

과제 1. 엔터주(와이지엔터테인먼트)로 구현하기

→ YG엔터 관련 시계열 feature 적용 + feature selection 거쳐서 모델 투입까지 구현함.

과제 2. 구글 Colab으로 코드 실행하기 run.ipynb

→ 코드가 실행은 되나 gpu 적용이 안 되는 것 같음.

<<< 과제 1. 엔터주(와이지엔터테인먼트)로 구현하기 >>>

: 지난주 In-Out 데이터와 거의 동일하나, feature의 변경이 있음.

1) 종목코드.csv (122870.csv)

차트 데이터 + 코스닥 YG 엔터와 관련된 주가지수 및 투자지표

※ 차트 데이터

지표명	표현
년/월/일	date
시가	open
고가	high
저가	low
종가	close
거래량	volume
거래대금	value
시가총액	stock_value
상장주식수	stock_volume

※ 투자지표

지표명	표현
주당순이익	eps
주가수익비율	per
주당 순자산가치	bps
주가순자산비율	pbr
주당배당금	dividend_per_stock
배당수익률	dividend_yeild_ratio
기관_매수량	volume_inst_buy
기관_매도량	volume_inst_sell
기관_순매수량	volume_inst_pure_buy
외국인_매수량	volume_fore_buy
외국인_매도량	volume_fore_sell
외국인_순매수량	volume_fore_pure_buy
기관_매수대금	value_inst_buy
기관_매도대금	value_inst_sell
기관_순매수대금	value_inst_pure_buy
외국인_매수대금	value_fore_buy
외국인_매도대금	value_fore_sell
외국인_순매수대금	value_fore_pure_buy

※ 주가지수 및 주가지수 관련 투자지표

구분	지수명(지표명)	표현
종합지수	코스닥	kosdaq_xxx
	(종가)	kosdaq_close
	(시가)	kosdaq_open
	(고가)	kosdaq_high
	(저가)	kosdaq_low
	(거래량)	kosdaq_volume
	(거래대금)	kosdaq_value
	(시가총액)	kosdaq_stock_value
	(배당수익률)	kosdaq_dividend_yield_ratio
	(주가수익비율)	kosdaq_per
	(주가순자산비율)	kosdaq_pbr
대표지수	코스닥 150	kosdaq150_xxx, (이하 동일)
섹터지수	코스닥 150 커뮤니케이션	kosdaq150_comm_xxx, (이하 동일)
산업별지수	오락,문화	kosdaq_enter_xxx, (이하 동일)
시가총액 규모별 지수	코스닥 대형주	kosdaq_large_xxx, (이하 동일)
소속부 지수	코스닥 우량기업부	kosdaq_super_xxx, (이하 동일)

2) 신경망 Input

전처리 통해 생성 (차트 데이터, 주가지수, 투자지표, 보조지표) + 에이전트 상태 (주식 보유 비율(ratio_hold), 포트폴리오 가치 비율(ratio_portfolio_value))

※ 보조지표

지표명	파라미터 값	표현
이평선 (이동평균, 지수평균, 가중평균)	5, 10, 20, 30, 60, 120	종가, 거래량, 기관 및 외국인 거래량에 적용 ma_xxx, ema_xxx, wma_xxx
볼린저밴드	20, 2	ubb, mbb, lbb
MACD (이동평균수렴확산)	12, 26	macd, macdsignal9, macdhist
RSI (상대강도지수)	14	rsi
스토캐스틱	5, 3 14, 5	slowk, slowd, fastk, fastd fastk_rsi, fastd_rsi
CCI (Commodity Channel Index)	14	cci
Williams'%R	14	willR
parabolic SAR		sar
ADX (Average Directional Movement Index)	14	adx
plusDI (Plus Directional Indicator)	14	plus_di
plusDM (Plus Directional Movement)	14	plus_dm
ATR (Average True Range)	14	atr
OBV (On Balance Volume)		obv
Variance	5, 1	var
Three Line Strike		line_str
Three Black Crows		blk_crw
Evening Star		evn_star
Abandoned Baby		abn_baby

train 인자

```
: --stock_code 122870 --rl_method a2c --net lstm --learning --num_steps 5  
--output_name c_122870 --num_epochs 100 --lr 0.001 --start_epsilon 1 --discount_factor  
0.9 --start_date 20180101 --end_date 20200630
```

test 인자

```
: --stock_code 122870 --rl_method a2c --net lstm --num_steps 5 --output_name  
c_122870 --num_epochs 1 --lr 0.001 --start_epsilon 0 --discount_factor 0.9  
--value_network_name a2c_lstm_value_c_122870 --policy_network_name  
a2c_lstm_policy_c_122870 --start_date 20180701 --end_date 20200918
```

3) Feature Selection Model

: 트리 기반 Regression/Classification 모델(Decision Tree, Random Forest, LightGBM)을 구축해서 feature Importance를 뽑아냄

(1) ta_practice.ipynb : Regression 모델

- 평가지표로 SMAPE 사용

(타겟 값이 0에 가까울 경우, 오차 값이 지나치게 커지는 것을 방지한 지표.

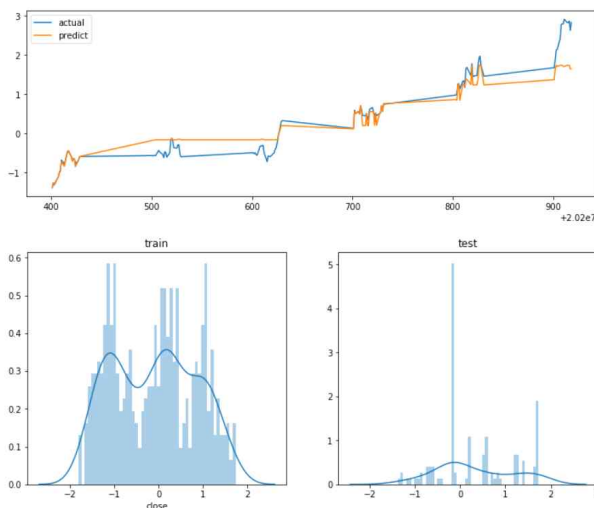
신경망 입력 데이터를 그대로 사용해서 변수가 스케일링 되어 있음.)

- ㄱ) KFold CV : 10개, train(2018.01.01.~2019.03.31)

- ㄴ) train_test_split : train(2018.06.29~2020.03.31), test(2020.04.01~2020.09.18)

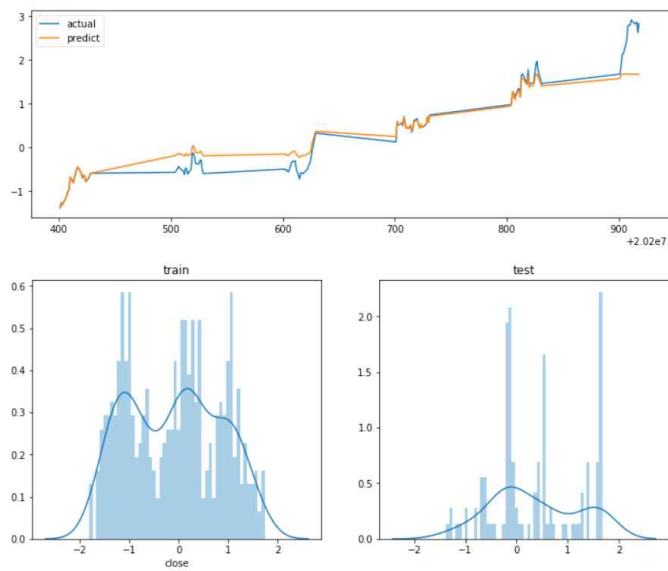
※ Decision Tree Score : val_SMAPE= 6.00146, test_SMAPE= 40.70313

- Test set : Actual vs Predict



※ Random Forest Score : val_SMAPE= 3.46538, test_SMAPE= 44.61635

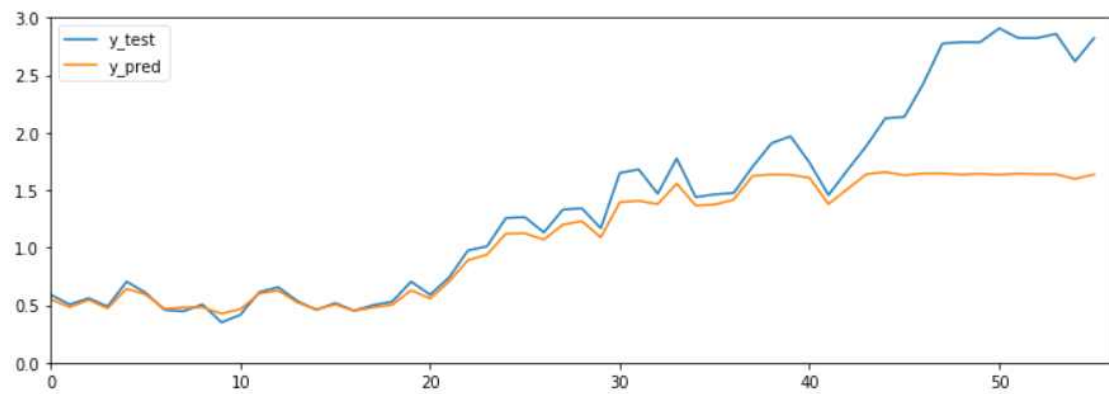
- Test set : Actual vs Predict



- Lightgbm train(80%) val(10%) test(10%)

※ Lightgbm Score : val_SMAPE= 59.99522833729, test_SMAPE= 16.070088410961063

- Test set : Actual vs Predict

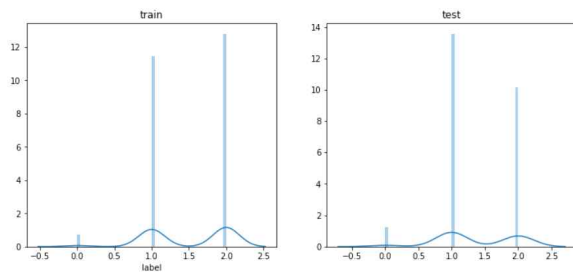


(2) ta_practice2.ipynb : Classification 모델

- 전일 대비 종가가 동일하면 0, 상승하면 1, 하락하면 2로 라벨링 진행
- 평가지표로 Accuracy 사용 (이하 동일)

※ Decision Tree Score : cross_val_acc= 80.80338, test_acc= 79.66102

- Test set : Actual vs Predict



※ Random Forest Score : cross_val_acc= 88.19239, test_acc= 88.98305

※ Lightgbm Score : val_acc= 89.09091, test_acc= 87.27273

confusion matrix :

```
[[ 0  1  1]
 [ 0 32  3]
 [ 0  2 16]]
```

(3) feature importance 기준으로, 강화학습 신경망 입력값으로 feature 선택함

- 변수간에 모든 corr는 고려하기가 힘들다 (너무 많음)
- Target과 상관계수가 낮은 피쳐만 제외시킴

case 1) 회귀모델에서 높은 중요도 피쳐만 사용

case 2) 분류모델에서 높은 중요도 피쳐만 사용

case 3) 회귀/분류모델에서 높은 중요도 피쳐만 사용

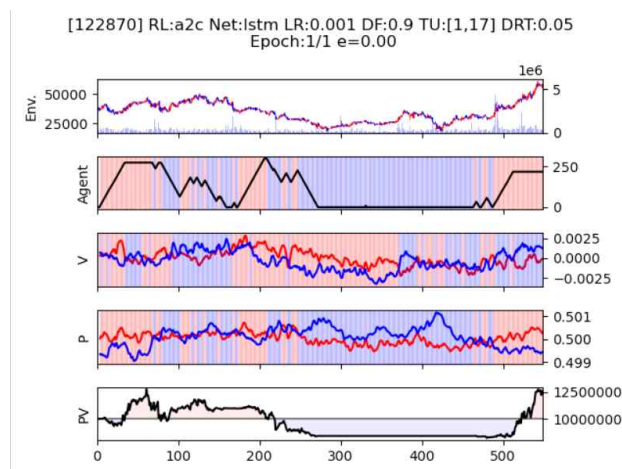
case 4) 모든 피쳐 사용

실험 결과 :

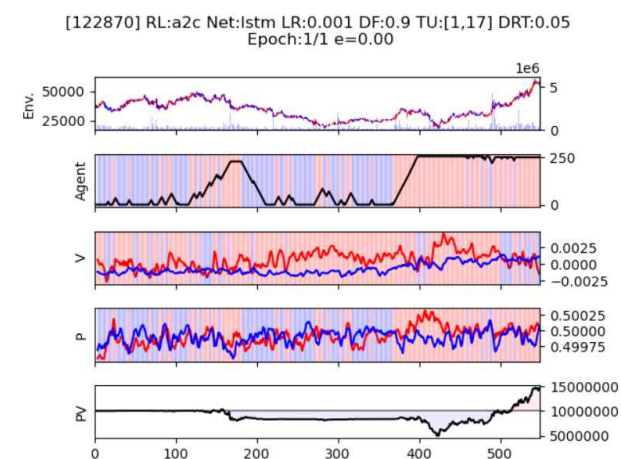
- 학습기간이 길다고 강화가 잘되는 것은 아님.
- 최종적으로는 수익이 났지만, case 3)과 case 4)에서 수익이 나는 구간이 달랐음.
- > 최적 조합을 찾는 실험이 더 필요해보임.

※ 학습기간(20180101~20200630), 평가기간(20200701~20200918)

case 4) 백테스팅



case 3) 백테스팅



감사합니다.