi2或者f_classif指标。

```
from sklearn.feature_selection import SelectKBest,chi2
X_new=SelectKBest(chi2,k=2).fit_transform(test_X,test_Y)
```

- 1. False Positive Rate，假阳性率
- 2. chi2,卡方统计量，X中特征取值必须非负。卡方检验用来测度随机变量之间的依赖关系。通过卡方检验得到的特征之间是最可能独立的随机变量，因此这些特征的区分度很高。

循环特征选取

不单独地检验某个特征的价值，而是检验特征集的价值。对于一个数量为n的特征集合，子集的个数为2的n次方减一。通过指定一个学习算法，通过算法计算所有子集的error，选择error最小的子集作为选取的特征。

RFE

- 1. 对初始特征集中每个特征赋予一个初始权重。
- 2. 训练，将权重最小的特征移除。
- 3. 不断迭代，直到特征集合的数目达到预定值。

ae5555 (http://blog.csdn....)

+关注

(http://blog.csdn.net/ae5555)

原创

粉丝

喜欢

码云未开通

143

11

1

(https://gitee.com/ae5555)

数据可视化

5GB

- 他的最新文章
- 更多文章 (http://blog.csdn.net/ae5555)
- xbgboost (http://blog.csdn.net/ae5555/article/details/73849647)

C++ 基础知识 (http://blog.csdn.net/ae5555/article/details/73656458)

tensorflow总结 (http://blog.csdn.net/ae5555/article/details/53129301)

Spark Sql (http://blog.csdn.net/ae5555/article/details/53089308)

ubuntu press S to skip (http://blog.csdn.net/ae5555/article/details/53089234)

- 文章分类
- 算法 (http://blog.csdn.net/ae5555/article/details/73849647)17篇

机器学习 (http://blog.csdn.net/ae5555/article/details/73656458)16篇

Hadoop&Spark (http://blog.csdn.net/ae5555/article/details/53129301)18篇

数学 (http://blog.csdn.net/ae5555/article/details/53089308)3篇

程序语言 (http://blog.csdn.net/ae5555/article/details/53089234)46篇
- 展开

```
from sklearn.svm import SVC
from sklearn.feature_selection import RFE
//X为样本集合，每个样本为一个数组，数组元素为各个特征值,Y样本的评分
svc=SVC(kernel="linear",C=1)
rfe=RFE(estimator=svc,n_features_to_select=5,step=1)
X_new=rfe.fit_transform(X,Y)
```

RFE CV

在RFE训练时，增加交叉验证。

1
L1-base

在线性回归模型中，每一个特征代表一个w，若得到的w系数等于或接近0，则说明这些特征不重要。

LinearSVC

参数C控制特征系数稀疏度，C的值越小，选择的特征数越少。



```
from sklearn.svm import LinearSVC
X_new=LinearSVC(C=0.01,penalty="l1",dual=False).fit_transform(x,y)
```



决策树特征选取

通过决策树可以计算特征的重要性，抛弃不太重要的特性。

```
from sklearn.ensemble import ExtraTreesClassifier
clf=ExtraTreesClassifier()
X_new=clf.fit(x,y).transform(x)
//各个特征重要性
print(clf.feature_importances)
```

sklearn分类数据

sklearn.datasets.make_classification用来随机产生一个多分类问题。

n_features=n_informative+n_redundant+n_repeated。

n_clusters_per_class 每个分类的集群数

```
import sklearn.datasets
(x,y)=make_classification(n_samples=,
n_features=,
n_informative=,
n_redundant=,
n_repeated=,
n_classes=,
random_state=,
shuffle=False
)
```

版权声明：本文为博主原创文章，未经博主允许不得转载。

文章存档

2017年6月 (http://blog.csdn.net/ae5555/article/details/49534263)	2篇
2016年11月 (http://blog.csdn.net/ae5555/article/details/49534263)	6篇
2016年9月 (http://blog.csdn.net/ae5555/article/details/49534263)	2篇
2016年7月 (http://blog.csdn.net/ae5555/article/details/49534263)	6篇
2016年5月 (http://blog.csdn.net/ae5555/article/details/49534263)	1篇

展开

他的热门文章

Python/scikit-learn机器学习库(特征选取)
(<http://blog.csdn.net/ae5555/article/details/49534263>)

9018

Spring常见问题 (<http://blog.csdn.net/ae5555/article/details/50822142>)

3838

Python/scikit-learn机器学习库(SVM支持向量机) (<http://blog.csdn.net/ae5555/article/details/49612215>)

3657

Python/gensim主题模型库 (<http://blog.csdn.net/ae5555/article/details/49488287>)

2690

spark/MLlib 协同过滤算法 (<http://blog.csdn.net/ae5555/article/details/46969417>)

2414



人脸识别算法



联系我们

网站客服 (<http://wpa.qq.com/msgrd?v=3&uin=2431299880&site=qq&menu=yes>)
vebmaster@csdn.net
(mailto:webmaster@csdn.net)

微博客服 (<http://e.weibo.com/csdnsupport/profile>)
100-660-0108

关于 招聘 广告服务 阿里云

©2018 CSDN 京ICP证09002463号

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

登录

注册



加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！



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/)



机器学习（一）：python三种特征选择方法

python中实现特征选择的三种方法：过滤型：选择与目标变量相关性较强的特征。缺点：忽略了特征之间的关联性。包裹型：基于线性模型相关系数以及模型结果AUC逐步剔除特征。如果剔除相关系数绝对值较小...

zhouwenyuan1015 2017年03月25日 17:54 2345

(<http://blog.csdn.net/zhouwenyuan1015/article/details/65938636>)



结合Scikit-learn介绍几种常用的特征选择方法



key_v 2015年08月26日 19:33 2876

此文系转载,原文地址:<http://chaoslog.com/te-zheng-xuan-ze.html> 特征选择(排序)对于数据科学家、机器学习从业者来说非常重要。好的特征选择能够提升模型的性...

(http://blog.csdn.net/key_v/article/details/48008725)

养猪不如学Python！想要百万年薪嘛？



Python入门：每天1小时，迈出人工智能第一步

机器学习系列(4)_机器学习算法一览，应用建议与解决思路

http://blog.csdn.net/han_xiaoyang/article/details/50469334 出处：http://blog.csdn.net/han_xiao...

zdy0_2004 2016年01月07日 13:55 4688

(http://blog.csdn.net/zdy0_2004/article/details/50475822)

机器学习-->特征选择



Mr_tyting 2017年06月18日 11:39 877

这篇博文将详细介绍sklearn中特征选择的相关内容，全文翻译自sklearn.Feature selection,并且加上了我自己的一些见解。特征选择主要有三种办法：其中第一种过滤型很少使用...

(http://blog.csdn.net/Mr_tyting/article/details/73413979)

结合Scikit-learn介绍几种常用的特征选择方法



Bryan__ 2016年06月07日 22:51 21957

特征选择(排序)对于数据科学家、机器学习从业者来说非常重要。好的特征选择能够提升模型的性能，更能帮助我们理解数据的特点、底层结构，这对进一步改善模型、算法都有着重要作用。特征选择主要有两个功能： ...

(http://blog.csdn.net/Bryan__/_/article/details/51607215)

养猪不如学Python！



Python入门，每天1小时，掌握通向未来的编程语言

Python机器学习库SKLearn的特征选择



cheng9981 2017年04月30日 17:10 1878



参考地址：http://scikit-learn.org/stable/modules/feature_selection.html#feature-selection sklearn.featur...

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](#)

(<http://blog.csdn.net/cheng9981/article/details/71023709>)

特征选择



 javastart 2017年08月09日 20:18  306

特征选择 作者 jacksu在简书 关注 2017.08.09 00:18 字数 1252 阅读 255评论 1喜欢 2 一个基本的数据挖掘场景如下: ...

(<http://blog.csdn.net/javastart/article/details/77015424>)



欢迎使用CSDN-markdown编辑器

 Erin_HH 2017年11月15日 11:35  47

scikit-learn:基于Python语言的机器学习工具scikit-learn(中文网站)是python中一个非常强大的开源工具包, 集合了机器学习经典的分类、回归、聚类算法及数据处理方法, 是数...

(http://blog.csdn.net/Erin_HH/article/details/78539097)

特征选择--scikit-learn

 a1368783069 2016年07月29日 12:28  10887

特征选择 (Feature Selection) :choosing a subset of all the features(the ones more informative)。最终得到的特征选是原来...

(<http://blog.csdn.net/a1368783069/article/details/52048349>)



【数据挖掘 特征选择】特征选择常用方法

 u013421629 2017年12月15日 11:21  91

下面列举一些最常用的方法 1、根据阈值过滤掉方差小的变量。 2、通过计算变量与标签的相关系数, 留下相关性高的特征。 3、根据决策树或者随机森林, 选择重要程度高的特征。 4、利用PCA等算法, 对...

(<http://blog.csdn.net/u013421629/article/details/78810947>)



机器学习算法一览, 应用建议与解决思路

 zwlq1314521 2016年01月14日 17:25  4753

作者: 寒小阳时间: 2016年1月。出处: http://blog.csdn.net/han_xiaoyang/article/details/50469334 声明: 版权所有, 转载请...

(<http://blog.csdn.net/zwlq1314521/article/details/50519077>)



一个更加强大的查壳工具, 更新版本

 stereohomology 2010年10月23日 15:42  13278

<http://pid.gamecopyworld.com/> 2003年的时候, 作者说: cdkillerMOV AL, 1 Join Date: Oct 2002 Posts: 1,...

(<http://blog.csdn.net/stereohomology/article/details/5960837>)

如何进行特征选择

 datoutong_ 2017年12月15日 14:59  100

如何进行特征选择原文地址: <http://blog.csdn.net/u010670689/article/details/73196546> 特征选择对于数据科学家、机器学习从业者来说非常重要。好...

(http://blog.csdn.net/datoutong_/article/details/78813233)



sklearn特征选择和分类模型

 linger2012liu 2015年08月24日 22:18  7869

sklearn特征选择和分类模型数据格式: 这里, 原始特征的输入文件的格式使用libsvm的格式, 即每行是label index1:value 1 index2:value2这种稀疏矩阵的格式。...

(<http://blog.csdn.net/linger2012liu/article/details/47960127>)

文本挖掘之特征选择(python 实现)

 lostinai 2015年12月13日 12:14  1288

<http://www.cnblogs.com/wangbogong/p/3251132.html> 机器学习算法的空间、时间复杂度依赖于输入数据的规模, 维度规约(Dimension...

(<http://blog.csdn.net/lostinai/article/details/50282003>)

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学习系列十九】基于scikit-learn库进行特征选择

场景：特征选择在模型训练前是非常有意义的，实际上就是先期对特征相关性进行分析。参考：<http://blog.csdn.net/fjssharpword/article/details/73550...>

 fjssharpword 2017年06月29日 17:20 1524

<http://blog.csdn.net/fjssharpword/article/details/73920694>


1

模式识别特征选择

2013年11月18日 22:35 143KB [下载](#)



5用scikit-learn进行FeatureSelection

 lming_08 2014年09月11日 19:45 4165

```
1. >>> from sklearn.datasets import load_iris >>> from sklearn.feature_selection import SelectKBe...
```

http://blog.csdn.net/lming_08/article/details/39210409


特征工程（sklearn）

 u012102306 2016年08月24日 13:09 3399

一、特征工程是什么 有这么一句话在业界广泛流传：数据和特征决定了机器学习的上限，而模型和算法只是逼近这个上限而已。那特征工程到底是什么呢？顾名思义，其本质是一项工程活动，目的是最大限度...

<http://blog.csdn.net/u012102306/article/details/52299427>

sklearn学习——特征工程(特征选择)

 sqiu_11 2017年03月02日 12:12 3868

##什么是特征工程？## 定义：特征工程是将原始数据转化为特征，更好表示预测模型处理的实际问题，提升对于未知数据的准确性。它是用目标问题所在的特定领域知识或者自动化的方法来生成、提取、删减或者组合变化...

http://blog.csdn.net/sqliu_11/article/details/59487177