

# 데이터분석

캡스톤디자인

“

소프트웨어융합학과  
2019102084 김윤희



## 로보어드바이저 알고리즘을 활용한 포트폴리오 구성



### 로보 어드바이저

: 로봇과 어드바이저의 합성어로 인간의 주관적 판단이나 개입 없이 수학적 규칙이나 알고리즘을 이용하여 투자자들에게 온라인 상으로 자산배분 포트폴리오를 제시하는 일종의 자산관리 로봇

- 시간적, 공간적 제약이 없고 맞춤형 포트폴리오를 추천해주기 때문에 많은 사람들이 이용
- 동일한 수익률에 낮은 위험을 원하는 사람에게 최적의 포트폴리오를 제공

코너해 문제

기대수익률이나 위험에 민감하게 반응

## 평균분산 모형



## 블랙리터만 모형

효율적 투자선 상의 포트폴리오를  
선택함으로써 효용 극대화

투자 상품들의 기대수익률,  
표준편차, 상관계수 이용

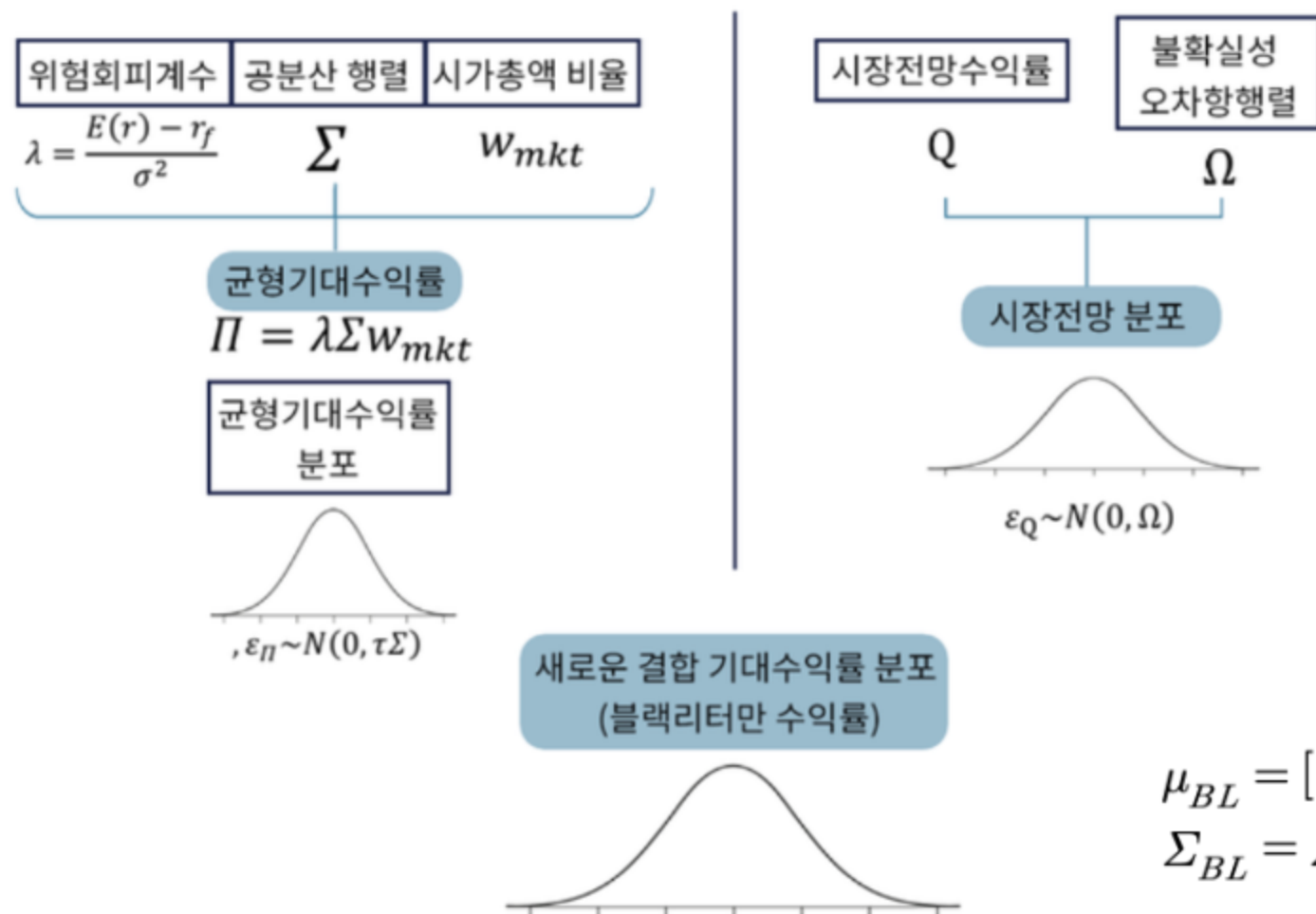
코너해 문제를 보완하면서 투자자의  
시장전망을 직관적으로 반영

기대 수익률  $\rightarrow$  내재균형기대수익률  
투자자 전망 이용

# 블랙 리터 만 모형

시장 포트폴리오  
 주식시장의 균형상태에서 전체  
 구성 종목의 시가총액의 비율로  
 구성된 포트폴리오  
 = 위험자산 포트폴리오 중 가장  
 효율적인 포트폴리오

## [ 균형기대수익률과 시장전망의 결합 ]



상대적 전망, 절대적 전망  
 전망에 대한 신뢰도 수준이 달리 질  
 수 있기에 이를 전망행렬에 반영

$$Q_t(k) = P_t \Pi_t(k) + \eta_k \sqrt{P_t \Sigma P_t'(k, k)}$$

$$\Omega = \alpha P(\tau \Sigma) P'$$

$$\mu_{BL} = [(\tau \Sigma)^{-1} + P' \Omega^{-1} P]^{-1} [(\tau \Sigma)^{-1} \Pi + P' \Omega^{-1} Q]$$

$$\Sigma_{BL} = \Sigma + [(\tau \Sigma)^{-1} + P' \Omega^{-1} P]^{-1}$$

=투자자전망을내재균형수익률과결합하여새로운사후확률분포인기대수익률벡터를계산

베이즈정리에의해사전에알고있던시장포트폴리오의확률분포로부터새로운투자자전망을반영한사후적변경결합확률분포를구하게되는것

## 데 이 터

국내 주식 데이터만을 이용해 포트폴리오 구성 - KOSPI 200을 구성하는 11개의 섹터지수

11개의 섹터지수: 커뮤니케이션비즈, 건설, 중공업, 철강/소재, 에너지/화학, 정보기술, 금융, 생활소비재, 경기소비재, 산업재, 헬스케어

2016~2018년 동안의 데이터를 학습해 얻은 포트폴리오 비중으로 2019년의 수익률을 예측

## 데 이 터 수 집

네이버 금융과 한국 거래소의 주가정보를 스크래핑하는 모듈인 pykrx를 사용하여 주식 데이터를 얻었다.

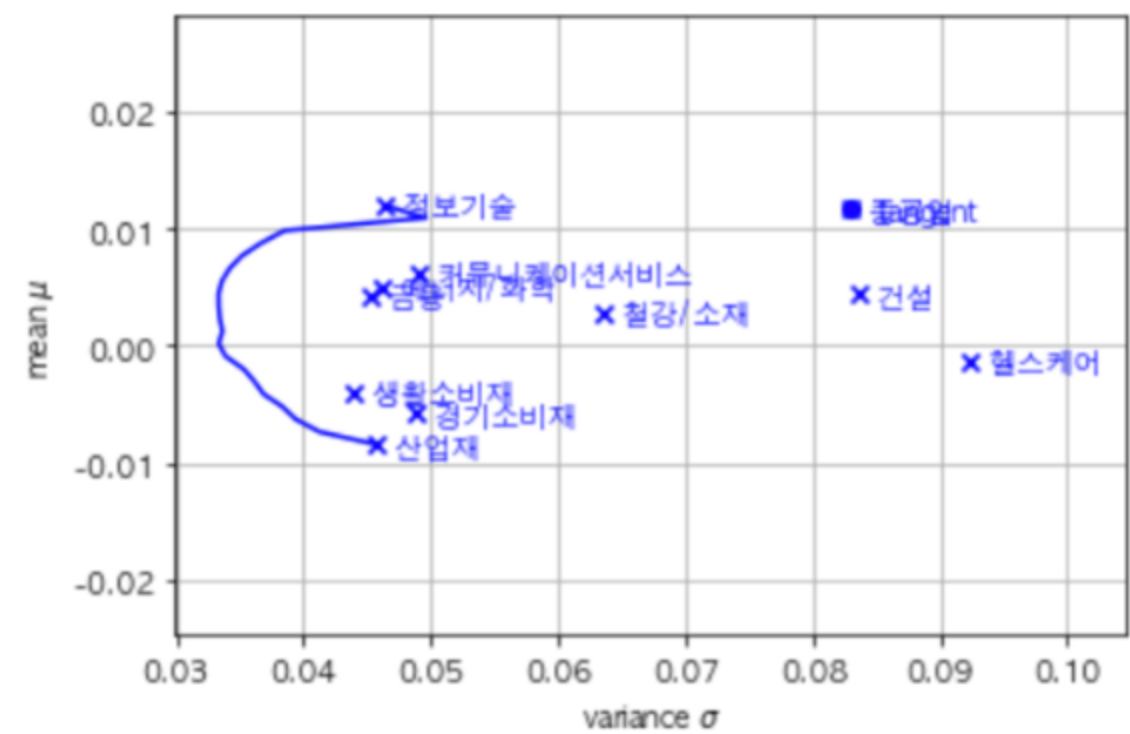
인덱스 조회 API로 코스피 인덱스를 조회해 코스피 200을 구성하고 있는 11개의 섹터지수에 대한 시가, 고가, 저가, 종가, 거래량에 대한 정보를 얻는다.

내재균형수익률을 구하기 위한 시장 포트폴리오의 시가총액에 대한 데이터는 pykrx에서는 제공하고 있지 않아 한국거래소에서 크롤링을 통해 시가총액 데이터를 얻었다.



# 평균 분산 모형

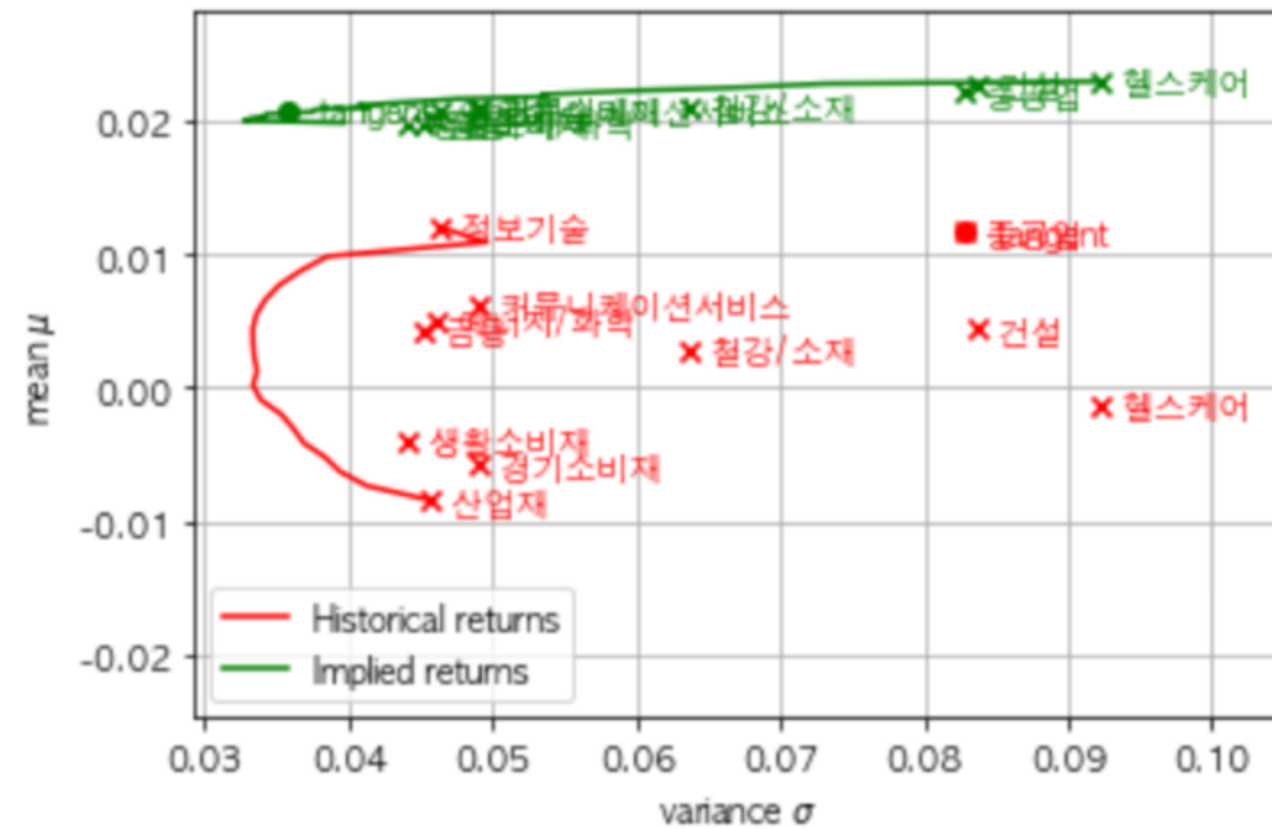
블랙리터만 모형과 비교하기 위해 평균 분산 모형으로 포트폴리오를 구성한 결과 효율적 투자선은 다음과 같고 포트폴리오 비중을 보면 앞서 말했던 특정 자산군에 몰리는 코너해 문제가 발생함을 알 수 있다.



	커뮤니케이션서비스	건설	중공업	철강/소재	에너지/화학	정보기술	금융	생활소비재	경기소비재	산업재	헬스케어
Weight	3.103854e-11	2.273887e-11	1.0	4.124469e-11	3.466807e-11	8.492871e-12	3.772688e-11	7.482418e-11	7.962232e-11	8.899085e-11	6.083805e-11

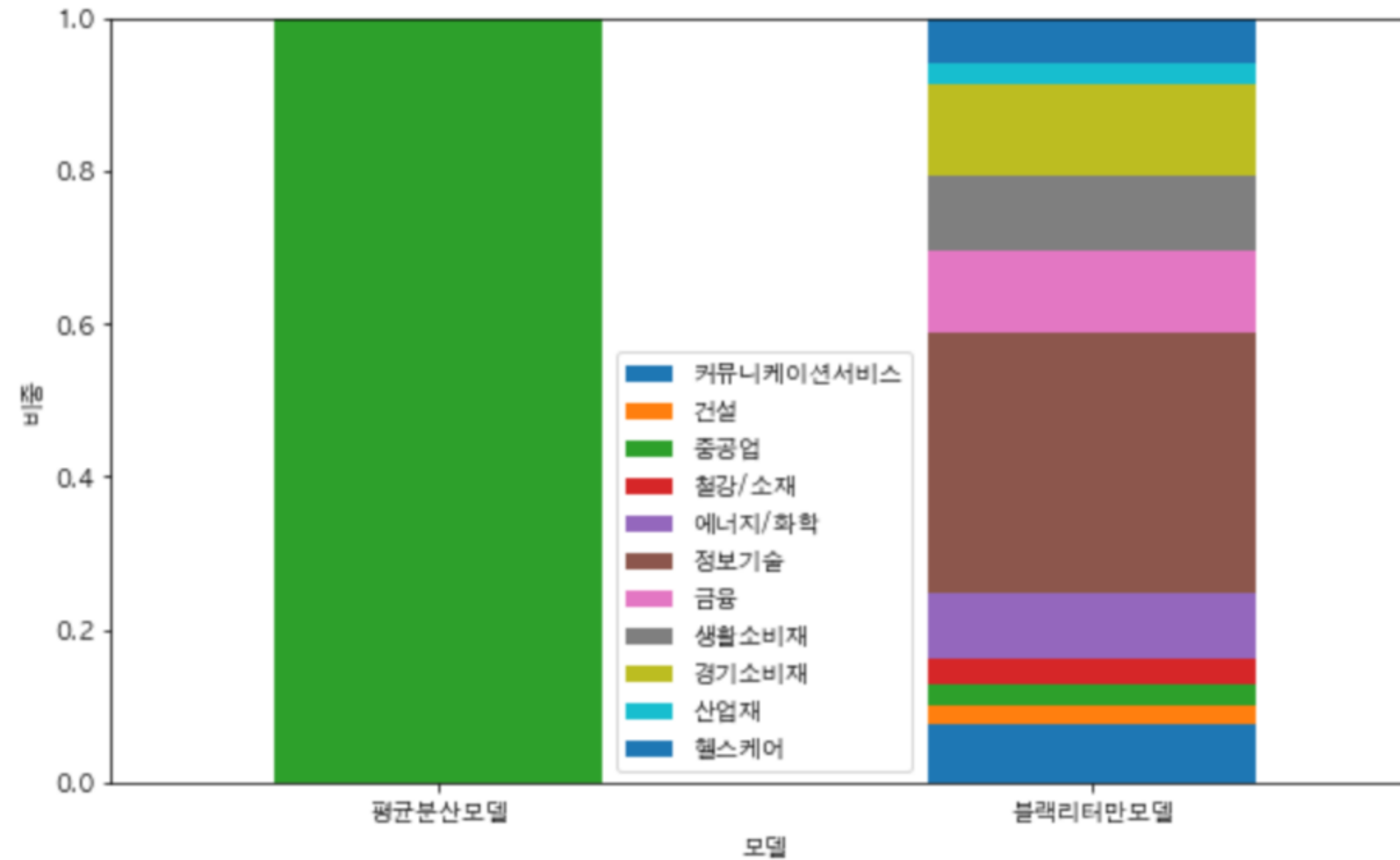
## 시장 포트폴리오를 활용한 내재 균형 수익률

평균 분산 모형의 문제를 해결하기 위해 시장 포트폴리오로부터 내재균형수익률을 계산하였을 때의 효율적 투자선은 초록색 선과 같고 포트폴리오 비중을 보면 코너해 문제를 해결하여 비중이 특정 자산군에 몰려있지 않은 것을 알 수 있다.



	커뮤니케이션서비스	건설	중공업	철강/소재	에너지/화학	정보기술	금융	생활소비재	경기소비재	산업재	헬스케어
Weight	0.076617	0.023941	0.028449	0.032978	0.086632	0.338472	0.107322	0.098032	0.121004	0.028266	0.058287

## 수익률 평균한 기대수익률 vs 내재 균형 수익률



따라서 수익률을 평균한 값을 기대수익률로 사용하는 평균 분산모델의 특정 자산군에 포트폴리오 투자비중이 몰려있는 코너해 문제를 시장 포트폴리오로부터 계산한 내재 균형 수익률을 사용하는 블랙리터만 모델이 문제를 해결한 것을 볼 수 있다.



# 투 자 자 전 망 예 측 1) SVM

input : 평균 수익률, 표준편차, Stochastic, 이격도

output : 수익률이 이전달보다 오르면 +1, 내려가면 -1, 그대로 유지한다면 0을 출력

## (1) 2016-2018년 3년 데이터를 학습하여 2019년 수익률 예측

예 측 값 을 담 은 행 렬 P

	커뮤니케이션서비스	건설	중공업	철강/소재	에너지/화학	정보기술	금융	생활소비재	경기소비재	산업재	헬스케어
2019/1	1	-1	1	-1	1	1	1	-1	1	1	1
2019/2	1	-1	-1	-1	1	1	1	-1	1	-1	-1
2019/3	1	-1	-1	-1	-1	1	1	-1	-1	1	1
2019/4	1	-1	-1	-1	1	-1	1	-1	-1	1	1
2019/5	1	-1	-1	-1	-1	1	1	-1	-1	1	-1
2019/6	1	-1	-1	-1	-1	1	1	-1	-1	1	-1
2019/7	1	-1	-1	-1	1	1	-1	-1	-1	1	1
2019/8	1	-1	-1	-1	1	1	-1	-1	-1	1	1
2019/9	1	-1	-1	-1	1	1	1	-1	-1	-1	1
2019/10	1	-1	-1	-1	1	1	1	-1	-1	1	1
2019/11	1	1	-1	-1	1	1	1	-1	-1	1	1
2019/12	1	1	-1	-1	1	1	1	-1	-1	1	1

$$Q_t(k) = P_t \Pi_t(k) + \eta_k \sqrt{P_t \sum P'_t(k, k)}$$

$$\mu_{BL} = [(\tau \Sigma)^{-1} + P' \Omega^{-1} P]^{-1} [(\tau \Sigma)^{-1} \Pi + P' \Omega^{-1} Q]$$

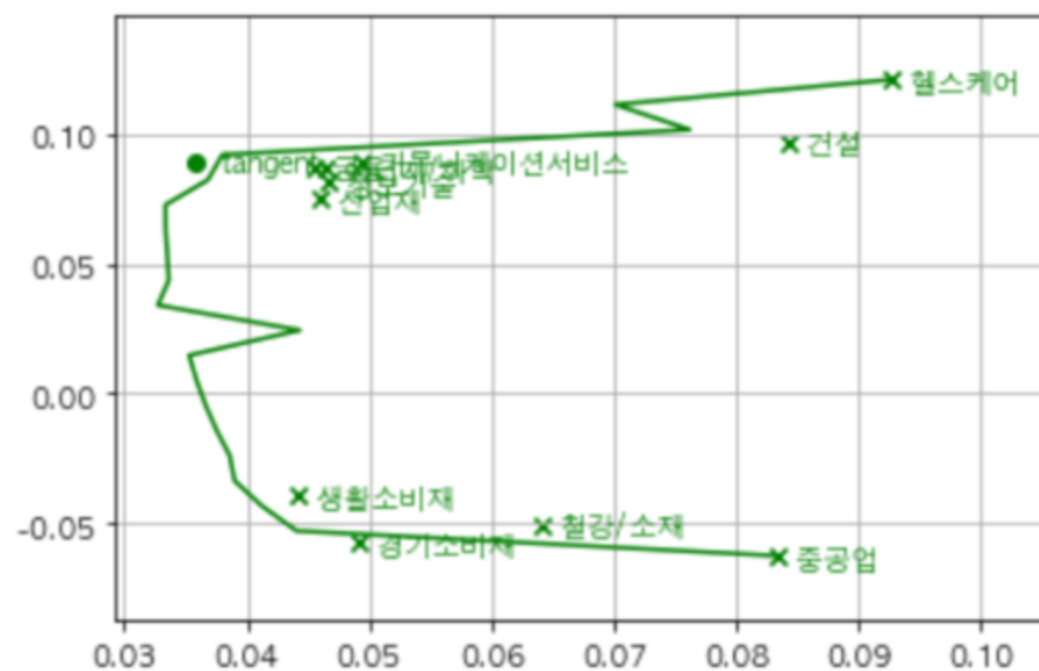
$$\Sigma_{BL} = \Sigma + [(\tau \Sigma)^{-1} + P' \Omega^{-1} P]^{-1}$$

$$\omega_{BL} = (\lambda \Sigma_{BL})^{-1} \mu_{BL}$$

# (1) 2016-2018년 3년 데이터를 학습하여 2019년 수익률 예측

포트폴리오의 각 자산별 투자비중

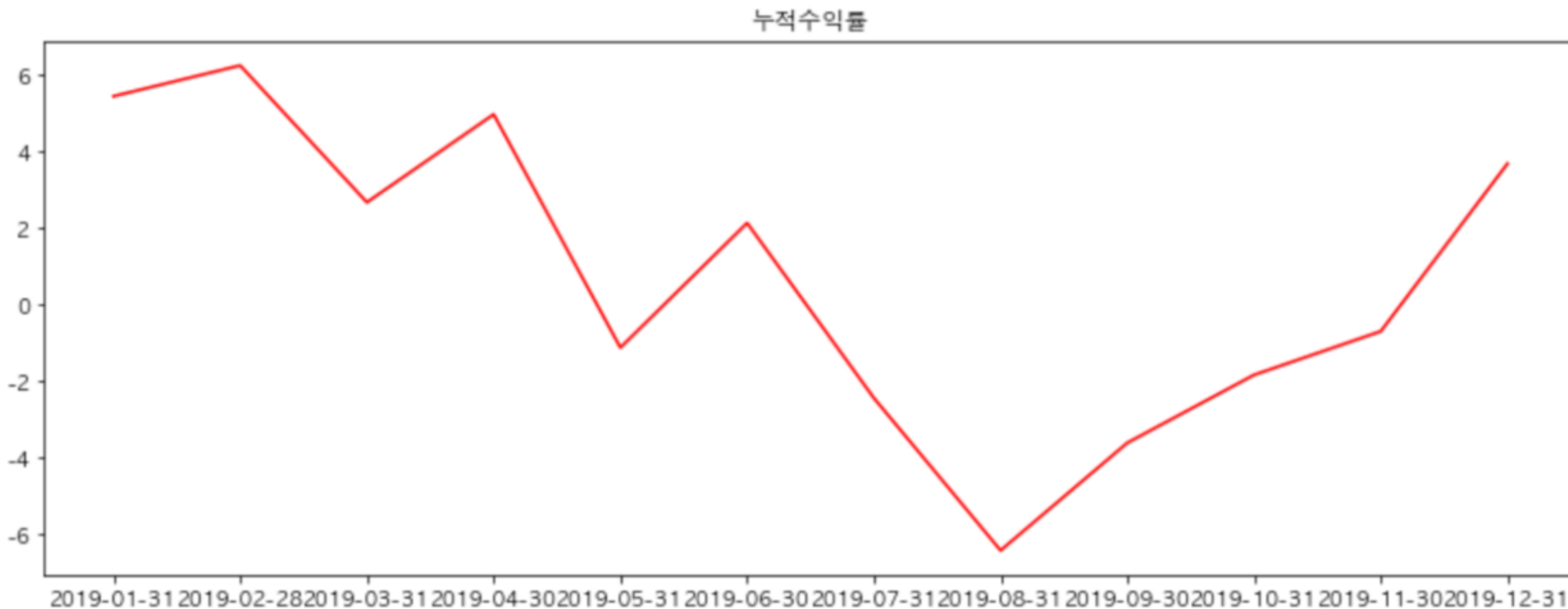
	커뮤니케이션서비스	건설	중공업	철강/소재	에너지/화학	정보기술	금융	생활소비재	경기소비재	산업재	헬스케어
Weight	0.099653	1.425455e-16	0.0	2.229391e-16	0.201286	0.22514	0.342025	0.0	0.0	0.030059	0.101837



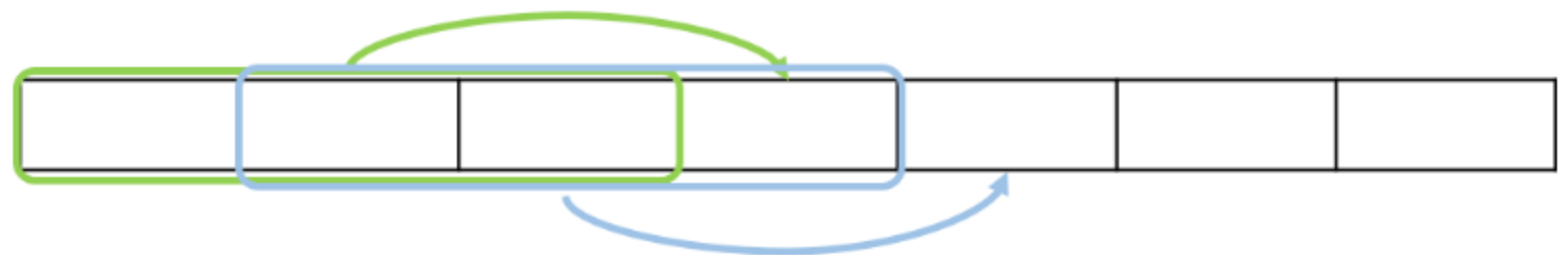
구한 투자 비중을 이용하여 2019년 월마다 수익률을 구해 누적 수익률을 나타낸 그래프이다. 예측한 투자자 전망을 결합한 블랙리터만 모델을 이용해 포트폴리오를 구성했을 때 2019년도의 **총 누적 수익률은 3.67%**가 된다.

Sharpe ratio: 0.91

이렇게 나온 포트폴리오의 투자 비중대로 했을 때 2019년 12월의 위험 대비 수익률의 지표인 샤프지수는 0.91이었다.



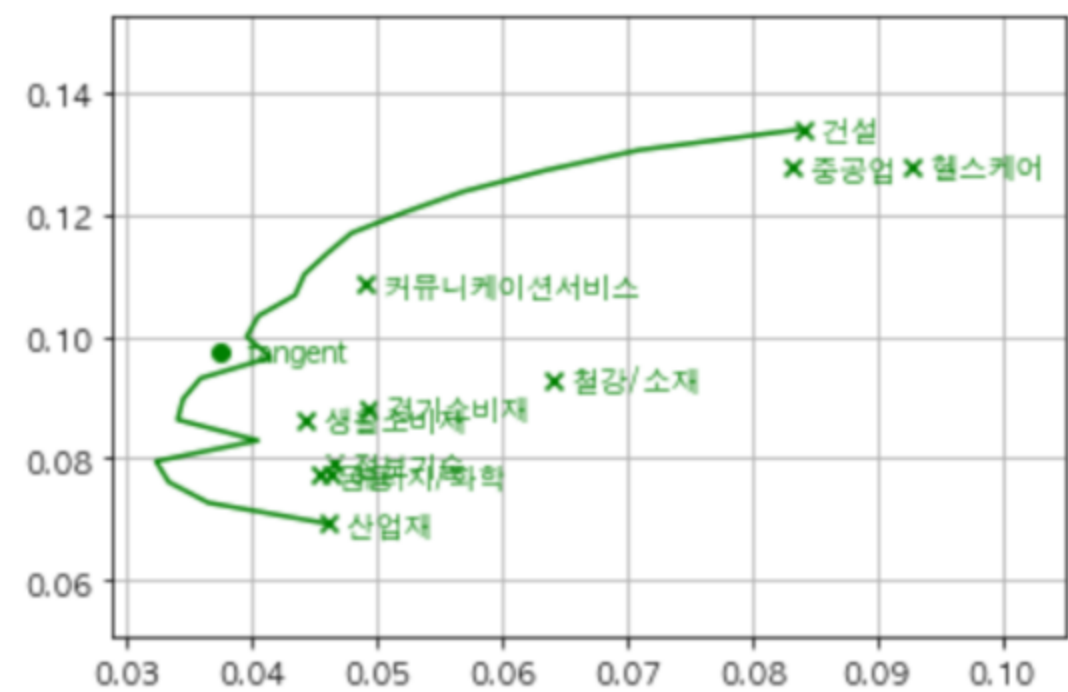
## (2) 예측값을 포함해 데이터로 다음 데이터를 예측



### ① 30개월 데이터 이용 다음 달 예측

포트폴리오 투자 비중

	커뮤니케이션서비스	건설	중공업	철강/소재	에너지/화학	정보기술	금융	생활소비재	경기소비재	산업재	헬스케어
Weight	0.349352	0.040971	0.078155	0.025036	0.125118	0.09377	0.054567	0.222921	4.065758e-19	2.710505e-19	0.01011



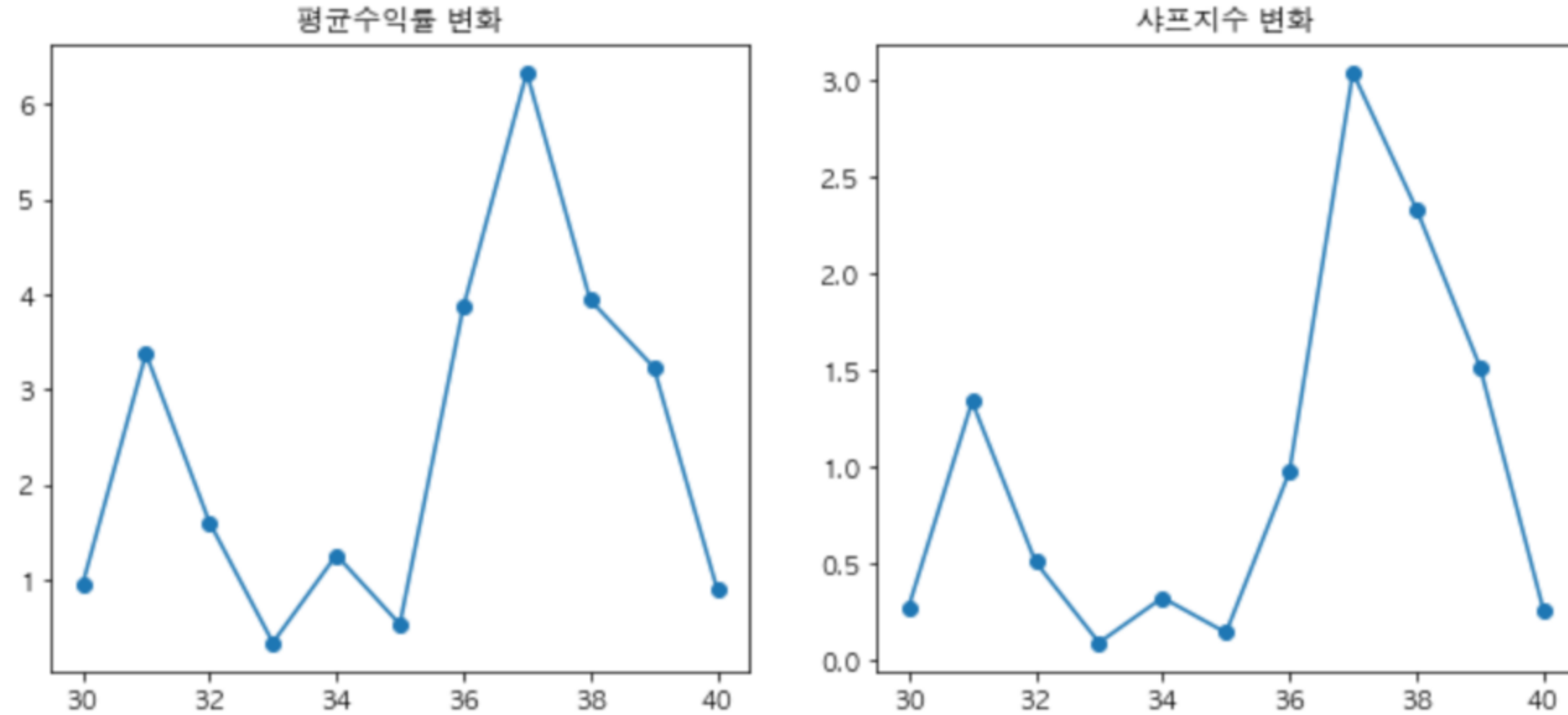
## ① 30개월 데이터 이용 다음 달 예측



30개월을 기준으로 예측하고 예측값을 이용해서 다음 값을 예측했을 때 구한 포트폴리오 투자 비중을 적용했을 때 2019년 누적 수익률의 변화는 다음과 같다. 2019년도의 **누적 총 수익률은 0.95%**이고 **샤프지수는 0.28**이다.

앞에서 실제 데이터를 이용해 예측한 것이 예측한 값으로 예측한 것보다 더 포트폴리오의 투자 성과는 좋은 것으로 확인되었다. 이 이유로는 학습 데이터를 이용해 예측한 값들은 오차가 더 많을 수 밖에 없기 때문이다.

## ② 기준 개월수 차이로 인한 수익률, 샤프지수 변화



기준을 30개월에서 40개월까지로 정했을때 나온 포트폴리오 투자 비중을 이용해 2019년의 평균수익률, 샤프지수의 변화를 나타낸 그래프이다.

보면 37개월을 예측사이즈로 정했을 때의 포트폴리오 투자 성능이 가장 높고 그 이후로 급히 떨어지는 것을 볼 수 있다. SVM을 예측기로 사용해 출력값이 -1, 0, 1이다보니 30개월 이전은 클래스에 대한 데이터의 개수가 차이가 많이 나 예측할 수 없었는데 37개월 전은 클래스에 대한 데이터의 수가 적절하지 않아 하락했다가 37개월에 클래스에 대한 분류가 잘 이루어지고 이후 너무 많은 개월수를 이용하여 예측하다보니 성능이 떨어지는 것을 볼 수 있다.

# 투자자 전망 예측 2) LSTM

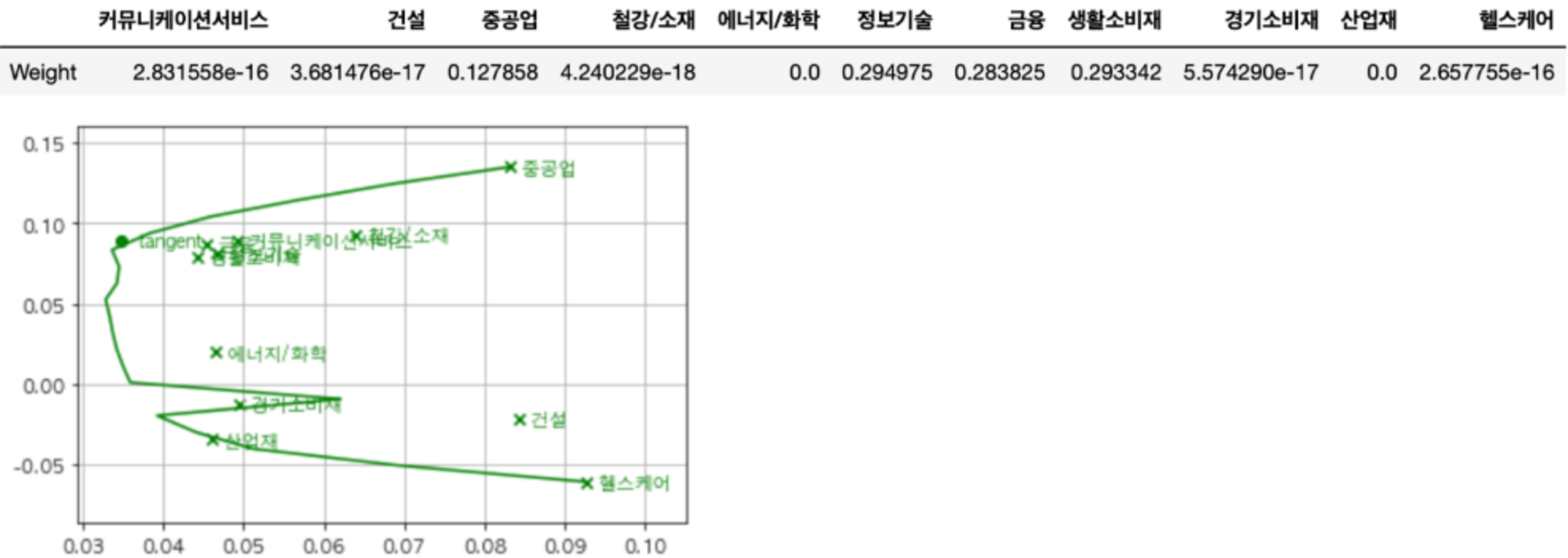
LSTM을 이용하여 투자자 전망을 예측할 때에는 수익률을 출력으로 예측된 수익률을 기준으로 상승, 하락, 유지로 -1, 0, 1로 나누어 행렬을 정하였다.

## ① window\_size를 6개월로 설정

	커뮤니케이션서비스	건설	중공업	철강/소재	에너지/화학	정보기술	금융	생활소비재	경기소비재	산업재	헬스케어
2019/1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2019/2	-1	1	-1	-1	1	-1	-1	-1	1	1	-1
2019/3	1	-1	1	1	-1	1	1	1	-1	-1	-1
2019/4	1	-1	1	-1	1	1	1	1	1	-1	1
2019/5	-1	0	1	-1	1	-1	-1	-1	1	1	-1
2019/6	1	-1	1	1	0	1	1	-1	-1	-1	-1
2019/7	1	1	-1	-1	1	1	1	1	1	-1	1
2019/8	-1	-1	1	-1	1	-1	-1	-1	1	1	-1
2019/9	1	-1	1	0	0	1	1	-1	-1	-1	-1
2019/10	1	1	-1	1	1	1	1	1	1	-	-
2019/11	-1	-1	1	-1	1	-1	-1	1	1	-	-
2019/12	1	-1	1	1	0	1	1	1	-1	-	-

P와 내재균형수익률을 이용해  
구한 포트폴리오 투자 비중

예측값을 담은 행렬 P





## ① window\_size를 6개월로 설정



LSTM을 이용해서 종가를 예측한 후 이를 상승, 유지, 하락으로 분류하여 투자자 행렬을 만들었다.

투자자 행렬을 이용해 얻은 투자자 전망과 내재균형수익률을 이용해 포트폴리오 투자 비중을 구한 결과  
2016년부터 2018년까지의 데이터를 학습 후 2019년은 이전 6개월치 종가를 이용해 다음 달의 종가를 예측했을 때  
2019년 **총 누적 수익률은 2.37%**가 나왔다. 또한 **샤프지수는 0.56**이 나왔다.

## 향 후 계 획

1. LSTM을 이용하여 SVM과 동일하게 window size별로 총수익률과 샤프지수의 변화 알아보기
  - 이 과정에서 오류가 나 문제를 해결 중에 있다
2. SVM, LSTM 이외에도 다른 예측기를 사용하여 포트폴리오를 구성해보기
3. 어떤 예측기가 가장 성능이 좋았고, 그 예측기를 사용했을 때 총수익률과 샤프지수는 어떤지 visualization

# 감사합니다!

---

