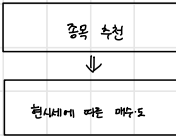


[전반적인 Trading Model]



→ 총 3개의 Step 존재 : 1) 사용자가 원하는 수익률·인입성 지표 (재무 데이터 기반)

↳ 각별 대비 통해 수익률, 전분기 대비 현황기 수익률, ROE, PER, PBR, 동종업계 PER, (외국인 호의율)

→ 시장 데이터에 적합하지 않나?

↳ 근미래에 어떤 현상이 필요 가능성

· 종목추천 알고리즘의 경우,

ML 모델 사용 예정 + back testing...

· 시세에 따른 매수도 타이밍은

강화학습 모델 사용...??

2) 시장데이터별로 다양한 분류별 그룹 : 이렇게 되면 원가 반영될 일수록 그룹 가능성: → 큰 권위 데이터 필요

↳ WTI(국제 원유 시세), WTI-rate (전월 대비 상승률), gold(국제 금 시세), copper(구리), dollar

: 우리가 결국 하고자 하는 바는 3개월 수익률 예측 → '기간: 3개월'로, 연할 ↑ 종목을까지 하나의 Group 형성?

3) 3개월 단위 수익률과 연관시켜서 상위 종목 추천 & 현시점에서의 3개월 수익률 예측

↳ 5월(1주), 20일(한달), 65일(한분기) 기준 평균거래량 대비 상승률, 수익률(종가 대비)

: 여가전 어정도 데이터면 충분하냐? 만약 상승추세를 넣고 싶으면, 차트 데이터 필요 (RSI, CCI...)

* PCA → 시장데이터 PCA 처리는 어떻게 (시장 데이터 영향 수 / ↓ 그룹 가능할 수도...?)

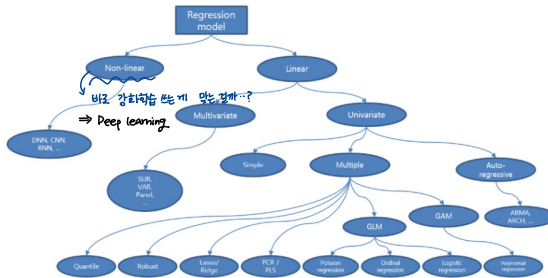
⇒ 동일일별 수익률도 feature로 넣으면 비슷한 성향끼리 되지 않겠나...?

* K-Means → 비파괴학습 군집화 방법

⇒ 이것도 지도학습으로 하면 그냥 하나의 조건수에 될 것 같은 개념이구나...

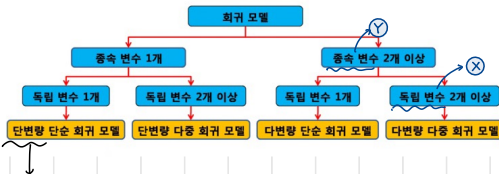
- 선택/비선택 → 독립 변수와 종속 변수의 관계 고려

(학계어문) 주석거절 (로그노일), 주석분포 (정규분포)



⇒ Peer learning

비교 강화학습은 이게 맞는 걸까...?



⇒ 그림 우리가 예측하고자 하는 건 종속변수니까

하서라 종속변수 + 여러 개의 독립변수인건...

⇒ 예시면, 예쁜 연인 DNN도 괜찮지 않을까...?

Linear regression model

: 오차항 평균 0, 정규분포 데이터

: 데이터에 아웃라이어 없음

: 독립변수·종속변수 (선형관계)

: 독립변수와 오차항은 서로 독립

: 오차항은 저기 상관성 없음

: 독립변수 간 서로 독립

over-fitting 가능성 ↑

- 독립 변수와 종속 변수가 선형 관계가 아닌 경우: Polynomial regression, Generalized Additive Model (GAM)
- 오차항의 확률분포가 정규분포가 아닌 경우: Generalized Linear Model (GLM)
- 오차항에 자기 상관성이 있는 경우: Auto-regression
- 데이터에 아웃라이어가 있는 경우: Robust regression, Quantile regression
- 독립변수 간에 상관성이 있는 경우(다중공선성): Ridge regression, Lasso regression, Elastic Net regression, Principal Component Regression (PCR), Partial Least Square (PLS) regression

overfitting 경감 ↓

- 독립 변수와 종속 변수가 선형 관계가 아닌 경우: Polynomial regression, Generalized Additive Model (GAM)
- 오차항의 확률분포가 정규분포가 아닌 경우: Generalized Linear Model (GLM)
- 오차항에 자기 상관성이 있는 경우: **Auto-regression**
- 데이터에 아웃라이어(이상치)가 있는 경우: **Robust regression**, Quantile regression \rightarrow Ridge + Rasso
- 독립변수 간에 상관성이 있는 경우(다중공선성): Ridge regression, **Lasso regression**, **Elastic Net regression**, Principal Component Regression (PCR), Partial Least Square (PLS) regression
 \rightarrow bias, covariance 문제 shrinkage parameter로 문제 해결 (즉 회귀 계수 축소)

⊕ **Stepwise regression** : 단계별 회귀분석, 한번에 하나씩 공변량 추가/삭제

- **forward selection** : 모델에서 가장 중요한 변수 시작 \rightarrow 단계별 변수 추가

- **Backward elimination** : 모든 변수 적용 후, 각 단계 중요 변수 제거 \Rightarrow 회귀 계수 예측 변수 예측 효율 최대화

* **XGBoost** : 선형 알고리즘 + 트리 알고리즘 \rightarrow 분류 회귀 모두 가능!

* **LightGBM** : 트리 기반 알고리즘만 사용
성능이 좋은 편이라 논문에서도 많이 쓰는 방법론