

阿里云天池牛年读书会

Python金融大数据分析

分享嘉宾：黄佳
《零基础学机器学习》作者

天池读书会

TIANCHI 天池

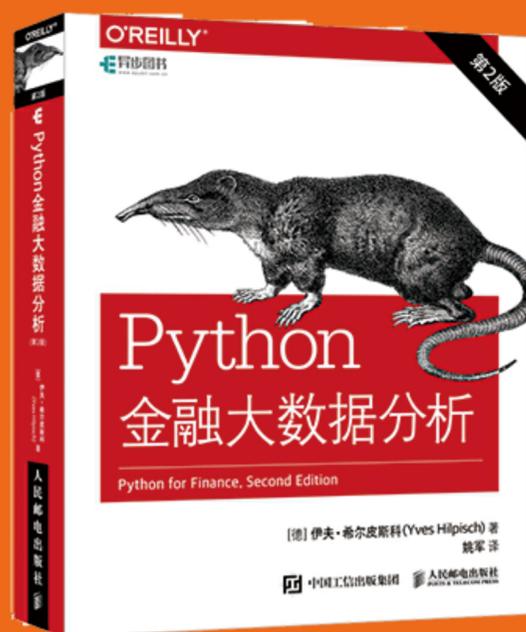
异步社区
人民邮电出版社

《Python金融大数据分析(第二版)》

向开发人员和量化分析人员介绍了如何使用Python程序库与工具完成金融数据科学、算法交易和计算金融任务的方法。

直播嘉宾：黄佳

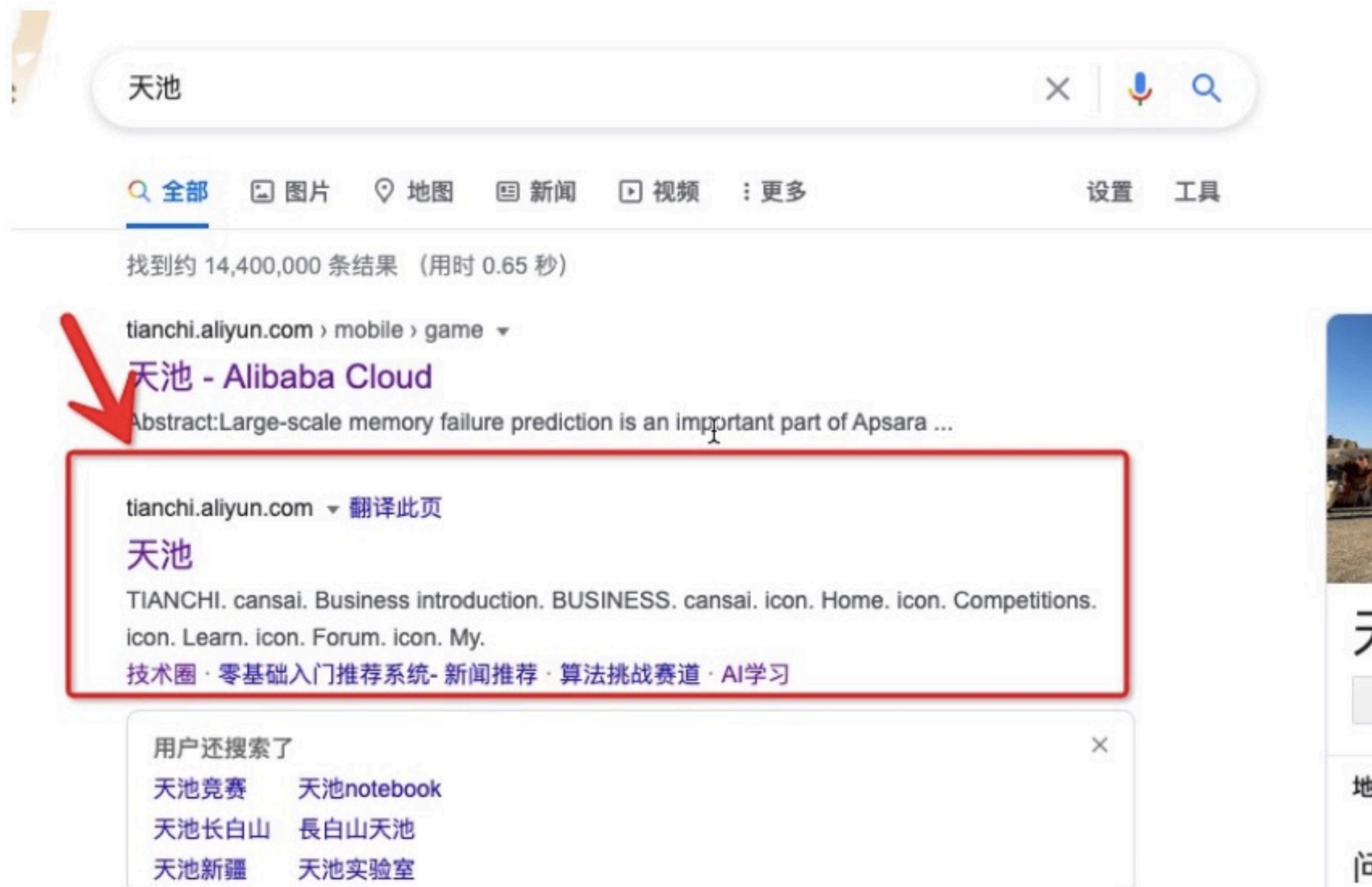
直播时间：4月30日20:00



扫码领取读书会配套学习资源



1) 首先需要进入天池官网，大家打开浏览器，搜索 天池，找到 tianchi.aliyun.com即可访问进入天池官



网；

2) 在天池官网，将鼠标移到 天池学习，即可出现下拉列表，点击 天池读书会，即可进入天池读书会的页面。



3) 在天池读书会页面，你可以对对应的读书会图书进行提问，优秀的提问还有机会获得赠书，还可以点击配套的训练营或者课程资源进入学习，还有点击实践代码获取读书会的项目实践的代码，跟着我一起进行项目实践和代码学习，同时还有很多其他的读书会，大家也可以观看举办过的读书会的回放，或者预约还没开始的读书会。



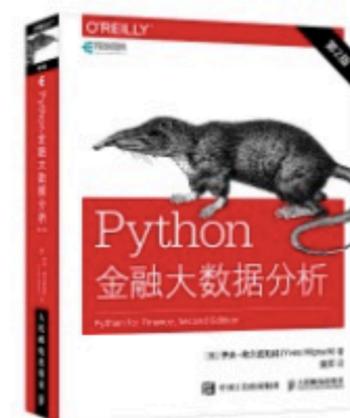
黄佳 机器学习专家

直播主题 《Python金融大数据分析(第二版)》

直播时间 2021年4月30日 20:00

学习资料 金融风控训练营

实践项目 比特币走势预测 (Prophet工具)



[🗨️ 提问](#) | [✍️ 学习训练营](#) | [🛒 购买地址](#) | [📄 PPT下载](#) | [👉 实践代码](#) | [🕒 预约直播](#)

1. 分享嘉宾简介

2. 图书简介

3. 项目实践-比特币走势预测（Prophet工具）

4. Q&A 答疑

分享嘉宾简介

TIANCHI 天池

黄佳

20年IT从业经验，

SAP 资深顾问，

机器学习，云计算专家，

参与实施公用事业、银行、能源、科技等多领域大型项目

代表作：《零基础学机器学习》 《SAP程序设计》



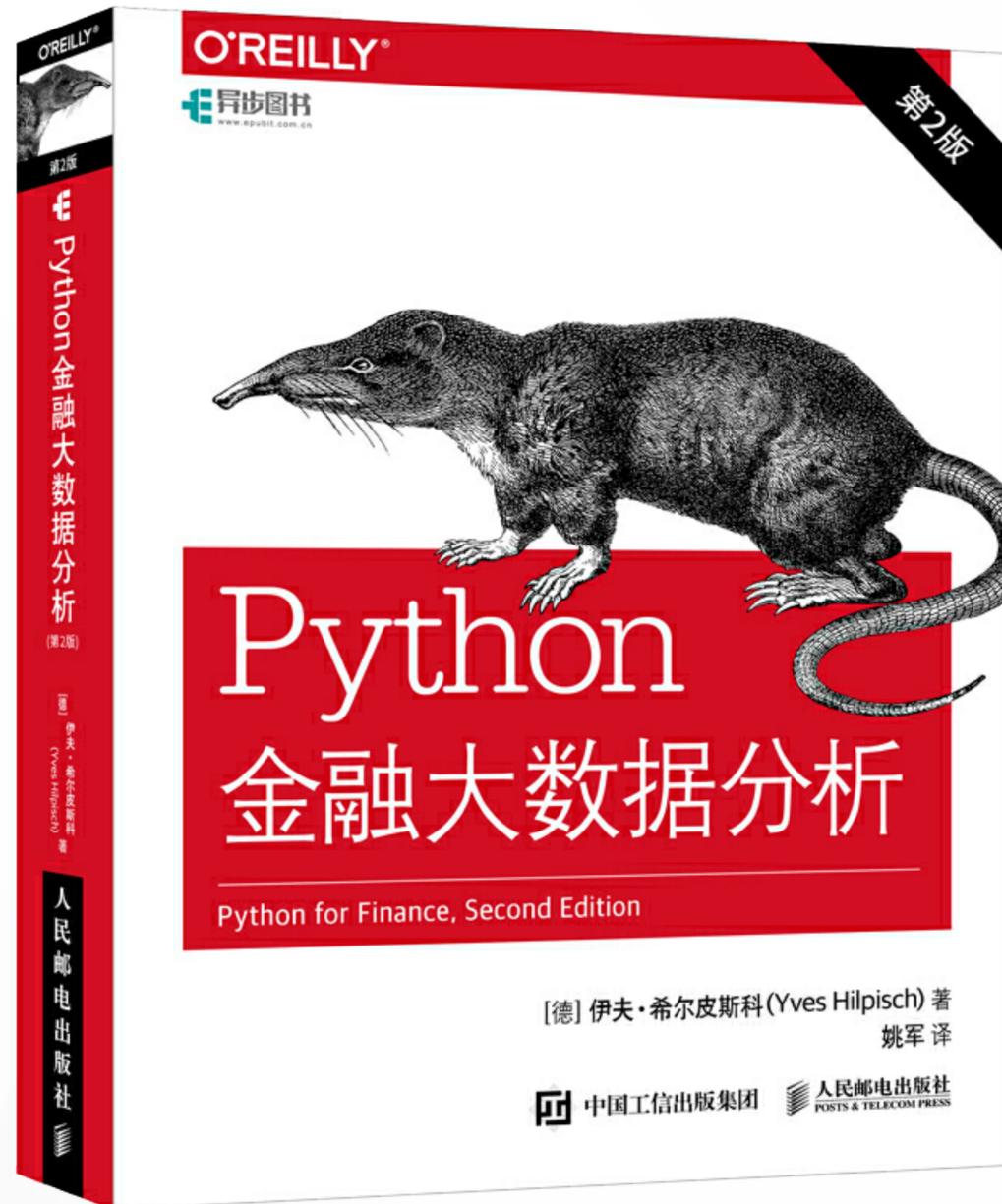
公众号：咖哥数据科学讲习所
(目前更新非常不频繁)

直播相关资料获取及回放查看地址：<https://tianchi.aliyun.com/specials/promotion/activity/bookclub>



图书简介

TIANCHI 天池

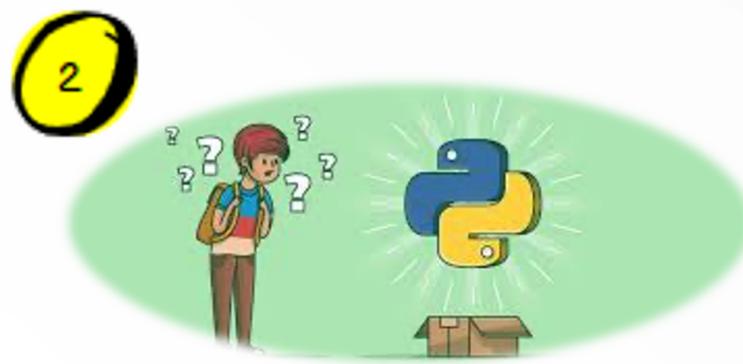


在这本书中，作者向金融数据分析系统的开发人员和量化分析人员介绍了使用Python程序库与工具，完成金融数据科学、算法交易和量化系统开发的方法。





Python与金融



Python编程基础



金融数据科学



算法交易



衍生品分析



Python在两个世界：即**高层次的抽象**和**严格的实现**中提供最佳的平衡。

```
In [16]: import numpy as np
import pandas as pd
from pylab import plt, mpl

In [17]: plt.style.use('seaborn')
mpl.rcParams['font.family'] = 'serif'
%matplotlib inline

In [18]: data = pd.read_csv('../source/tr_eikon_eod_data.csv',
index_col=0, parse_dates=True)
data = pd.DataFrame(data['.SPX'])
data.dropna(inplace=True)
data.info()
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 2138 entries, 2010-01-04 to 2018-06-29
Data columns (total 1 columns):
.SPX    2138 non-null float64
dtypes: float64(1)
memory usage: 33.4 KB

In [19]: data['rets'] = np.log(data / data.shift(1))
data['vola'] = data['rets'].rolling(252).std() * np.sqrt(252)
In [20]: data[['.SPX', 'vola']].plot(subplots=True, figsize=(10, 6))
```





Python数据结构

- 元组 (tuple) 任意对象的集合。
- 列表 (list) 任意对象的集合 (较为灵活)。
- 字典 (dict) 键-值存储对象。
- 集合 (set) 其他独特对象的无序集合对象。

对列表切片：

```
In [110]: l[2:5] ①  
Out[110]: [2.5, 1.0, 1.5]
```

Numpy和ndarray数组

```
In [1]: v = [0.5, 0.75, 1.0, 1.5, 2.0]
```

```
In [2]: m = [v, v, v] ①  
        m ②  
Out[2]: [[0.5, 0.75, 1.0, 1.5, 2.0],  
         [0.5, 0.75, 1.0, 1.5, 2.0],  
         [0.5, 0.75, 1.0, 1.5, 2.0]]
```

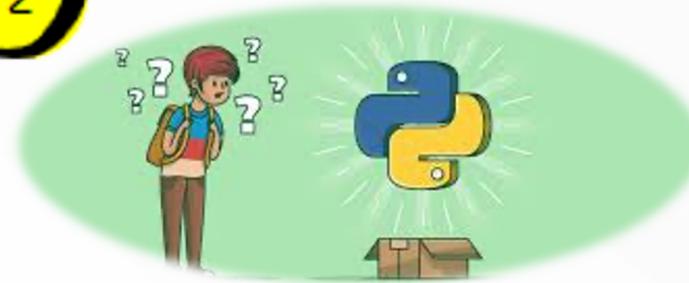
向量化

```
In [129]: np.random.seed(100)  
         r = np.arange(12).reshape((4, 3)) ①  
         s = np.arange(12).reshape((4, 3)) * 0.5 ②  
  
In [130]: r ③  
Out[130]: array([[ 0,  1,  2],  
                [ 3,  4,  5],  
                [ 6,  7,  8],  
                [ 9, 10, 11]])  
  
In [131]: s ④  
Out[131]: array([[0. , 0.5, 1. ],  
                [1.5, 2. , 2.5],  
                [3. , 3.5, 4. ],  
                [4.5, 5. , 5.5]])  
  
In [132]: r + s ⑤  
Out[132]: array([[ 0. ,  1.5,  3. ],  
                [ 4.5,  6. ,  7.5],  
                [ 9. , 10.5, 12. ],  
                [13.5, 15. , 16.5]])
```

Python编程基础



Python编程基础



TIANCHI天池

Pandas

从ndarray类到DataFrame对象

```
In [25]: import numpy as np
In [26]: np.random.seed(100)
In [27]: a = np.random.standard_normal((9, 4))
In [28]: a
Out[28]: array([[ -1.74976547,  0.3426804 ,  1.1530358 , -0.25243604],
 [ 0.98132079,  0.51421884,  0.22117967, -1.07004333],
 [-0.18949583,  0.25500144, -0.45802699,  0.43516349],
 [-0.58359505,  0.81684707,  0.67272081, -0.10441114],
 [-0.53128038,  1.02973269, -0.43813562, -1.11831825],
 [ 1.61898166,  1.54160517, -0.25187914, -0.84243574],
 [ 0.18451869,  0.9370822 ,  0.73100034,  1.36155613],
 [-0.32623806,  0.05567601,  0.22239961, -1.443217 ],
 [-0.75635231,  0.81645401,  0.75044476, -0.45594693]])
```

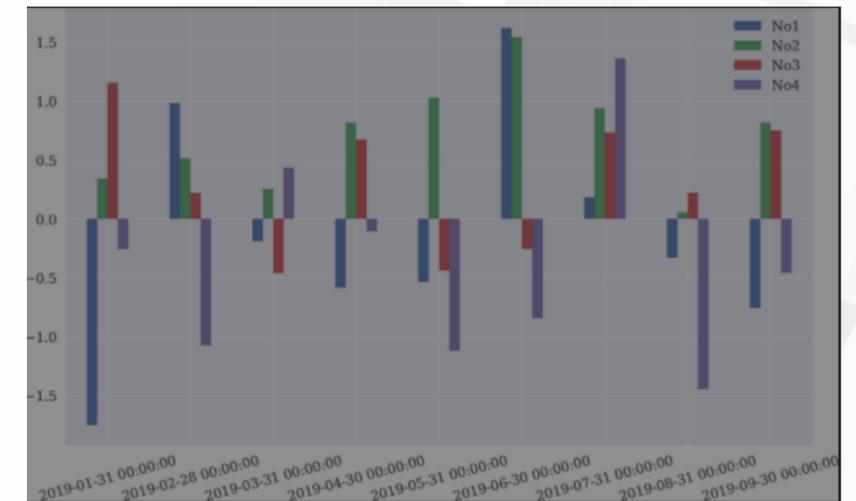


```
In [29]: df = pd.DataFrame(a)
In [30]: df
Out[30]:
```

	0	1	2	3
0	-1.749765	0.342680	1.153036	-0.252436
1	0.981321	0.514219	0.221180	-1.070043
2	-0.189496	0.255001	-0.458027	0.435163
3	-0.583595	0.816847	0.672721	-0.104411
4	-0.531280	1.029733	-0.438136	-1.118318
5	1.618982	1.541605	-0.251879	-0.842436
6	0.184519	0.937082	0.731000	1.361556
7	-0.326238	0.055676	0.222400	-1.443217
8	-0.756352	0.816454	0.750445	-0.455947

DataFrame可视化

```
df.plot.bar(figsize=(10, 6), rot=15);
```



数据可视化

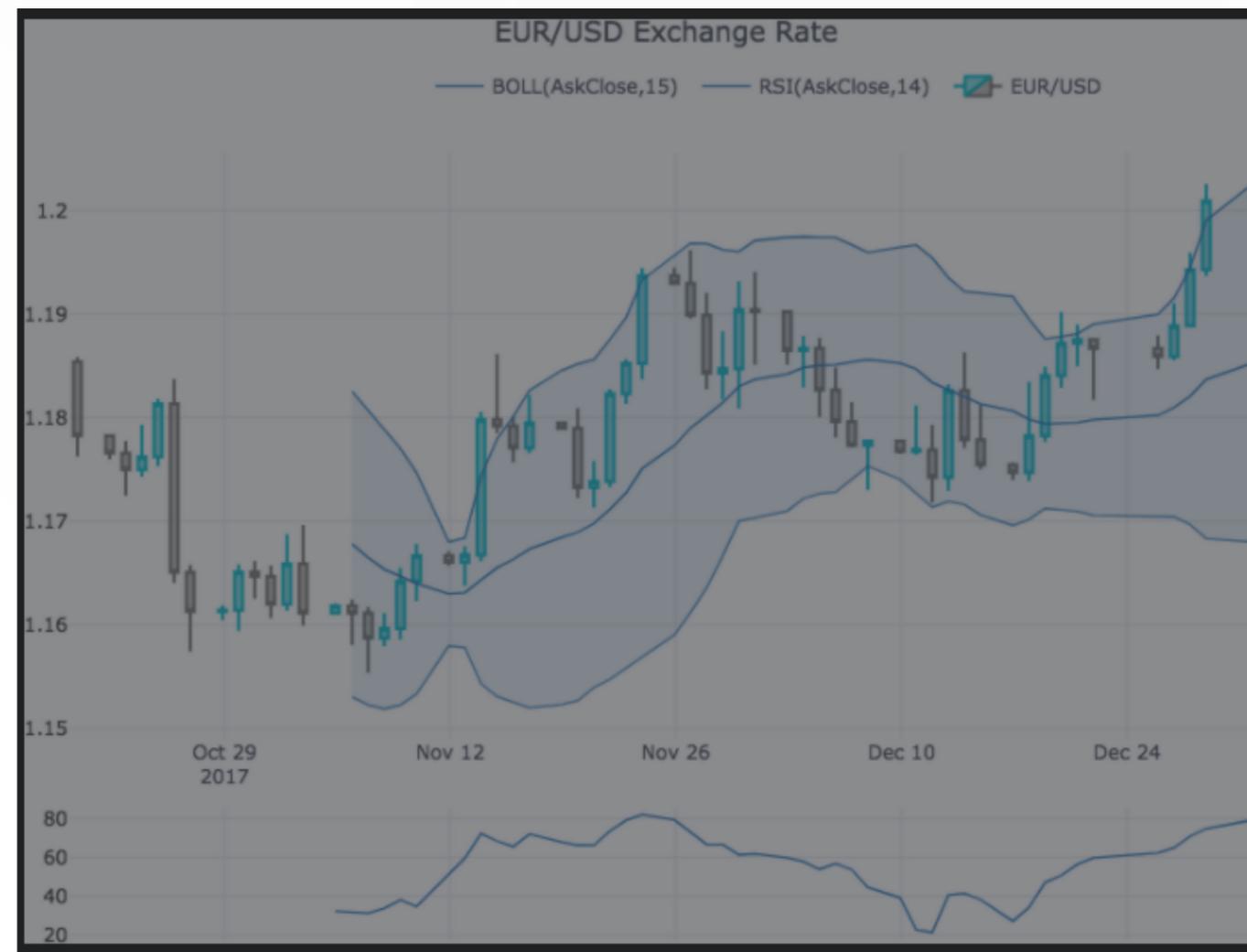
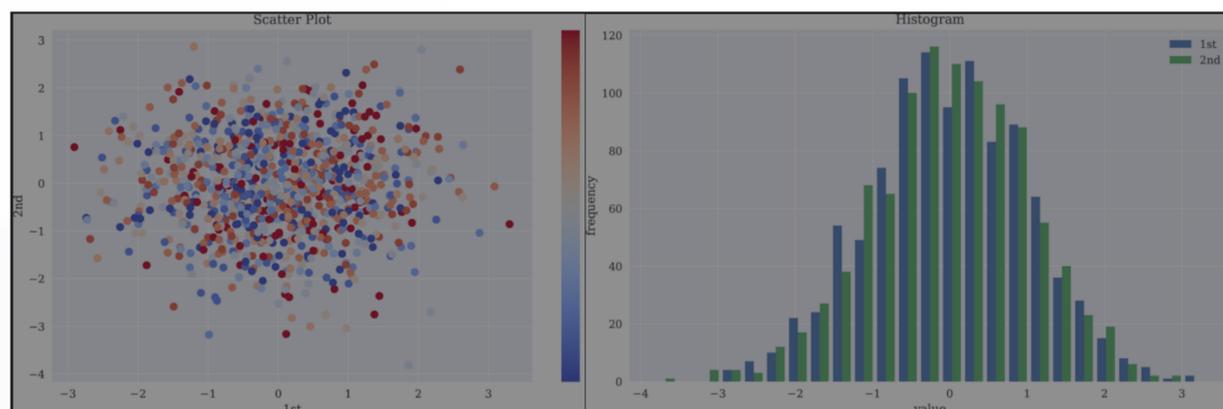
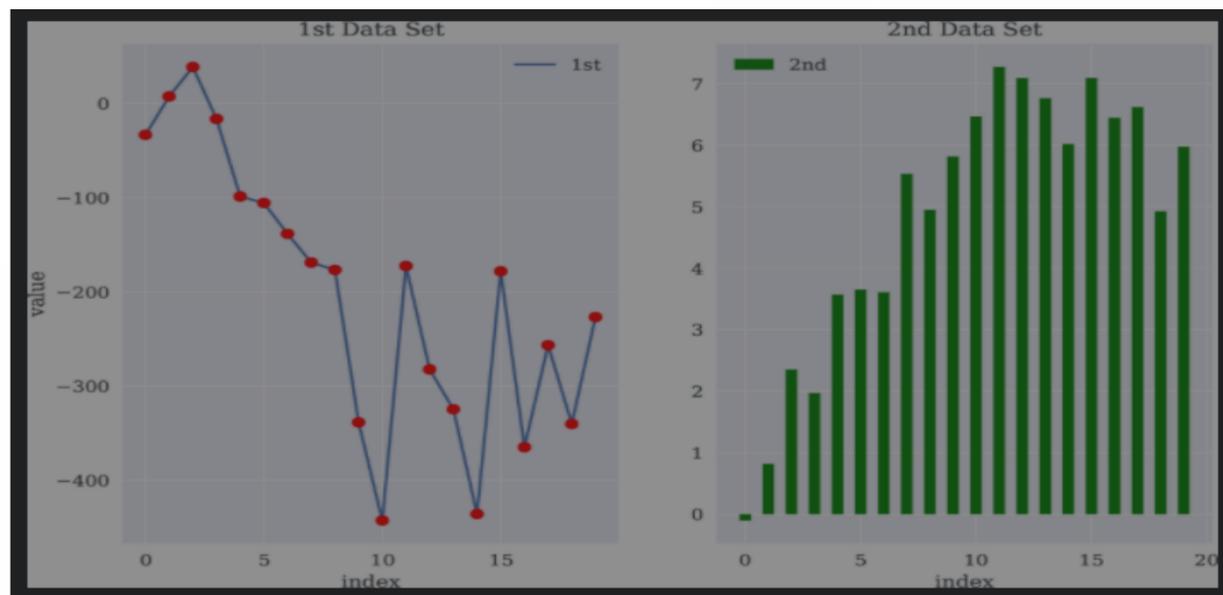
3

金融数据科学

TIANCHI 天池

Matplotlib

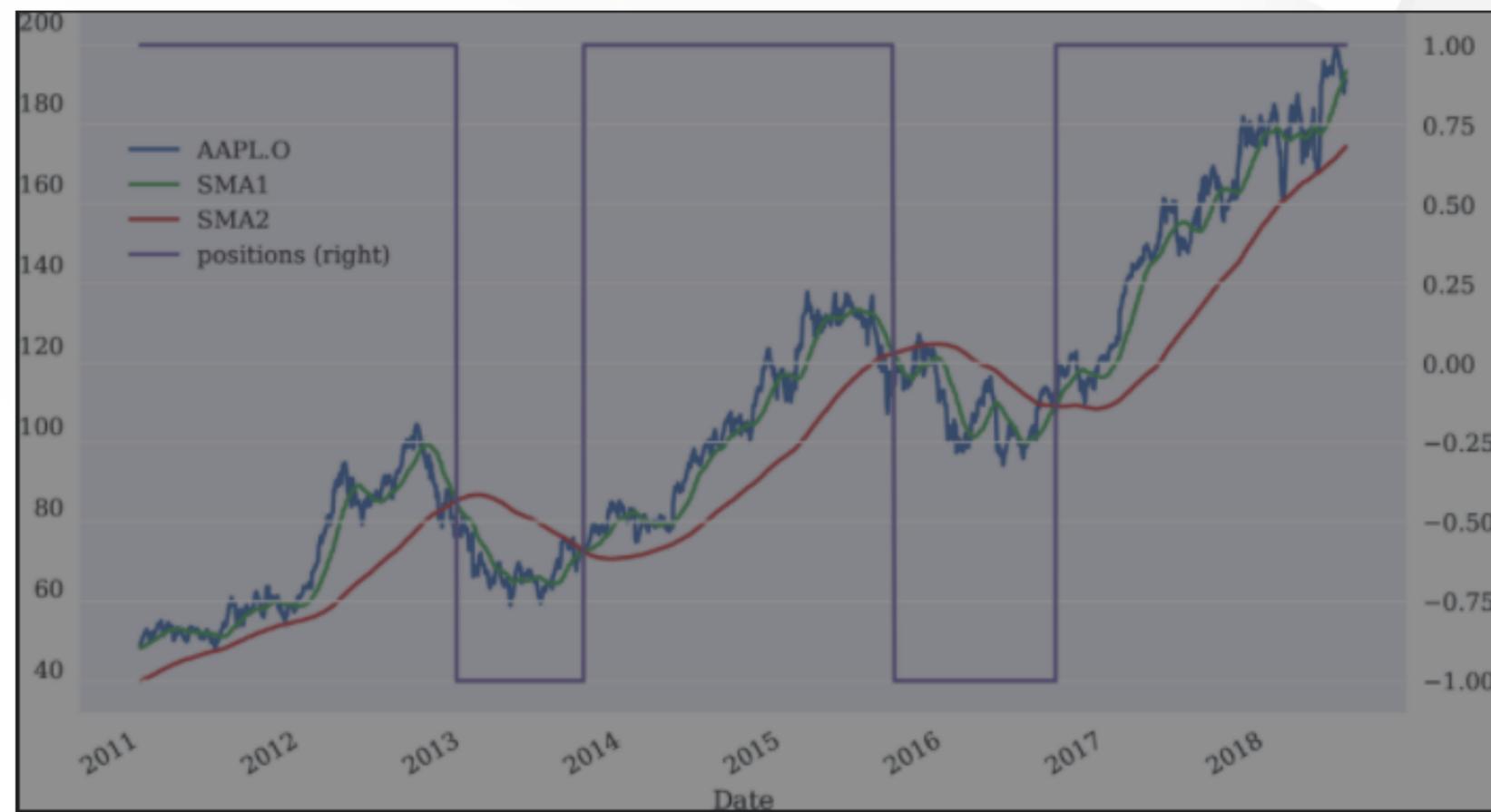
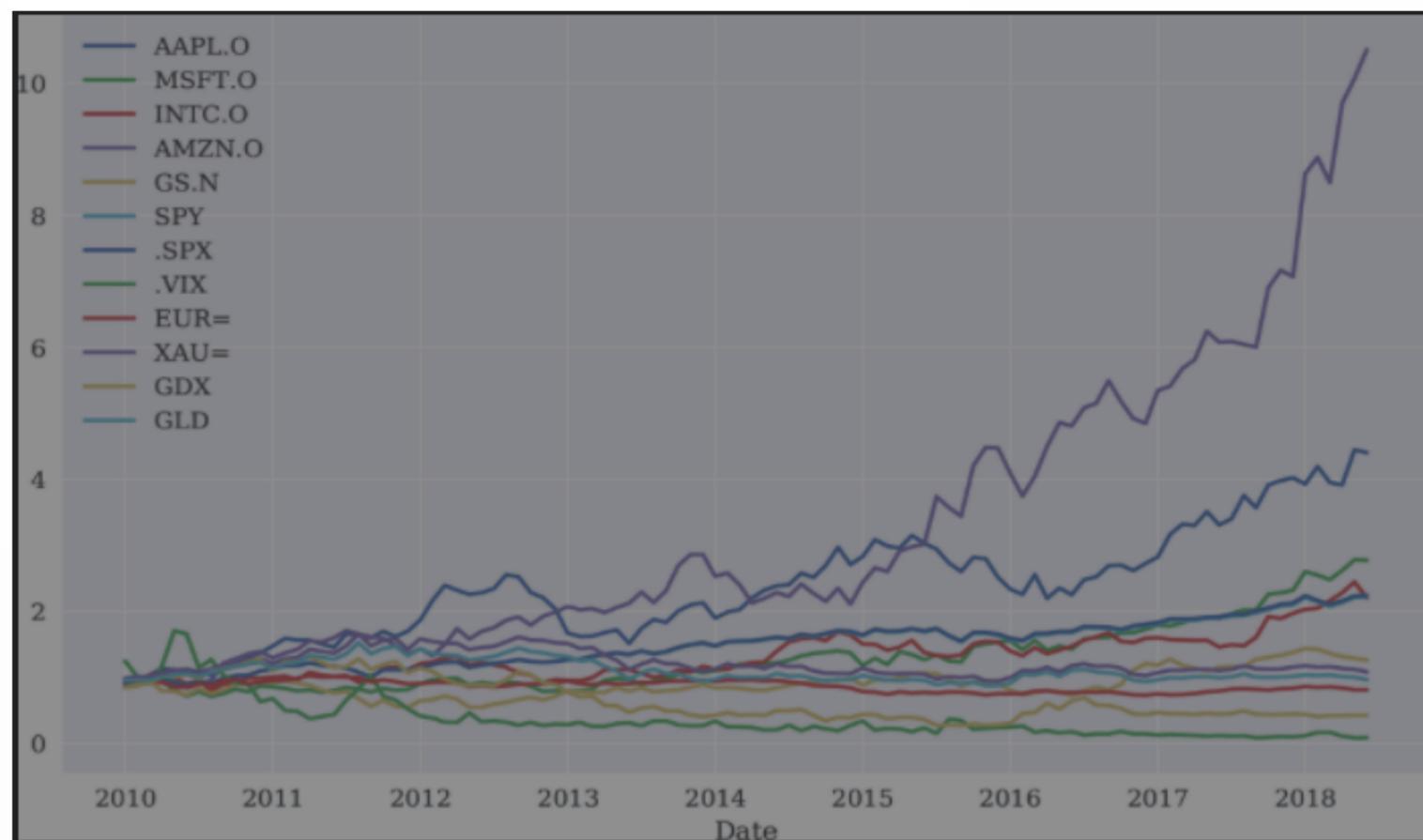
Plotly



时间序列



TIANCHI 天池



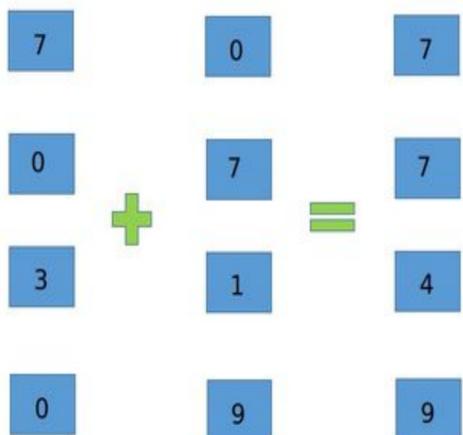
性能优化

金融数据科学

TIANCHI 天池

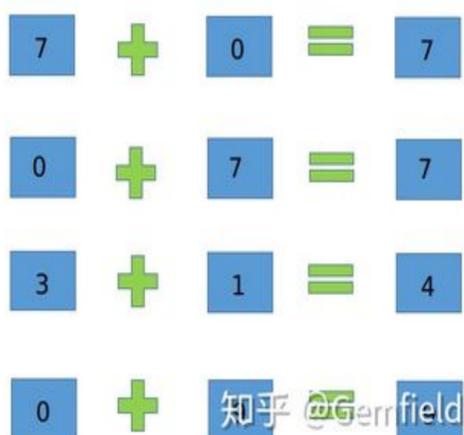
1

Vectorization



gemfield.org

For loop



知乎 @Gemfield

2



3



何为算法交易



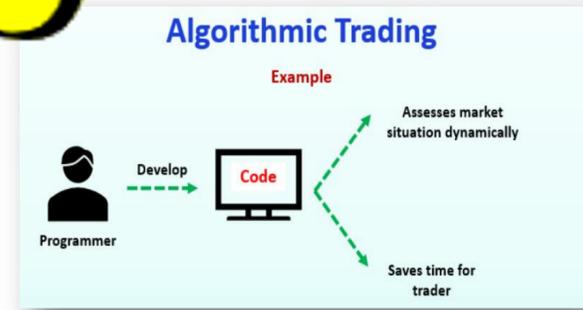
TIANCHI 天池

越来越多的交易平台和经纪商允许客户使用各种工具（如REST API），**以编程的方式读取历史数据或流数据，或者下单买入及卖出**。大型金融机构长期以来占据的这块“领地”，现在居然向使用算法交易的散户和业余**宽客**敞开了大门。

FXCM交易平台的REST API

算法交易

4



TIANCHI 天池

1 安装Python包管理器库fxcmpy

```
pip install fxcmpy
```

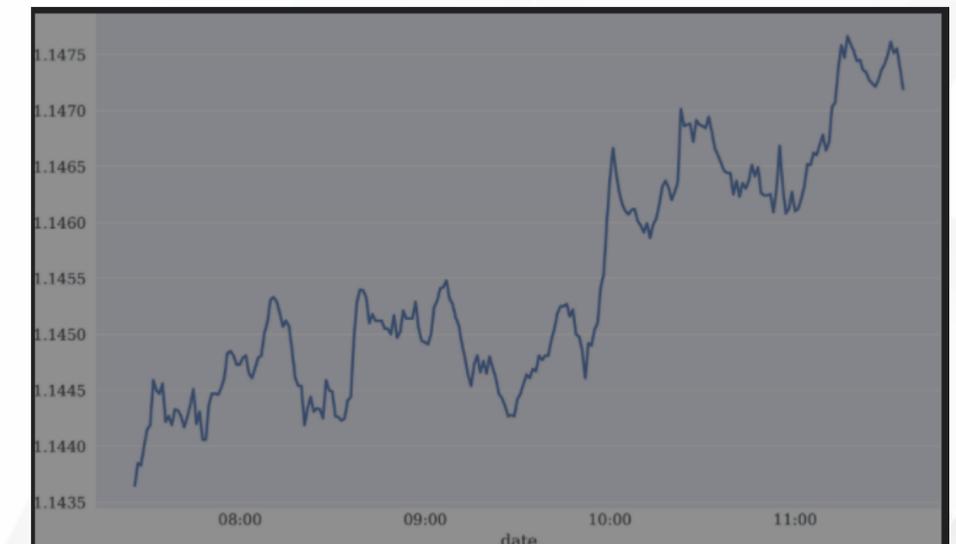
2 打开到API的连接

```
import fxcmpy
api = fxcmpy.fxcmpy(access_token=YOUR_FXCM_API_TOKEN, log_level='error')
```

3 读取最近可用的一分钟价格

```
candles = api.get_candles('EUR/USD', period='m1', number=250)
candles['askclose'].plot(figsize=(10, 6))
```

4 显示欧元/美元历史收盘买入价（分钟线）



直播相关资料获取及回放查看地址：<https://tianchi.aliyun.com/specials/promotion/activity/bookclub>

选择交易算法

- 简单移动平均数
- 随机游走假设
- 线性回归
- 聚类
- 频率方法
- 分类
- 深度神经网络算法交易



1

```
In [101]: train[cols_bin].head()
Out[101]:
```

Date	lag_1_bin	lag_2_bin	lag_3_bin	lag_4_bin	lag_5_bin
2010-01-12	3	3	0	2	1
2010-01-13	1	3	3	0	2
2010-01-14	2	1	3	3	0
2010-01-15	1	2	1	3	3
2010-01-20	1	0	1	2	1

```
In [115]: train, test = train_test_split(data, test_size=0.5,
                                         random_state=100)

In [116]: train = train.copy().sort_index()

In [117]: test = test.copy().sort_index()

In [118]: model = MLPClassifier(solver='lbfgs', alpha=1e-5, max_iter=500,
                               hidden_layer_sizes=3 * [500], random_state=1)

In [119]: %time model.fit(train[cols_bin], train['direction'])
CPU times: user 2min 26s, sys: 1.02 s, total: 2min 27s
Wall time: 1min 31s

Out[119]: MLPClassifier(activation='relu', alpha=1e-05, batch_size='auto',
                        beta_1=0.9,
                        beta_2=0.999, early_stopping=False, epsilon=1e-08,
                        hidden_layer_sizes=[500, 500, 500], learning_rate='constant',
                        learning_rate_init=0.001, max_iter=500, momentum=0.9,
                        n_iter_no_change=10, nesterovs_momentum=True, power_t=0.5,
                        random_state=1, shuffle=True, solver='lbfgs', tol=0.0001,
                        validation_fraction=0.1, verbose=False, warm_start=False)

In [120]: test['pos_dnn_sk'] = model.predict(test[cols_bin])

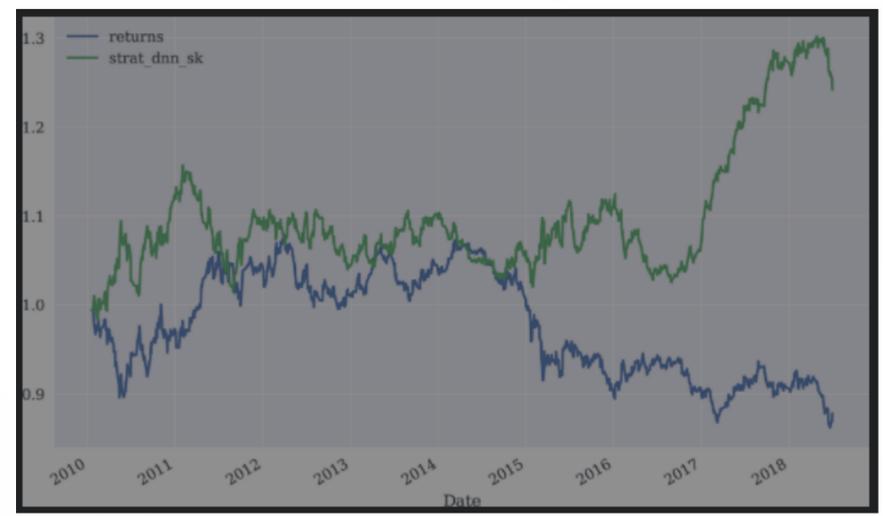
In [121]: test['strat_dnn_sk'] = test['pos_dnn_sk'] * test['returns']

In [122]: test[['returns', 'strat_dnn_sk']].sum().apply(np.exp)
Out[122]: returns      0.878078
          strat_dnn_sk  1.242042
          dtype: float64

In [123]: test[['returns', 'strat_dnn_sk']].cumsum(
          ).apply(np.exp).plot(figsize=(10, 6));
```

2

Scikit-learn中的MLPClassifier算法 (ANN)



3



最近50年里，美国股票获得了大致平均9%的真实收益。当中仅有大约1%的收益归功于利率，而**剩余的8%是持有风险所获得的溢价**。不确定性，或者风险修正促使资产定价成为吸引人的、而且富于挑战性的领域。

蒙特卡洛模拟

5

衍生品分析



TIANCHI 天池

- 1 建模
- 2 模拟
- 3 估值



实战：比特币价格预测



时间序列具有一些关键特征于时间的延展相关联，而**预测**是根据过去和现在的数据对未来进行预测的过程。

在这次实战中，我们尝试用Facebook的开源机器学习工具**Prophet**对历史比特币价格数据进行时间序列分析，对走势进行预测。



机器学习，是利用计算机的运算能力从大量的数据中发现一个“函数”或“模型”，并通过它来模拟现实世界事物间的关系，从而实现预测或判断的功能。

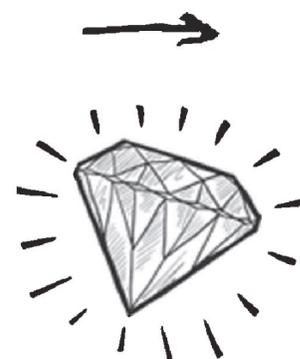
特征：

大小 (X_1)

重量 (X_2)

颜色 (X_3)

密度 (X_4)



标签：

价格 (Y)



机器学习

≈

从数据中
寻找一个函数

预测票房的函数：

$$f(x, x \text{ 为制作成本、演员、广告等数据}) = 500 \text{ 万! (票房数字)}$$

人脸识别的函数：

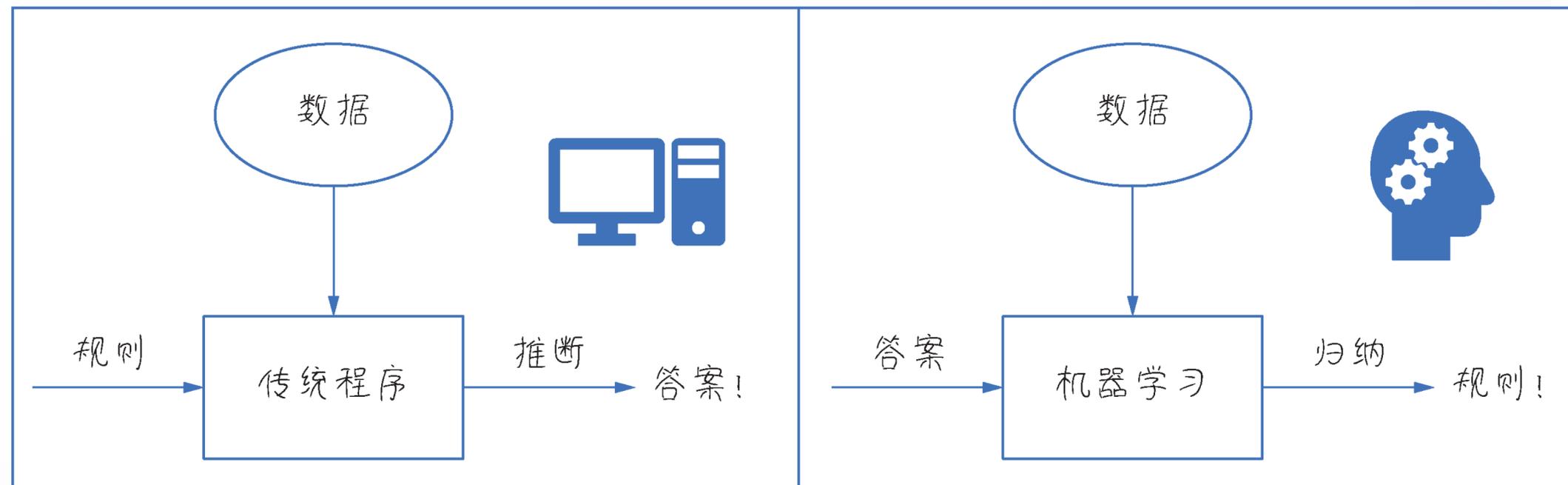
$$f(\text{人脸}) = \text{咖哥! (脸的主人)}$$

玩游戏的函数：

$$f(\text{游戏画面}) = \text{大力跳! (下一步指令)}$$

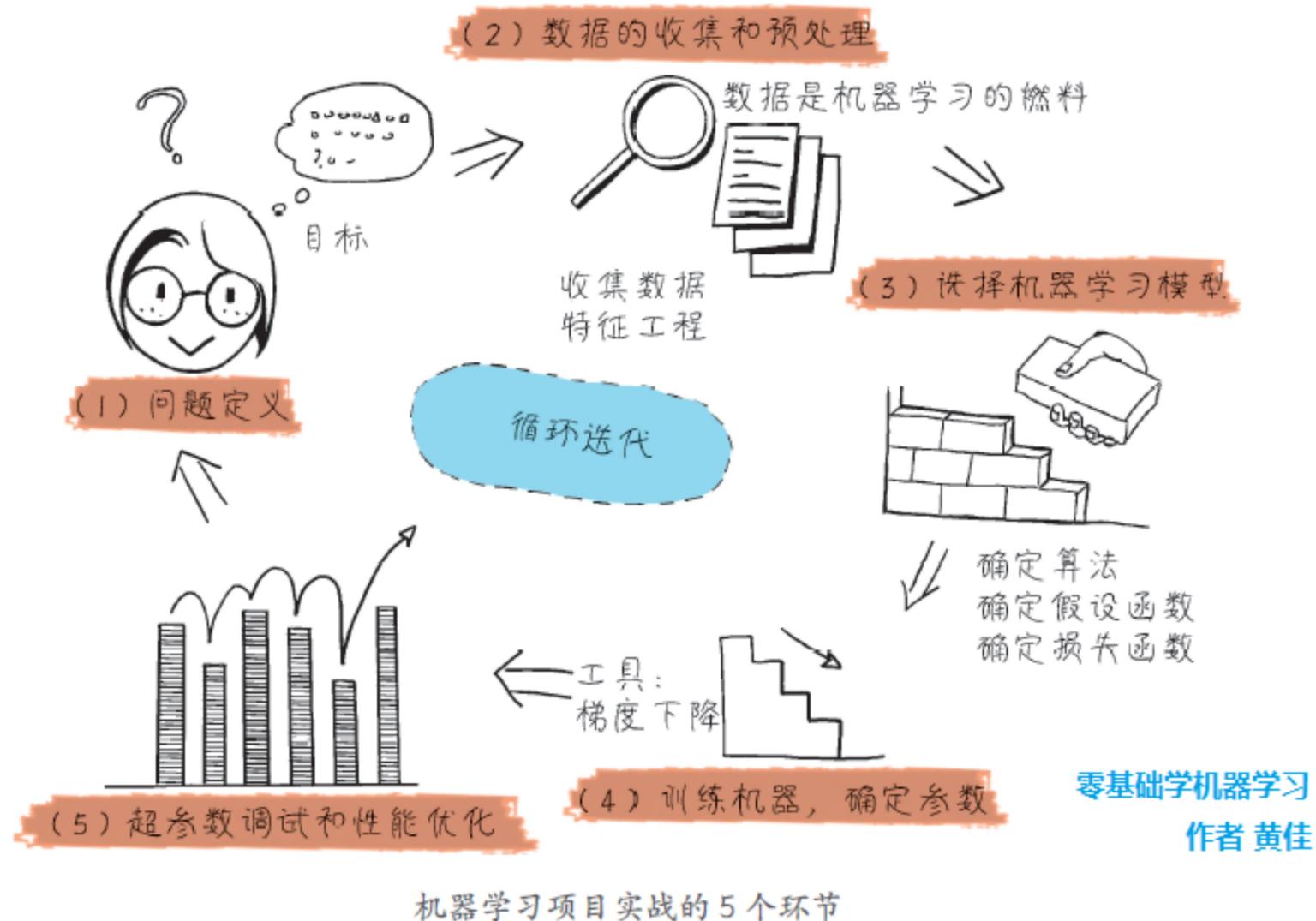
机器学习和传统程序的区别

机器学习是突破传统的学习范式，是从数据中发现规则，而不是接受人类为它设定的规则。



机器学习项目实战五部曲

TIANCHI 天池



更多机器学习
基础知识可以
参考我之前的分享
《零基础学机器学习》

实战演示、互动交流

TIANCHI天池

阿里云 | TIANCHI天池 异步社区
人民邮电出版社

天池读书会 Python金融大数据分析

分享嘉宾：黄佳 SAP 资深顾问

直播时间：4月30日 20:00

直播通道：@B站达摩院扫地僧
@天池读书会



扫码
观看直播



Python和机器学习方法已经广泛应用于交互式金融分析及金融应用开发，在高层次的抽象和严格的算法实现中提供最佳的平衡。

- 01 Python编程基础与金融应用
- 02 金融数据科学与算法交易基础知识
- 03 机器学习在金融数据科学中的应用

如何将机器学习方法应用于金融数据科学实战，黄佳老师将带来易于上手的实战案例分享。

大家可以使用电脑访问下方地址进入天池读书会页面，点击今天读书会中的 **实践代码** 和我一起进行项目实践学习，天池为大家准备好了代码和运行环境，非常方便。

<https://tianchi.aliyun.com/specials/promotion/activity/bookclub>



黄佳 机器学习专家

直播主题 《Python金融大数据分析(第二版)》

直播时间 2021年4月30日 20:00

学习资料 金融风控训练营

实践项目 比特币走势预测 (Prophet工具)

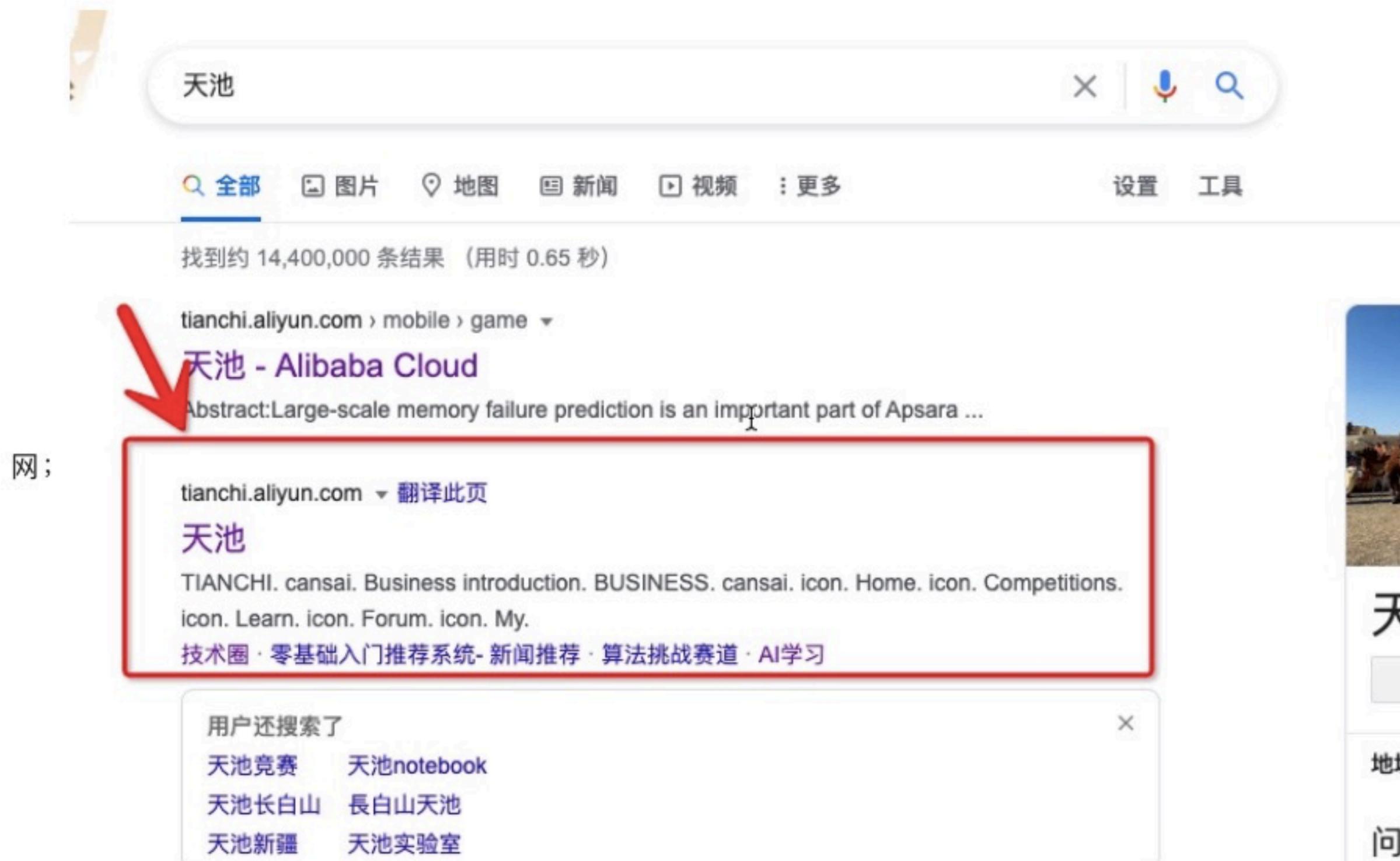


[🗨️ 提问](#) | [📖 学习训练营](#) | [🛒 购买地址](#) | [📄 PPT下载](#) | [👉 实践代码](#) | [🕒 预约直播](#)

直播相关资料获取及回放查看地址：<https://tianchi.aliyun.com/specials/promotion/activity/bookclub>

Q & A

1) 首先需要进入天池官网，大家打开浏览器，搜索 天池，找到 tianchi.aliyun.com即可访问进入天池官



网；

2) 在天池官网，将鼠标移到 天池学习，即可出现下拉列表，点击 天池读书会，即可进入天池读书会的页面。



3) 在天池读书会页面，你可以对对应的读书会图书进行提问，优秀的提问还有机会获得赠书，还可以点击配套的训练营或者课程资源进入学习，还有点击实践代码获取读书会的项目实践的代码，跟着我一起进行项目实践和代码学习，同时还有很多其他的读书会，大家也可以观看举办过的读书会的回放，或者预约还没开始的读书会。



黄佳 机器学习专家

直播主题 《Python金融大数据分析(第二版)》

直播时间 2021年4月30日 20:00

学习资料 金融风控训练营

实践项目 比特币走势预测 (Prophet工具)



[🗨️ 提问](#) | [📖 学习训练营](#) | [🛒 购买地址](#) | [📄 PPT下载](#) | [👉 实践代码](#) | [🕒 预约直播](#)

谢谢观看

TIANCHI 天池

直播相关资料获取及回放查看地址：<https://tianchi.aliyun.com/specials/promotion/activity/bookclub>