

ML для портфельной теории Марковица

Полузёров Тимофей Дмитриевич

Структура работы

- 1 Портфельная теория
 - Постановка задачи
 - Оптимизационная задача
- 2 Методы оценки характеристик доходностей
 - Линейная регрессия
 - Случайный лес
 - ARIMA
- 3 Проверка стратегий на реальных данных
 - Обзор данных
 - Оценка средних доходностей
 - Оценка ковариаций доходностей
 - Результаты оценки стратегий
- 4 Выводы

Однопериодная задача инвестирования

Доступны N активов. S_i^0, S_i^1 — цены i -го актива в моменты времени $t = 0$ и $t = 1$ соответственно.

Доходность актива за период

$$r_i := \frac{S_i^1 - S_i^0}{S_i^0}, i = \overline{1, N}$$

Необходимо сформировать портфель

$$b = (b_1, \dots, b_N)$$

где b_i — число приобретаемых активов i -го типа.

Портфель покупается в момент $t = 0$ и продается в момент $t = 1$ по рыночным ценам.

Доходность портфеля [4, 5, 6, 8]

Пусть инвестор имеет капитал x .

Перейдем к долям инвестирования капитала x в доступные активы.

$$\omega_i := \frac{b_i S_i^0}{x}, i = \overline{1, N}$$

Цена портфеля в момент времени $t = 0$

$$X^0 = x$$

В момент $t = 1$

$$X^1 = (1 + R)X^0 = \sum_{i=1}^N \omega_i r_i = \omega^T r$$

где R — доходность портфеля

Характеристики доходности портфеля

Предположим, что случайные величины доходностей $r_i, i = \overline{1, N}$ известны, тогда матожидание и дисперсия случайной величины доходности портфеля равны

$$\mu_X := \mathbb{E}[R] = \sum_{i=1}^N \omega_i \mathbb{E}[r_i] = \omega^T \mu$$

$$\sigma_X^2 := \sum_{i=1, j=1}^N \omega_i \omega_j \text{Cov}(r_i, r_j) = \omega^T \Sigma \omega$$

где $\mu = \mathbb{E}[r]$, а Σ — ковариационная матрица случайного вектора r .

Интерпретация характеристик доходности портфеля

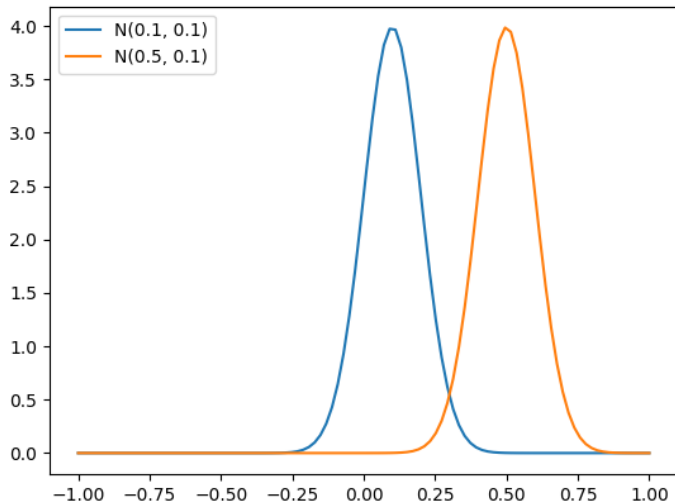


Рис.. Предпочтения в средней доходности портфеля

Дисперсия как мера риска

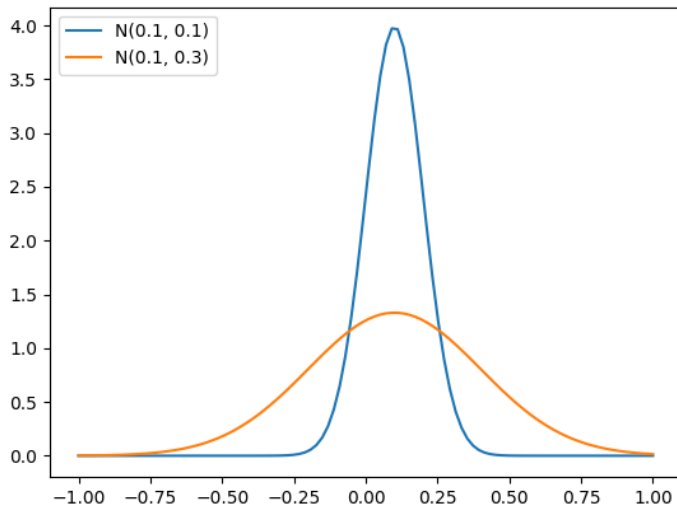


Рис.. Предпочтения в риске портфеля

Оптимизационная задача

Введем параметр $\tau \in [0, +\infty)$ — толерантность к риску и сформулируем критерий оптимальности

$$f(\mu_X, \sigma_X^2) = \tau \omega^T \Sigma \omega - \omega^T \mu \rightarrow \min_{\omega}$$

Оптимизационная задача имеет вид

$$\begin{cases} \tau \omega^T \Sigma \omega - \omega^T \mu \rightarrow \min_{\omega} \\ \omega^T e = 1 \\ \omega \geq 0 \end{cases}$$

При $\tau = 0$ имеем портфель максимальной доходности, при $\tau \rightarrow +\infty$ — минимального риска.

Пример

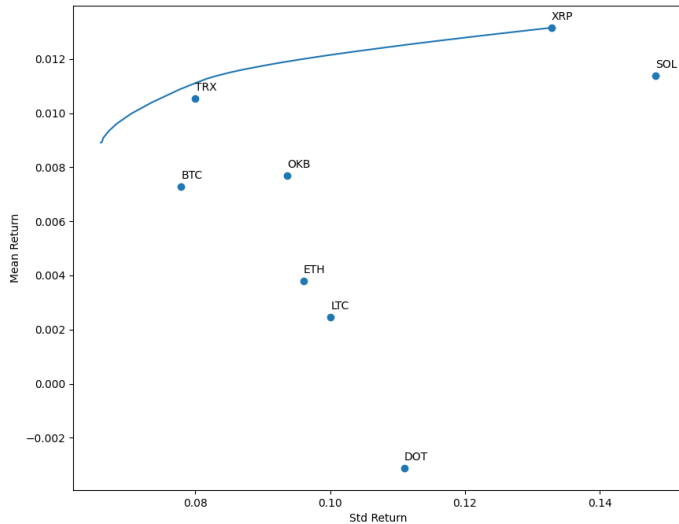


Рис.. Множество оптимальных портфелей

Подходы к оценке характеристик

На практике в момент времени $t = 0$ случайные величины доходностей r неизвестны.

Необходимо оценить характеристики $\hat{\mu}_X$ и $\hat{\Sigma}$.

Известные подходы:

- Модель CAPM (Capital Asset Pricing Model), У. Шарп, Дж Линтер [9, 13]
- Теория APT (Arbitrage Pricing Theory), С. Росс, Р. Ролл. [11, 12]

Предлагается иной подход — применить методы машинного обучения и модели временных рядов для решения задачи регрессии по историческим данным.

Линейная регрессия [1], [2]

Пусть X — матрица объектов-признаков, где признаки это лаги ряда, y — соответствующие истинные доходности. $\theta = (\theta_0, \dots, \theta_k)$ — параметры.

$$a(x) = \theta_0 + \theta_1 x_1 + \dots + \theta_k x_k$$

Оптимальные веса θ^* определяются методом наименьших квадратов

$$Q(a) = \|X\theta - y\|^2 \rightarrow \min_{\theta}$$

Случайный лес [2, 7]

Дерево решений — это бинарное дерево. Вершины двух типов:

- внутренние — содержат предикат $b_v : \mathbb{X} \rightarrow \{0, 1\}$
- листовые — хранят выходное значение $c_v \in \mathbb{Y}$

Обработка входящего объекта x :

- 1 старт из корня
- 2 вычисляем значение текущего предиката $b_v(x)$
- 3 если значение $b_v(x) = 0$ делаем шаг в левое поддереву, иначе — в правое
- 4 пока не дошли до листа повторяем шаги 2 и 3
- 5 возвращаем значение c_v из текущего листа

Случайный лес [2, 7]

Случайный лес есть усреднение набора решающих деревьев

$$a(x) = \frac{1}{k} \sum_{i=1}^k b_i(x)$$

Среднеквадратичную ошибку можно разложить на слагаемые:

$$Q(a) = \textit{bias}(a) + \textit{variance}(a) + \textit{noise}$$

где

$$\textit{bias}(a) := f(x) - \mathbb{E}_X [a(x, X)]$$

$$\textit{variance}(a) := \mathbb{E}_X [a(x, X)] - \mathbb{E}_X [a(x, X)]^2$$

$$\textit{noise} := \mathbb{E}_X \left[\mathbb{E}_\varepsilon \left[(y(x, \varepsilon) - f(x))^2 \right] \right]$$

Случайный лес [2, 7]

Смещение композиции равно смещению базового алгоритма

$$bias(a) = bias_X(b)$$

Разброс композиции определяется размером композиции и коррелированностью базовых алгоритмов

$$variance(a) = \frac{1}{k^2} \sum_{i=1}^k variance_X(b_i) + \frac{1}{k^2} \sum_{i \neq j} Cov(b_i, b_j)$$

ARIMA [5]

Модель $ARMA(p, q)$ объединяет модели авторегрессии $AR(p)$ и скользящего среднего $MA(q)$.

$$\begin{aligned}x_n = & (a_0 + a_1x_{n-1} + \cdots + a_px_{n-p}) + \\& + (b_1\varepsilon_{n-1} + b_2\varepsilon_{n-2} + \cdots + b_q\varepsilon_{n-q}) \\& + \sigma\varepsilon_n\end{aligned}$$

где $\varepsilon = (\varepsilon_n)$ — белый шум.

Разности процесса

$$\Delta x_n = x_n - x_{n-1}$$

Символически $ARIMA(p, d, q)$ выражается через модель $ARMA(p, q)$

$$\Delta^d ARIMA(p, d, q) = ARMA(p, q)$$

Входные данные

Общие характеристики данных:

- Криптовалютная биржа OKX [3]
- 8 наиболее популярных активов
- Временной период с 1 января 2022 г. по 1 января 2025 г.
- Дневной таймфрейм
- Период инвестирования — 1 неделя

Проведение тестирования:

- Данные до 1 января 2024 г. используются для подбора числа лагов и гиперпараметров моделей
- Тестирование проводится на данных за 2024 год



Рис.. Цены активов

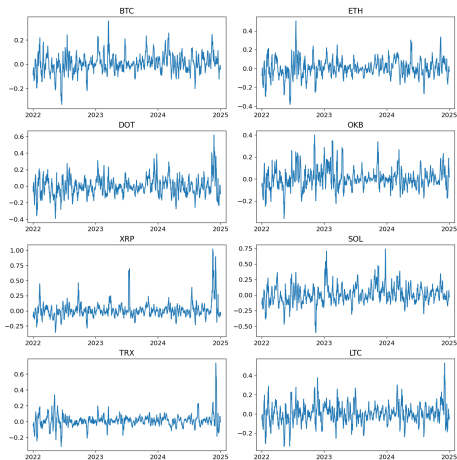


Рис.. Доходности активов

Таблица. Доходности активов

	BTC	ETH	DOT	OKB	XRP	SOL	TRX	LTC
mean	0.0073	0.0038	-0.0031	0.0077	0.0132	0.0114	0.0105	0.0025
std	0.0779	0.0961	0.1110	0.0936	0.1328	0.1482	0.0800	0.1001
min	-0.3328	-0.3830	-0.3925	-0.3591	-0.3596	-0.6018	-0.3162	-0.3392
25%	-0.0358	-0.0477	-0.0724	-0.0401	-0.0506	-0.0765	-0.0236	-0.0513
50%	0.0026	-0.0013	-0.0084	-0.0016	-0.0003	-0.0034	0.0106	0.0001
75%	0.0446	0.0555	0.0573	0.0494	0.0430	0.0906	0.0373	0.0546
max	0.3566	0.5056	0.6188	0.4003	1.0235	0.7409	0.7407	0.5294

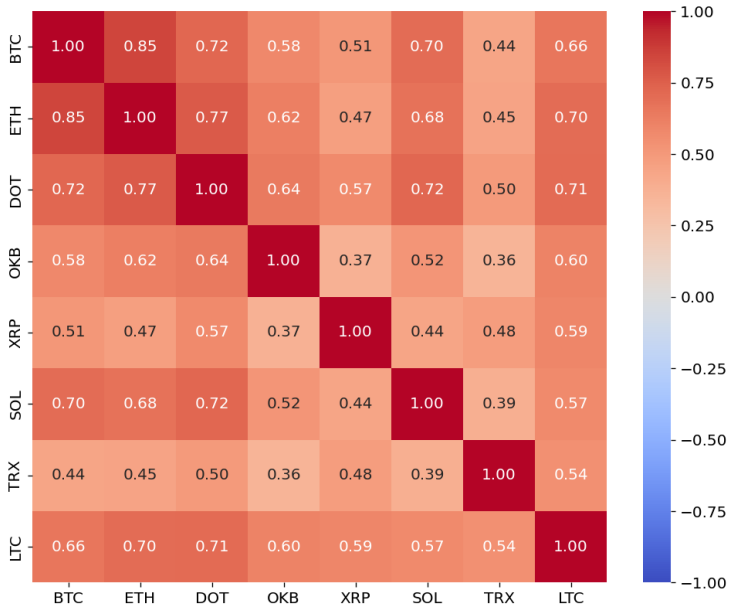


Рис.. Корреляции доходностей активов

Модели оценки средней доходности $\hat{\mu}$

- 1 NAIVE — среднее выборочное
- 2 MARTINGAL — последнее наблюдаемое
- 3 ARIMA — модель ARIMA
- 4 LR — линейная регрессия
- 5 RF — случайный лес

Качество прогнозирования оценивается по метрике

$$MSE = \frac{1}{n} \sum_{i=1}^n (\mu_i - \hat{\mu}_{ii})^2$$

где μ_i - истинное значение доходности, а $\hat{\mu}_{ii}$ - прогнозное значение модели на i -м объекте тестовой выборки.

Таблица. Качество прогнозирования, $MSE \cdot 10^4$

	NAIVE	MARTINGAL	LR	ARIMA	RF
BTC	5.63	1.20	1.58	1.62	2.07
ETH	8.00	1.99	4.47	3.70	5.05
DOT	16.51	3.89	4.09	3.98	5.28
OKB	6.06	1.56	1.77	1.95	2.05
XRP	24.33	5.04	6.98	5.71	6.53
SOL	21.19	4.47	11.86	6.16	5.31
TRX	7.96	1.85	4.19	5.30	6.04
LTC	8.53	2.66	2.24	3.22	5.11

Оценка матрицы ковариаций $\hat{\Sigma}$

Предполагается стационарность ковариаций во времени.

Имея r_t - вектор-столбец доходностей в момент времени t , по истории наблюдений r_1, \dots, r_n выборочная ковариация $\hat{\Sigma}$ рассчитывается как

$$\hat{\Sigma} = \frac{1}{n} \sum_{t=1}^n (r_t - \bar{r}) \cdot (r_t - \bar{r})^T$$

где $\bar{r} = \frac{1}{n} \sum_{t=1}^n r_t$.

Стратегия — принцип по которому в каждый момент времени формируется портфель.

Стратегии Марковица определяются моделью, лежащей в основе оценок $\hat{\mu}$ и $\hat{\Sigma}$, а так же зависят от параметра τ .

Тривиальные стратегии:

- ① UNIFORM — равномерное инвестирование во все активы
- ② MOST RISKY — наиболее рискованный
- ③ LESS RISKY — наименее рискованный
- ④ BEST RETURN — с наибольшей доходностью
- ⑤ WORST RETURN — с наименьшей доходностью

Доходность стратегии **ROI** (Return On Investment) определяется аналогично доходности портфеля.

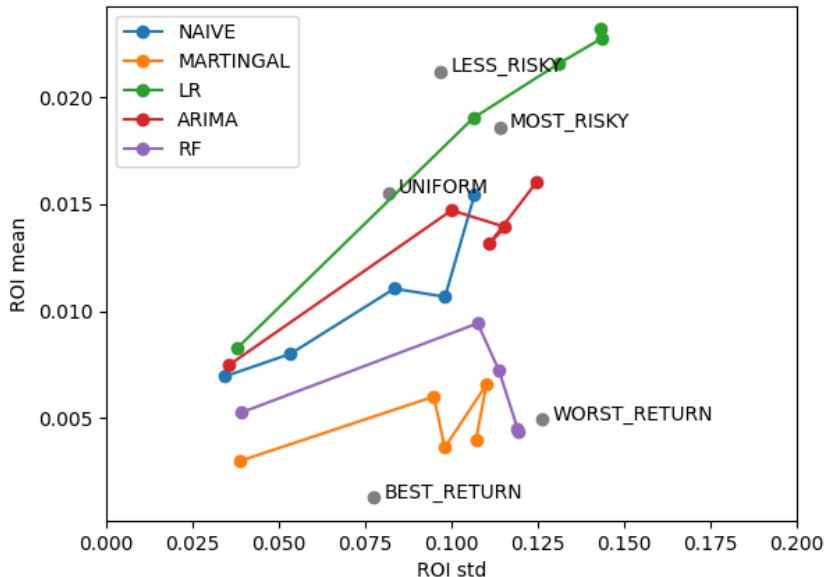


Рис.. Результаты тестирования стратегий

Таблица. Тривиальные портфели

	mean ROI $\cdot 10^3$	std ROI $\cdot 10^2$
UNIFORM	15.5372	8.1775
MOST RISKY	18.5904	11.3970
LESS RISKY	21.1635	9.6881
BEST RETURN	1.2790	7.7329
WORST RETURN	4.9217	12.6324

Таблица. Средние ROI $\cdot 10^3$

	0.01	0.25	0.50	0.75	1.00
NAIVE	6.9451	8.0025	11.0462	10.6657	15.4227
MARTINGAL	2.9859	6.0019	3.6178	6.5656	3.9942
LR	8.2600	19.0185	21.5537	22.7442	23.1668
ARIMA	7.4648	14.7066	13.9296	13.1520	15.9971
RF	5.2633	9.4236	7.2398	4.3777	4.5050

Таблица. Стандартное отклонение ROI $\cdot 10^2$

	0.01	0.25	0.50	0.75	1.00
NAIVE	3.4328	5.3375	8.3427	9.8131	10.6650
MARTINGAL	3.8577	9.4886	9.8051	11.0111	10.7128
LR	3.7892	10.6314	13.1000	14.3570	14.3071
ARIMA	3.5454	10.0037	11.5506	11.1001	12.4518
RF	3.9215	10.7564	11.3710	11.9397	11.9118

Выводы

- активы имеют сильную положительную корреляцию
- стремление сформировать портфель с большей доходностью влечет большие риски
- диверсификация действительно позволяет снижать риск портфеля
- как правило, формирование портфеля доминирует над инвестированием в отдельные активы
- линейная модель авторегрессии показала лучшее качество для оценки средней доходности

Дальнейшие шаги

- рассмотреть другие классы методов прогнозирования временных рядов
- помимо авторегрессионных признаков, учесть влияние внешних факторов на формирование цен
- расширить рассматриваемый набор активов
- исследовать другие таймфреймы и периоды инвестирования

Использованные источники

- ❶ Библиотека Python для прогнозирования временных рядов с использованием моделей машинного обучения [Электронный ресурс]. — Режим доступа: <https://skforecast.org/0.15.1/>. — Дата доступа 25.05.2025
- ❷ Библиотека машинного обучения с открытым исходным кодом. [Электронный ресурс]. — Режим доступа: <https://scikit-learn.org/stable/>. — Дата доступа 25.05.2025
- ❸ Криптовалютная биржа с расширенными финансовыми предложениями [Электронный ресурс]. — Режим доступа: <https://www.okx.com/>. — Дата доступа: 25.05.2025
- ❹ Шарп, У.Ф. Инвестиции : учебник : пер. с англ. / У.Ф. Шарп, Г.Д. Александер, Д.В. Бэйли. — Москва : ИНФРА-М,, 2022. — 1028 с.
- ❺ Ширяев, А. Н. Основы стохастической финансовой математики: Т.1: Факты, модели / А. Н. Ширяев – МЦНМО, 2016. – 440 с.
- ❻ Ширяев, А. Н. Основы стохастической финансовой математики: Т.2: Теория / А. Н. Ширяев – МЦНМО, 2016. – 464 с.

Использованные источники

- 7 Breiman, L. Random Forests. Machine Learning / Leo Breiman — Statistics Department University of California Berkele, CA, 2001, 33 p.
- 8 Markowitz, H. Portfolio selection / H. Markowitz — The Journal of Finance, March 1952, 77-91 p.
- 9 Linter J. The valuation of risky assets and the selection of risky investments on stock portfolios and capital budgets / John Linter — Review of Economics and Statistics, February 1965, 13-34 p.
- 10 Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System / Satoshi Nakamoto — Japan, 2008, 9 p.
- 11 Roll R., Ross S. A. An emperical investigation of the arbitrage pricing theory / Richard Roll, Stephen Ross — Journal of Finance, 1980, 1073-1103 p.
- 12 Ross S. A. The arbitrage theory of capital asset pricing / Stephen A. Ross — Journal of Finance, 1989, 1-18 p.
- 13 Sharpe W. F. Capital asset prices: A theory of market equilibrium under conditions of risk / William F. Sharp — Journal of Finance,

Спасибо за внимание.