

博客 (//blog.csdn.net/)

学院 (//edu.cs(///wwwt).csd示藏t)(http://download.csdn.net)

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

ďЪ





證录 (https://pass/part.csdn/het/socouth/hog/ship://etihorktcfp://passport.csdn.net/account/mobileregister?action=mobileRegister) /postedit) /new/gitchat

三机器学习的特征选择**(feature selection)**

2016年11月30日 22:11:37

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

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

෯ matlab (http://so.csdn.net/so/search/s.do?q=matlab&t=blog) /

特征选择 (http://so.csdn.net/so/search/s.do?q=特征选择&t=blog) /

scikit (http://so.csdn.net/so/search/s.do?q=scikit&t=blog)

905

写在开头:刚开始接触机器学习,选择了周志华教授的《机器学习》作为入门教材,很不错的书籍,受益 良多。没有使用matlab去编写代码,而是选择了使用python的scikit-learn的开发包,大致看了一下开发包 的特征选择方法,结合周志华教授的书,想先总结总结feature selection的方法

特征选择

filter(过滤式)

思路

通过评估自变量与目标变量之间的关联,也就是量化两者之间的关系,我们暂且叫做量化值为"相关统计 量", 然后根据相关统计量, 找出相关性较强的k个自变量。

如何量化两者的关联

通常有相关系数,卡方检验 (http://blog.csdn.net/idatamining/article/details/8564981),信息增益,和互

结合scikit-learn

1.Removing features with low variance

移除变量中变化较小的,一般用于自变量为离散型时,某个变量在n个样本中取相同的值的频率大于既定的 阈值的时候,我们直接删除

2.Univariate feature selection

单变量特征选择,其中涉及到的函数最主要的两个是SelectKBest,SelectPercentile,一个选择前k个,一个 选择前k%个。其中涉及到的量化方式主要有:chi2(卡方检验),f_classif,f_regression前两个用于分类 , 后一 加入仓务内回与受更精准的内容推荐,与500万程序员共同成长!

selous (http://blog.csdn.... +关注 (http://blog.csdn.net 码云 /selous) 粉丝 原创 喜欢 未开通 79 8 10 (https://gite



他的最新文章

更多文章 (http://blog.csdn.net/selous)

ubuntu更新cmake到3.0之后的最新版本 (http://blog.csdn.net/selous/article/details/ 79080077)

ubuntu分区 (http://blog.csdn.net/selous/a rticle/details/79074317)

Adversarial Sample misleading the Model (生成对抗样本迷惑模型) (http://blog.csd n.net/selous/article/details/78815393)

ubuntu nohup命令 (http://blog.csdn.net/s elous/article/details/78330991)

关于svm的自己一直比较困惑的点 (http:/ /blog.csdn.net/selous/article/details/7829 6142)

文章分类

web服务器开发 (http://blog.cs... 网络 (http://blog.csdn.net/selo... 1篇 python (http://blog.csdn.net/s... 9篇 云计算 (http://blog.csdn.net/s... 机器学习 (htt 加 log.csdn.net 注册 22篇 🗙

第1页 共7页

还有几个看不懂的SelectFpr,SelectFdr,SelectFwe还没用到就没有深究。

wrapper(包裹式)

思路

好的特征子集。

如何产生特征子集

通常使用的方法有,完全搜索,启发式搜索,随机搜索。

结合scikit-learn

[.Recursive feature elimination(RFE算法)

。 园归特征估计,类似与启发式搜索,产生特征子集。

First, the estimator is trained on the initial set of features and weights are assigned to each one of them. Then, features whose absolute weights are the smallest are pruned from the current set features. That procedure is recursively repeated on the pruned set until the desired number of features to select is eventually reached.

也就是说一开始使用一个选择器在原始数据上面训练,然后除掉权重最小的那个特征,然后继续递归训 练,直到剩下的特征满足要求的时候结束。有点贪心思想的意味在里面。

RFE算法一般要控制筛选之后的特征的数目,不一定是最优的

2.RFECV算法是结合交叉验证的RFE,会通过调查验证的方式,去寻找最优的特征数目。

embedding(嵌入式)

思路

学习器自身就具有特征选择的能力

常见的选择器

lasso,引入l1正则化项;ridge regression,引入l2正则化项;决策树的算法,通过信息增益和信息熵等方 式,一边筛选,一边建模。

结合scikit-learn

1. Feature selection using SelectFromModel 从模型中实现特征选择,其中需要注意的就是其中prefit参数的设置

Whether a prefit model is expected to be passed into the constructor directly or not. If True, transform must be called directly and SelectFromModel cannot be used with cross val score, GridSearchCV and similar utilities that clone the estimator. Otherwise train the model using fit and then transform to do feature selection.

加入CSDN,享受更精准的内容推荐,与500万程序员共同成长!

展开~

文章存档

2018年1月 (http://blog.csdn.n... 2篇 2017年12月 (http://blog.csdn.... 1篇 2017年10月 (http://blog.csdn.... 2篇 2017年9月 (http://blog.csdn.n... 1篇 2017年8月 (http://blog.csdn.n... 7篇

展开~

┃他的热门文章

tensorflow使用tensorboard实现数据可视 化 (http://blog.csdn.net/selous/article/det ails/71637384)

10383

解决pip install国外源太慢的问题 (http://b log.csdn.net/selous/article/details/533097 24)

\$5545

python实现LBP方法提取图像纹理特征 实现分类 (http://blog.csdn.net/selous/arti cle/details/69486823)

Python numpy实现二维数组和一维数组 拼接 (http://blog.csdn.net/selous/article/d etails/72650426)

windov(ksttp) 并bTo使使sderclinests。到线ks/abt服务 器后台頃副記憶13979)sp+javabean+ser **3747**



联系我们

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

/csdnsupport/profile) 100-660-0108

> 招聘 广告服务 [-] 阿里豆 关干

🝘 经营性网站备案信息

(http://www.hd315.gov.cn/beian

©2018 CSDN 京ICP证09002463号

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

/view.asp?bianhao=010202001032100010) 🥑 网络110报警服务 (http://www.cyberpolice.cn/)

也就是prefit如果为true,模型就直接训练好了,不用显示的去调用 fit函数,也就是不能使用 cross_val_score和GirdSearchCV这些并行验证的方式了(这两个我还没有看到,应该是交叉验证之类的 在拟合过程中使用到的函数)

当然也还有estimator变量,要求

 L^{L}

estimator that has a coef_ or feature_importances_ attribute after fitting.



用于筛选特征。其他的可以多看看开发文档 (http://scikit-learn.org/0.17/modules /feature_selection.html)

2. Feature selection as part of a pipeline



这个感觉还没怎么看懂,看起来不难,等懂了再补充。



刚开始做机器学习,这篇文章也只是第一篇,写出来希望大家能一起学习。非监督的还没接触,所以没有 及,之后再补充!有错的欢迎大家指出,一起进步!——2016.11.30

版权声明: 本文为博主原创文章, 未经博主允许不得转载。

Д

机器学习之文本特征选择



십 FlySky1991 2017年02月15日 20:51 🕮 610

1.基本概念1.1 定义从给定的特征集合中选择出相关特征子集的过程, 称为"特征选择"。特征选择是从特征集T={t_1,...,t_ s}中选择一个真子集T'={t_1,...,t_(s')},满足(s≪s')...

(http://blog.csdn.net/FlySky1991/article/details/55222934)

机器学习中的特征选择



rui307 2016年04月25日 17:24 □ 8365

首先声明,本人个人观点,仅供交流。本人欠专业人士,并不了解显示实践中的特征工程。特征选择是一个重要的数据 预处理过程,获得数据之后要先进行特征选择然后再训练模型。主要作用:1、降维2、去除不相关特...

(http://blog.csdn.net/rui307/article/details/51243796)

斯坦福大学机器学习——特征选择(Feature selection) - CSDN博客

2018-1-18

类似前向/后向搜索方法,称为封装模型特征选择算法(wrapper model feature selection algorithm),因为在该方法执行过程中, 不断重复的运行学习算法去评估属性集的好坏。...

(http://blog.csdn.net/linkin1005/article/details/43018827?>)

机器学习的特征选择(feature selection) - CSDN博客

2018-1-15

写在开头:刚开始接触机器学习,选择了周志华教授的《机器学习》作为入门教材,很不...一下开发包的特征选择方法,结合 周志华教授的书,想先总结总结feature selection的加入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

第3页 共7页

2018/1/22 上午10:15

(http://blog.csdn.net/selous/article/details/53414578?>)

机器学习基础(二十五)—— Feature Selection - CSDN博客

2018-1-10

(2)与目标问题不相关特征(irrelevant feature),保险类型与最终的癌症预测。原始...您举报文章:机器学习基础(二十五)——Feature Selection 举报原因: 色情 政治...

##p://blog.csdn.net/lanchunhui/article/details/50986909?>)

台大机器学习听课笔记---RF 10-3 Feature Selection - CSDN博客

2018-1-10

*Feature Selection 从庞大的特征集合中筛选特征(例如1W维选出300维)---来剔除...您举报文章:台大机器学习听课笔记---R F 10-3 Feature Selection 举报原因: ...

(http://blog.csdn.net/findsd1989/article/details/46273167?>)

پ

Python scikit-learn机器学习:feature selection模块 - CSDN博客

2018-1-17

条金: python / 机器学习 / 算法 /1260 编辑 删除 sklearn.feature_selection模块的作用是feature selection,关键在于选择合适的变量达到降维的目的,而feature ...

(http://blog.csdn.net/sinat_29508201/article/details/50771051?>)



特征选择(feature selection) - 天泽28的专栏 - CSDN博客

2018-1-18

Feature Selection for Classification[J]. Intelligent Data Analysis, 1997,1(1-4):131–156. [2]周志华.机器学习[M].北京:清华大学出版社,2016:252-253....

(http://blog.csdn.net/u012328159/article/details/53954522?>)

台大机器学习听课笔记---RF 10-3 Feature Selection - CSDN博客

2017-11-6

对于Decision Tree,在切刀的时候就有FeatureSelection的过程了。 ---...Coursera台大机器学习课程笔记3 – 机器学习的可能性 提纲: 机器学习为什么可能? 引入计算橙...

(http://blog.csdn.net/findsd1989/article/details/46273167?locationNum=2&fps=1?>)

特征选择 (feature selection) - CSDN博客

2018-1-17

当数据预处理完成后,我们需要选择有意义的特征输入机器学习的算法和模型进行训练...在此代码片段中,将 sklearn.svm.L inearSVC 和 sklearn.feature_selection.Select...

(http://blog.csdn.net/u010670689/article/details/73196546?>)

Java机器学习库ML之二Feature Selection(特征选择) - ..._CSDN博客

2018-1-16

机器学习中训练模型的前提必备工作就是特征选择,起到降维和降低计算开销的目的,...import net.sf.javaml.featureselectio n.ranking.RecursiveFeatureEliminationSVM; impor...

(http://blog.csdn.net/fjssharpsword/article/details/73549542?>)

Scikit-learn:Feature selection特征选择和学习 - 皮皮... CSDN博客

2018-1-1

[机器学习中,有哪些特征选择的工程方法? - 知乎] 皮皮blog sklearn.feature_selection 模块中的类能够用于数据集的特征选择/降维,以此来提高预测模型的准确率或改善...

(http://blog.csdn.net/pipisorry/article/details/40200111?>)

斯坦福大学机器学习——特征选择(Feature selection)

特征选择是一种及其重要的数据预处理方法。假设你需要处理一个监督学习问题,样本的特征数非常大,但是可能仅仅有少部分特征会和对结果产生影响。甚至是简单的线性分类,如果样本特征数超过了n,但假设函数的VC维... 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 🗙



🎒 linkin1005 2015年01月22日 14:06 🖺 13714

(http://blog.csdn.net/linkin1005/article/details/43018827)

scikit-learn(工程中用的相对较多的模型介绍): 1.13. Feature selection

旁考: http://scikit-learn.org/stable/modules/feature_selection.html The classes in the sklearn.fea...

mmc2015 2015年08月07日 09:03 🚨 2585

(http://blog.csdn.net/mmc2015/article/details/47333579)

斯坦福大学机器学习——特征选择(Feature selection)

萨征选择是一种及其重要的数据预处理方法。假设你需要处理一个监督学习问题,样本的特征数非常大,但是可能仅仅 有少部分特征会和对结果产生影响。甚至是简单的线性分类,如果样本特征数超过了n,但假设函数的VC维...



🥌 linkin1005 2015年01月22日 14:06 🕮 13714

(http://blog.csdn.net/linkin1005/article/details/43018827)

猪不如学Python!



Python入门,每天1小时,掌握通向未来的编程语言

机器学习中的特征选择



rui307 2016年04月25日 17:24 □ 8365

首先声明,本人个人观点,仅供交流。本人欠专业人士,并不了解显示实践中的特征工程。特征选择是一个重要的数据 预处理过程,获得数据之后要先进行特征选择然后再训练模型。主要作用:1、降维2、去除不相关特...

(http://blog.csdn.net/rui307/article/details/51243796)

机器学习知识点(三十七)特征选择方法总结 🐠 fissharpsword 2017年06月21日 16:21 🛱 2080 在模型训练中,特征选择是非常重要。样本有限的情况,特征过多计算开销过大;通过特征选择去掉冗余和不相关特征 ,可提高模型性能降低计算开销。两个字:降维。当然降维后,模型性能应该是提升的。特征选择的最终目的...

(http://blog.csdn.net/fjssharpsword/article/details/73550337)

Java机器学习库ML之二Feature Selection (特征选择)

机器学习中训练模型的前提必备工作就是特征选择,起到降维和降低计算开销的目的,当然在获取尽可能小的特征子集 的同时,应不显著降低分类精度、不影响类分布、保持特征子集的稳定适应性强等。 ML库提供了特征选择...



🕡 fjssharpsword 2017年06月21日 15:25 🕮 2131

(http://blog.csdn.net/fjssharpsword/article/details/73549542)

总结 特征选择(feature selection)算法笔记 🌄 adore1993 2017年01月02日 16:50 🖫 2057 什么是特征选择特征选择也称特征子集选择,或者属性选择,是指从全部特诊中选取一个特征子集,使构造出来的模型 更好。为什么要做特征选择在机器学习的实际应用中,特征数量往往较多...

(http://blog.csdn.net/adore1993/article/details/53980327)

特征选择方法学习笔记(二)



🆚 zhuchengzhang 2013年12月11日 19:30 🕮 1339

之前看过了mRMR的方法了,从特征与目标的最大相似度和特征与特征间的最小冗余出发来寻找特征子集。这样的方法

登录 注册

(http://blog.csdn.net/zhuchengzhang/article/details/17268663)

特征选择之基于相关性的特征选择(CFS)

参 littlely_II 2017年05月10日 16:28 □ 1357

此为本人学习笔记,转载请劳烦告知!特征选择特征抽取整合原始特征,这样可能产生一些新的特征,而特征选择是去除无关紧要或庸余的特征,仍然还保留其他原始特征。特征提取主要用于图像分析,信号处理和信息检索领域...

##p://blog.csdn.net/littlely_ll/article/details/71545929)

机器学习之(四)特征工程以及特征选择的工程方法

(http://blog.csdn.net/boon 228/article/details/51749646)

6

总结 特征选择(feature selection)算法笔记

总结 特征选择(feature selection)算法笔记原创 2017年01月02 🔞 kebu12345678 2017年11月09日 07:52 🖺 68 日 16:50:18 标签:算法 /数据 /特征选择 996 ...

(http://blog.csdn.net/kebu12345678/article/details/78485093)

机器学习中的特征——特征选择的方法以及注意点

转载自: http://blog.csdn.net/google19890102/article/details/40019271 关于机器学习中的特征我有话要说 在这次校园招聘的过程中, 我学到了...

🔞 u012856866 2015年10月07日 10:42 🕮 2034

(http://blog.csdn.net/u012856866/article/details/48947289)

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 □ 21955 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)

周志华《Machine Learning》学习笔记(13)--特征选择与稀疏学习

上篇主要介绍了经典的降维方法与度量学习,首先从"维数灾难"导致的样本稀疏以及距离难计算两大难题出发,引出了降维的概念,即通过某种数学变换将原始高维空间转变到一个低维的子空间,接着分别介绍了KNN、MD...

◎ u011826404 2017年06月04日 18:51 **◎** 1698

(http://blog.csdn.net/u011826404/article/details/72860607)

机器学习中的特征——特征选择的方法以及注意点

关于机器学习中的特征我有话要说一、特征选择和降维二、特征选择的目标三、...

google19890102 2014年10月12日 17:23
□ 25352

(http://blog.csdn.net/google19890102/article/details/40019271)

登录 注册 🗙

(http://blog.csdn.net/LegenDavid/article/details/51656508)

sklearn特征选择类库使用小结

(kancy110 2017年06月01日 20:42 □ 243

sklearn.feature_selection模块实现了特征选择算法,目前主要包括:单变量特征选择和递归特征消除。该模块的类主要用于特征选择或样本集上降维,提高算法的准确率或者提升在高位数据集上表...

(http://blog.csdn.net/kancy110/article/details/72835050)

机器学习之 特征选择

qq_23968185 2016年09月12日 19:13 🕮 495

机器学习过程中,特征选择是避免维数灾难和减低学习难度的有效途径。特征选择的过程,即为选择对当前学习任务有用的属性(相关特征)和去除无用属性(冗余特征)的过程。那么,如何选择有用的特征子集以及如何评价相...

(http://blog.csdn.net/qq_23968185/article/details/52510992)







