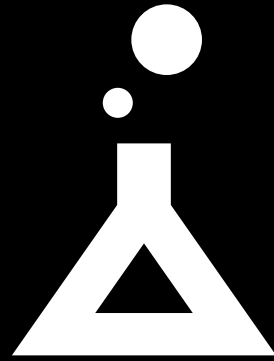
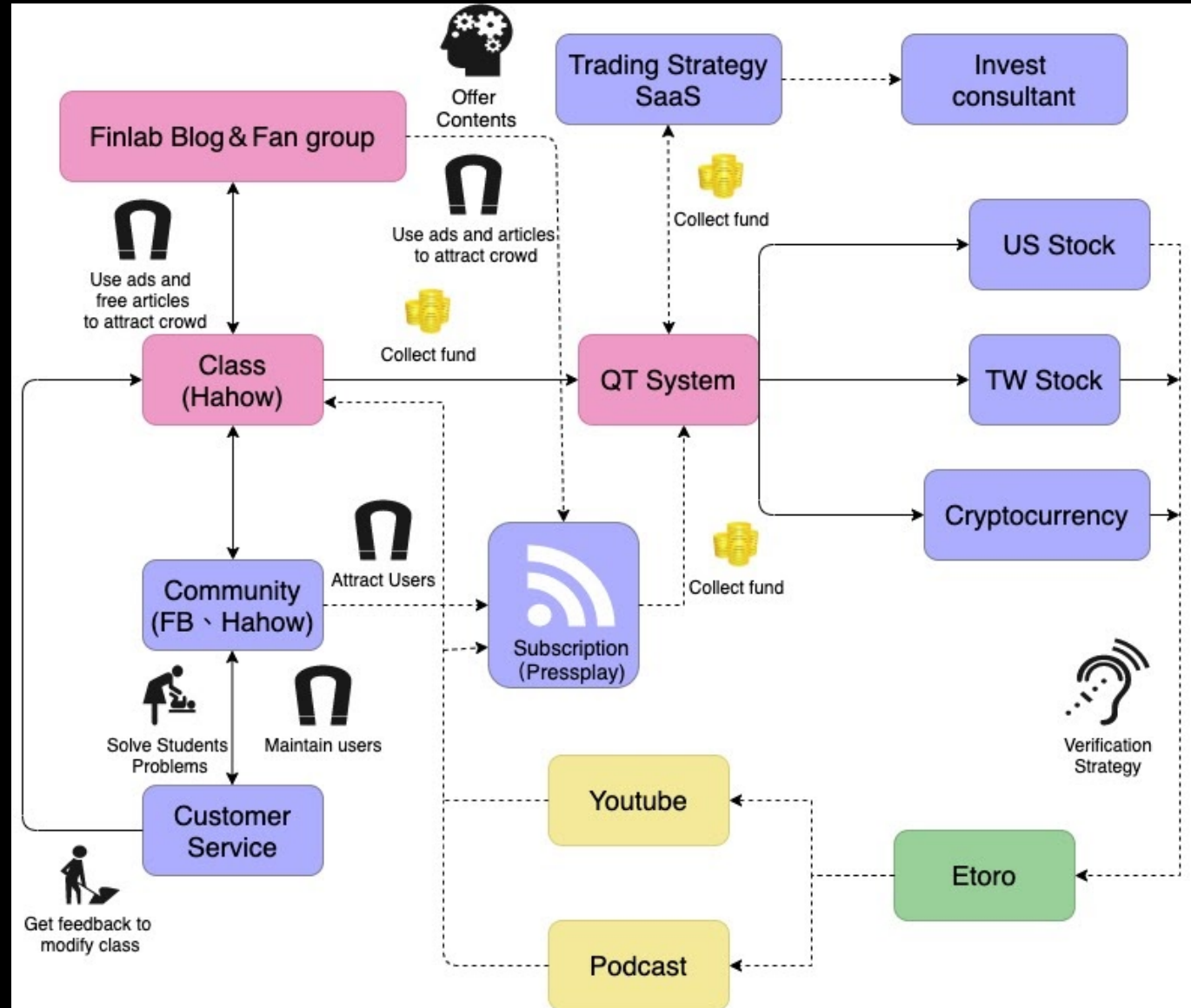


用 Python 理財 機器學習交易訊號

9/6 PyCon 韓承佑



FinLab



■ Trading programming language

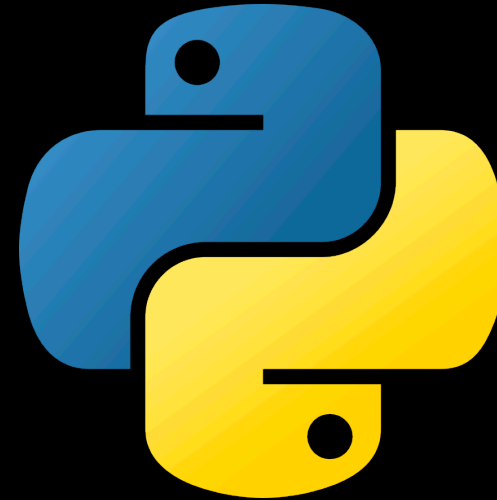


- Easy learning curve for the beginners
- Integrated with language editor in platforms
- Can be extend by external DLL
- Most of the functions are encrypted or the source code is not provided
- Does not support statistic analysis or machine learning toolkit

Trading programming language

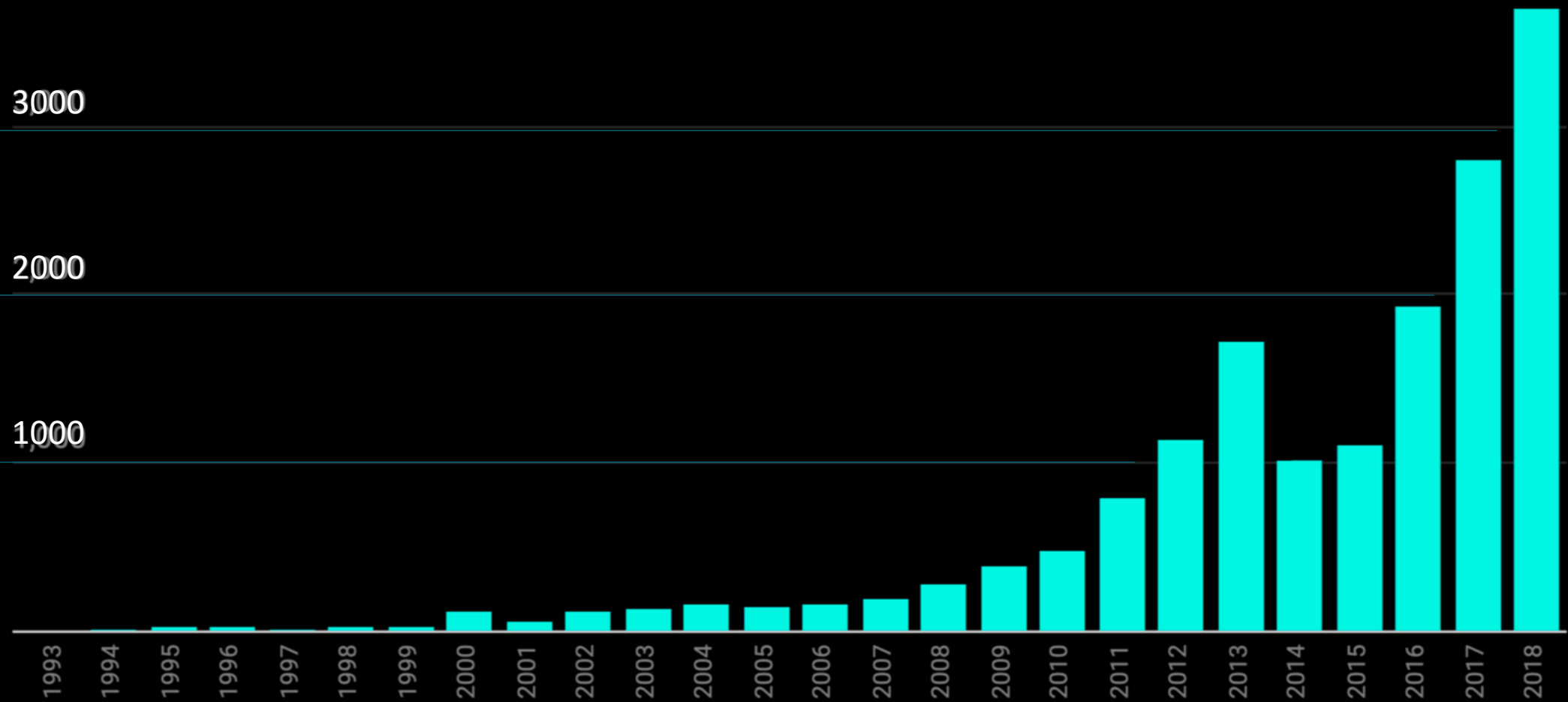


- Friendly statistic toolkit



- Friendly statistic toolkit
- Strong community and widely applied
- Easy to deploy (Flask/Django/...)
- More innovative data science applications

Artificial Intelligence papers



All of the papers available in the “artificial intelligence” section (arXiv)

Outline




Financial Data	Features
	Labels
Machine Learning Models	NN
	LSTM
	CNN
Evaluation	Backtesting
	Purged K-fold

ML algorithms in finance?



Supervised Machine Learning




Training

features			labels
Color	Weight	Age	Category
	3.2 kg	2	cat
	4.2 kg	5	cat
	6.2 kg	4	dog



ML
Model

Testing

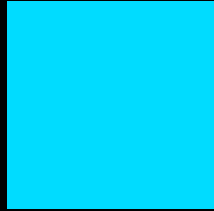
features		
Color	Weight	Age
	3.2 kg	2
	4.2 kg	5
	6.2 kg	4



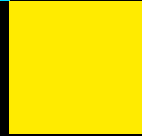
ML
Model



labels	
True Answer	Prediction
cat	cat
cat	dog
dog	dog



Financial Data (Features)



Financial Data Structures

Fundamental data

Focusing on creating a portrait of a company

- Useful to combine other data types
- Difficult to confirm data release date
- Missing data is often backfilled
- Consider multiple correction

Trading data

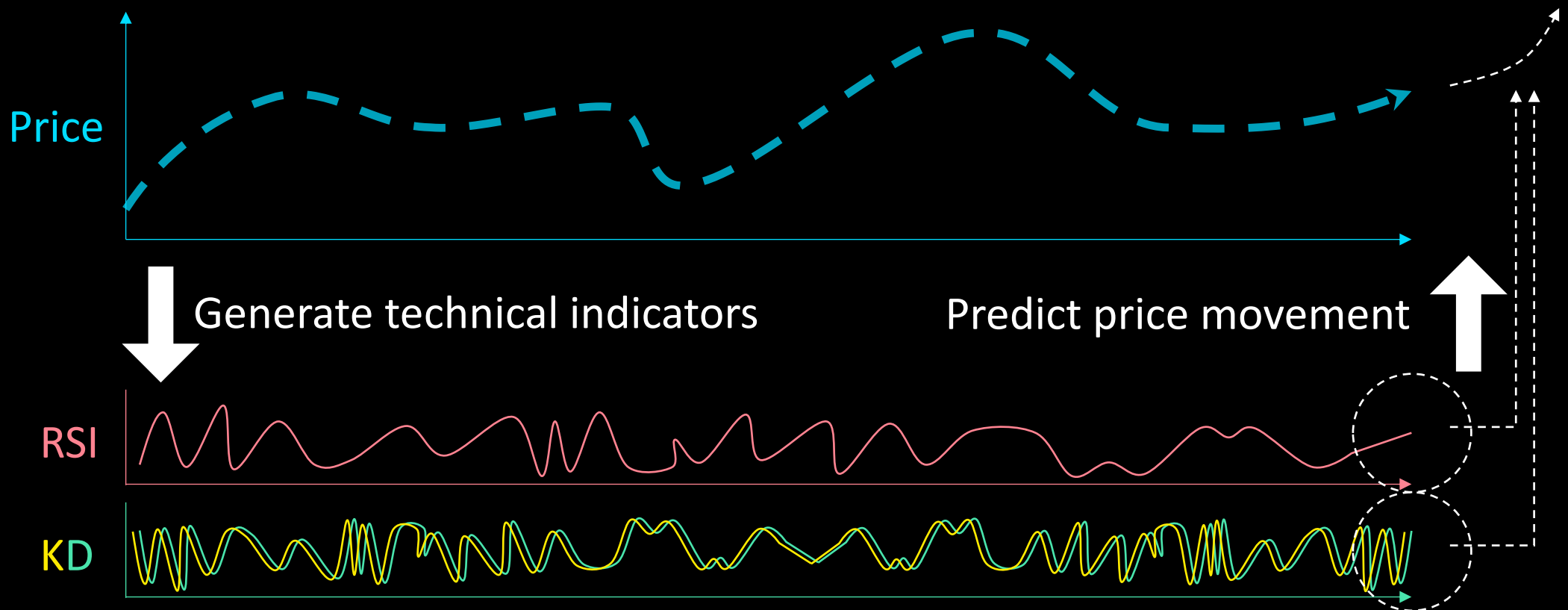
Market participant characteristic footprint

Trading book, price, broker trading summary...etc

- Data often with timestamp
- Generate extra features (ex: technical indicators)
- Massive amount of data generated in one day
- Some of the data is difficult to obtain

Creating Technical indicators

Price historical data

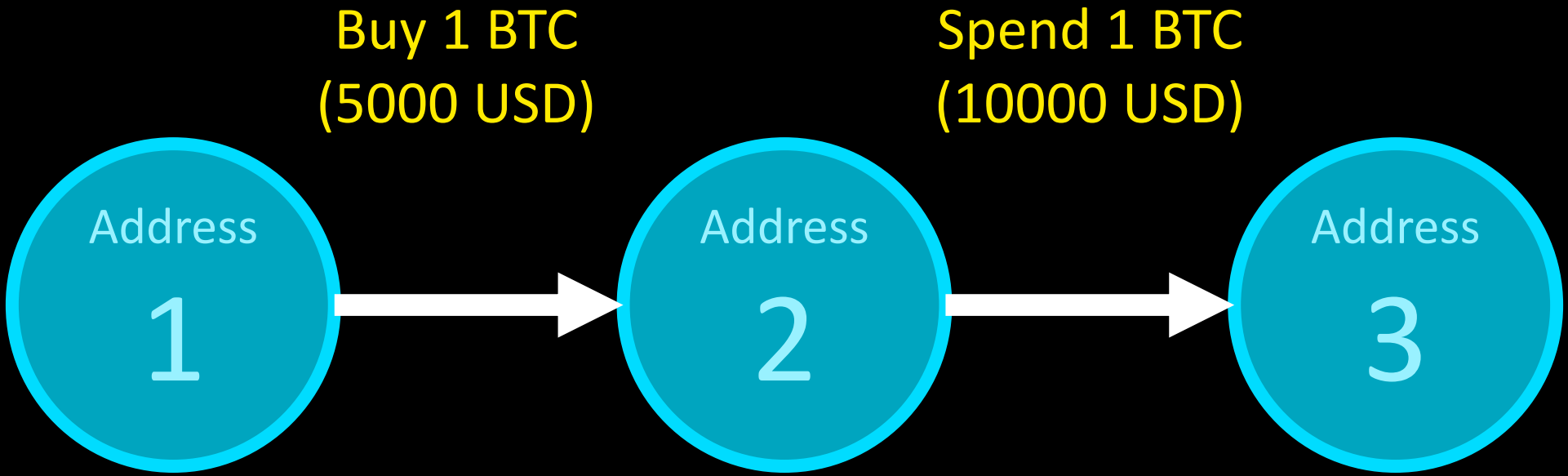




■ Fundamental Indicators

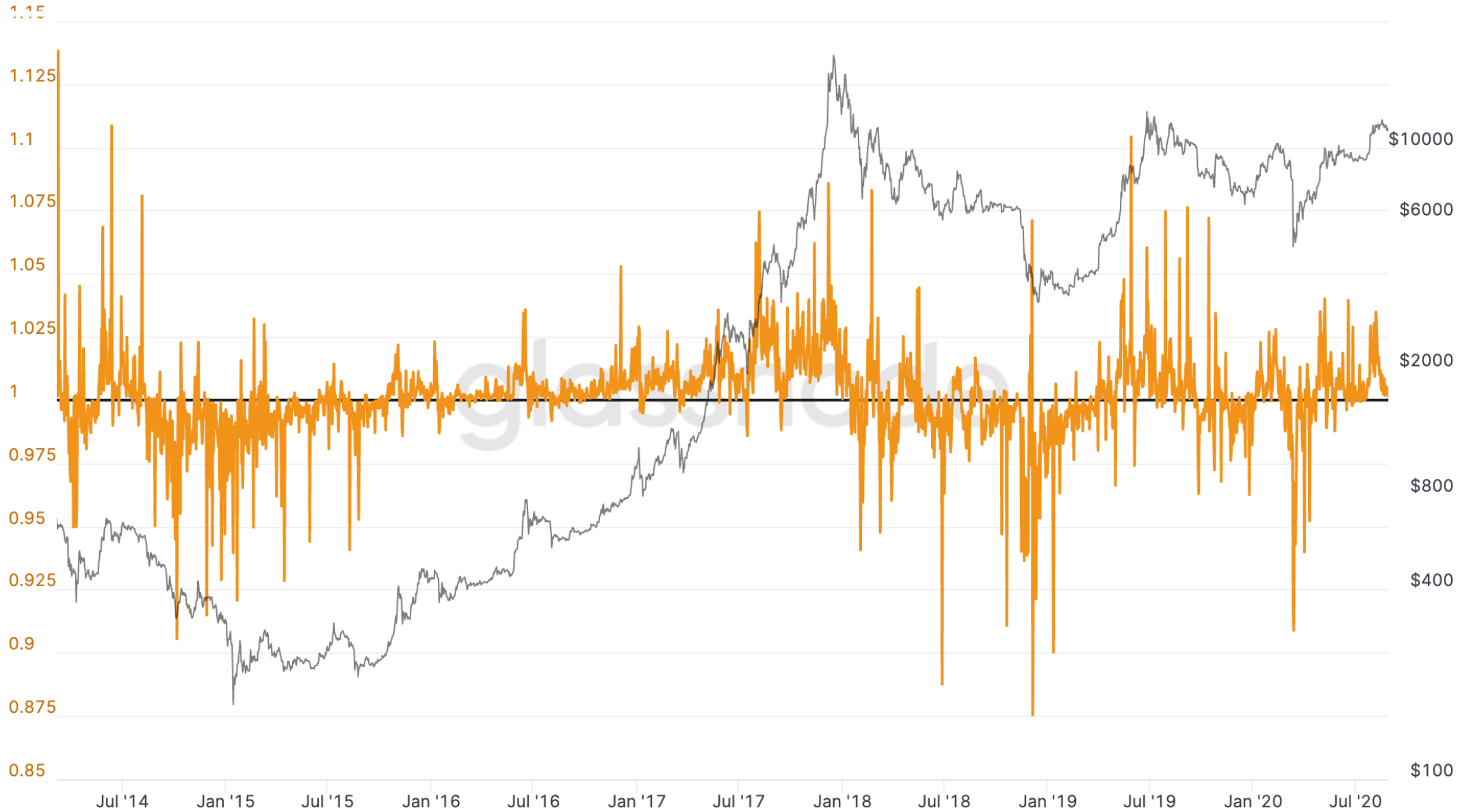
- Spent Output Profit Ratio
- Network value to transaction ratio
- Transfers volume to exchanges

Spent Output Profit Ratio



$$\text{SOPR} = 10000 / 5000 = 2$$

Spent Output Profit Ratio



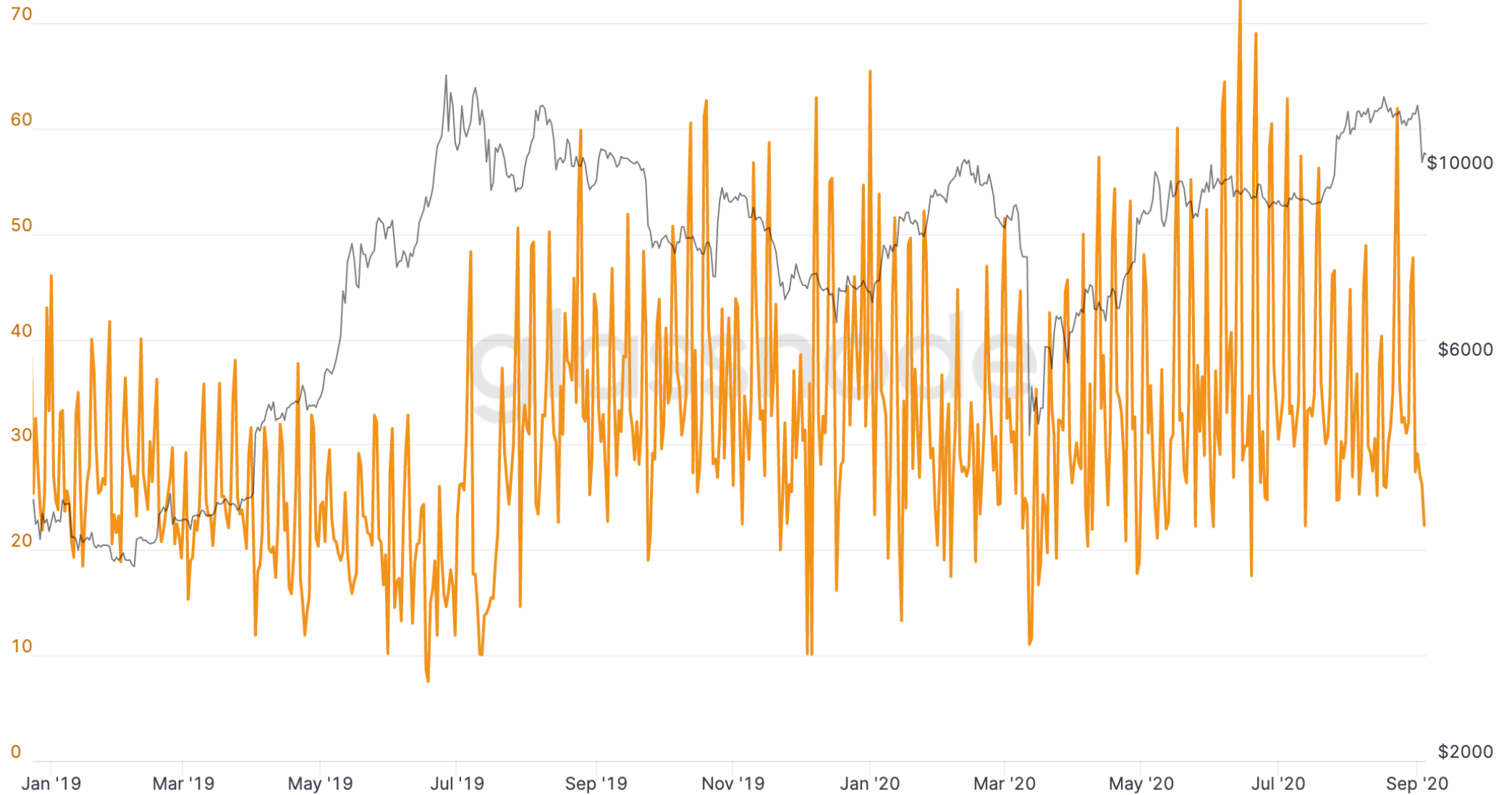
■ Network Value to transaction ratio

$$\text{本益比} = \frac{\text{總市值}}{\text{總盈餘}}$$

■ Network value to transaction ratio

$$\text{NVT} = \frac{\text{總市值}}{\text{總金流}}$$

Network value to transaction ratio

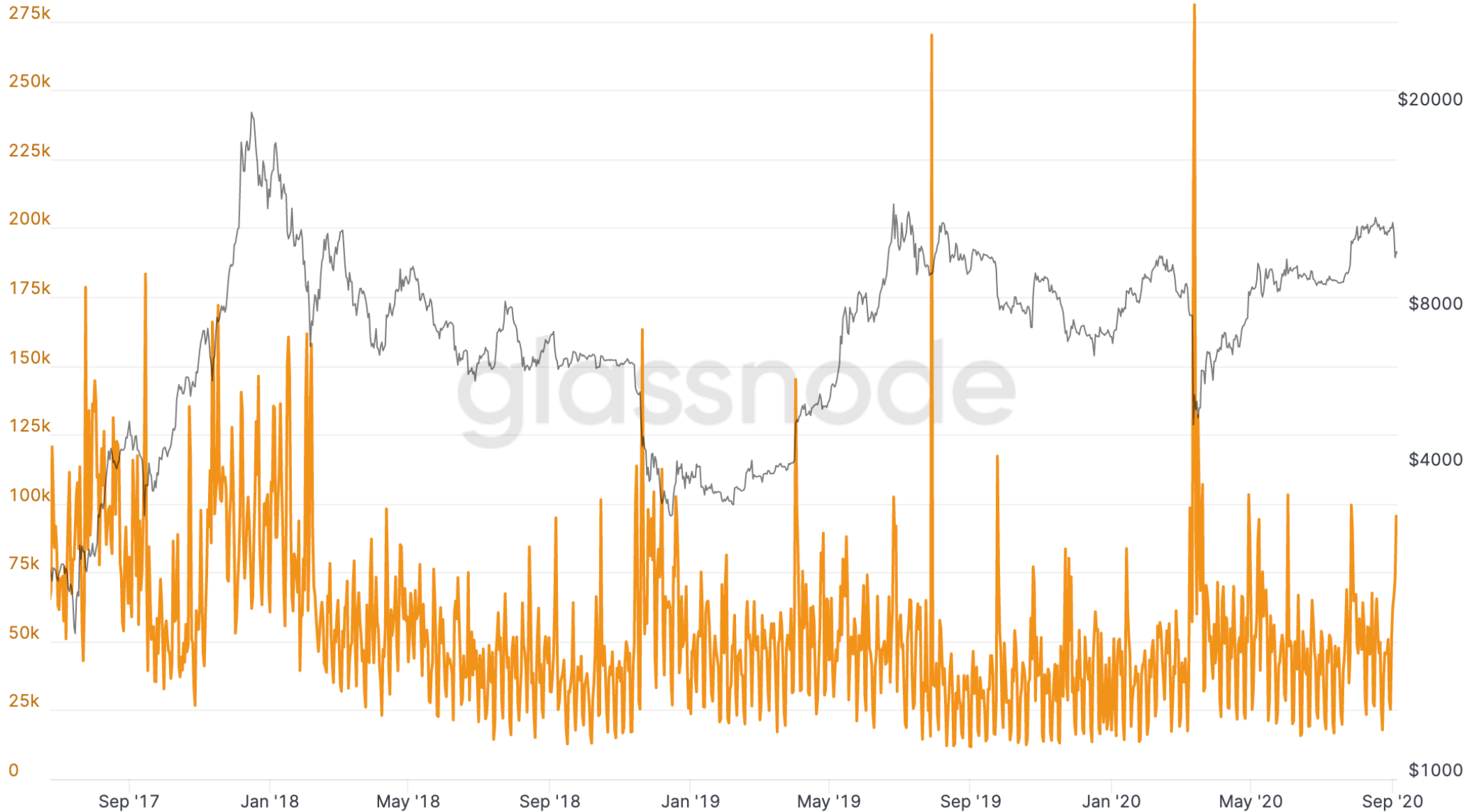


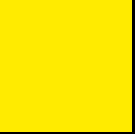
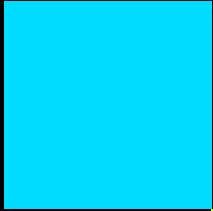
Network Value to transaction ratio

address
address
address



Network Value to transaction ratio





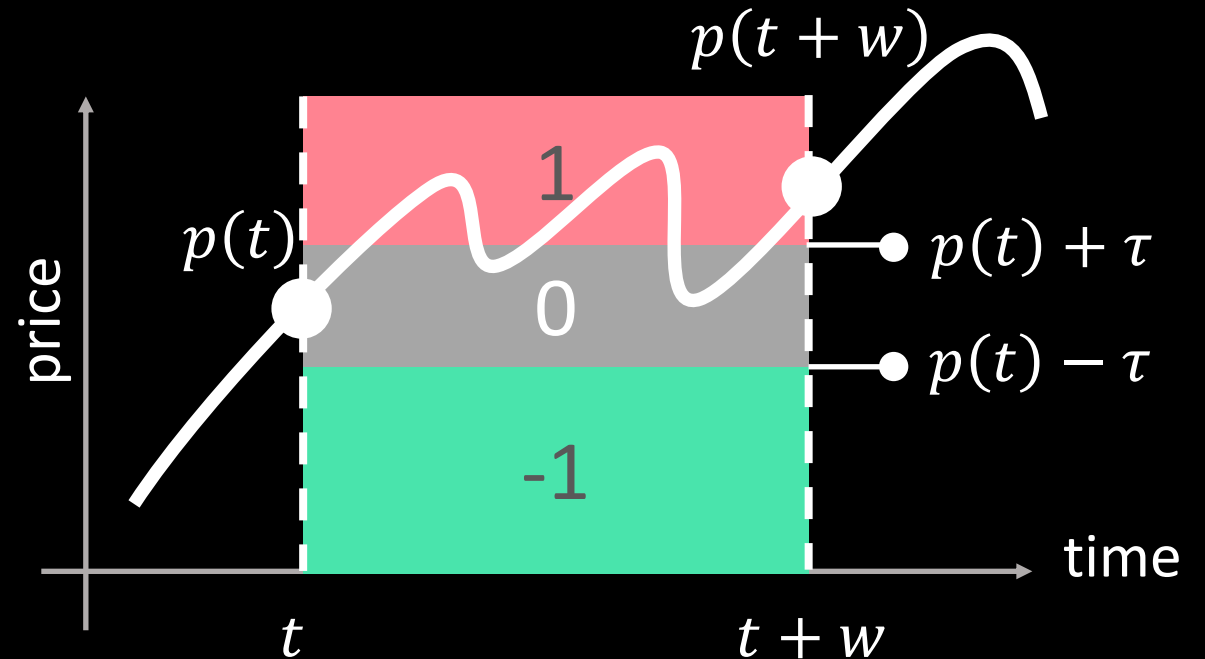
Labeling

■ Challenging of Labeling the data

Fixed time horizon

A popular method in the literature

- τ is a constant
- Do not have stop-loss limits





Label Generation Methods

- Triple barrier [Prado 2018]
- Continuous trading signals [Dash 2016]
- Trading Point decision [Chang 2009]

[Prado 2018] Advances in Financial Machine Learning

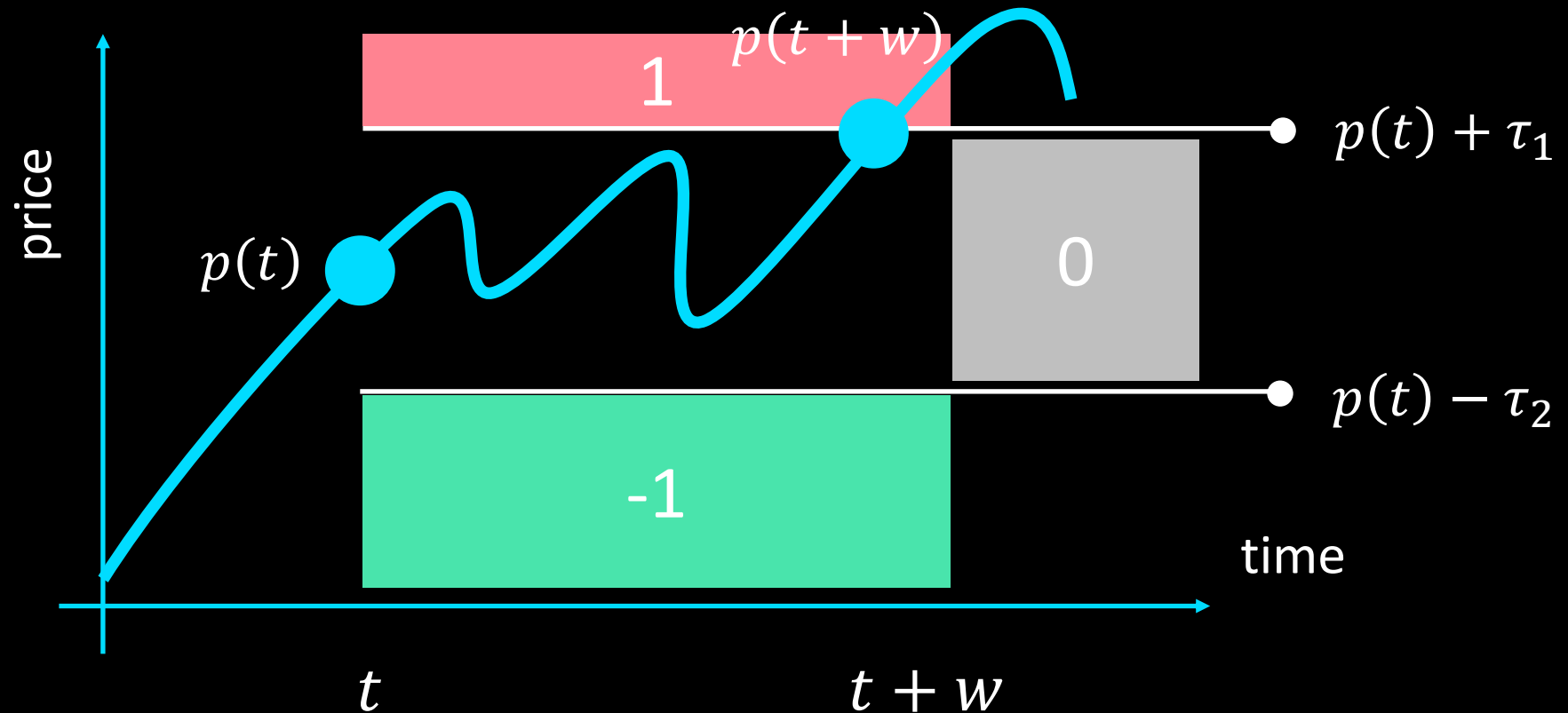
[Tsantekidis 2017] Using Deep Learning to Detect Price Change Indications in Financial Markets

[Dash 2016] A hybrid stock trading framework integrating technical analysis with machine learning techniques

[Chang 2009] Integrating a Piecewise Linear Representation Method and a Neural Network Model for Stock Trading Points Prediction

■ Triple barriers [Prado 2018]

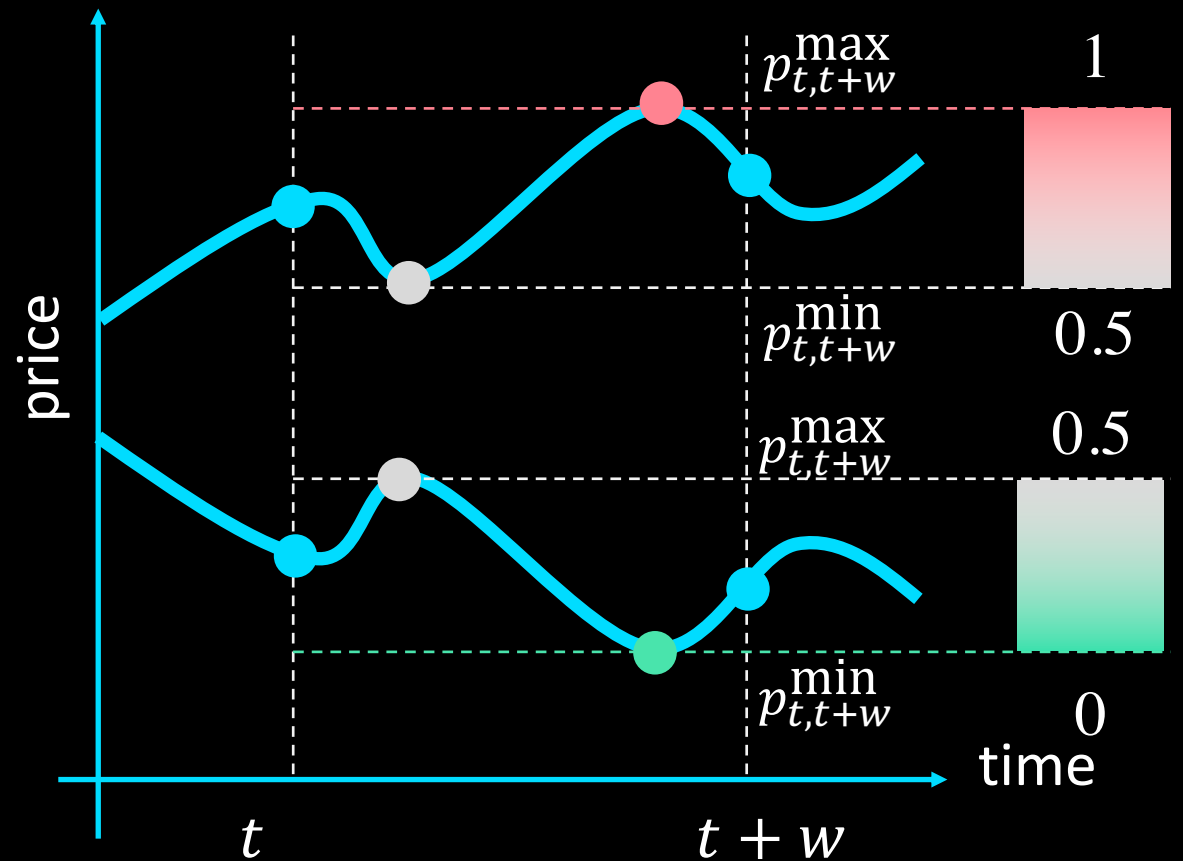
- Horizontal barriers are defined by profit-taking and stop-loss limit
- τ_1 and τ_2 are dynamic according to estimated volatility



Continuous trading signals [Dash 2016]

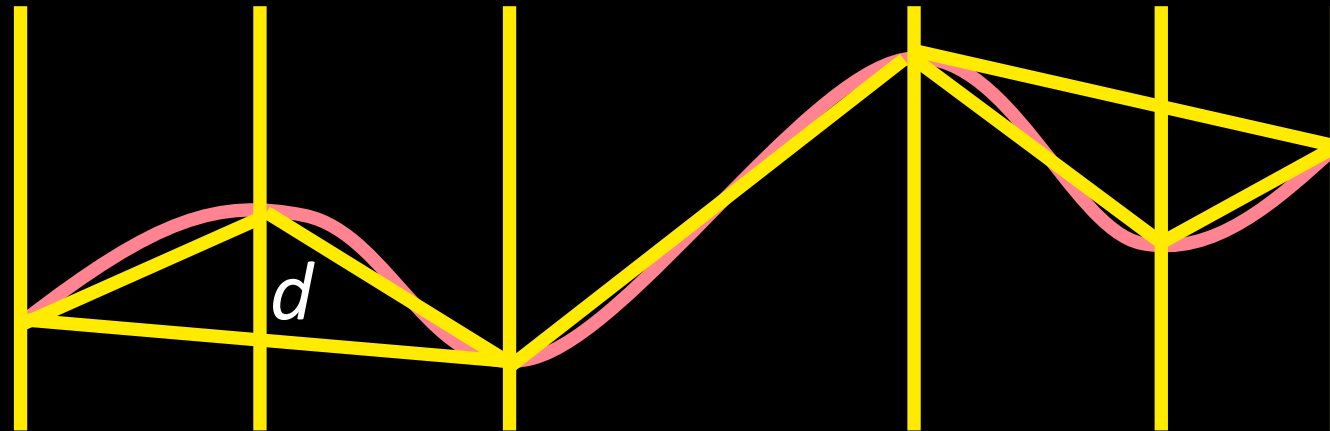
- Using **momentum** of the stock price
- $y(t)$'s are **continuous**
- Provides more detailed information

$$y(t) = \begin{cases} \frac{p_{t+w} - p_{t,t+w}^{\min}}{p_{t,t+w}^{\max} - p_{t,t+w}^{\min}} & \text{if } p_{t+w} > p_t \\ 0.5 \left(1 - \frac{p_{t+w} - p_{t,t+w}^{\min}}{p_{t,t+w}^{\max} - p_{t,t+w}^{\min}} \right) & \text{else} \end{cases}$$



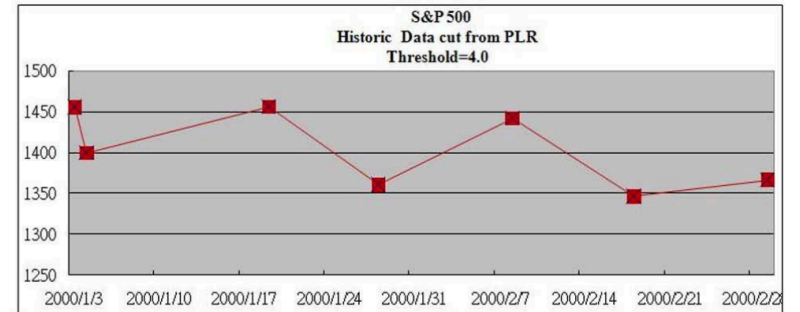
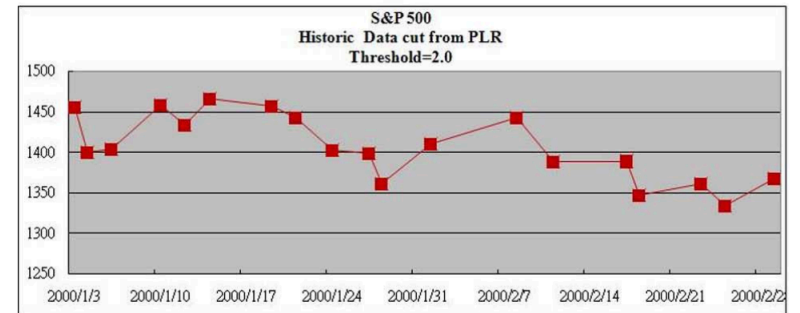
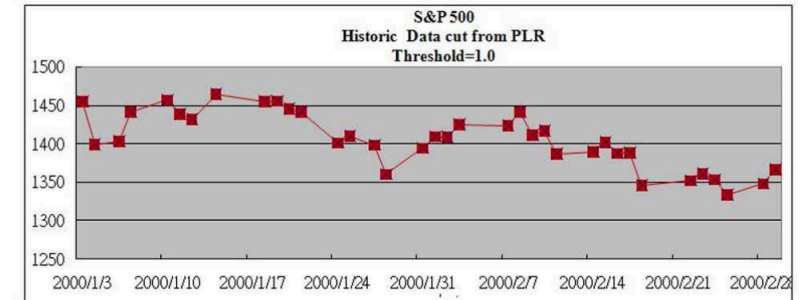
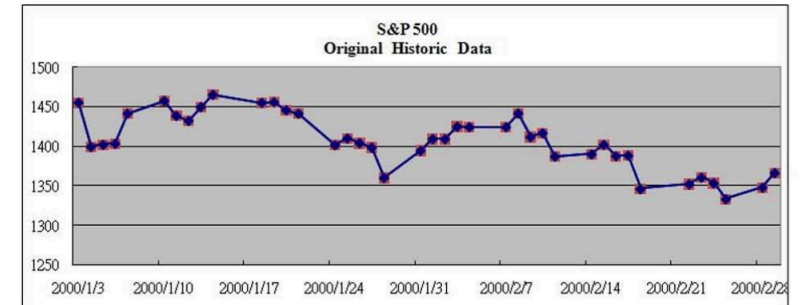
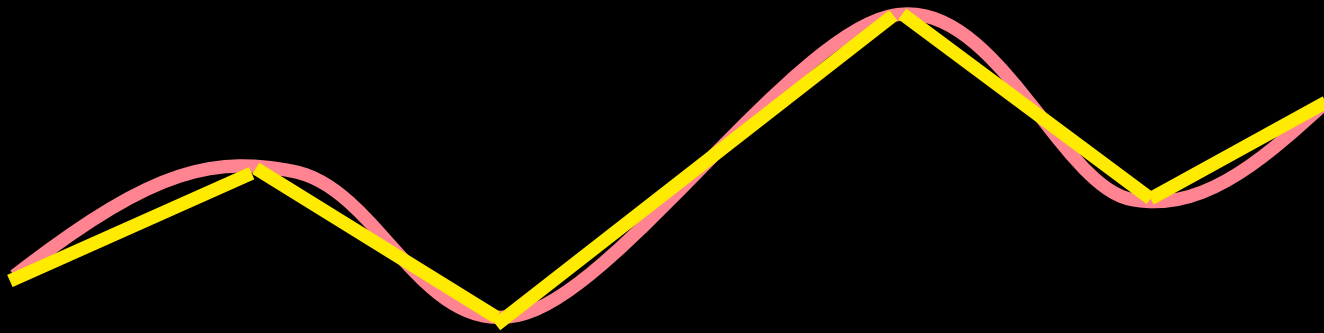
Trading point decision

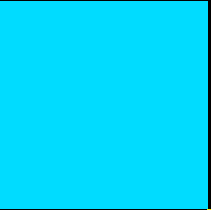
- Find the local minimum and maximum points
- Divide the time series into subsegments
- Threshold value d \rightarrow length of trend



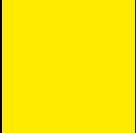
Trading point decision

- Find the local minimum and maximum points
- Divide the time series into subsegments
- Threshold value $d \rightarrow$ length of trend





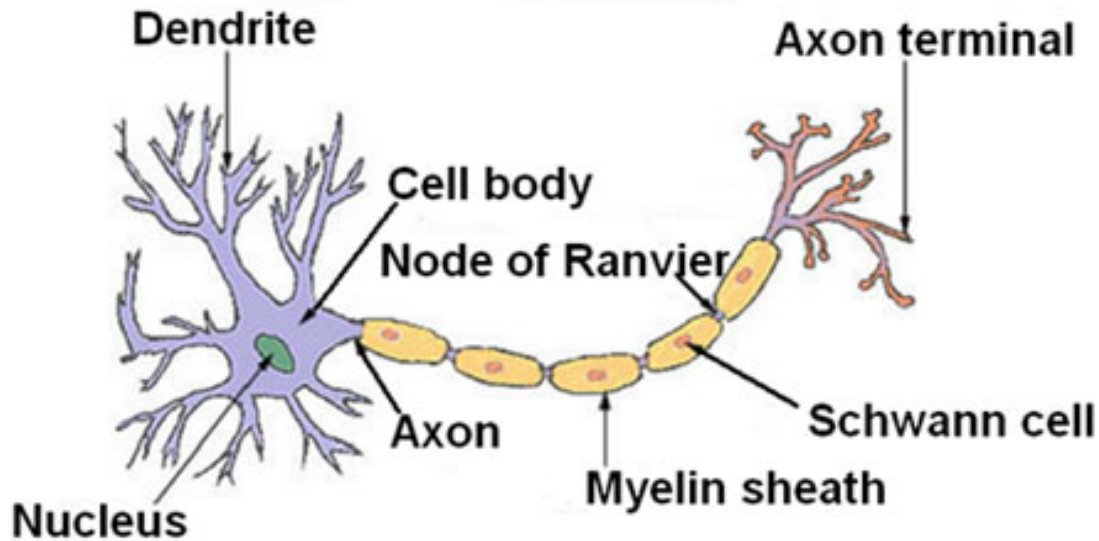
ML Models



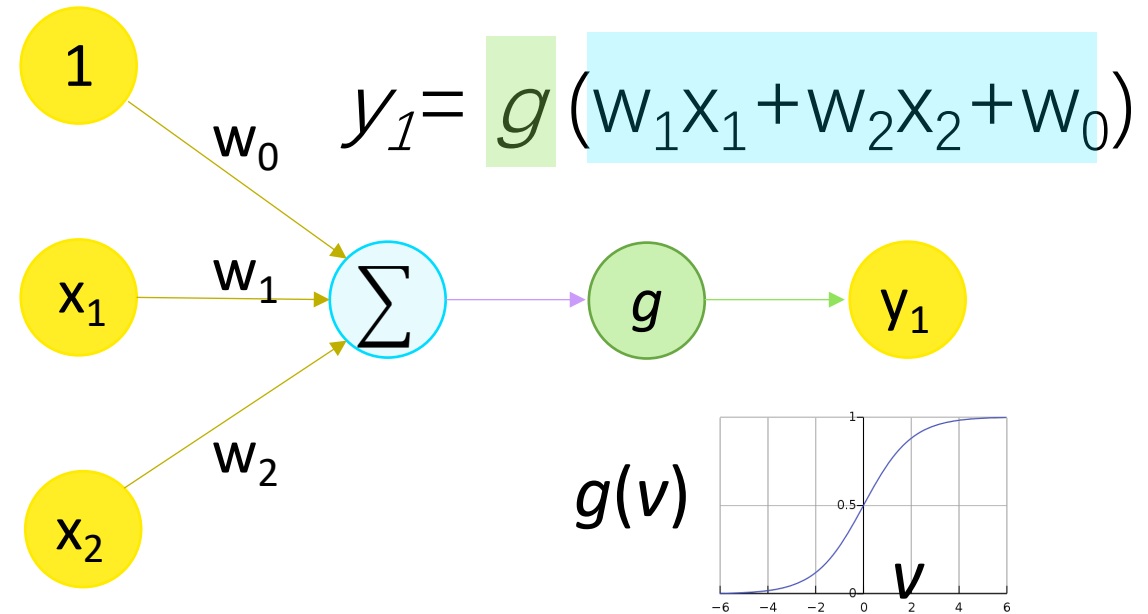
Neural Network

- Built to model the human brain
- interpret numeric data through a kind of machine perception

Human neuron structure

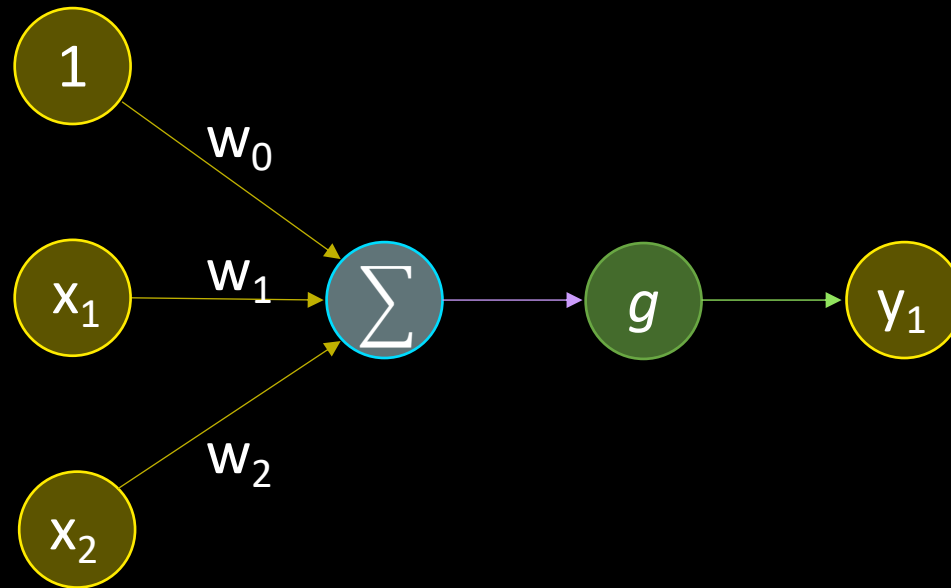


Single neuron model



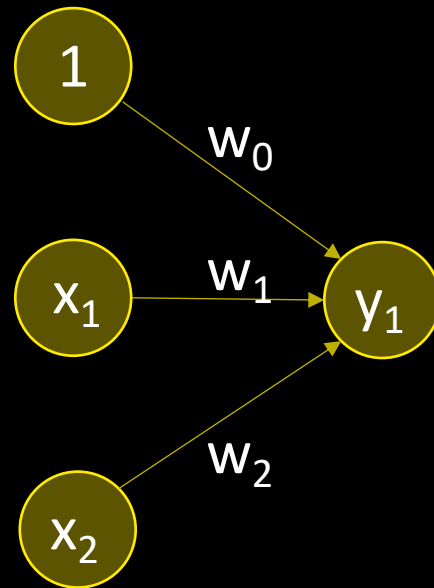
Neural Network

Single node in neural network



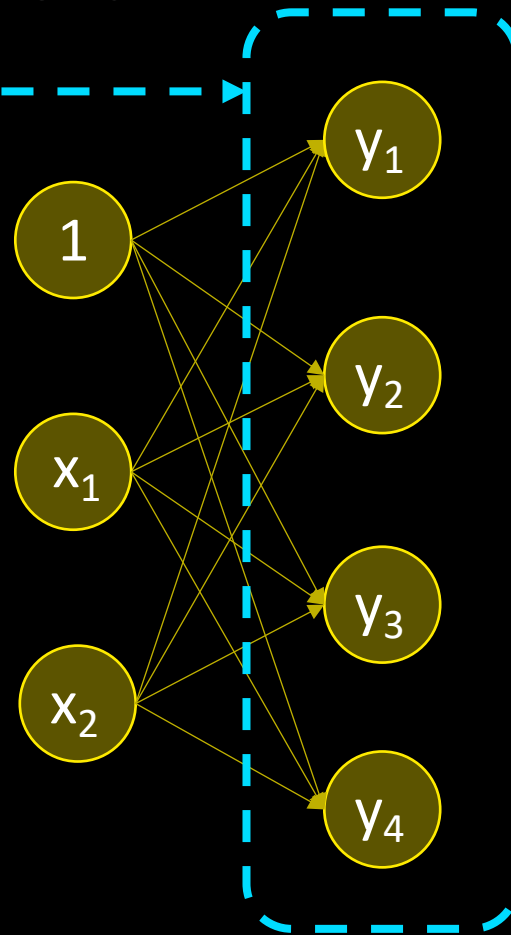
Neural Network

Simplified expression



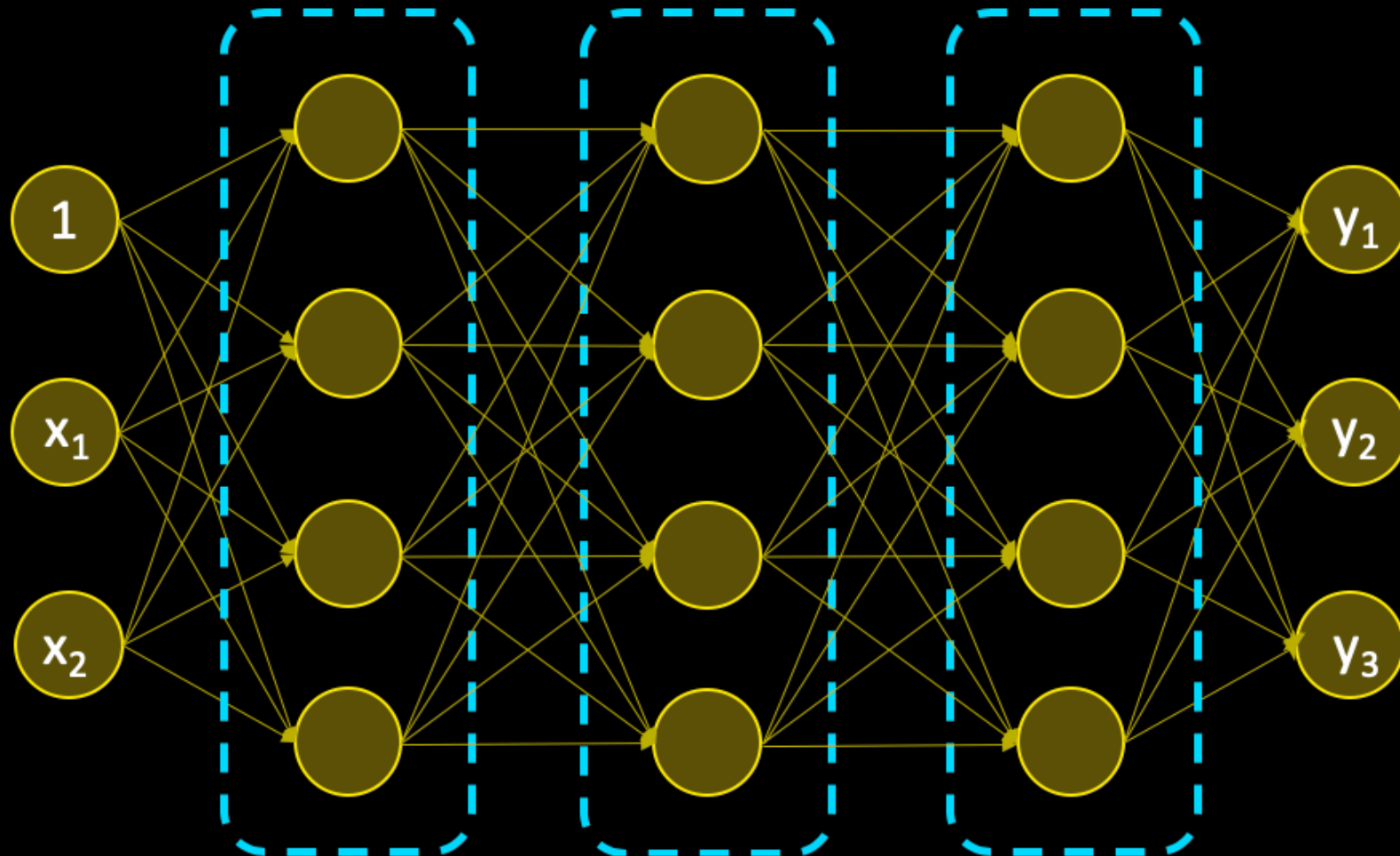
Neural Network

A layer contain multiple neurons



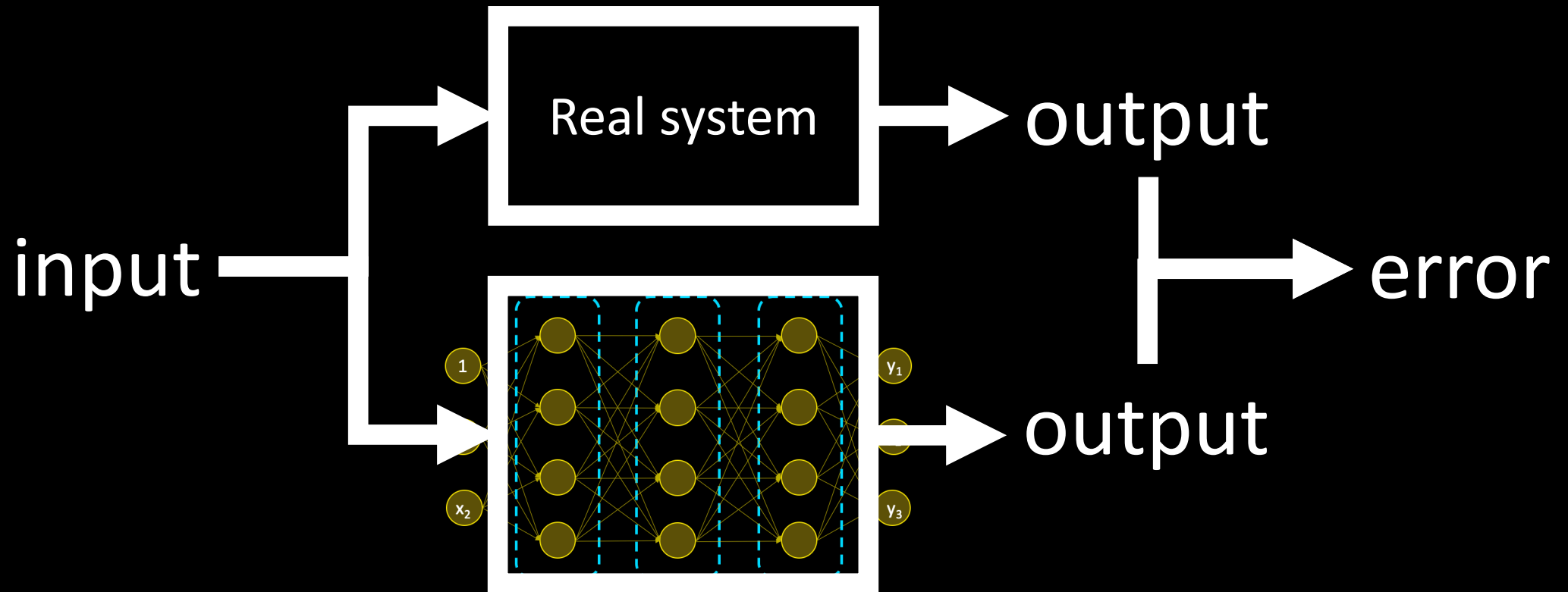
Deep Neural Network

Multi-layer deep neural network



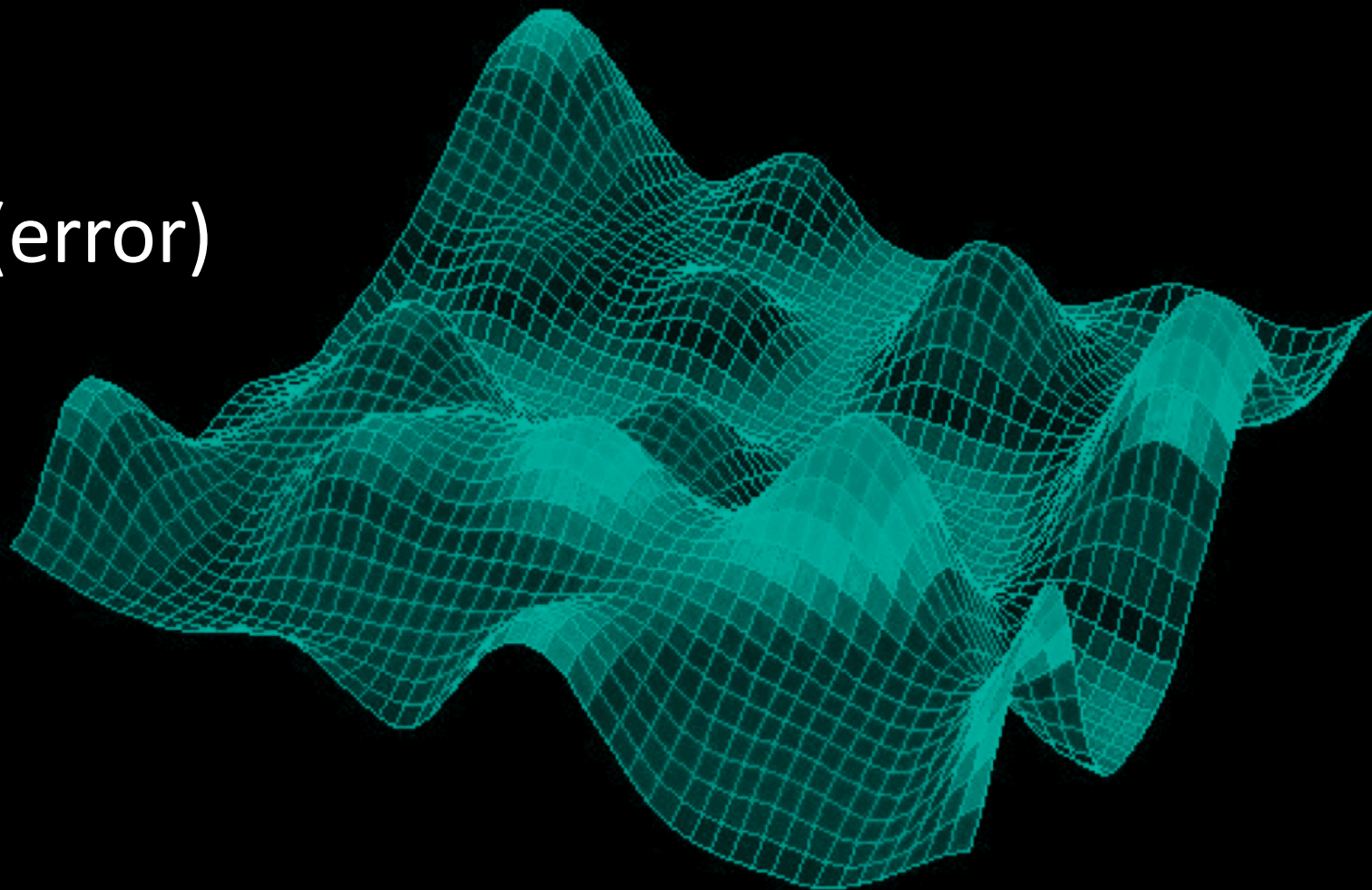
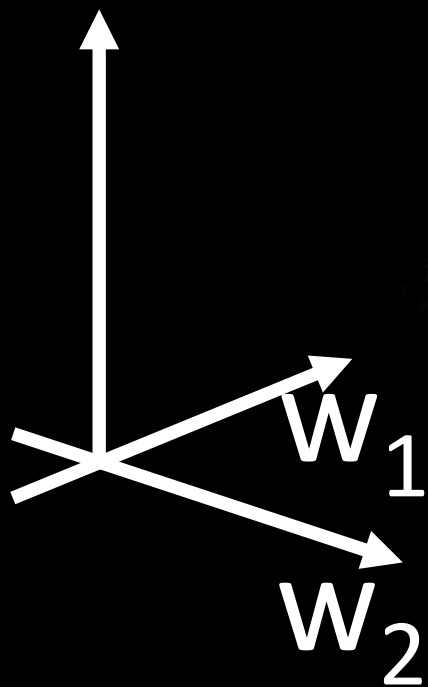
■ Deep Neural Network

Multi-layer deep neural network



Neural Network Optimization

Cost function (error)



Deep Neural Network Training Result

Asset

Taiwan Capitalization
Weighted Stock Index

Data split

Train

Validate

Backtest

2006 ~ 2014

2015

2016 ~ 2019-3-1

Features

Scaled Technical Indicators

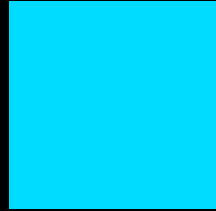
Labels

Fixed time horizon

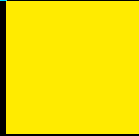
2018-1-1

2019-7-1





Model Interpretation



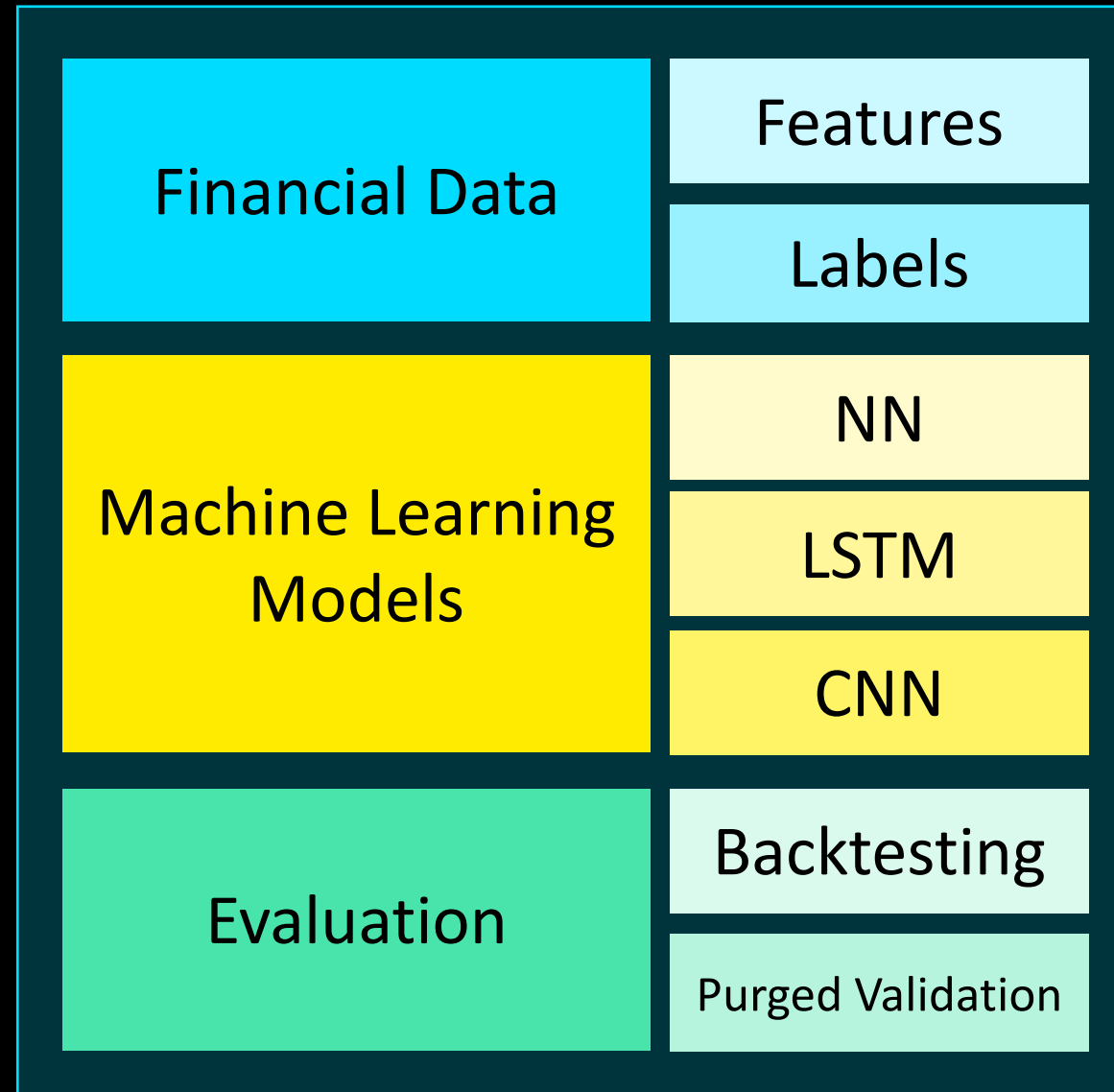


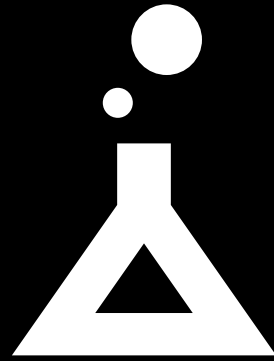
Backtest

- Survivor bias, lookahead bias, transaction cost, outlier, overfitting
- Finding the lottery tickets that won the last game
- Solutions
 - Develop model for entire asset or classes
 - Use Bootstrap aggregating
 - Record every backtest conducted
 - Resist the temptation of reusing a failed strategy

Conclusion

Machine Learning





FinLab