返回主题列表

(/community?fromPost=1054)

【研究】运用HMM模型的择时策略



(/user/d975b9103aa2dae30c383146e6020b0e)

陈小米。

发布于2016-04-09 回复 35 浏览 12277 ♡ 141

(/user/d975b9103aa2dae30c383146e6020b0e)

分享到: ● 微信 💣 微博 () 🔞 雪球

好久没写研究贴了,这篇贴出来也只是这两天的一个阶段成果。最近好几个地方看见讨论基 $oldsymbol{0}$ 于HMM模型做策略,所以自己也来JQ的平台上写一写。

当然,问题嘛还有蛮多的,也是之后要解决的。不多说了全在研究里了。

欢迎大家批评指正。

ps: 更正-1

因为每次运行每个状态的编号是随机的,最后一段文字里提到的状态编号要进行更正。

回测结果截图贴在这。回撤和收益都看起来很漂亮。虽然我也没太懂为什么跟可卖空的结果差不多==,可能是单纯算收益率的伪回测不太 准吧。第二是状态5(浅蓝色)【原文为:状态1(绿色)】有涨有跌,卖空它也有亏钱的时候。

ps:更正-2

原文中第三部分实证第二张图的解释需要更换(同样因为状态的编号改变)

更正为:

上图可以看出:

状态0--蓝色--震荡下跌

状态1--绿色--小幅的上涨

状态2——红色——牛市上涨

状态3--紫色--牛市下跌

状态4--黄色--震荡下跌

状态5--浅蓝色--牛市下跌





欢迎关注: 神秘的宽客们

量化研究,量化学习资料 神策alpha信号每日播报 更多精彩,等你发现

E%8B.ipynb¬ebookReport=/research/users/20883955875/share2/2ee8a1b5c6d3116847b582b3ab585591_new.html&postId=1054)



(/user**/duses/dom/1753base**12

陈小光练小(光线ser//dustets

【研究】研究】HMM科 探析资金流资金流验价 资金流模型的模型的研 多因子模型所模型组织 【研究】研究】证量机造量处设

查看更多的/d

userld#sd9765-b79775692120

HMM模型

之前做过相关方面的研究,针对的是国外的期货市场。也有朋友建议说拿A股来试试看。所以有了下面的这篇东西。

一、从大奖章讲起

Renaissance & Medallion(文艺复兴科技和大奖章)量化圈都非常熟悉了。Simons一群物理学家和数学家碰撞在一起,1989年到 2008年的yearly return达到35.6%。文艺复兴大概一百多个员工,AUM: 50亿美金,在全球金融危机的08年,大部分对冲基金都亏损,而大奖章的return高达80%。

神秘的文艺复兴科技和神秘的大奖章基金,到底一群数学家和物理学家聚在一起搞出了什么赚钱利器?外界猜测众说纷纭。而隐马尔科夫模型也由于一些原因被推举出来。

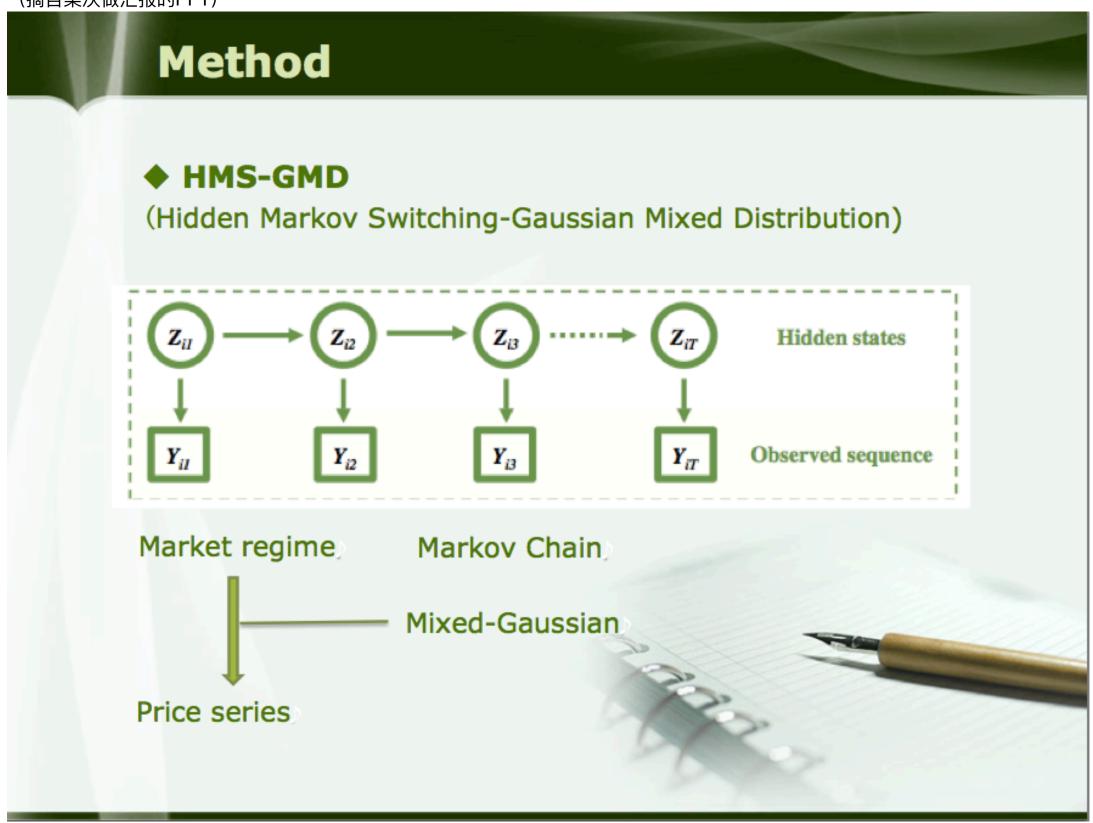
成立初期的创始人中,有一位科学家发明了广泛应用在语音识别等领域的鲍姆-威尔士算法,用来确定不可确知的变量可能出现的概率。今天要介绍的HMM模型,也是在语音识别中运用非常成功的模型,最早是由鲍姆等人提出的。

之前拜读人大的一位教授14年写的一本书,解密复兴科技:基于隐蔽马尔科夫模型的时序分析方法(亚马逊链接) (https://www.amazon.cn/gp/product/B00LBITP1K/ref=oh_aui_detailpage_o01_s00?ie=UTF8&psc=1), 书中介绍了为什么觉得 HMM是Renaissance使用的模型,并且含有详细的公式推导和运算,以及附上了一些些实证结果(个人认为还有很多有待补充和粗 糙的地方)。感兴趣推荐研读。

二、认识模型

HMM模型,又叫隐马尔科夫模型。要正确的理解和搞懂模型,教材里有很多经典的例子。我从自己理解后的角度尽量浅显的给大家做一个解释,方便大家快速理清概念投入应用。

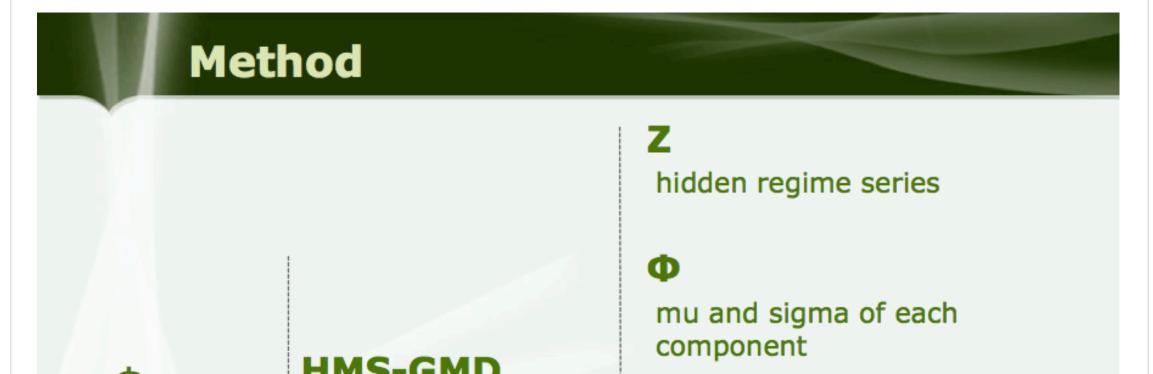
(摘自某次做汇报的PPT)



我们能观测到的序列\$\$ Y1,...,Yn \$\$ 称为可观测序列,如股价,成交量,资金净额等等。

而每一个可观测值的产生对应着市场状态序列\$\$Z1,...,Zn\$\$,每个状态通过不同的分布函数来产生观测值。

通过HMM模型,可以用简单的输入,来得出对目前市场状态的判断,从而帮助我们进行择时选择。因为市场状态不是显性可观测的,属于隐藏状态,我们通过对可观测变量的处理来进行推测。



P: Price Series

Q: Num of Regime M: Num of Component

Mixmat

weight of each component in certain regime

Transmat

transition probability of each regime

这里对HMM模型进行了扩展得到HMS-GMD模型,因为收益率序列尖峰厚尾的特性导致的非正态分布,引入了混合高斯分布作为状态到观测值之间产生关系的分布函数。

将HMM模型看作一个黑箱子,这个黑箱子可以利用极其方便、简洁的数据,处理后得出:

- 1. 每个时刻对应的状态序列;
- 2. 混合分布的均值和方差矩阵;
- 3. 混合分布的权重矩阵;
- 4. 状态间转移概率矩阵。

而黑箱子需要事先给定两个参数:状态数目、混合高斯分布的成分数目。

当然,输入这里是拿单一的价格序列举例。输入也可以是并行的数据矩阵,比如从价格、成交量、资金净额等多个角度来看。

总结一下,使用模型需要在初期设定:

- 1. 隐藏状态数目
- 2. 输入的观测变量
- 3. 混合高斯分布成分数目

三、A股市场实证

下面拿A股市场来做检验。

模型的设定如下:

- 1. 隐藏状态数目: 6
- 2. 输入变量: 当日对数收益率, 五日对数收益率, 当日对数高低价差(其他备选因素成交量、成交额等大家可以自行尝试)
- 3. 混合高斯分布成分数目: 1 (为了简便,假定对数收益率服从单一高斯分布)

HMM模型的算法使用hmmlearn模块。简单介绍一下函数的各个参数意思。

```
from hmmlearn.hmm import GMMHMM, GaussianHMM
In [2]:
        import datetime
        import numpy as np
        import pandas as pd
        import seaborn as sns
        from matplotlib import cm
        from matplotlib import pyplot
        startdate = '2012-06-01'
        enddate = '2016-04-07'
        df = get price(['000300.XSHG'], start date=startdate, end date=enddate, frequency='da
        ily', fields=['close','volume','high','low'])
        close = df['close']['000300.XSHG']
        high = df['high']['000300.XSHG'][5:]
        low = df['low']['000300.XSHG'][5:]
        volume = df['volume']['000300.XSHG'][5:]
        money = df['volume']['000300.XSHG'][5:]
        datelist = pd.to datetime(close.index[5:])
        logreturn = (np.log(np.array(close[1:]))-np.log(np.array(close[:-1])))[4:]
        logreturn5 = np.log(np.array(close[5:]))-np.log(np.array(close[:-5]))
        diffreturn = (np.log(np.array(high))-np.log(np.array(low)))
        closeidx = close[5:]
        X = np.column_stack([logreturn,diffreturn,logreturn5])
        len(X)
```

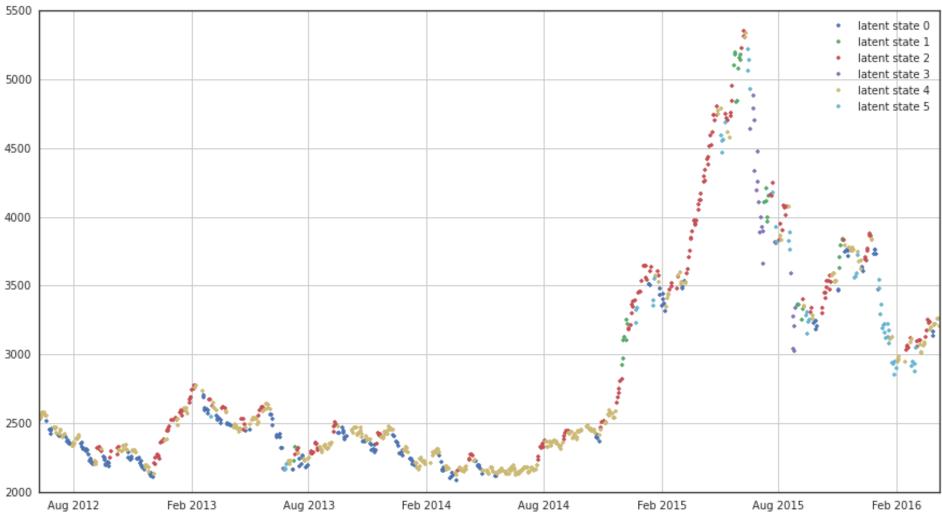
Out[2]: 931

```
In [3]: hmm = GaussianHMM(n_components = 6, covariance_type='diag',n_iter = 5000).fit(X)
    latent_states_sequence = hmm.predict(X)
    len(latent_states_sequence)
```

Out[3]: 931

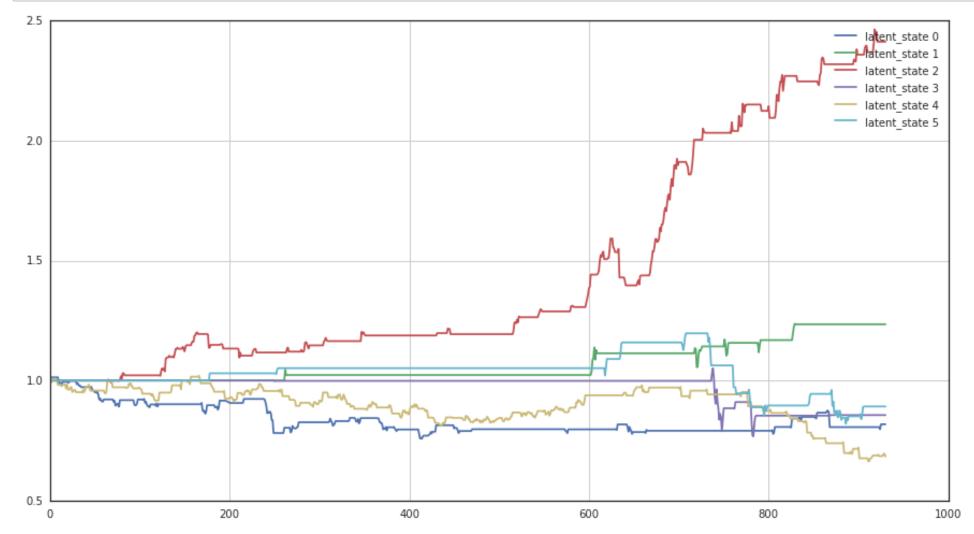
```
In [4]: sns.set_style('white')
  plt.figure(figsize = (15, 8))
  for i in range(hmm.n_components):
     state = (latent_states_sequence == i)
```

```
plt.plot(datelist[state],closeidx[state],'.',label = 'latent state %d'%i,lw = 1)
plt.legend()
plt.grid(1)
```



以上。我们看到了六个状态的HMM模型输出的市场状态序列。需要注意的是:HMM模型只是能分离出不同的状态,具体对每个状态赋予现实的市场意义,是需要人为来辨别和观察的。

下面我们来用简单的timming策略来识别6种latent_state所带来的效果。



上图可以看出:

- 1. 状态0--蓝色--牛市上涨
- 2. 状态1--绿色--牛市下跌
- 3. 状态2——红色——牛市下跌
- 4. 状态3——紫色——小幅的上涨
- 5. 状态4——黄色——震荡下跌
- 6. 状态5——浅蓝色——震荡下跌

以上的意义归结是存在一定主观性的。因为HMM模型对输入的多维度观测变量进行处理后,只负责分出几个类别,而并不会定义出 每种类别的实际含义。所以我们从图形中做出上述的判断。

四、择时策略

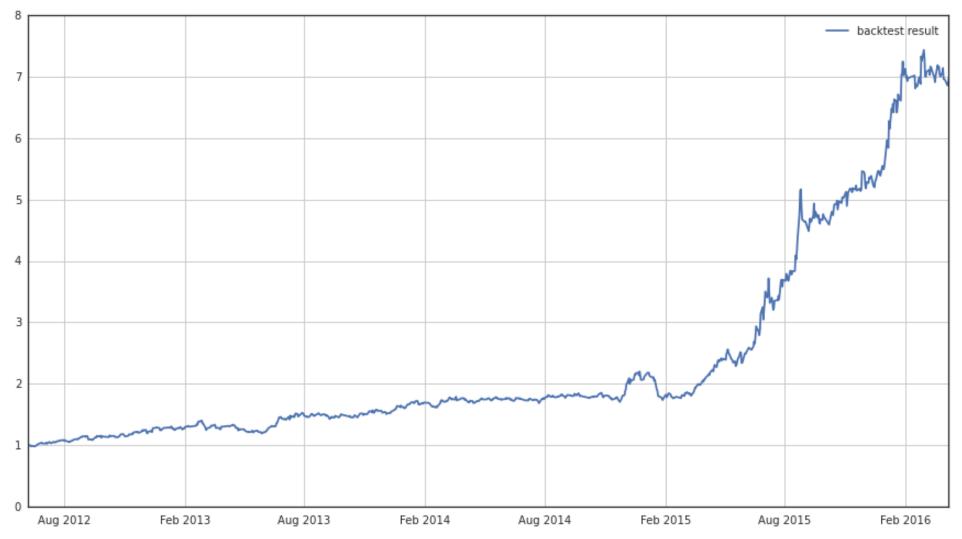
我们根据模拟出来的隐藏状态,来进行择时。

1. 理论版:股指期货可卖空。

策略是这样设计的:

- 1. 当天处在状态0, 3时, 买入指数基金;
- 2. 当天处在状态1, 2, 4, 5时, 卖空股指期货;

我们来看一下收益效果:



2. A股版

鉴于卖空指数对散户来说没什么可操作性,我们单看能做多的A股市场。选择嘉实沪深300基金来复制沪深300指数。

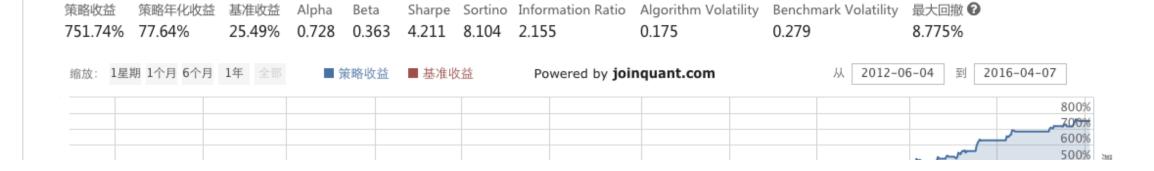
策略是这样设计的:

- 1. 当天处在状态0,3时,买入指数基金;
- 2. 当天处在状态1, 2, 4, 5时, 空仓;

鉴于研究模块自己搭伪回测算出来的收益曲线很难考虑到滑点、交易规则限制等各方面的要求,所以如下的策略回测只贴上最后的图。代码具体见策略。

我采取的方式是把研究模块得出的状态序列导出,导入到回测模块使用。

回测结果截图贴在这。回撤和收益都看起来很漂亮。虽然我也没太懂为什么跟可卖空的结果差不多= =,可能是单纯算收益率的伪回测不太准吧。第二是状态1(绿色)有涨有跌,卖空它也有亏钱的时候。



感谢作者



立即体验



道法自然 (/user/85132aa9b3c13a6104e9e206dc637e22)

谢谢,收藏研究。

2016-04-09



(/user/ed90e94ee201fbca8e0d55abe17e5443)

幻音 (/user/ed90e94ee201fbca8e0d55abe17e5443)

叼 哈哈 学习大神思路 我发现大神在数据优化上想法很多哈

2016-04-10



(/user/ed90e94ee201fbca8e0d55abe17e5443)

幻音 (/user/ed90e94ee201fbca8e0d55abe17e5443)

这个不行的 我觉的源码里不对的 我也研究过这个是优矿上的 实际做回测的时候只能根据前面的几日指标来进行买和卖 达不到那个 效果的 大神可以把代码发出来看下

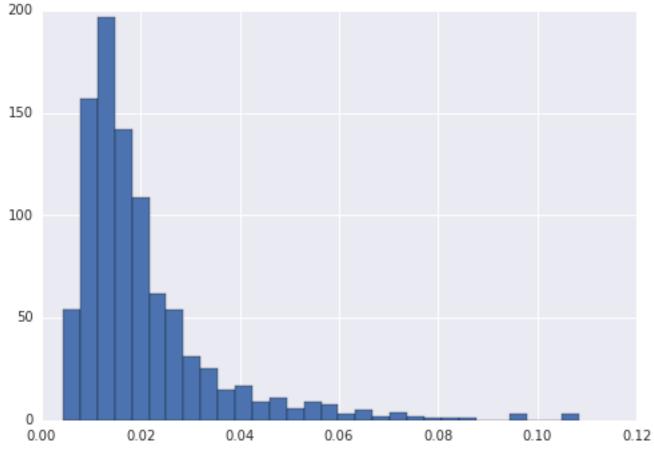
2016-04-10



(/user/f19e826797a5844926b44c3bd3c688ed)

yang (/user/f19e826797a5844926b44c3bd3c688ed)

膜拜下大神!另外有些疑问, diffreturn = (np.log(np.array(high))-np.log(np.array(low)))的分布看起来不像高斯分布。



是不是改成(np.log(np.array(high)-np.array(low)))更合适?

2016-04-10



(/user/ec97b14e136b14aac476051038ecdf07)

sharkblue (/user/ec97b14e136b14aac476051038ecdf07)

看来平台终于更新HMM库了。感谢小米大师的文章。

2016-04-11



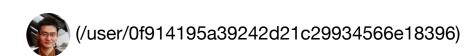
(/user/0f914195a39242d21c29934566e18396)

韭菜Hulk (/user/0f914195a39242d21c29934566e18396)

赞一个

请教两个细节一点的问题:

- 1 训练HMM是每天每天都做训练再预测次日状态吗 还是只拿某一段时间训练好之后不变了呀? 我之前有试过每天每天训练再预测次日状态,但是viterbi训练结果非常不稳定,因为状态转移的矩阵不是stationary的。
- 2 还有那个观测分布是Mixture of Gaussian的话,prior distribution要怎么设呀,在训练的时候怎么更新呢?



韭菜Hulk (/user/0f914195a39242d21c29934566e18396)

@yang 恩 这个diffreturn可以先做个fisher transform把他整成Gaussian

2016-04-11



(/user/0f914195a39242d21c29934566e18396)

韭菜Hulk (/user/0f914195a39242d21c29934566e18396)

楼主可以分享一小部分回测代码不,我看研究上训练用的数据是2012-2016年的

startdate = '2012-06-01'

enddate = '2016-04-07'

df = get_price(['000300.XSHG'], start_date=startdate, end_date=enddate, frequency='daily', fields=['close','volume','high','low'])

然后后面回测区间也是2012-2016年,要是每天预测用的是2012-2016年的训练出来的状态就不太准确了,比如今天是2013-01-01 应该只能用2013-01-01之前的数据训练训练 而且你如果把每天训练出来的那个6*6的转移矩阵打印出来会发现差别好大。。。

还有在sklearn里使用观测分布是mixture of Gaussian的话得调用GMMHMM,不是GaussianHMM,否则是认为三个观测变量是相互独立的高斯分布。

2016-04-11



(/user/d975b9103aa2dae30c383146e6020b0e)

陈小米。(/user/d975b9103aa2dae30c383146e6020b0e)

@韭菜Hulk

韭菜果然好懂。你提的建议是对的,这个问题确实存在。如果用GaussianHMM的话,一段时间做的状态图是固定的,而用mixture of Gaussian的话,还涉及几个高斯分布的估计,训练出来每次结果差异都比较大,所以没有用GMMHMM。

回测还没有加hmmlearn模块,我是导出状态序列做的回测,其实是有未来函数的。代码就不给了免得误导大家。之后还是要试试把 未来拨出去做回测。

2016-04-11



(/user/f19e826797a5844926b44c3bd3c688ed)

yang (/user/f19e826797a5844926b44c3bd3c688ed)

@韭菜Hulk 我感觉这里用价差的对数更合适一些而不是对数价差,因为这个算法是通过价差的变化模式作为输入之一训练模型来预测未来涨跌.

2016-04-11



(/user/0f914195a39242d21c29934566e18396)

韭菜Hulk (/user/0f914195a39242d21c29934566e18396)

@陈小米。 我刚下单买了你推荐的书,刚翻了两页电子版,感觉很不错,不少建模细节都有说。

期待分享后续大作。

2016-04-11



(/user/2be576585828842a9ba9baf10271e319)

柿子 (/user/2be576585828842a9ba9baf10271e319)

@韭菜Hulk 哪本书呀? 也想买来拜读一下

2016-04-11



宏观经济占卜师 (/user/e10cad1361a9e943edb301c3f1482cbd)

@韭菜Hulk 啥书? 同问

2016-04-11



(/user/0f914195a39242d21c29934566e18396)

韭菜Hulk (/user/0f914195a39242d21c29934566e18396)

@柿子 是《解密复兴科技》,陈小米贴了亚马逊的链接。。。从第八章开始就不仅仅是课堂上讲的HMM了,有很多干货。

2016-04-11



(/user/ed90e94ee201fbca8e0d55abe17e5443)

幻音 (/user/ed90e94ee201fbca8e0d55abe17e5443)

@陈小米。 大哥 期待你去掉导出序列做的 你导出序列做的回测就像 让你把股市先看一般 分辨下哪里是什么状态分为5类 然后什么 类买什么 如果你按之前出现的状态来预测结果会大不一样 这个HMM模型识别状态性能很好

2016-04-11



(/user/d975b9103aa2dae30c383146e6020b0e)

陈小米。(/user/d975b9103aa2dae30c383146e6020b0e)

@幻音 我这两天在尝试。有结果会再发出来的:)

2016-04-11



(/user/ed90e94ee201fbca8e0d55abe17e5443)

幻音 (/user/ed90e94ee201fbca8e0d55abe17e5443)

@陈小米。 期待 我做我一个策略 就是提取前面一段时间的序列 然后取最后一个值 用他来判断买入还是卖出 可是效果不理想 还有一个想法是取前面出现的数据的序列结果 然后取最后一段的几天 然后设定一个参数 当序列几出现几次然后作为买入或者卖出的策略 这个方法还没尝试过

2016-04-11



(/user/d975b9103aa2dae30c383146e6020b0e)

陈小米。(/user/d975b9103aa2dae30c383146e6020b0e)

@幻音

要注意,没办法直接取值用,因为每次出来的状态数值究竟代表哪种状态含义是不确定的。

2016-04-11



(/user/ed90e94ee201fbca8e0d55abe17e5443)

幻音 (/user/ed90e94ee201fbca8e0d55abe17e5443)

并不是的哦 出来的值是确定的好像 看前面模型设定的参数 如果实在不行 那就只能取前面固定时间段出现的值让他当标准 上升期的值当状态几就当作几 好像我做策略的时候 直接取值当买入卖出 结果是稳定不变的

2016-04-11



(/user/d975b9103aa2dae30c383146e6020b0e)

陈小米。(/user/d975b9103aa2dae30c383146e6020b0e)

@幻音

拿我研究里举例,状态数目确定后,出来的状态曲线图是一样的,但是颜色会不同,也就是每种状态的那个数值是随机给的。你说的后者的思路比较可行。先跑一段时间把状态值和市场意义联系上,然后固定参数来估计后面的状态值来进行择时操作。

2016-04-11

1 (/post/1054) 2 (/post/1054?page=2) 下一页 (/post/1054?page=2) 尾页 (/post/1054?page=2)

您尚未登录,请 登录 (/user/login/index?redirect=/post/1054) 或者 注册 (/user/register/index) 聚宽发表回复。

关于

数据

帮助

@2015 @joinquant.com | 京ICP备13039866号-3

(/)