Dacon AI Bit Trader Competition

Day trading with ARIMA

Presenter: Sungguk Cha

MINDs Lab

Abstract

1. We show using past information is important with non-autoregressive CATBOOST.

2. Using a simple linear model Autocorrelation Integrated Moving Average (ARIMA), we show that understanding longer time range helps forecasting.

3. We ranked **29**/138 (top 21%) in Dacon AI Bit Trader Competition.

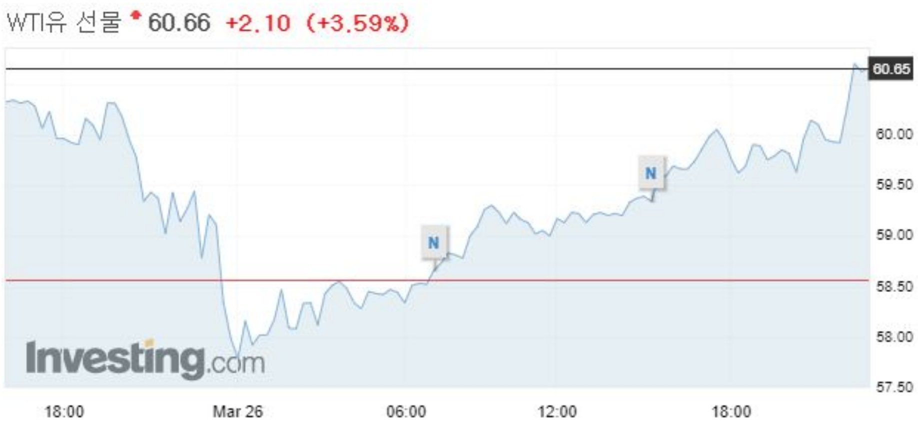
MINDs Lab

Contents

MINDs Lab

Introduction

Forecasting (time series analysis)



MINDs Lab

Introduction

Our experiment on Dacon with CATBOOST shows that
non-autoregressive approach works poorer than chimpanzee.

Dacon: a bitcoin trading competition

CATBOOST: gradient boosting deep model, it forecasts with a current state only.

Autoregressive:

A statistical model is autoregressive if it predicts future values **based on past values**.

MINDs Lab

Introduction

In our following experiments,
autoregressive models
outperform chimpanzee
and ranked 29/138 (top 21%) in Dacom competition.

MINDs Lab

Approach: Statistical models

c is a constant. ϵ is a Normal(0, 1) white noise at time t. ψ and θ are parameters. μ is the expectation.

Autocorrelation (AR(p))

$$X_t = c + \sum_{i=1}^{p} \psi_i X_{t-i} + \epsilon_t$$

Moving Average (MA(q))

$$X_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

$$\mu = \sum_{i}^{t-1} w_i X_i$$

Autoregressive Moving Average (ARMA(p, q))

$$X_t = c + \epsilon_t + \sum_{i=1}^{p} \psi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

MINDs Lab

Approach: ARIMA

Autoregressive Integrated Moving Average (ARIMA(p, d, q))

Backshift operator B.

$$BX_t = X_{t-1}$$

$$B(BX_t) = B^2 X_t = X_{t-2}$$

$$X_t' = X_t - X_{t-1} = X_t - BX_t = (1 - B)X_t$$

$$X_t'' = X_t - 2X_{t-1} + X_{t-2} = (1 - 2B + B^2)X_t = (1 - B)^2 X_t$$

$$X_t^d = (1 - B)^d X_t$$

ARIMA is modeled as following.

$$X_t^d = c + \psi_1 X_{t-1}^d + \ldots + \psi_p X_{t-p}^d + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t$$

MINDs Lab

Experiments

Implementation detail

ARIMA

Used default setting of *statsmodels* library in python to train.

ARIMA(p, 1, 1) for p in [8, 16, 32, 64] are used.

ARIMA X denotes ARIMA(X, 1, 1) model.

MINDs Lab

Experiments

Implementation detail

Dacon dataset

- 10종류의 암호화폐
  - 비트코인 포함 (종류 모름)
- 529개의 샘플
  - 각 샘플은 coin index(0~9) 와 23시간 분봉 (open, high, low, close, volume, quate_av, trades, tb_base_av, tb_quote_av)
- 각 샘플에 대해서 23시 시점에 [0, 1]개 만큼 사서 2시간 안[0, 119]에 팔아야함
- 10,000 원시작
- 대회에 제출하여 채점

MINDs Lab

Experiments

Results

Approach	Score
ARIMA(16, 1, 1)	10976.97647
ARIMA(64, 1, 1)	10952.20946
ARIMA(32, 1, 1)	10914.36141
ARIMA(8, 1, 1)	10797.23666
ARIMA(4, 1, 1)	10797.23666
Nothing	10000.0
Baseline-Chimpanzee2	8820.57
Baseline-Chimpanzee1	6640.34
CATBOOST	5222.54

#	팀	팀 멤버	점수	제출수	등록일
29	지아봇		10976.97646	7	2분 전
1	씩씩한오리너구리		26,085.52387	11	한 시간 전
2	Tinys		22,591.49509	7	하루 전
3	이산		20,717.95157	21	14시간 전
4	킹남지컴퍼니		19,640.13688	14	2시간 전
5	ToBigs		16,960.78941	19	한 시간 전
6	ModuWay		16,891.40409	11	5일 전
7	정재민		15,993.97509	20	2일 전
8	titanumm		15,758.45809	8	2일 전
9	버터		15,358.30881	23	3일 전
10	stat17_hb		15,282.79101	4	3일 전

Total 138 entries
Ranked top 21%

MINDs Lab

Conclusion

- We need time information. (I.e., autoregressive model is essential.)
- To capture longer time distance, we need to use a ***deep*** *model*.
- Simple linear model (ARIMA) is effective.

MINDs Lab