

평균회기론을 사용한 주가 방향 예측

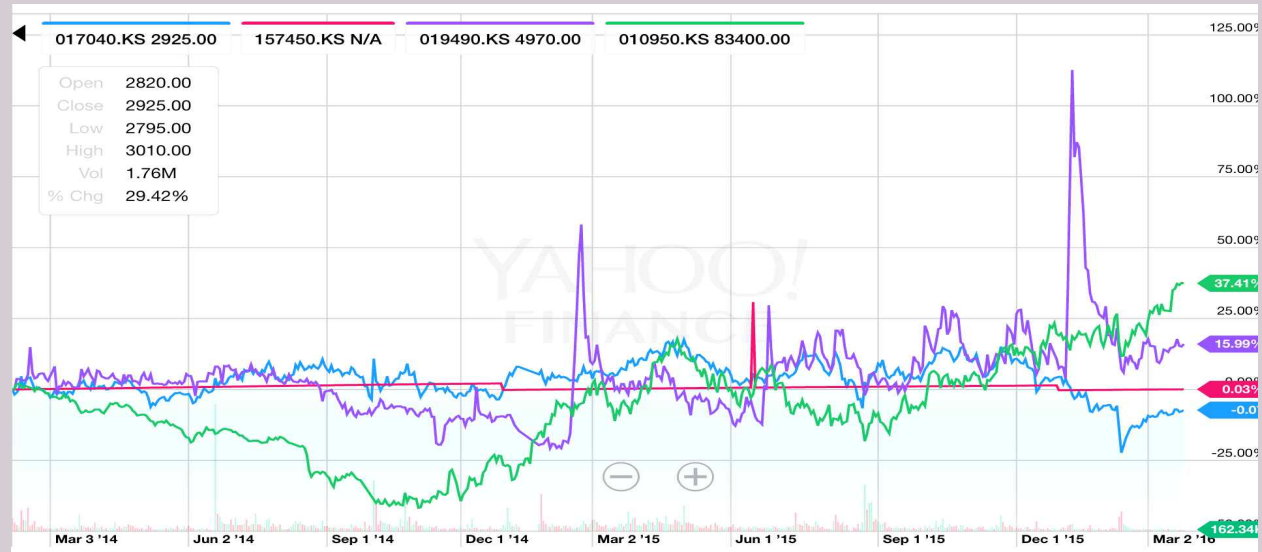
2011104027 안재성

평균회귀



평균회귀(2014-01-01 ~ 2016-12-10)

rank_score 상위권

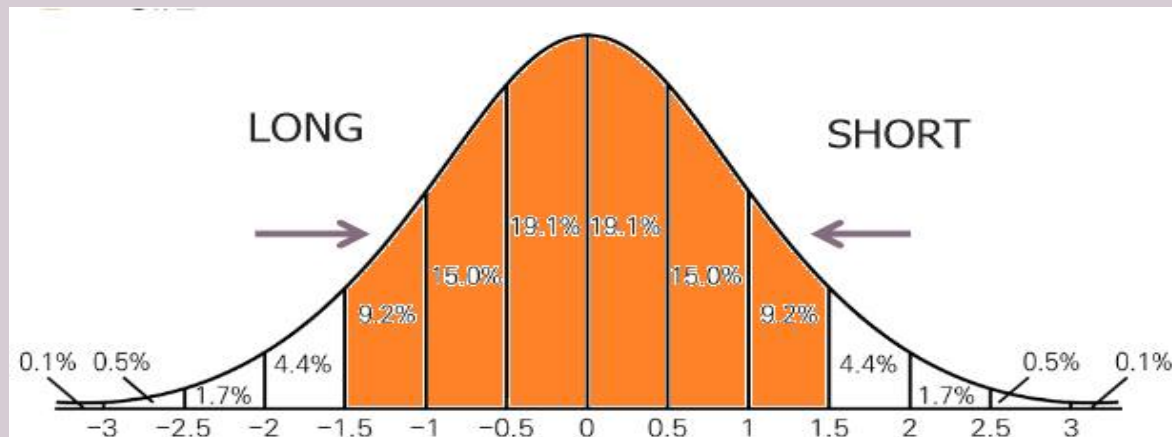


rank_score 하위권



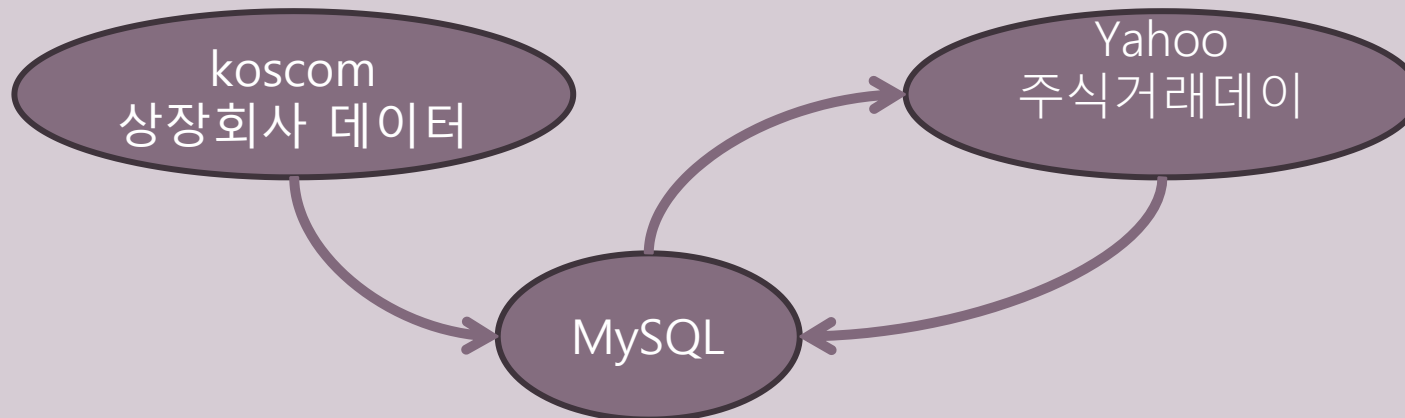
주식 시계열 데이터 분석, 방향 예측

1. 데이터들의 평균회귀 성향을 확인하는 테스트
2. Augmented Dickey-Fuller Test, 허스트 지수, Half-Life 를 사용해 1/3비율로 합산한 점수로 평균회귀성향 순위 획득
3. LONG /SHORT = 상승/하강 방향성 파악
4. 방향성 다음날에 맞았는 지 예측 결과값 반영
5. 2016-01-01~2016-12-10예측 결과를 결과물로 보임



작동방식 (매일)

1. 주식관련 데이터 업데이트
2. 지난날의 예측 결과값 DB에 반영



3. ADF, Hurst, Half-life 이용해
거래데이터의 평균회귀 성향 파악
4. 평균 회귀 성향이 있는
주식의 방향 예측.
5. 다음날의 예측 결과값 반영 대기

모듈

runDownloadStockData : koscom,yahoo에서 주가데이터
다운로드

runCheckMeanReversion : MeanReversionTest진행,
rank, direction생성

runCreateStockPrediction : direction으로 예측결과생성

runShowStockPrediction : 예측결과 보여줌

PortfolioBuilder : 주가 방향 예측결과내기, 보기용

DataReadWriter : MySql, Koscom, yahoo 데이터관련

MeanReversionModel : 평균회귀성향 파악용 계산모델

StockData : 상장회사 정보, 주식일일거래 정보 수집용

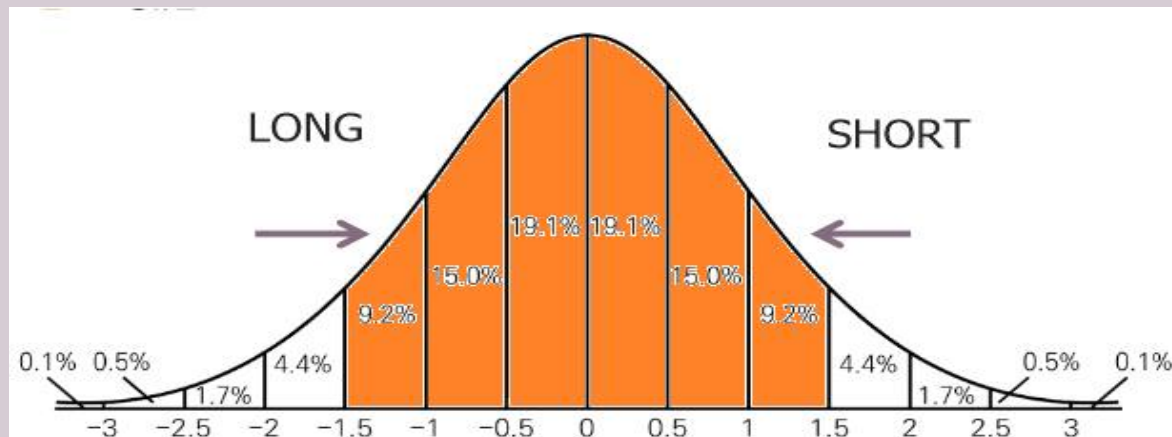
PortfolioBuilder

determineMeanReversionDirection : 평균회기모델로
방향성 파악.

rankMeanReversion: ADF, Hurst, Half-life 각각 평가된
점수를 토대로 DataFrame 생성

buildUniverse : 평가된 순위중 상위 주식목록 리턴

showHitRatio : 일정 기간내에 얼마나 방향을 맞췄는지
예측결과 보여줌.



MeanReversionModel

calcADF : satsmodels.tsa.stattools를 이용해 결과값, 1,5, 10%의 기각값을 리턴. $\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$,

calcHurst : 결과로 Hurst지수가 나와서 0에 가까울 수록 평균회귀성향을 띤. 1에 다가갈 수록 발산성향을 띤.

calcHalflife : 오른스타인-우렌벡 과정 참고해 평균회귀하는데 걸리는 시간을 구함. 시간이 적을 수록 평균회귀 성향을 띤다.

$$R/S = \frac{R(\tau)}{S(\tau)} \quad S(\tau) = \sqrt{\frac{1}{\tau} \sum_{t=1}^{\tau} \{\xi(t) - \langle \xi \rangle_{\tau}\}^2}$$

$$E \left[\frac{R(n)}{S(n)} \right] = Cn^H \quad \text{as } n \rightarrow \infty$$

$$E \left[\frac{R(n)}{S(n)} \right] = Cn^{0.5} \quad \text{as } n \rightarrow \infty$$

$$ret_{t-\Delta, t} = \log(P_t) - \log(P_{t-\Delta})$$

$$ret_{t-\Delta, t} = \frac{(P_t - P_{t-\Delta})}{P_{t-\Delta}}$$

$$dx_t = \lambda(\mu - x_t)dt + \sigma dW_t$$

$$f(t) = y_0 e^{-\lambda t}$$

$$f(t_{1/2}) = \frac{f(t)}{2}$$

$$Half - life, t_{1/2} = -\frac{\ln 2}{\lambda}$$

DB Tables

```
mysql> describe codes;
```

Field	Type	Null	Key	Default	Extra
last_update	datetime	NO		NULL	
code	char(20)	NO	PRI	NULL	
full_code	char(200)	NO		NULL	
market_type	int(1)	YES		0	
company	char(100)	NO		NULL	

5 rows in set (0.01 sec)

```
mysql> describe prices;
```

Field	Type	Null	Key	Default	Extra
last_update	datetime	NO		NULL	
price_date	datetime	NO	PRI	NULL	
code	char(20)	NO	PRI	NULL	
price_open	double	YES		NULL	
price_close	double	YES		NULL	
price_low	double	YES		NULL	
price_high	double	YES		NULL	
price_adj_close	double	YES		NULL	
volume	int(1)	YES		NULL	

9 rows in set (0.00 sec)

```
mysql> describe directions;
```

Field	Type	Null	Key	Default	Extra
last_update	datetime	NO		NULL	
price_date	datetime	NO	PRI	NULL	
code	char(200)	NO	PRI	NULL	
company	char(200)	NO		NULL	
target_column	char(30)	NO	PRI	NULL	
direction	char(10)	NO		NULL	
rank_score	double	YES		NULL	

7 rows in set (0.00 sec)

```
mysql> describe countPrediction;
```

Field	Type	Null	Key	Default	Extra
last_update	datetime	NO		NULL	
code	char(20)	NO	PRI	NULL	
company	char(100)	NO		NULL	
target_column	char(30)	NO	PRI	NULL	
count_true	int(1)	YES		NULL	
count_false	int(1)	YES		NULL	
count_all	int(1)	YES		NULL	
money_diff	int(1)	YES		NULL	

8 rows in set (0.00 sec)

runCheckMeanReversion (1일치)

```
target_index = 519
adf_statistic: -3.19060690549, adf_1기 각 값: -3.44301269337, adf_5기 각 값: -2.8671249839, adf_10기 각 값: -2.56974459023
hurst exponent: 0.388522703933, half life: 17.3531585548
... 521 of 1306 : Testing Mean Reversion Model 2016-02-15 on 015540 핫텍
target_index = 504
adf_statistic: -1.48980344593, adf_1기 각 값: -3.44344373198, adf_5기 각 값: -2.86731468755, adf_10기 각 값: -2.56984568848
hurst exponent: 0.421661492618, half life: 92.70763602
... 522 of 1306 : Testing Mean Reversion Model 2016-02-15 on 015590 대경기계
target_index = 521
adf_statistic: -0.633426026308, adf_1기 각 값: -3.44308668477, adf_5기 각 값: -2.86715755002, adf_10기 각 값: -2.56976194535
hurst exponent: 0.356080491059, half life: 92.7282134964
... 523 of 1306 : Testing Mean Reversion Model 2016-02-15 on 015760 한국전력
target_index = 521
adf_statistic: -1.50934472942, adf_1기 각 값: -3.44306192508, adf_5기 각 값: -2.86714665253, adf_10기 각 값: -2.56975613785
hurst exponent: 0.292538009773, half life: 46.9956392289
... 524 of 1306 : Testing Mean Reversion Model 2016-02-15 on 015860 일진홀딩스
```

-adf, hurst, half-life값을 계산하고 rank_score를 구함(50분)

-rank_score 상위 90%로 direction을 예측하여 저장(10분)

runCreateStockPrediction (1일치)

특정 주식의 등향표

	price_date	code	company	target_column	direction \
0	2016-02-15	230480	KOSEF 미국달러선물인버스2X(합성)	price_close	HOLD
1	2016-02-12	230480	KOSEF 미국달러선물인버스2X(합성)	price_close	HOLD
2	2016-01-22	230480	KOSEF 미국달러선물인버스2X(합성)	price_close	SHORT

rank_score

0 2.409716

1 2.373669

2 1.871359

date : 2016-02-15

diff price: -130.0, direction: HOLD

date : 2016-02-12

diff price: 65.0, direction: HOLD

date : 2016-01-22

diff price: 65.0, direction: SHORT

Correct Ratio: 0.0, all_count: 1, money_diff: -703, code: 230480, target_column: price_close

-direction예측 값으로 1306개의 예측 적중율, 투자금의 변동액을 알아냄 (8분)

runShowStockRecommendation

```
(tensorflow) AnJaeSeongui-MacBook-Pro:workspaceTensorflow AJSMac$ python runShowStockPrediction.py
Total Prediction Show
[2016-01-04 ~ 2016-xx-xx] correct_ratio:55.814, count_true:120, count_all:215
                benefit_ratio:1.01, money_diff:216655, use_money:21500000
search Recommend Direction By Date. Input Date: (xxxx-xx-xx):2016-01-20
(하강 예측)
rank_score:3.45630756357, target_column:price_close, code:157450, company:TIGER 유동자금
rank_score:2.48094651026, target_column:price_close, code:214390, company:경보제약
rank_score:2.11023092308, target_column:price_close, code:155660, company:DSR
rank_score:1.82814777726, target_column:price_close, code:225030, company:TIGER S&P500;인버스선물(H)
(상승 예측)
rank_score:3.56080489024, target_column:price_close, code:002410, company:범양건영
rank_score:2.64914827228, target_column:price_close, code:017040, company:광명전기
rank_score:2.52691501108, target_column:price_close, code:007110, company:일신석재
rank_score:2.475773078, target_column:price_close, code:001550, company:조비
```

-전체 날짜의 예측결과 통계 (예시 : 16-01-04 ~ 16-01-22)

-원하는 날짜의 동향, 추천주식 보기

원하는 날짜 동향 예측 보기

search Recommend Direction By Date. Input Date: (xxxx-xx-xx)2016-01-08

(하강 예측)

rank_score:2.45122656149, target_column:price_close, code:001550, company:조 비

rank_score:2.43942614266, target_column:price_close, code:003560, company:IHQ

rank_score:2.05075227101, target_column:price_close, code:081660, company:휠 라 코 리 아

(상승 예측)

rank_score:3.18847088614, target_column:price_close, code:001440, company:대 한 전 선

rank_score:2.44361177576, target_column:price_close, code:001560, company:제 일 연 마

rank_score:3.54326670923, target_column:price_close, code:002410, company:범 양 건 영

rank_score:1.92892110966, target_column:price_close, code:002550, company:KB손 해 보 험

rank_score:2.3251874238, target_column:price_close, code:006490, company:인 스 코 비

rank_score:2.25299966785, target_column:price_close, code:007610, company:선 도 전 기

rank_score:2.72372738187, target_column:price_close, code:017040, company:광 명 전 기

rank_score:2.36884889938, target_column:price_close, code:028670, company:팬 오 셴

rank_score:1.93479998818, target_column:price_close, code:18064K, company:한 진 칼 우

사용자가 짠 투자 전략

- 평균회귀 성향이 강한(상위 10%) 주식만 투자진행
- 동향예측 1회당 100,000원 일괄 투자
- 사용자가 입력한 기간(35일, 9일) 뒤의 주식데이터로 예측 적중률, 순 이익률을 평가

모의 투자 결과 분석(35일 뒤, 9일 뒤)

16-01-04~	correct_ ratio	count_ true	count_ all	benefit_ ratio	money_ diff	money_ used
16-01-22 (35d)	55.814%	120	215	101.0%	216,655	21,500,000
16-02-16 (35d)	56.757%	231	407	102.2%	907,326	40,700,000
16-03-08 (35d)	52.264%	277	530	101.3%	677,690	53,000,000
16-01-22 (9d)	53.814%	127	236	101.1%	267,335	23,600,000
16-02-16 (9d)	57.385%	237	413	101.6%	660,788	41,300,000
16-03-08 (9d)	52.857%	296	560	100.9%	505,997	56,000,000

실제 주가 방향 적중률이 55% 수준으로 이득을 볼 수 있을것이라
예상이 가지만 실제 투자금 반환율이 101% 수준으로 현재의 투자
전략으로는 이득을 볼 수 없다.

사용 라이브러리

pandas : 금융 데이터 라이브러리

numpy : 과학, 수학 계산 라이브러리

BeautifulSoup : 데이터 수집용 라이브러리 html가져옴

MySQLdb : mysql연결용 MySQL-python에 동봉된 모듈

Statsmodels : 통계용 라이브러리

수학관련 참고자료 , 프로젝트 Git주소

MeanReversionTest

<https://www.quantstart.com/articles/Basics-of-Statistical-Mean-Reversion-Testin>

ADF

https://en.wikipedia.org/wiki/Augmented_Dickey%E2%80%93Fuller_test

<http://kostat.go.kr/attach/journal/10-1-8.pdf>

hurst exponent

http://www.bearcave.com/misl/misl_tech/wavelets/hurst/

<http://www.stator-afm.com/hurst-exponent/>

half-life

<http://www.investinganswers.com/financial-dictionary/debt-bankruptcy/half-life-6199>

Ornstein–Uhlenbeck_process

<http://mediatum.ub.tum.de/doc/1072652/1072652.pdf>

프로젝트 Git

<https://github.com/AnJaeSeongS2/StockRecommendationML>