



Algorithm Trading

Week 1



The background of the slide is a blurred image of a financial market display. It features a grid of dashed lines in blue and red, with various data series represented by lines and bars. A bright light source in the top left corner creates a lens flare effect. Two white L-shaped corner brackets are positioned on the left and top right sides of the slide.

Algorithm Trading

What is Algorithm Trading?

- 컴퓨터를 사용하여 인간(트레이더)이 할 수 없는 속도와 주기로 수익을 내도록 거래를 하는 일련의 명령들을 수행하는 프로그램
- 정의된 규칙들은 타이밍, 가격, 수량 혹은 다른 수학적 모델에 기초를 두고 있음
- 트레이더에게 수익을 낼 기회를 줄 뿐 아니라, 시장을 좀 더 유동적으로 만들고 트레이딩 활동에서 인간의 감정적인 영향을 없애 좀 더 체계적으로 거래를 할 수 있음
- Reference : '알고리즘 트레이딩의 기초 : 개념과 예시'

What is Algorithm Trading?

트레이더가 다음의 단순한 거래 규칙을 따른다고 가정합시다:

- 50일 이동평균선이 200일 이동평균선 **위로 올라가면** 주식 50주를 **매수**한다.
- 50일 이동평균선이 200일 이동평균선 **아래로 내려가면** 주식 50주를 **매도**한다.

이 두 가지의 간단한 규칙을 사용하면, 주식 가격(이동평균선 포함)을 자동으로 모니터링하고 매수 및 매도 주문을 넣는 컴퓨터 프로그램을 작성하는 것은 그리 어렵지 않겠죠. 트레이더는 더이상 실시간 가격과 그래프를 계속 보면서 직접 주문을 할 필요가 없습니다. 알고리즘 트레이딩 시스템은 자동으로 거래 기회를 포착하고 주문을 넣어주죠.



Stock Price Data

Backtesting

Alternative Data

Risk Management

Companies to Predict

Weight Agnostic NN

Timeline

내용	참고	실명의 팀원	2019.10.31(목)	2019.11.8(금)	2019.11.14(목)	2019.11.21(목)	2019.11.28(목)	2019.12.5(목)
데이터 수집 (대체 데이터)	에어코리아 + 빅콘 팀한테 데이터 구걸							
업종 선정	회사 리스트 뽑기 + 나름의 이유	이은진, 김민석						
데이터 수집(추가 데이터 인프라)	회사 리스트 정해져야 가능	이세희, 김미라, 유현우						
선정된 업종 검증	뽑힌 회사 리스트로 EDA	이은진, 김민석						
Backtesting Metr	패키지(zipline) 사용방법, 여러 method 정리해서 공유	유현우, 양수형, 이세희, 김민석						
Risk Managemer	여러 방법, 개념 정리해서 공유	현예성, 유현우						
WANN + 문제 정의	논문읽기, 선행논문 있으면 간단하게 정리, 깃허브 한번 보고 튜토리얼까지?	이세희, 김미라, 양수형, 유현우						
RL	보류							
모델 검증	백테스팅							
...								

Stock Price Data, Alternative Data

1. 주가 데이터 수집

- [pykrx](#): 주가 받아오는 패키지 (일단위 주가)
- 증권사API (분단위 주가)
- 직접 크롤링 (일단위 주가): [GitHub](#)

2. 대체 데이터 수집

- 대기오염 수치 데이터 (에어코리아)

측정자료(수치)

날짜 (월-일-시)	PM ₁₀ (μg/m ³)		PM _{2.5} (μg/m ³)		오존 (ppm)		이산화질소 (ppm)		일산화탄소 (ppm)		이황산화스 (ppm)	
	1시간	1시간	1시간	1시간	1시간	1시간	1시간	1시간	1시간	1시간	1시간	1시간
10-31-13	●	48	●	19	●	0.031	●	0.035	●	0.5	●	0.005
10-31-12	●	36	●	16	●	0.013	●	0.049	●	0.6	●	0.005
10-31-11	●	38	●	19	●	0.006	●	0.050	●	0.7	●	0.004
10-31-10	●	35	●	16	●	0.004	●	0.051	●	0.8	●	0.004
10-31-09	●	33	●	14	●	0.003	●	0.049	●	0.7	●	0.004
10-31-08	●	30	●	14	●	0.002	●	0.047	●	0.7	●	0.004
10-31-07	●	28	●	15	●	0.002	●	0.045	●	0.6	●	0.003
10-31-06	●	27	●	14	●	0.002	●	0.044	●	0.6	●	0.003
10-31-05	●	27	●	14	●	0.002	●	0.046	●	0.6	●	0.003
10-31-04	●	29	●	14	●	0.002	●	0.049	●	0.8	●	0.004
10-31-03	●	28	●	13	●	0.001	●	0.053	●	0.8	●	0.003
10-31-02	●	29	●	14	●	0.002	●	0.053	●	0.7	●	0.003
10-31-01	●	30	●	14	●	0.002	●	0.056	●	0.7	●	0.003

PyKrx

본 모듈은 다양한 웹 사이트 ([Naver](#) / [KRX](#))에서 주가 정보를 스크래핑(scraping) 합니다. API 사용에 대한 책임은 사용자 본인에게 있습니다. 도의적으로도 무분별한 API 호출은 자제해 주시기 부탁드립니다. 또한 결과물은 참고용으로만 사용해야하며, 투자에 대한 책임은 사용자에게 있습니다.

1 환경설정

1.1 Installation

pykrx는 pip를 사용해서 쉽게 설치할 수 있습니다.

```
pip install pykrx
```

1.2 Import

pykrx는 유가 증권시장의 주식 정보를 스크래핑 합니다. 모듈의 사용은 import부터 시작합니다.

```
from pykrx import stock
```

2. API 설명

2.1 Stock 모듈

2.1.1 MarketData API

코스피 종목의 주가 관련 정보를 얻는 API 입니다.

2.1.1.1 Ticker

Companies to Predict

종목	관련주	이유
전기차	코스모신소재, 예코프로, 일진머티리얼즈, 포스코캠텍, 대주전자재료, 솔브레인, 씨아이에스, 후성, 토타, SK이노베이션, LG화학, 삼성SDI	미세먼지 저감조치에 따른 석유 및 석탄 등의 화력발전 비용 감소와 신재생에너지의 사용량 증가
제약	휴비츠, JW중외제약, 디에이치피코리아, 안국약품, 삼일제약, 보령제약, 조아제약, 국제약품, 삼천당제약, 이연제약, 케이피엠텍, 바이넥스, 휴온스	대기오염은 안구질환 및 호흡기질환을 유발하며, 발암물질을 포함하고 있음. 따라서 제약 산업은 대기오염과 관련이 있다.
마스크	오공, 모나리자, 웰크론, 케이엠, 카스, 동아쏘시오홀딩스	대기오염, 특히 미세먼지를 막기 위해 일반인들이 가장 많이 구매하는 제품 중 하나가 마스크임.
대기환경설비	나노, KC그린홀딩스, 하츠, 코스코ICT, 한국테크놀로지, 비디아이, 누리플랜, 에어릭스, 선도전기, 예코프로	대기오염이 늘어남에 따라 대기환경을 정화할 수 있는 설비에 대한 관심도 높아짐.
공기청정기	LG전자, 대유위니아, 위닉스, 코웨이, 오텍, 신일산업, 쿠쿠홈시스, 쿠쿠홀딩스, 하이마트, 성호전자, 에스피지, 크린앤사이언스, 성장오토텍	실내 공기를 정화하기 위해 일반인들의 구매가 늘어난 제품 중 하나.
천연가스	한국가스공사, 지역난방공사, 대강솔루션, 동성화인텍, 삼강엠엔티, 삼천리	화력발전에 대한 규제가 심해지고, 원자력은 탈원전의 영향으로 축소되는 시점에서, 현실적인 화력발전의 대안책으로 각광받는 자원
헬스케어	소프트센, 인피니트헬스케어, 솔고바이오	미세먼지가 각종 질병의 원인이 될 수 있다는 것이 과학적으로 밝혀지고 있음과 동시에, 헬스케어 산업에 대한 관심 또한 증가
여과설비	KC코트렐, 포스코ICT, S&TC 두산중공업, 한국전력기술	미세먼지 저감대책 일환으로 공장 미세먼지 배출 허용치 규제가 강화되는 시점에서, 여과설비에 대한 관심이 증가 중임.
수소차	뉴인텍, 이엠코리아, 미코, 삼화콘덴서, 세종공업, 제이엔케이히터, 유니크, 효성중공업, 삼화전자, 대명코퍼레이, 에스퓨얼셀	미세먼지 저감조치에 따른 석유 및 석탄 등의 화력발전 비용 감소와 신재생에너지의 사용량 증가, 수소 관련 정부정책 또한 시행되고 있다

선정 회사 리스트

공기청정기	071840	하이마트
공기청정기	045520	크린앤사이언스
공기청정기	284740	쿠쿠홈시스
공기청정기	192400	쿠쿠홀딩스
공기청정기	044340	위닉스
공기청정기	071460	위니아딤채
공기청정기	021240	웅진코웨이
공기청정기	067170	오텍
공기청정기	058610	에스피지
공기청정기	002700	신일사업
공기청정기	043260	성호전자
공기청정기	080470	성창오토택
공기청정기	066570	LG전자
대기환경설비	053590	한국테크놀로지
대기환경설비	066130	하츠
대기환경설비	086520	에코프로
대기환경설비	007610	선도전기
대기환경설비	148140	비디아이
대기환경설비	069140	누리플랜

대기환경설비	187790	나노
대기환경설비	009440	KC그린홀딩스
마스크	042040	케이피엠테크
마스크	016920	카스
마스크	000640	동아쏘시오홀딩스
수소차	298040	효성중공업
수소차	126880	제이엔케이히터
수소차	095190	이엠코리아
수소차	011320	유니크
수소차	288620	에스퓨얼셀
수소차	033530	세종공업
수소차	011230	삼화전자
수소차	059090	미코
수소차	007720	대명코퍼레이
수소차	012340	뉴인텍
여과설비	052690	한국전력기술
여과설비	022100	포스코ICT
여과설비	034020	두산중공업
여과설비	100840	S&TC
여과설비	119650	KC코트렐
전기차	093370	후성
전기차	009520	포스코앰텍
전기차	108230	툭텍
전기차	005070	코스모신소재

선정 회사 리스트

전기차	020150	일진머터리얼즈
전기차	086520	에코프로
전기차	222080	씨아이에스
전기차	036830	솔브레인
전기차	089980	상아프론테크
전기차	001820	삼화콘덴서
전기차	006400	삼성SDI
전기차	078600	대주전자재료
전기차	096770	SK이노베이션
전기차	051910	LG화학
전기차	066970	엘앤에프
제약	243070	휴온스
제약	065510	휴비츠
제약	083550	케이엠
제약	034940	조아제약
제약	102460	이연제약
제약	065950	웰크론
제약	045060	오공
제약	001540	안국약품
제약	000250	삼천당제약
제약	000520	삼일제약
제약	003850	보령제약
제약	053030	바이넥스
제약	012690	모나리자

제약	131030	디에이치피코리아
제약	002720	국제약품
제약	001060	JW중외제약
천연가스	036460	한국가스공사
천연가스	071320	지역난방공사
천연가스	004690	삼천리
천연가스	100090	삼강엠앤티
천연가스	033500	동성화인텍
천연가스	096350	대창솔루션
헬스케어	071200	인피니트헬스케어
헬스케어	043100	솔고바이오
헬스케어	032680	소프트센

Backtesting Method

트레이딩 알고리즘 검증

트레이딩 알고리즘 검증을 위해 활용할 수 있는 지표

- Profit/Loss
- Hit Ratio
- Drawdown
- Sharpe Ratio
- ...

[Backtesting.py](#)

- backtesting을 위한 python package

PyAlgoTrade

- 마지막 버전이 2년 전이고 개발 중단

Backtrader

- ask and bid prices만 사용하는 거래 시뮬레이션

Backtesting Method

What is Backtesting?

- 새로운 전략개발 또는 기존 전략의 검증에 위해 과거 데이터를 가지고, 실험하고자 하는 전략의 성과를 측정해 보는 과정이다.
- 이 투자 전략을 과거에 똑같이 적용했다면 성과가 어땠을까?

1. Profit / Loss

- 개발한 시스템에 특정 기간의 데이터를 입력해 trading을 실시한다. 그때 발생하게 되는 Profit과 Loss를 평가한다.



- <https://programmingfbf7290.tistory.com/entry/6-%EC%95%8C%EA%B3%A0%EB%A6%AC%EC%A6%98-%EC%A0%84-%ED%9B%84-%EC%88%98%EC%9D%B5%EC%9C%A8-%ED%85%8C%EC%8A%A4%ED%8A%B8-%ED%82%A>

2. Hit Ratio(적중률)

- 모델을 이용해 예측한 결과의 정확도를 의미한다.
- Hit Ratio = 수익횟수/전체거래횟수
- 즉, 모델에 따라 거래를 했을 때 전체거래에서 수익을 발생시킨 비율.
- If Hit Ratio=1, 모든 거래에서 수익을 발생시켰다. vs Hit Ratio = 0.4 , 10번의 거래 중 4번 수익을 냈다.

model A : 수익률 15%, 적중률 10%.

model B : 수익률 9%, 적중률 55%

=> 이 경우 적중률이 높은 B 모델이 안정적이고 예측가능 하다.

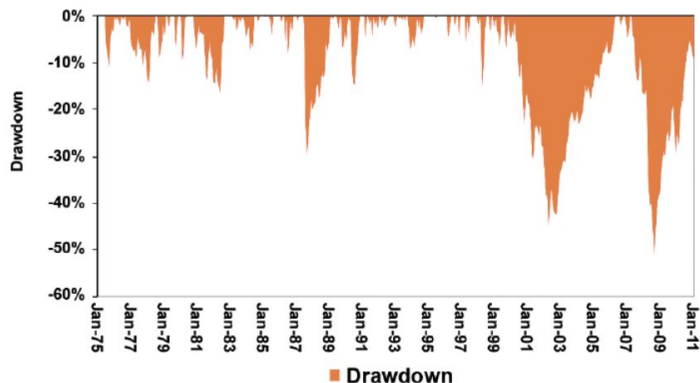
- 따라서 모델의 성능을 평가하는 데에 단순 Profit(수익률)보다 효과적이다.
- 뿐만 아니라 손실이 지속된다면 투자모델을 바꾸어야 하는지, 일시적 현상인지 고민을 해야 하는데, 만약 적중률이 높은 모델이라면 상대적으로 높은 신뢰도를 가지고 있기 때문에 유지할 수 있다. 하지만 그렇지 않은 모델의 경우 다시 투자모델을 생성해야 할지도.
- 과거 데이터를 이용하기 때문에 머신러닝 모델의 경우 학습에 사용되었던 데이터와 테스트에 사용한 데이터가 중복되지 않도록 주의한다.

Backtesting Method

3. Drawdown

- Drawdown(손실률)은 상품(investment), 펀드 또는 선물에서 특정 기간동안 발생한 시세 고점에서 저점까지의 하락을 의미한다.
- 잘못된 예측이 얼마나 되는지, 잘못된 예측의 기간이 얼마인지를 알려준다.
- 따라서, 금융 리스크와 파산의 위험성에 대한 정보를 제공한다.

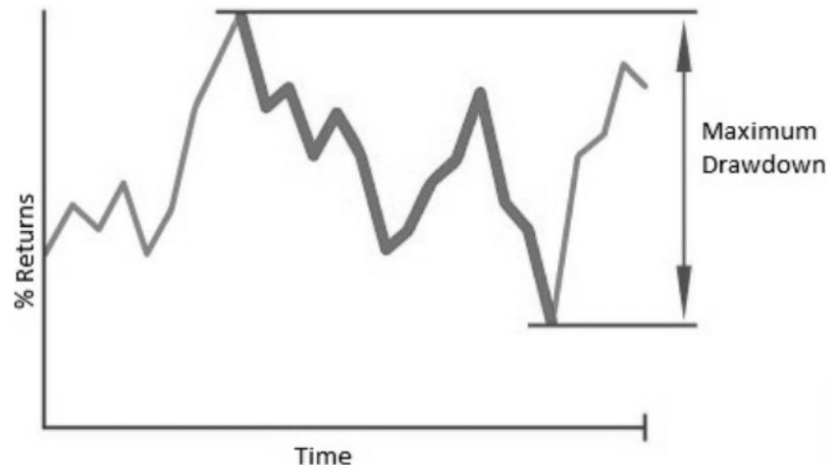
Figure 17: Underwater Chart



- 최대 손실률(Maximum Drawdown), 손실 지속 기간(Drawdown duration), 손실 경향(Drawdown Curve) 등을 알 수 있다.

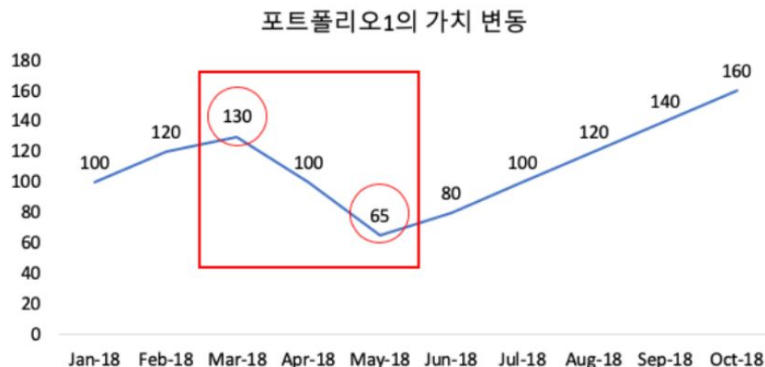
- MDD(Maximum Drawdown) : 시세가 새 고점에 도달하기 전에 그 이전 고점에서 저점까지의 최대 손실을 의미한다. 계산 방식은 다음과 같다.

$$(\text{저점값} - \text{고점값}) / \text{고점값} * 100 \%$$

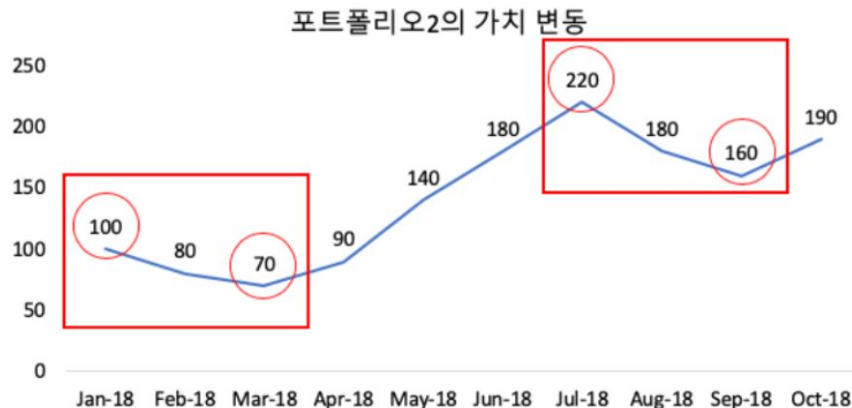


Backtesting Method

- Understanding MDD



포트폴리오 1에서는 drawdown(낙폭)이 1번 발생했다. 따라서 MDD는 $(65-130)/130 = -50\%$ 로 계산된다.



포트폴리오 2에서는 Drawdown(낙폭)이 2번 발생했다. 이 두개의 $(70-100)/100 = -30\%$ 와 $(160-220)/220 = -27\%$ 중 MDD는 둘 중 큰 값인 -30% 가 된다.

- 수익률과 더불어 알고리즘 트레이딩을 지속해야 할 지 말지를 결정하는 데에 중요한 역할을 한다.

Backtesting Method

4. Sharpe Ratio

- 위험을 고려한 성능을 나타내는 수치이다.
- Sharpe Ratio란 위험 자산에 투자함으로써 얻은 초과 수익의 정도를 나타내는 지표이다.
- Sharpe Ratio = 초과수익률/초과수익률의 표준편차 (1994년 기준)

$$S_a = \frac{E[R_a - R_b]}{\sigma_a} = \frac{E[R_a - R_b]}{\sqrt{\text{var}[R_a - R_b]}}$$

분자 : 기준지표 수익률 대비 자산의 초과수익률

분모 : 자산수익률의 표준편차

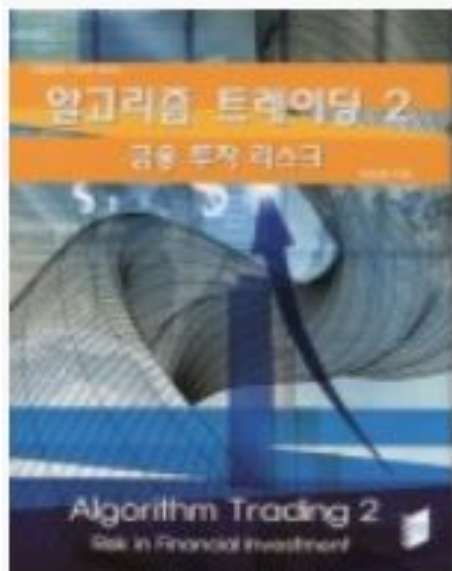
초과수익률은 펀드 수익률에서 무위험 수익률(국공채 수익률)을 차감하여 계산한다.

주식형 펀드의 경우 종합주가지수 또는 주식 수익률을 기준 수익률로 삼는다.

Backtesting Method

- Sharpe Ratio가 높으면 감수한 위험 대비 수익이 좋다는 의미이기 때문에 Sharpe Ratio가 높은 펀드가 좋은 펀드라고 볼 수 있다.
- 일반적인 펀드 영역에서는 Sharpe Ratio가 1 정도면 그럭저럭 괜찮은 펀드, 2정도면 쓸만한 펀드, 3 이상이면 아주 괜찮은 펀드로 볼 수 있다.
- 주식형 펀드는 괜찮은 성능을 내고 있지만 Sharpe Ratio가 1에 미치지 못하는 경우가 많다. 이는 주식형 펀드의 경우 기준으로 삼는 수익률은 종합주가지수 또는 주식 수익률이지만 샤프 지수 계산 식에서 초과 수익률을 계산할 때는 국공채 수익률을 이용하는 것과 관련되어있다.
- 따라서 Sharpe Ratio를 해석할 때 1보다 크거나 작다를 따지기 보다는 다른 펀드와의 상대적인 비교 용도로 사용하는 것이 좋다. 즉, 두 자산을 공동의 기준지표와 비교할 경우, 더 높은 Sharpe Ratio를 나타내는 자산이 동일 위험에 대해 더 높은 수익률을 제공한다.
- 주의할 점은, 샤프 비율은 모든 변동성을 동일하게 취급한다. 즉, 상승편차(upside volatility)를 갖는 (높은 수익률을 갖는) 전략에 불이익을 주게 된다.

Risk Management



알고리즘 트레이딩. 2 : 금융 투자 리스크 / 장순용 지음

Risk Management

리스크의 척도

1. 표준편차 : 리스크의 주된 척도

Risk Management

리스크의 척도

2.VaR(Value As Risk) : 어느 확률 안에서 발생할 수 있는 손실의 최저 한도

$$\text{VaR}(\alpha) = -(\text{현재의 자산가}) \times Q(\alpha)$$

$Q(\alpha)$: quantile function

hyper parameter : $\Delta t, \alpha$

Δt : 데이터의 시간주기에 이미 내포됨

α : 손실의 확률

Risk Management

ex) $\text{VaR}(\alpha=0.05) = 100$

-> 5%의 확률로 100이상의 손실을 볼 수 있다.

단점

- 1. 손실의 기댓값을 계산한 것이 아님 -> ES
- 2. Δt 보다 긴 시간의 손실을 나타낼 수 없음 -> 누적손실

Risk Management

리스크의 척도

3.ES(Expected Shortfall) : 손실의 기댓값

$$ES(\alpha) = \frac{\int_0^\alpha VaR(x) dx}{\alpha}$$

Risk Management

리스크의 척도

4.누적손실 (accumulated loss) : $\Delta t = m \times \Delta t$ 를 기반으로 VaR과 ES계산

$$\text{VaR}'(\alpha) = m^{1/n} \text{VaR}(\alpha)$$

$$\text{ES}'(\alpha) = m^{1/n} \text{ES}(\alpha)$$

Risk Management

리스크 관리의 원리

단일지표모델 (선형모델)

$$Y_i - Y_0 = \alpha_i + \beta_i(Y_M(t) - Y_0) + \varepsilon_i(t)$$

$Y_i(t)$: 특정 금융자산의 수익률(종속변수)

Y_M : 시장의 수익률(독립변수)

Y_0 : 리스크 없는 자산의 수익률

Risk Management

선형 회귀법을 적용해서 α, β 와 전체 수익률의 표준편차를 구할 수 있다.

$\alpha_i < 0$: 리스크 노출을 감안할 때 수익률 미달

$\alpha_i = 0$: 리스크 노출을 감안할 때 수익률 적절

$\alpha_i > 0$: 리스크 노출을 감안할 때 수익률 초과 달성

$$\beta_i = \frac{\text{cov}(Y_i, Y_M)}{\text{Var}(Y_M)}$$

$\beta_i < 1$: 시장 변동에 둔감하다.

$\beta_i = 1$: 시장과 동일한 변동성을 갖는다.

$\beta_i > 1$: 시장 변동에 민감하다.

$$\sigma_p \cong \left| \frac{\sum_{i=1}^N \beta_i}{N} \right| \sigma_M$$

Risk Management

리스크 관리방법(리스크의 종류에 따라 다름)

1. 개별자산의 고유 리스크 -> 포트폴리오의 다각화

2. 업종에 적용되는 리스크 -> 같은 업종에 속하는 다른 자산과의 공분산이 작은 자산을 선택

3. 시장전체에 적용되는 리스크 -> 베타계수들이 낮은 자산을 중심으로 투자

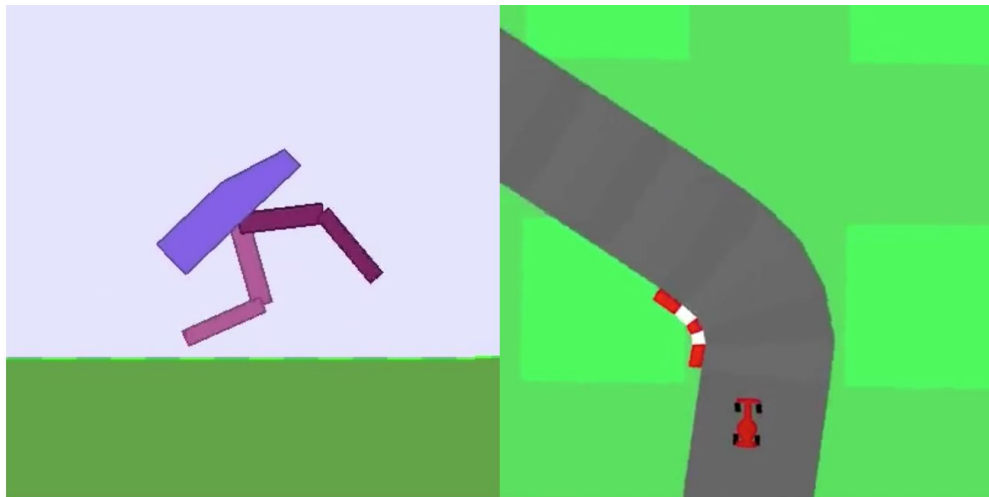
Weight Agnostic Neural Network (WANN)

[Introduction]

- 랜덤으로 weight가 초기화 되어 있어도 특정한 task를 수행할 수 있는 뉴럴넷 구조를 찾는 방법론
- 동물이 태어나자마자 걸을 수 있다는 것에서 영감을 받은 연구

[Idea]

- 가중치를 학습하지 않고 한번 랜덤하게 설정한 후에, 네트워크 아키텍처를 탐색함으로써 좋은 퍼포먼스를 내고자 하는 것



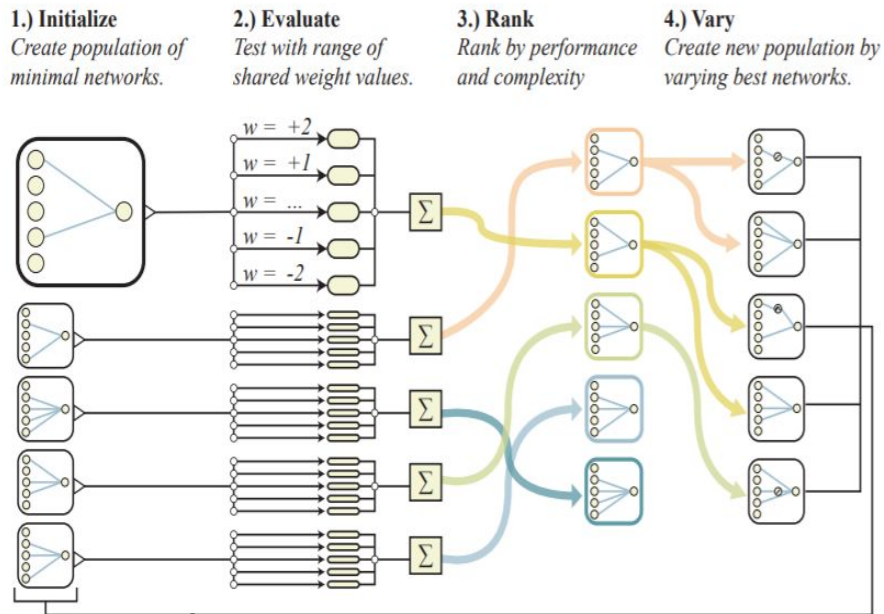
<https://arxiv.org/pdf/1906.04358.pdf>

WANN

- weight training이 아닌 **weight random sampling**
- 퍼포먼스는 weight와는 무관하게 **네트워크의 형태(topology)**에 따라 결정됨
- 아키텍처 스스로 문제를 푸는 과정을 좀 더 효과적으로 만들기 위해, weight를 최소화하고자 → **하나의 공유된 weight** → computational cost 감소 → 계산이 더욱 빨라짐.
- 즉, weight training을 통해 optimal한 **weight**를 찾는 과정이 **costly**하다는 문제점을 개선하고자 함.
- 아키텍처가 weight에 관대하도록 진화하게 됨.
- WANN에서 **simple**의 정의
 - soft-weight sharing을 통해 모델을 단순화
 - weight의 정보량을 감소
 - weight의 search space를 단순화한다.

WANN

1. 작은 단위의 뉴럴 네트워크를 형성.
2. 각 네트워크가 서로 다르지만 공유된 weight값들로 테스트됨
3. 퍼포먼스와 모델의 복잡도에 따라 네트워크들의 순위가 매겨짐
4. 가장 높은 랭크를 차지한 네트워크 모형이 변형되어 새로운 뉴럴 네트워크가 만들어짐
5. 2번부터의 과정 반복
6. test 결과의 양상들을 통해 퍼포먼스 향상

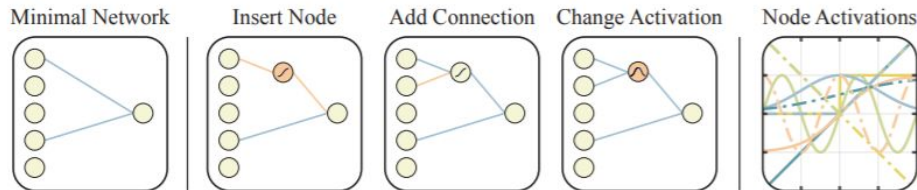


<https://arxiv.org/pdf/1906.04358.pdf>

WANN

[Topology Search]

- 초기 population은 네트워크 연결 적음.
- hidden node의 갯수 적으며, input과 output 사이의 관계의 '일부'만이 드러남.
- 앞선 process를 거치며 3가지 operator로 네트워크를 조금씩 변형 → 새로운 population 형성 → 앞선 프로세스 반복.
- 3 operators: insert node, Add connection, change activation



- 1) **insert node:** 새로운 hidden node를 지나도록 connection을 2개로 나눈다. new node에서 활성화 함수는 random하게 할당된다.
- 2) **add connection:** new connection이 네트워크의 feed forward를 통해 이전에 연결되지 않았던 node와 hidden node의 연결 형성.
- 3) **change activation:** hidden node의 활성화 함수를 다시 random하게 할당된다.

<https://arxiv.org/pdf/1906.04358.pdf>