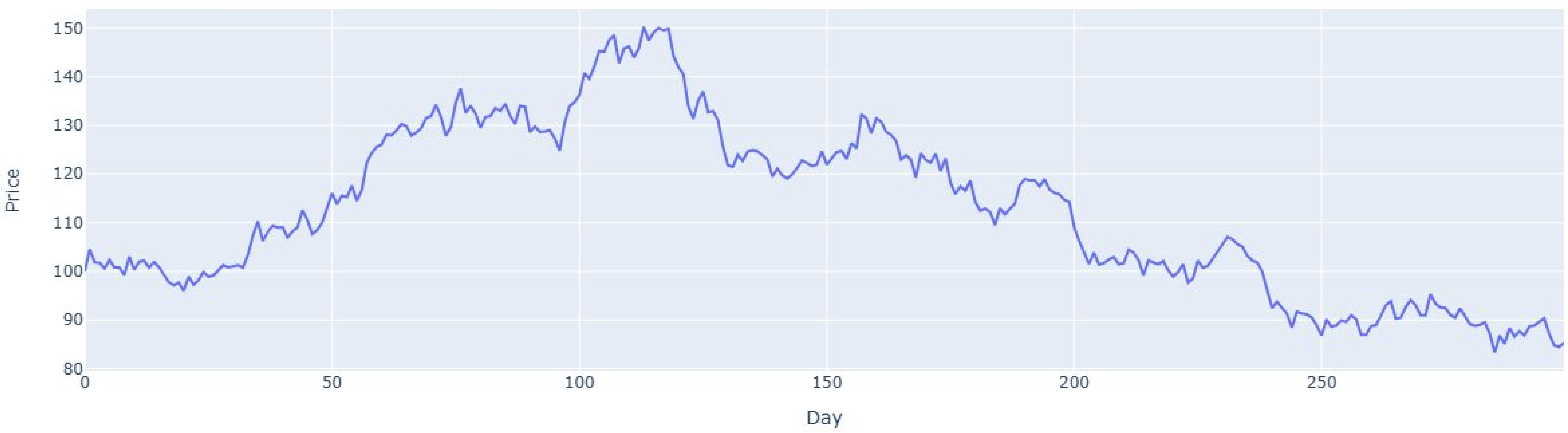
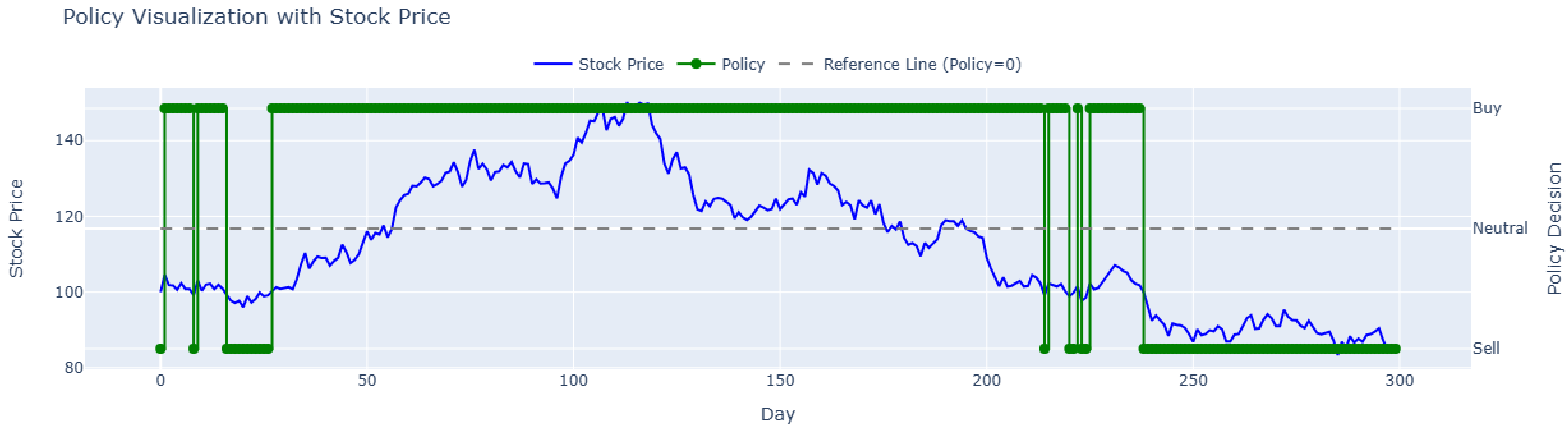
Environment - Stock prices

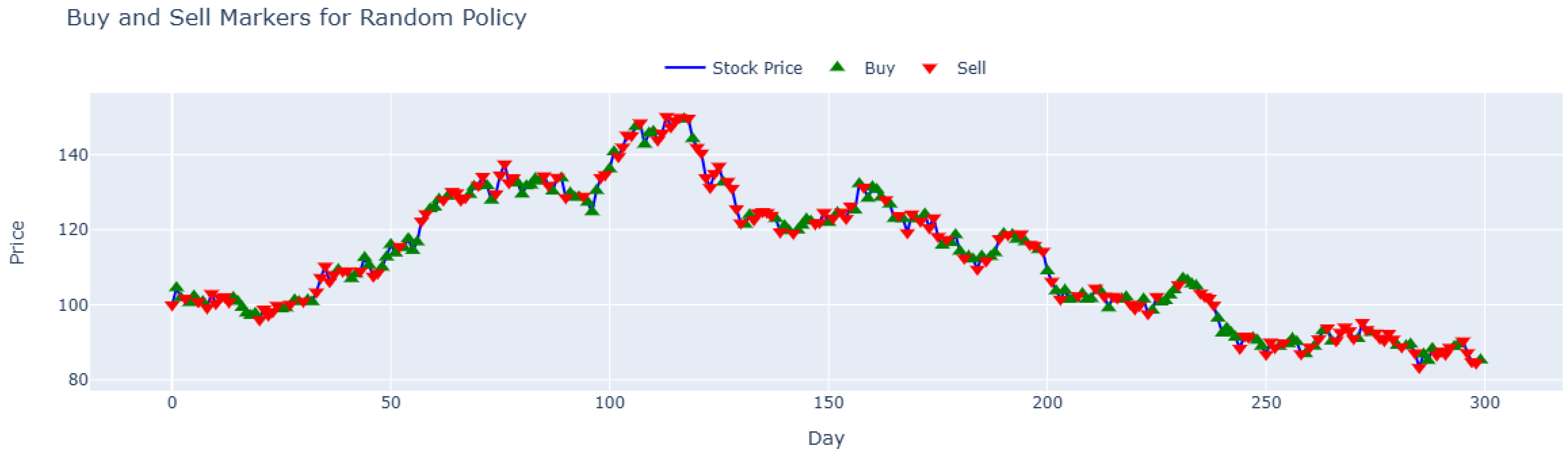


Stock Price and State Visualization
Stock Price
State
Stock Price
140
120
100
80
0
50
100
150
200
250
Day
1
0.5
0
State (1: Above 100, 0: Below 100)

Stock Price and Actions with Buy/Sell Markers
Stock Price
Buy
Sell
Stock Price
140
120
100
80
0
50
100
150
200
250
300
Day

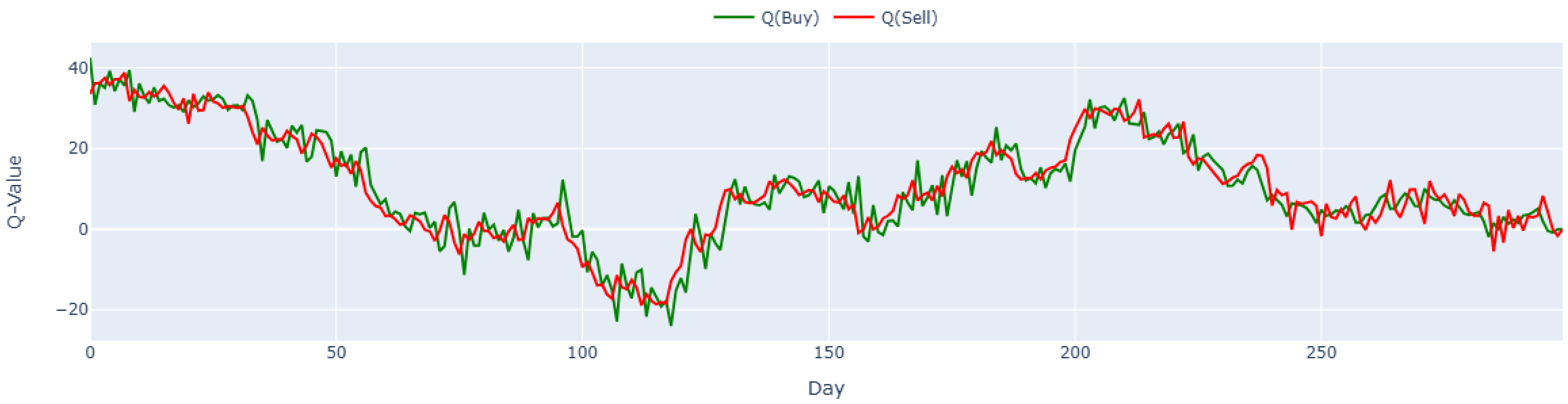

Reward Visualization: Daily and Cumulative
Cumulative Reward
Daily Reward
60
40
20
0
Reward Value
0
50
100
150
200
250
Day


Policy Visualization with Stock Price
Stock Price
Policy
Reference Line (Policy=0)
Stock Price
Policy Decision
Buy
Neutral
Sell
140
120
100
80
0
50
100
150
200
250
300
Day

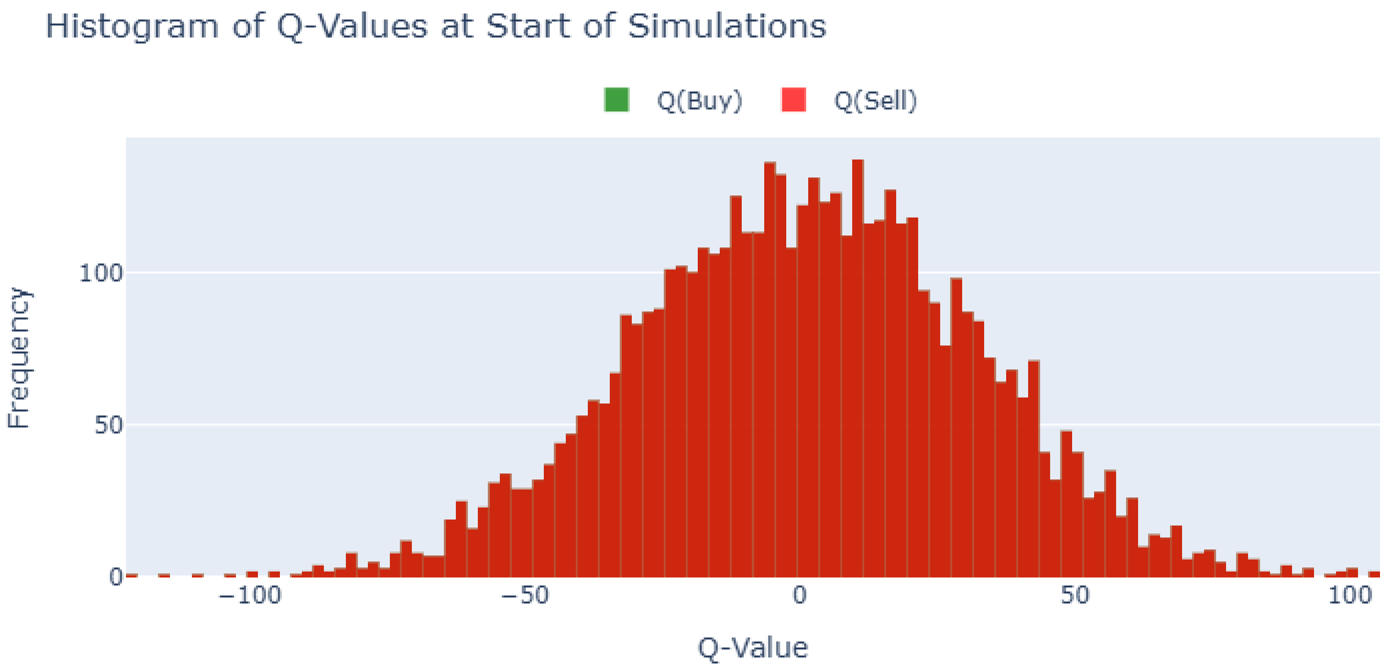


Buy and Sell Markers for Random Policy
Stock Price
Buy
Sell
Price
140
120
100
80
0
50
100
150
200
250
300
Day

Q-Value Visualization for Buy and Sell Actions




Q-Value Simulation (5000 Runs) and Average
Avg Q(Buy)
Avg Q(Sell)
100
50
0
−50
−100
Q-Value
20
40
60
80
100
Day


Histogram of Q-Values at Start of Simulations
Q(Buy)
Q(Sell)
Frequency
100
50
0
−100
−50
0
50
100
Q-Value

Deep Reinforcement Learning for Finance

Agenda

Financial Markets

Electronic markets and Market Microstructure

Optimal Trade Execution

Option pricing and hedging

Electronic Market making

Robo-advising

Smart-order routing

Artificial Intelligence – Neural Networks

What is it?

Data

The Mathematics

Computing Power

The universal approximation theorem

AI and its applications in Finance

Regulatory considerations

Fully connected NNs

RNNs

CNNs

Deep Reinforcement Learning

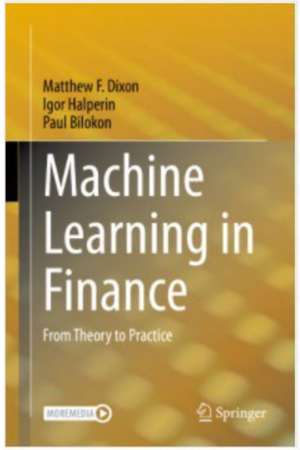
SARSA

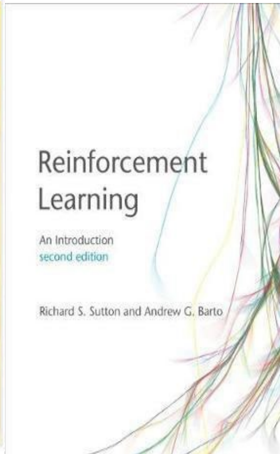
Q-Learning

Market Impact in Finance

Market Making in Finance

Finance meets Reinforcement Learning meets Deep Learning



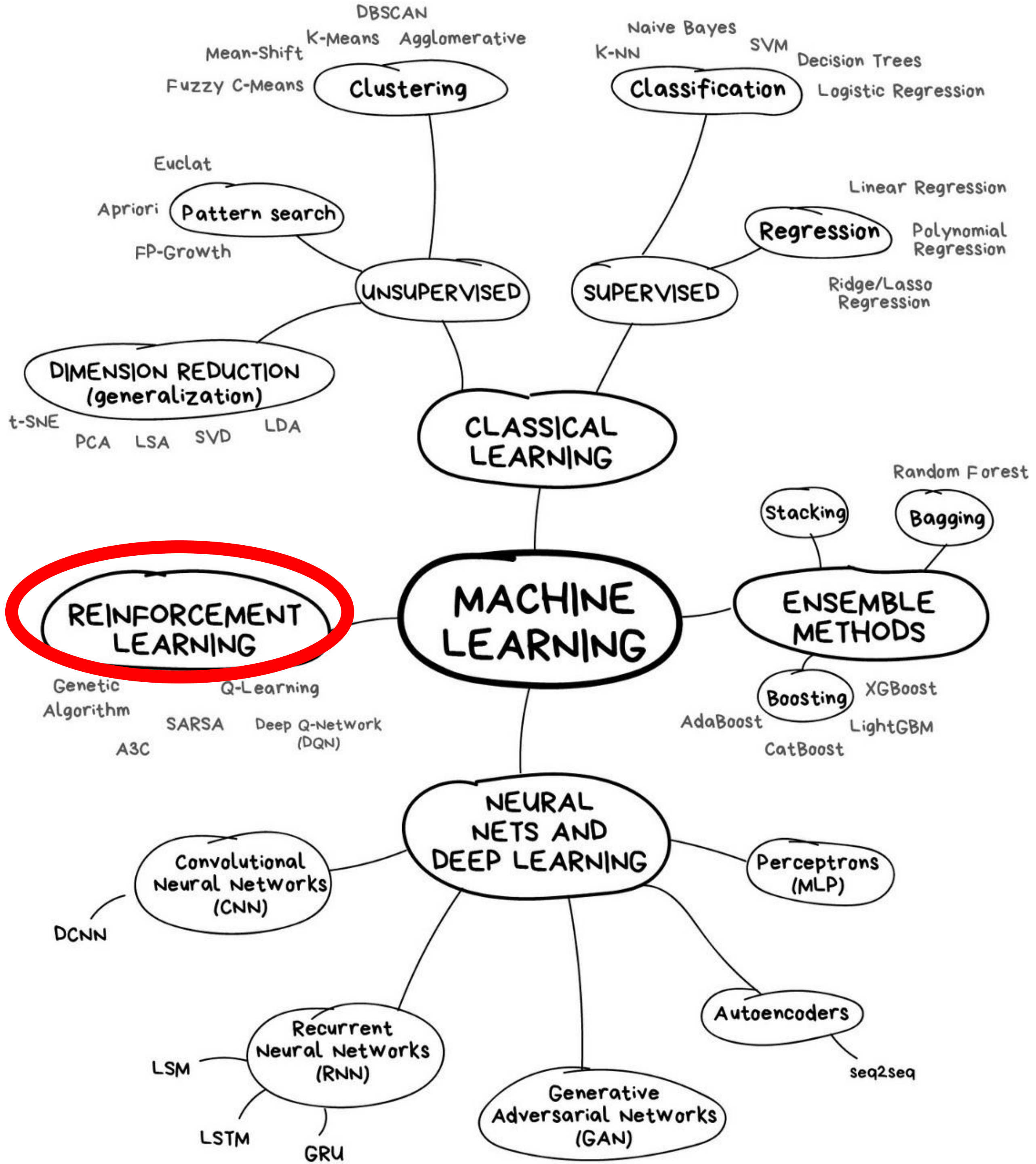


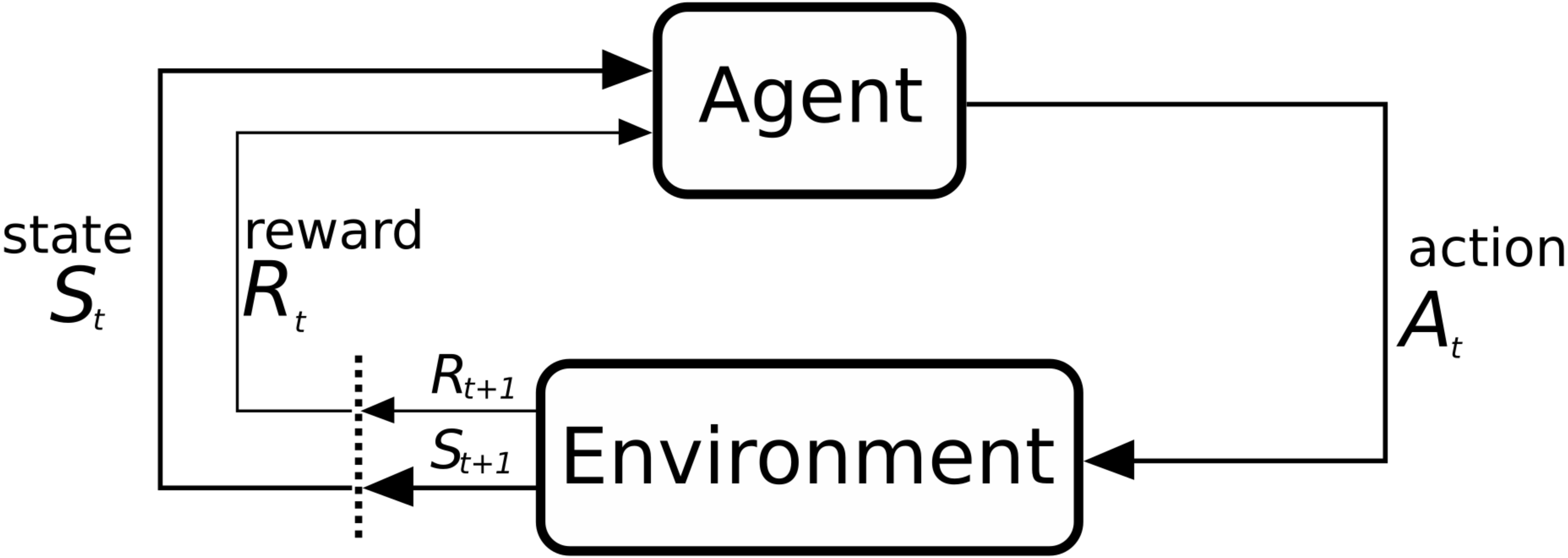
What you will learn

Goals:

- Get an understanding of the basic concepts of RL
- Understand the intuition behind standard RL algorithms
- Understand how to implement the RL workflow in python
- Understand how RL is currently being applied in Finance

A detour – Machine Learning




Agent
Environment
state
S_t
reward
R_t
R_{t+1}
S_{t+1}
action
A_t

Elements of Reinforcement Learning

Environment

An (evolving) data source which is characterized at time t by its state, S_t

Policy

$\pi_t(S_t)$ gives a rule by which the agent should act at time t given the state S_t of the environment

Rewards

A reward function, $r_t(S_t, A_t)$, determines the goal of a reinforcement learning problem.

Value Function

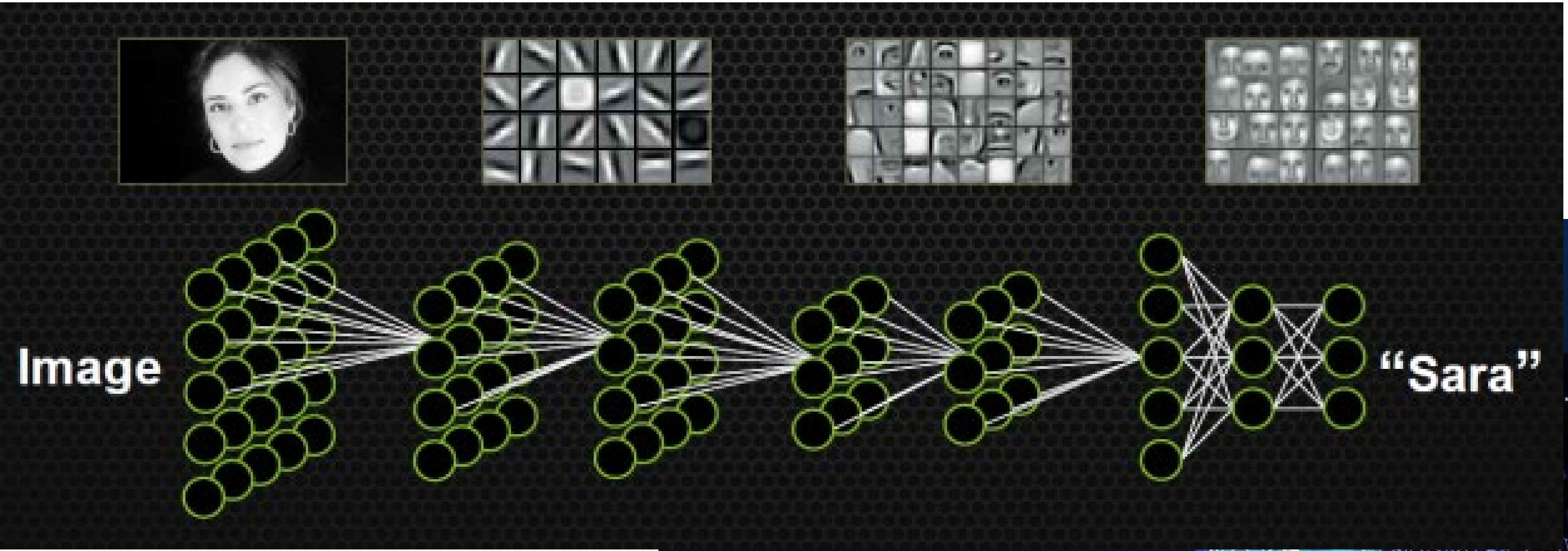
$V_\pi(S_t)$ gives the cumulative reward when following policy π and being in state S_t.

Elements of Reinforcement Learning - Example

Environment	Policy	Rewards	Value Function
An (evolving) data source which is characterized at time t by its state, S_t	$\pi_t(S_t)$ gives a rule by which the agent should act at time t given the state S_t of the environment	A reward function, $r_t(S_t, A_t)$, determines the goal of a reinforcement learning problem.	$V_\pi(S_t)$ gives the cumulative reward when following policy π and being in state S_t.
Stock price of ING at time t plus all previous stock prices	Buy one unit if current stock price is above 100 Sell otherwise	Price tomorrow – price today if we bought one unit. Vice versa, if we sold one unit	Cumulative profit

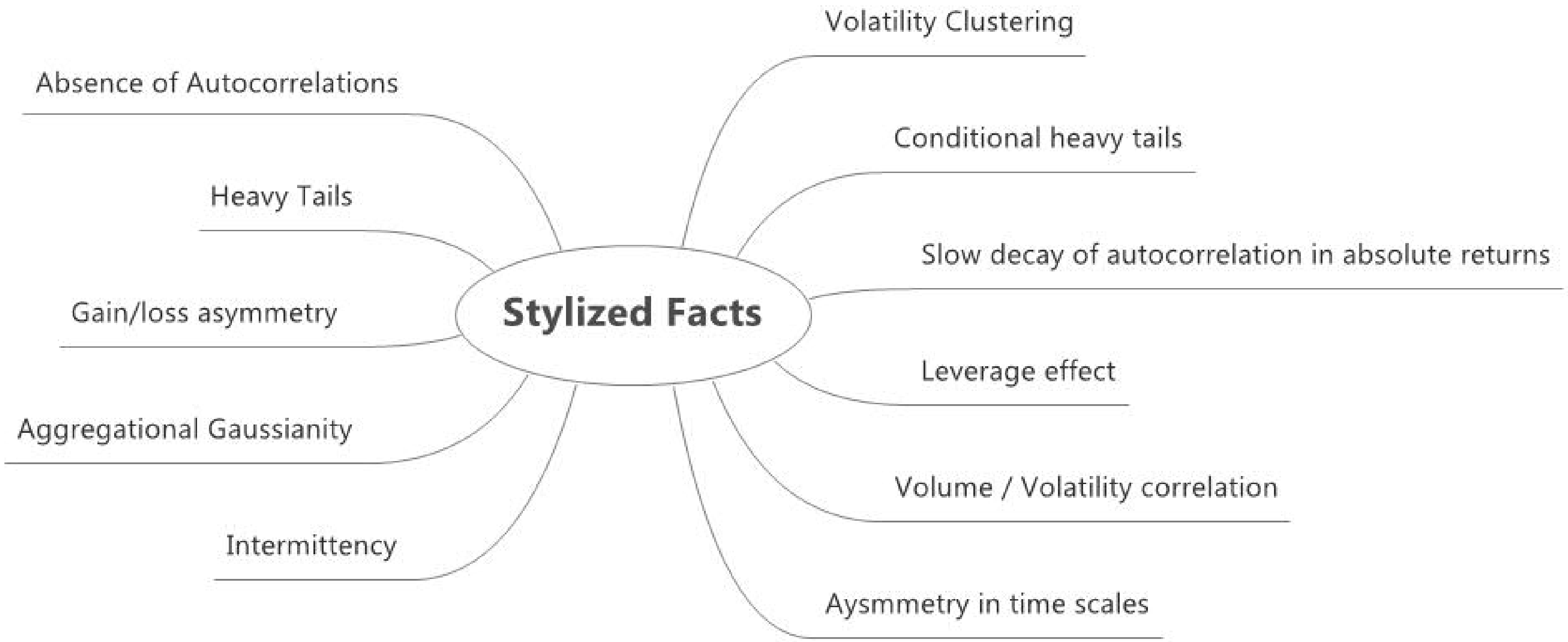
Recap – Finance

Face recognition is easy ?!



The key stylized facts of financial time series

- Absence of autocorrelations
- Fat-tailed distributions
- Volatility clustering
- Gain/loss asymmetry
- Aggregational Gaussianity



Financial time-series data is non-stationary, non-markovian, with non-parametric distributions

https://radhakrishna.typepad.com/rks_musings/2014/06/stylized-facts.html

Applications of Reinforcement Learning in Finance

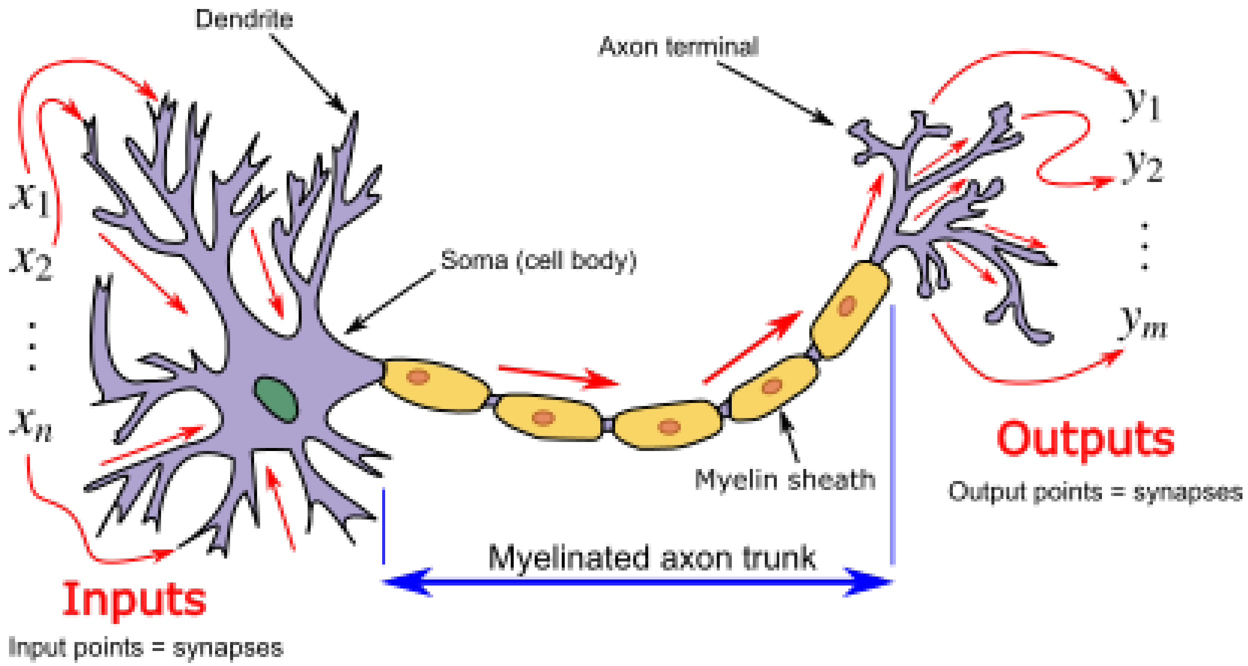
Foundations

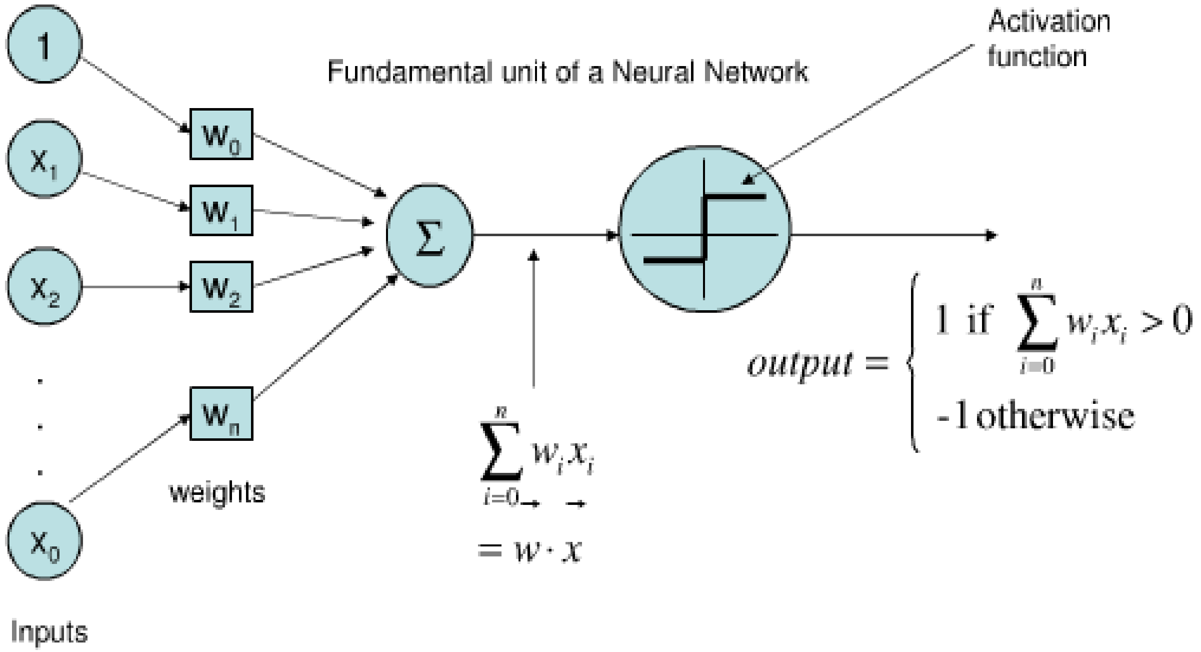
- Electronic markets
- OTC Markets
- Market microstructure
- Market Participants
- Limit Order Books

Applications

- Optimal Execution
- Portfolio Optimization
- Option Pricing and Hedging
- Market Making
- Robo-Advising
- Smart Order Routing

Recap – Neural Networks

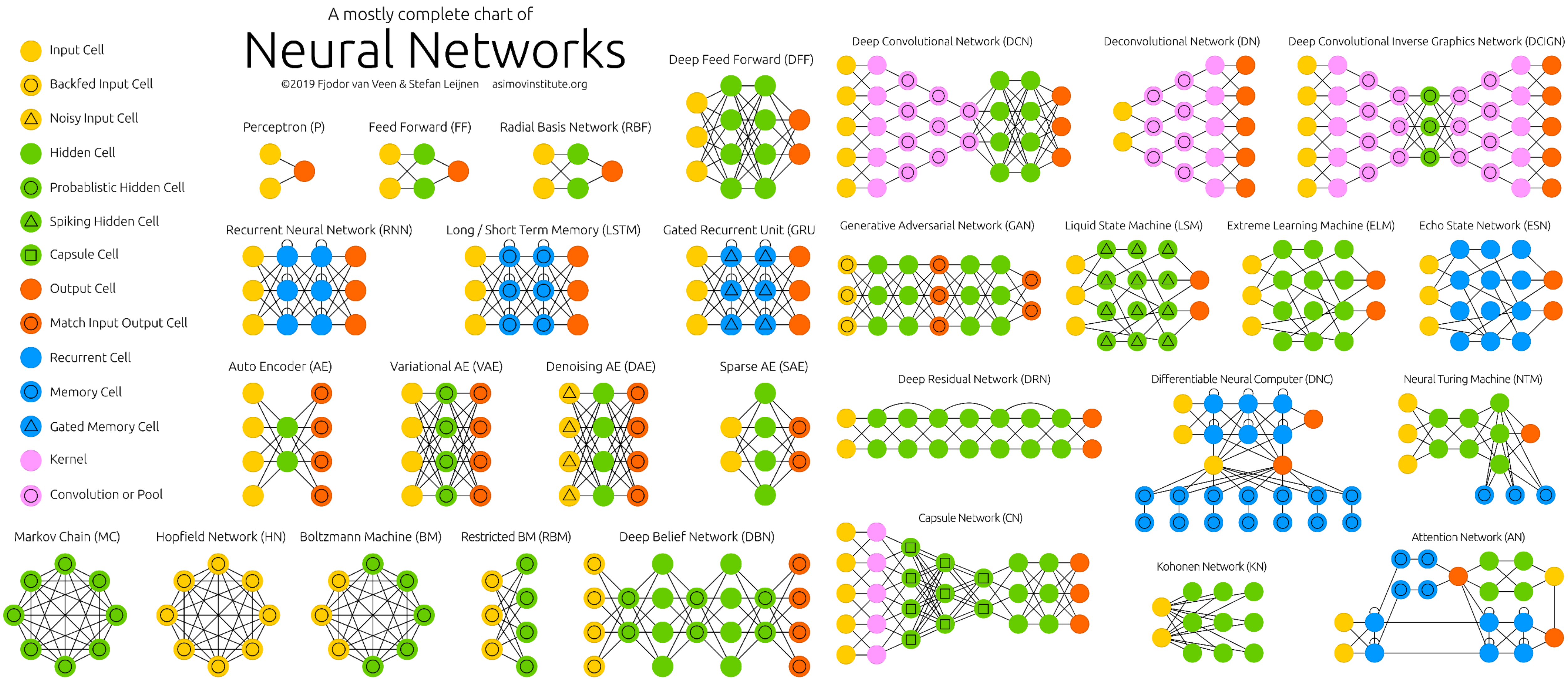




$$f : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

Neural networks (Artificial Intelligence) are functions

Neural networks



Geometric Brownian Motion

Recap – Bellman equations for Markov Decision Processes

$$V^{\pi}(s) = R(s,\pi(s)) + \gamma \sum_{s'} P(s'|s,\pi(s)) V^{\pi}(s').$$

This equation describes the expected reward for taking the action prescribed by some policy.

The equation for the optimal policy is referred to as the Bellman optimality equation:

$$V^{\pi*}(s) = \max_{a} \left\{ R(s,a) + \gamma \sum_{s'} P(s'|s,a) V^{\pi*}(s') \right\}$$

The equation above describes the reward for taking the action giving the highest expected return.

Source: Wikipedia

The importance of global financial markets

Derivatives

https://www.bis.org/statistics/derstats.htm

Global banking review

https://www.mckinsey.com/industries/financial-services/our-insights/global-banking-annual-review

Financial sector statistics

https://data.worldbank.org/topic/7

Global financial markets are one of the main drivers of our global economy

Electronic markets and Market Microstructure

Electronic Markets

Electronic markets have emerged as popular venues for the trading of a wide variety of financial assets

Limit Order Books

An **LOB** is a list of orders that a trading venue uses to record the interest of buyers and sellers in a particular financial asset.

Limit buy (sell) order with a preferred price for a given volume of the asset

Market buy (sell) order with a given volume which will be immediately executed with the best available limit sell (buy) orders

Over-the-counter Markets

OTC trading is done directly between two parties, without the supervision of an exchange.

The electronification process is dominated by Multi-dealer-to-client (MD2C) platforms enabling clients to send the same request for a quote (RFQ) to several dealers simultaneously and therefore put the dealers into competition with one another

Market Participants

Fundamental (or noise or liquidity) traders: those who are driven by economic fundamentals outside the exchange.

Informed traders: traders who profit from leveraging information not reflected in market prices by trading assets in anticipation of their appreciation or depreciation.

Market makers: professional traders who profit from facilitating exchange in a particular asset and exploit their skills in executing trades

A multitude of markets and market participants interacting with each other – theoretical models still need improvements

The Limit Order Book

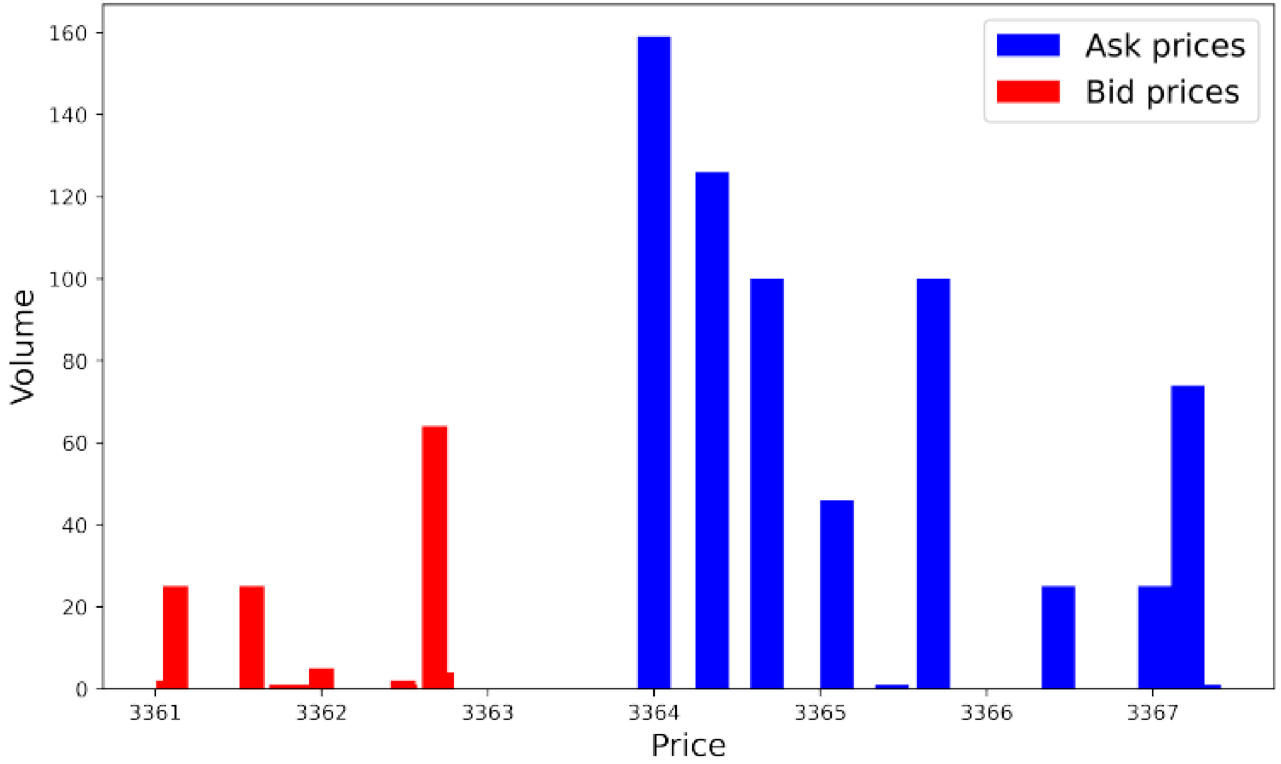
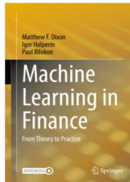


Figure 1: A snapshot of the LOB of AMZN(Amazon) stock at 10:12:52.509 am on 16 April, 2021 with ten levels of bid/ask prices.



Optimal Trade Execution – Market Impact

Market Impact

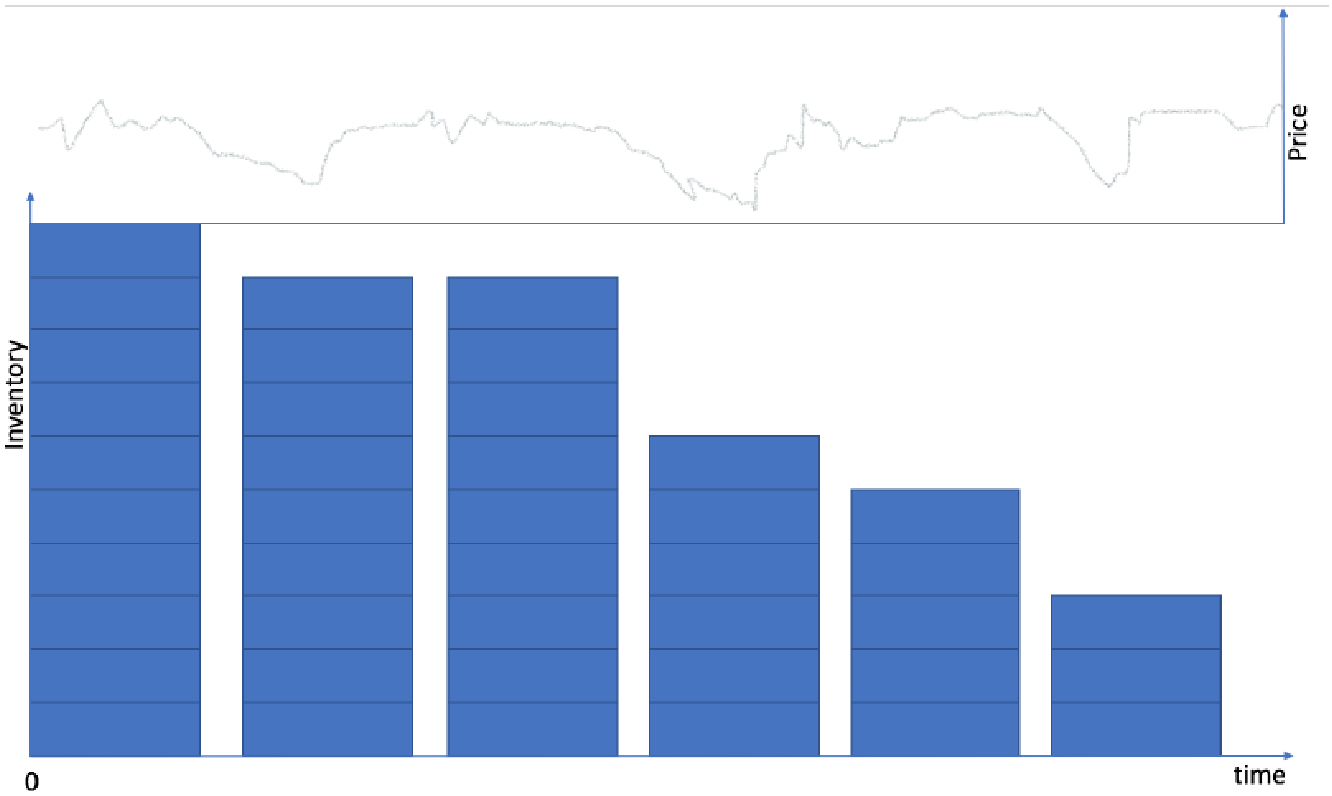
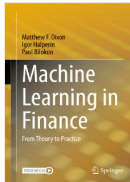


Figure: The optimal execution problem: how to break up large market orders into smaller orders with lower market impact. In the finite MDP formulation, the state space is the inventory, shown by the number of blocks, stock price and time. In this illustration, the agent decides whether to sell $\{0, 1, 2, 3\}$ blocks at each time step. The problem is whether to retain inventory, thus increasing market risk but reduces the market impact, or quickly sell inventory to reduce exposure but increase the market impact.



Optimal Trade Execution

What is it?

- Buy or sell a given amount of a single asset within a given time period
- Minimize the cost of the execution of the transaction
- Benchmarks: VWAP, TWAP, Implementation Shortfall, Submit and Leave policy
- Model of price dynamics needed (permancent and temporary price impact)

Existing models

Theoretical:

Model price dynamics

Given functional form of temporary and permanent price impact

Empirical:

Calibrate temporary and permanent price impact to order execution data for a given functional form

Solution to the optimization problem

Model of the price impact

VWAP

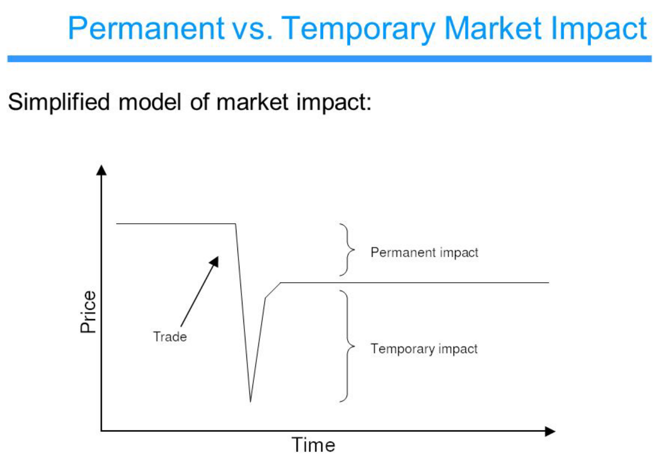
Almgren-Chris Model

Solution depends on functional form of temporary and permanent impact

The precise functional form of permanent and temporary price impact is still unknown; a sound theoretical foundation is also missing

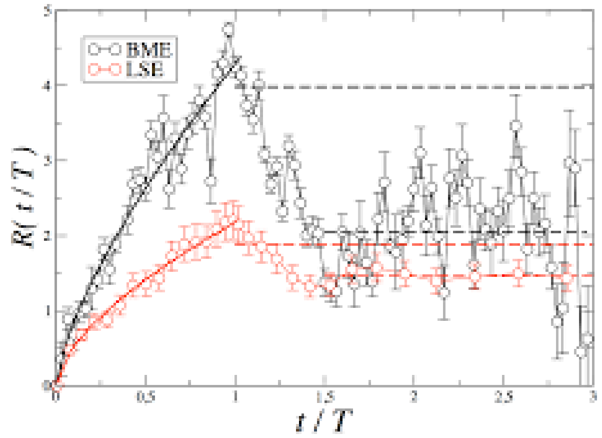
Market impact

Market Impact

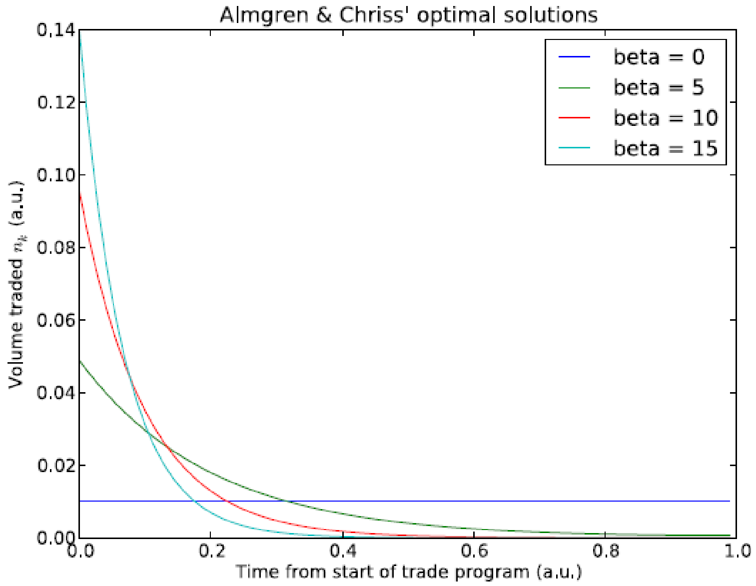


Execution puzzle

Jim Gatheral: https://studylib.net/doc/18172785/the-execution-puzzle--how-and-when-to-trade-to-minimize-cost



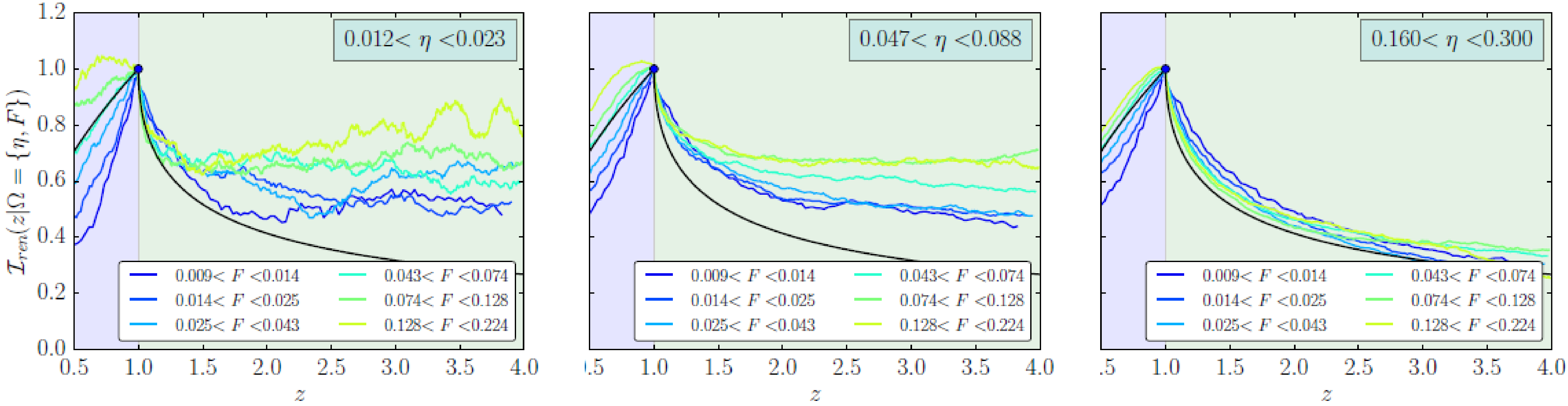
Almgren-Chriss



Any theoretical optimal strategy is highly dependent on the underlying model of the market impact

The price impact trajectory - after the execution

Decay of temporary price impact after the execution of the meteorder. We follow the normalised price impact as a function of the rescaled variable z = v/F.



Source: Lillo, Imperial College

Optimal Trade Execution

Environment/ State

Market data: midprice, spread, traded volume, bid/ask volume, time-of-day, past returns

Order data: Remaining volume to trade, inventory process

Market/ order data: execution performance so far

Benchmark

Time-Weighted Average Price (TWAP)

Volume-Weighted Average Price (VWAP)

Submit and Leave (SnL) policy where a trader places a sell order for all shares at a fixed limit order price, and goes to the market with any unexecuted shares remaining at time T.

Almgren-Chriss solution

Rewards

Performance measures:

Sharpe ratio, portfolio returns, cumulative returns

Rewards:

cash inflow or outflow (depending on whether we sell or buy), implementation shortfall, Profit, Sharpe ratio, Return, PnL

Actions/ Control variables

Volume to trade right now (using market orders)

Volume to post as limit orders (at what price level)

An area currently heavily researched by international investment banks

Portfolio optimization

What is it?

A trader needs to select and trade the best portfolio of assets in order to maximize some objective function, which typically includes the expected return and some measure of the risk

The benefit of investing in such portfolios is that the diversification of investments achieves higher return per unit of risk than only investing in a single aksset

Existing models

Markowitz mean-variance optimization

Y. Sato, Model-free reinforcement learning for financial portfolios: A brief survey, arXiv preprint arXiv:1904.04973, (2019).

Multi-period portfolio optimization, maximing terminal utility (mean – risk aversion * variance)

Risk Parity

Kelly Criterion

Joint asset price dynamics needed

Solution to the optimization problem

Optimal frontier

The Markowitz mean-variance framework has been in use for more than 50 years – despite its fundamental shortcomings

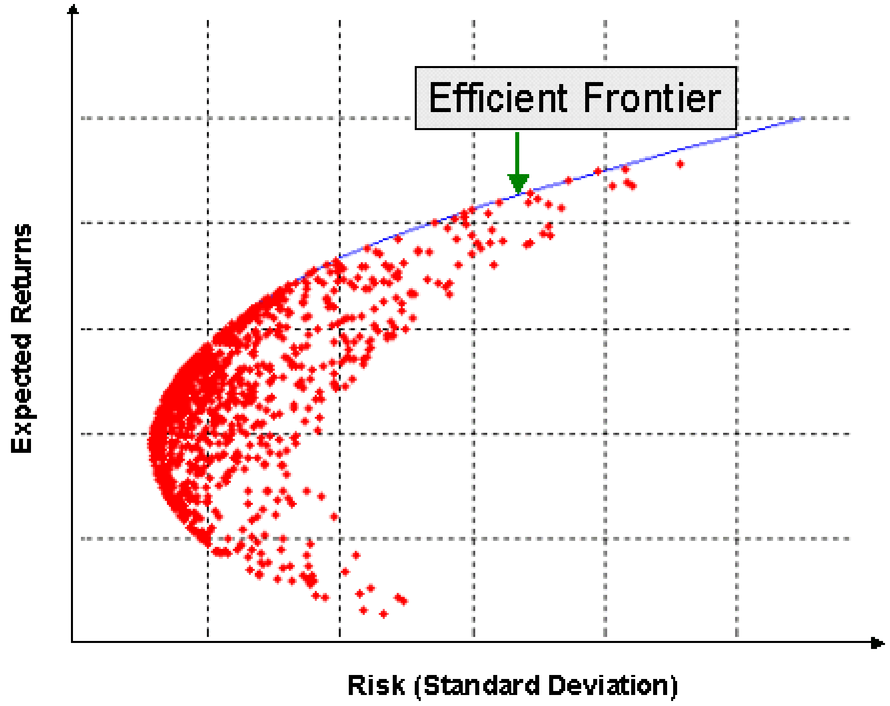
Optimal Frontier of Portfolio Optimization – Mean-Variance Framework

Modern Portfolio Theory

Modern portfolio theory (MPT), or mean-variance analysis, is a mathematical framework for assembling a portfolio of assets such that the expected return is maximized for a given level of risk. It is a formalization and extension of diversification in investing, the idea that owning different kinds of financial assets is less risky than owning only one type. Its key insight is that an asset's risk and return should not be assessed by itself, but by how it contributes to a portfolio's overall risk and return. It uses the variance of asset prices as a proxy for risk.

Economist Harry Markowitz introduced MPT in a 1952 essay, for which he was later awarded a Nobel Memorial Prize in Economic Sciences;

Efficient Frontier



Capital Asset Pricing Model

$$\mathrm{E}(R_i) = R_f + \beta_i(\mathrm{E}(R_m) - R_f)$$

Modern Portfolio Theory has been with us since 1952 – Times have changed

Portfolio optimization

Environment/ State

Market data: Time, asset prices, asset past returns,

Potfolio data: current holdings of assets, remaining balance

Benchmark

Benchmarks:

- Constantly Rebalanced Portfolio
- Buy and-hold strategy

Rewards

Performance measures:

Sharpe ratio, portfolio returns, cumulative returns

Rewards:

portfolio return, (differential) Sharpe ratio, profit

Actions/ Control variables

amount/proportion of wealth invested in each component of the portfolio

Option pricing and hedging

What is it?

A financial derivative is a contract that derives its value from the performance of an underlying entity.

Understanding how to price and hedge financial derivatives is a cornerstone of modern mathematical and computational finance due to its importance in the finance industry.

Existing models

Black-Scholes model

F. Black and M. Scholes, The pricing of options and corporate liabilities, Journal of Political Economy, 81 (1973), pp. 637–654.

R. C. Merton, Theory of rational option pricing, The Bell Journal of Economics and Management Science, (1973), pp. 141–183.

Solution to the optimization problem

Hedging via replication in complete markets

Arbitrage-free bounds in incomplete markets

One of the most active quantitative finance research areas over the last 50 years

Option pricing and hedging

Environment/ State

Asset price, current positions, option strikes, time remaining to expiry

Benchmark

Benchmarks:

- Black-Scholes Merton model
- Binomial option pricing model
- Heston model

Rewards

- (Risk-adjusted) expected wealth/return (as in the mean-variance portfolio optimization)
- Option payoff
- (Risk-adjusted) hedging cost

Performance measures

- (Expected) hedging cost/error/loss
- PnL
- Average payoff

Actions/ Control variables

Change in holdings

Practical considerations (transaction costs, position constraints, limits on trading size) are equally important

Electronic Market making

What is it?

A market maker in a financial instrument is an individual trader or an institution that provides liquidity to the market by placing buy and sell limit orders in the LOB for that instrument while earning the bid-ask spread.

Instead of profiting from identifying the correct price movement direction, the objective of a market maker is to profit from earning the bid-ask spread without accumulating undesirably large positions (known as inventory)

Existing models

LOB dynamics are modeled directly by some stochastic process and an optimal market making strategy that maximizes the market maker's expected utility can be obtained by solving the Hamilton–Jacobi–Bellman equation.

Solution to the optimization problem

Strong model assumptions are made about the prices or about the LOB or both.

This requirement of full analytical specification means these papers are quite removed from realistic market making, as financial markets do not conform to any simple parametric model specification with fixed parameters

Electronic Market Making – Sources of risk

Inventory risk

The inventory risk refers to the risk of accumulating an undesirable large net inventory, which significantly increases volatility due to market movements.

Execution risk

The execution risk is the risk that limit orders may not get filled over a desired horizon.

Adverse selection risk

The adverse selection risk refers to the situation where there is a directional price movement that sweeps through the limit orders submitted by the market marker such that the price does not revert back by the end of the trading horizon. This may lead to a huge loss as the market maker in general needs to clear their inventory at the end of the horizon

Risk – reward trade-off important

Electronic Market Making

Environment/ State

Bid and ask prices, current holdings of assets, order-flow imbalance, volatility, and some sophisticated market indices

Agent states include inventory level and active quoting distances Market states include market (bid-ask) spread, midprice move, book/queue imbalance, signed volume, volatility, relative strength index.

Benchmark

Zero profit

Quote one share 100% of the time

Rewards

PnL with inventory cost or Implementation Shortfall with inventory cost

Linear combination of:

- profit (to maximize)
- inventory risk (to minimize)
- Market qualities (to maximize)

Actions/ Control variables

spread to post a pair of limit buy and limit sell orders

sizes of the limit buy/sell orders

Multi-agent RL algorithms are also used to improve the strategy for market making with a particular focus on the impact of competition from other market makers or the interaction with other types of market participant.

Electronic Market Making – A deep dive

Source:

Electronic Market Making

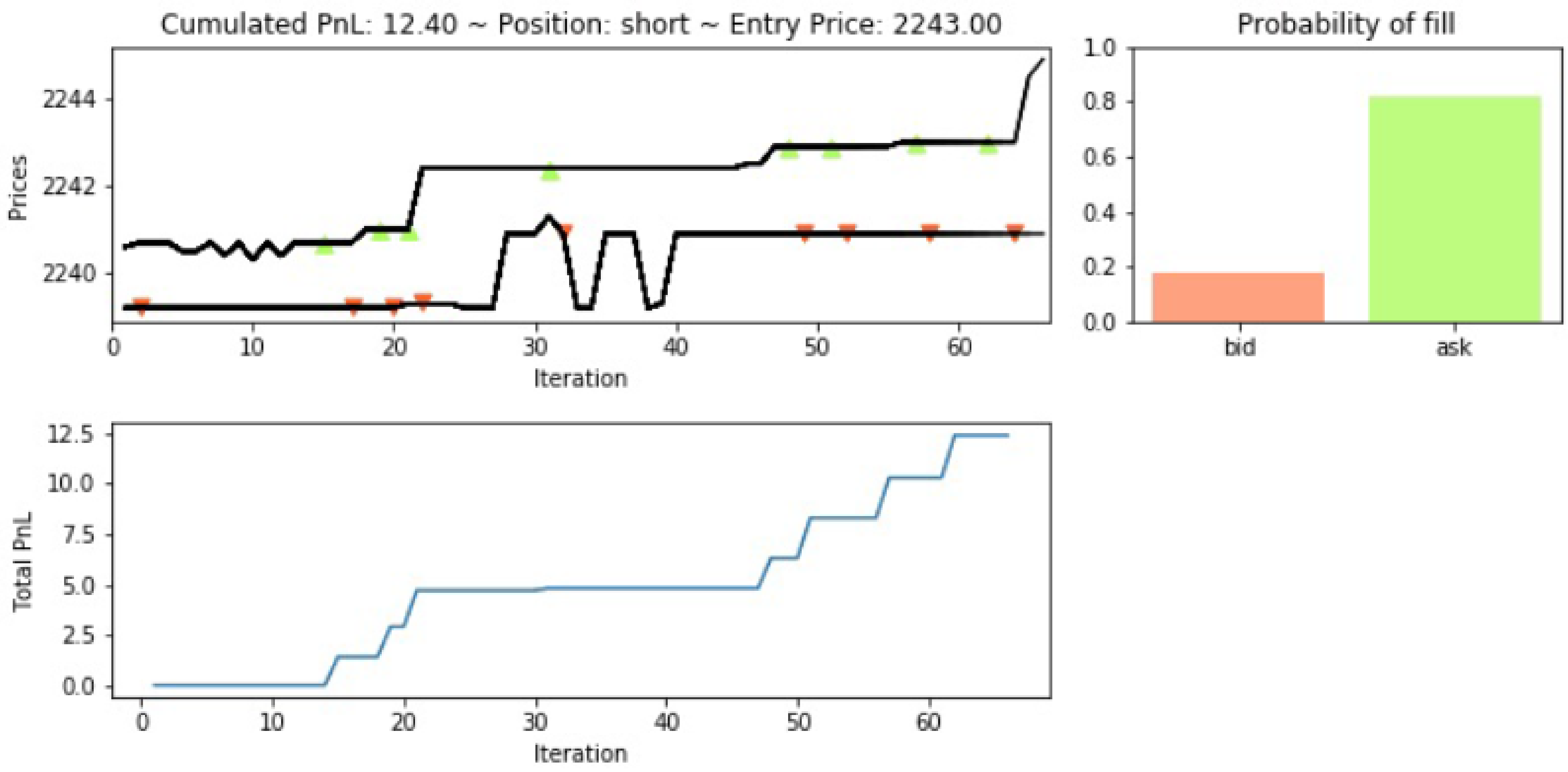
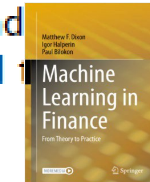


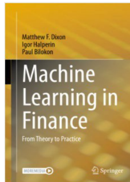
Figure: The market making problem requires the placement of bid and ask quotes to maximize P&L, while maintaining a position within limits. For each limit order book update, the agent must anticipate which quotes shall be filled to capture the bid-ask spread Transaction costs, less than a tick, are imposed to penalize trading without capturing at least a tick. A simple model is introduced t determine the fill probabilities and the state space is the product of the position and gridded fill probabilities.



The limit order book

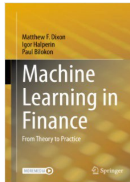
- The limit order book $\mathcal{X}_t$ consists of all depths and prices at time t (Level 2 data).
- Inventory $I_t \in \mathcal{I} \subset \mathcal{Z}$ represents the current position.
- At any time, our agent can place (or cancel and replace) ask and bid limit orders of size L at level i, L_i^a and L_i^b or place market orders M^b and M^s
- Encode this information as

$$\begin{array}{ll} \text{States:} & (\mathcal{X}_t, I_t) \in \mathbb{S} \\ \text{Actions:} & (\{L_i^a\}_{i=1}^d, \{L_i^b\}_{i=1}^d, M^b, M^s) \end{array}$$



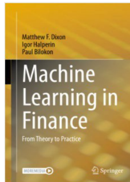
Toy Example: Electronic Market Making

- Assume that a market maker seeks to capture the bid-ask spread by simply placing one lot best bid and ask limit orders only.
- They are required to strictly keep their inventory between -1 and 1.
- The problem is when to optimally bid to buy ('b'), bid to sell ('s') or hold ('h'), each time there is a limit order book update.
- For example, sometimes it may be more advantageous to quote a bid to close out a short position if it will almost surely yield an instantaneous net reward, other times it may be better to wait and capture a larger spread.
- The agent uses the liquidity imbalance in the top of the order book as a predictor of price movement and, hence, fill probabilities.
- The example does not use market orders, knowledge of queue positions, cancellations and limit order placement at different levels of the ladder.



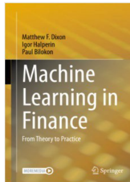
Toy Example: Electronic Market Making

- At each non-uniform time update, t, the market feed provides best prices and depths $\{p_t^a, p_t^b, q_t^a, q_t^b\}$.
- The state space is the product of the inventory, $I_t \in \{-1, 0, 1\}$, and gridded liquidity ratio $\hat{R}_t = \lfloor \frac{q_t^a}{q_t^a + q_t^b} N \rfloor \in [0, 1]$, where N is the number of grid points and q_t^a and q_t^b are the depths of the best ask and bid. $\hat{R}_t \to 0$ is the regime where the mid-price will go up and an ask is filled.
- Conversely for $\hat{R}_t \to 1$. The dimension of the state space is chosen to be $3 \times 10 = 30$.



Toy Example: Electronic Market Making

- A bid is filled with probability $\epsilon_t := \hat{R}_t$ and an ask is filled with probability $1 - \epsilon_t$.
- The rewards are chosen to be the expected total P&L.
- If a bid is filled to close out a short holding, then the expected reward $r_t = -\epsilon_t(\Delta p_t + c)$, where Δp_t is the difference between the exit and entry price and c is the transaction cost.
  - For example, if the agent entered a short position at time $s < t$ with a filled ask at $p_s^a = 100$ and closed out the position with a filled bid at $p_t^b = 99$, then $\Delta p_t = 1$.
- The agent is penalized for quoting an ask or bid when the position is already short or long respectively.



Robo-advising

What is it?

Robo-advisors, or automated investment managers, are a class of financial advisers that provide online financial advice or investment management with minimal human intervention. They provide digital financial advice based on mathematical rules or algorithms. Robo-advisors have gained widespread popularity and emerged prominently as an alternative to traditional human advisers in recent years

Existing models

Challenges:

- client's risk preference may change over-time and may depend on the market returns and economic conditions.
- dilemma when it comes to investing according to the client's risk preferences or going against the client's wishes in order to seek better investment performance.
- trade-off between the rate of information acquisition from the client and the accuracy of the acquired information.

Solution to the optimization problem

Stochastic control framework with four components [1]:

- a regime switching model of market returns
- a mechanism of interaction between the client and the robo advisor
- a dynamic model (i.e., risk aversion process) for the client's risk preferences
- an optimal investment criterion.

[1] A. Capponi, S. Olafsson, and T. Zariphopoulou, Personalized robo-advising: Enhancing investment through client interaction, Management Science, (2021).

Robo-advising

Environment/ State

Set of various market environments of interest

Benchmark

Learning the risk preferences from investment portfolios using an inverse optimization technique.

Measure time varying risk preferences directly from market signals and portfolios. This approach is developed based on two methodologies: convex optimization based modern portfolio theory and learning the decision-making scheme through inverse optimization.

Rewards

Portfolio return

Actions/ Control variables

In each period, the robo-advisor places an investor's capital into one of several pre-constructed portfolios which can be viewed as the action space

Each portfolio decision reflects the robo-advisor's belief concerning the investor's true risk preference from a discrete set of possible risk aversion parameters

Smart-order routing

What is it?

In order to execute a trade of a given asset, market participants may have the opportunity to split the trade and submit orders to different venues, including both lit pools and dark pools, where this asset is traded. This could potentially improve the overall execution price and quantity. Both the decision and hence the outcome are influenced by the characteristics of different venues as well as the structure of transaction fees and rebates across different venues.

Existing models

Dark Pools vs. Lit Pools.

R. Cont and A. Kukanov, Optimal order placement in limit order markets, Quantitative Finance, 17 (2017), pp. 21–39.

Allocation Across Dark Pools.

B. Baldacci and I. Manziuk, Adaptive trading strategies across liquidity pools, arXiv preprint arXiv:2008.07807, (2020).

Solution to the optimization problem

For market orders: Best price given available liquidity

For limit orders: (Over) allocate to exchanges with higher historical fill rates; better historical prices

Smart-order routing

Environment/ State

Exercise 1

Benchmark

Exercise 1

Rewards

Exercise 1

Actions/ Control variables

Exercise 1

A traditionally heuristics-based approach moves to a modern approach

Further Developments for Mathematical Finance and Reinforcement Learning

Risk-aware or Risk-sensitive RL

Offline Learning and Online Exploration

Learning with a Limited Exploration Budget

Learning with Multiple Objectives

Learning to Allocate Across Lit Pools and Dark Pools

Robo-advising in a Model-free Setting

Sample Efficiency in Learning Trading Strategies

Transfer Learning and Cold Start for Learning New Assets

Recap

Reinforcement Learning

Environment

Agents

Rewards

States

Actions

Financial Markets

Electronic markets and Market Microstructure

Optimal Trade Execution

Option pricing and hedging

Electronic Market making

Robo-advising

Smart-order routing

Recap: Financial Applications

Optimal Execution of Orders in the market

- When buying or selling assets, orders must be placed into the market.
- If orders are large, they will affect prices
    - To avoid this, they must be split across multiple child orders so that to not have too much impact on the price of the asset.
    - Optimal trade execution requires a balance between getting the best price without influencing the market too much.

Can you identify examples of the components of RL in this use case? (i.e.: action, reward, states)

Recap: Financial Applications

Hedging Derivatives

- A derivative is a financial product; its value is dependent on another, underlying asset.
- Example: Put- or Call options.
- Fundamental questions for sell side bank:
- How can we **hedge the risk** of it from the bank's books?
- Hedging requires balancing a **trade off between rebalancing often (to adjust the hedge) and trading costs**.

Can you identify examples of the components of RL in this use case? (i.e.: action, reward, states)

Recap: Financial Applications

Investment Portfolio Allocation

- Managing a portfolio requires sequential decisions about what to buy and what to sell at each given point in time.
- As we will see later, RL can be applied for instance for problems like Goal Based Wealth Management (GBWM)
- GBWM requires **repeated rebalancing of assets in order to fulfill a required, future path of wealth** (for instance: saving for university, etc.)

Can you identify examples of the components of RL in this use case? (i.e.:
action, reward, states)

Agenda

Financial Markets

Electronic markets and Market Microstructure

Optimal Trade Execution

Option pricing and hedging

Electronic Market making

Robo-advising

Smart-order routing

Artificial Intelligence – Neural Networks

What is it?

Data

The Mathematics

Computing Power

The universal approximation theorem

AI and its applications in Finance

Regulatory considerations

Fully connected NNs

RNNs

CNNs

Deep Reinforcement Learning

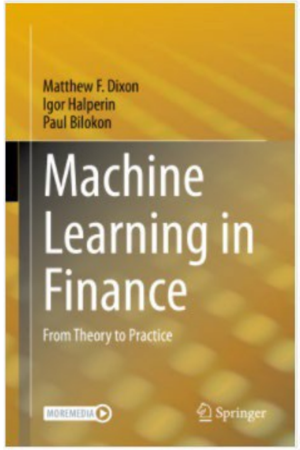
SARSA

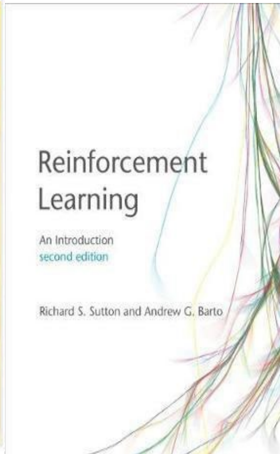
Q-Learning

Market Impact in Finance

Market Making in Finance

Finance meets Reinforcement Learning meets Deep Learning





Homework assignment

Deep RL in finance

1. Find one example of a Deep RL algorithm on github (in Python) for one application in financial markets.

2. Find one academic paper describing that application using deep RL.

3. Write a short summary (1-3 pages, plus 3 ppt slides), focusing in particular on

1) Problem statement
2) Deep RL algorithm
3) Solution
4) Shortcoming of the method

Comments

Please run the python code.

The exam could have a question on extending a specific deep RL implementation.

Please describe the components of the RL setup as precisely as possible (rewards, states, environment, …)

Examples you can choose

Market Making
https://github.com/mfrdixon/ML_Finance_Codes/blob/master/Chapter9-Reinforcement-Learning/ML_in_Finance_Market Making.ipynb

Market Impact
https://github.com/mfrdixon/ML_Finance_Codes/blob/master/Chapter9-Reinforcement-Learning/ML_in_Finance_Market_Impact.ipynb