

СИНГУЛЯРНОЕ РАЗЛОЖЕНИЕ И ЛИНЕЙНАЯ РЕГРЕССИЯ В ЗАДАЧЕ ПРОГНОЗИРОВАНИЯ ДИНАМИКИ ПОВЕДЕНИЯ ЦЕННЫХ БУМАГ

SINGULAR VALUE DECOMPOSITION AND LINEAR REGRESSION IN THE TASK OF PREDICTING STOCK MARKET DYNAMICS

A. Solobuto
V. Pavlov

Summary. The study explores the application of singular value decomposition (SVD) as a tool for analyzing the linear dependence of the predicted stock price on a set of market indicators. Singular value decomposition is employed to decompose the feature matrix, enabling the identification of the most significant components that describe data variability and determining the degree of linear dependence [1] between the indicators and the target variable—the stock price. This approach facilitates dimensionality reduction, mitigates multicollinearity, and highlights key factors influencing the price.

Additionally, the study addresses the regression task for predicting stock prices in the short term based on the extracted features. Machine learning methods, such as linear regression, regularized models (e.g., Ridge or Lasso), or more complex algorithms like gradient boosting applied depending on the data characteristics. The analysis includes data preprocessing steps, such as normalization, overseeing missing values, and feature selection. Predictions evaluated using metrics such as mean squared error (MSE), mean absolute error (MAE), or the coefficient of determination (R^2) [2], allowing for a quantitative assessment of the model's accuracy.

The study also discusses the limitations of the approach, including assumptions of linear dependence, market volatility, and the impact of external factors not captured in the dataset. To improve prediction accuracy, the integration of additional data, such as news feeds, macroeconomic indicators, or market sentiment analysis. The results obtained can be utilized for developing trading strategies or supporting investment decision-making in the context of short-term trading.

Keywords: singular value decomposition, linear regression, linear dependence, stocks, indicators.

Введение

Линейная регрессия — один из наиболее универсальных и широко используемых методов машинного обучения благодаря своей простоте и высокой интерпретируемости. Этот подход находит применение в самых разных областях, включая прогнозирование биржевых показателей. В настоящее время крупные финансовые институты — банки, хедж-фонды

Солобуто Алексей Викторович
аспирант, Московский финансово-юридический
университет МФЮА
29352138@s.mfua.ru

Павлов Валерий Анатольевич
к.э.н., доцент, Московский финансово-юридический
университет МФЮА
29359332@s.mfua.ru

Аннотация. В работе исследуется применение сингуллярного разложения (SVD) как инструмента для анализа линейной зависимости прогнозируемой цены акций от набора биржевых индикаторов. Сингуллярное разложение используется для разложения матрицы признаков, что позволяет выявить наиболее значимые компоненты [1], описывающие вариации в данных, и определить степень линейной зависимости между индикаторами и целевой переменной — ценой акций. Этот подход способствует снижению размерности данных, устранению мультиколлинеарности и выделению ключевых факторов, влияющих на цену.

Кроме того, в работе решается задача регрессии для прогнозирования стоимости акций в краткосрочной перспективе на основе выделенных признаков. Для этого применяются методы машинного обучения, такие как линейная регрессия, регуляризованные модели (например, Ridge или Lasso) или более сложные алгоритмы, такие как градиентный бустинг, в зависимости от характеристик данных. Для оценки работы регрессионных моделей применяются следующие метрики: среднеквадратичная ошибка (MSE), средняя абсолютная ошибка (MAE), коэффициент детерминации (R^2) [2].

Полученные результаты могут быть использованы для разработки торговых стратегий или поддержки принятия инвестиционных решений в условиях краткосрочной торговли.

Ключевые слова: сингуллярное разложение, линейная регрессия, линейная зависимость, ценные бумаги, индикаторы.

и компании, занимающиеся высокочастотной торговлей (HFT) — активно разрабатывают и внедряют собственные аналитические инструменты для предсказания рыночной динамики. Такие инструменты позволяют оперативно принимать решения на основе анализа больших массивов данных, что особенно важно в условиях высокой волатильности и неопределенности на финансовых рынках.

В данной работе рассматриваются ключевые преимущества и ограничения линейной регрессии в контексте эконометрических задач, связанных с прогнозированием биржевых показателей. Особое внимание уделяется вопросам повышения устойчивости модели и снижению рисков, связанных с её потенциальными недостатками.

Одной из типичных проблем при построении регрессионных моделей является мультиколлинеарность между признаками. Для минимизации её негативного влияния в работе предлагается использовать сингулярное разложение (SVD, Singular Value Decomposition) для отбора признаков. Этот метод позволяет выделить из исходного набора только те переменные, которые обладают выраженной линейной зависимостью с целевой переменной, тем самым повышая стабильность и качество модели.

Таким образом, основная цель данной работы — исследовать применимость линейной регрессии в задачах прогнозирования биржевых показателей, а также продемонстрировать эффективность использования SVD для оптимизации модели. Полученные результаты могут быть полезны как для теоретического анализа, так и для практического применения в аналитике финансовых рынков.

В качестве набора данных для обучения и тестирования линейной регрессии были выбраны исторические данные по бумагам, содержащимся в индексе мосбиржи с 2014 по 2023 годы.

Анализ актуальных исследований в области

Применение линейной регрессии рассматривается в статье «Прогнозирование цены акции с помощью метода регрессионного анализа», где авторы применяют модель линейной регрессии для нескольких бумаг [3]. В данной работе применяются скользящие различные скользящие средние и цена за текущий период. В качестве модификации предлагается первоначально сделать расчёт дополнительных биржевых индикаторов [4], а также сделать ранжирование их важности с использованием сингулярного разложения.

Преимущества и недостатки модели линейной регрессии

У линейной регрессии, несмотря на свою простоту есть несколько преимуществ по отношению к другим моделям машинного обучения:

- 1) Скорость обучения — для вычисления коэффициентов линейной регрессии не требуются такие временные затраты, как для более сложных моделей машинного обучения.
- 2) Интерпретируемость — итоговые значения коэффициентов модели можно записать в виде функции.

Также следует выделить недостатки:

- 1) Все параметры должны обладать линейной зависимостью с целевой переменной
- 2) Необходимо избегать мультиколлинеарности признаков — линейной зависимости между признаками

Из-за вышеперечисленных недостатков возникла необходимость строгого отбора признаков, чтобы минимизировать ошибку [5].

Отбор признаков с помощью сингулярного разложения

Достоинство сингулярного разложения в том, что оно работает с совершенно любыми матрицами: и вырожденными, и невырожденными, и квадратными, и прямоугольными. Секрет в том, что для расчётов используется не сама матрица, а произведение её с транспонированной. Эти произведения обязательно будут квадратными матрицами [6]. Для анализа значимости признаков необходимо отранжировать строки по значению собственных чисел, которые хранятся в левой матрице сингулярного разложения.

Для данных по ценным бумагам было принято решение выбрать первые 6 самых значимых признаков на цену, поскольку значения собственных чисел остальных признаков имели малую линейную зависимость, значения полученных признаков будут перечислены в таблице 1. Остальными признаками было решено пренебречь, т.к. значения их собственных чисел были малы относительно перечисленных.

Таблица 1.

Значения собственных чисел

Название признака	Значение собственного числа
RSI	8.9345
ADX	6.2185
PP0	5.3333
TRIX	4.8248
OBV	4.5454
ATR	3.1111

Описание признаков

По итогам отбора признаков, были выбраны следующие:

- 1) Индекс относительной силы (RSI) — индикатор технического анализа, определяющий силу тренда и вероятность его смены [7].
- 2) Индикатор ADX (Average Directional Index, средний индекс направленного движения) — тех-

нический индикатор, используемый в анализе финансовых рынков для оценки силы тренда, независимо от его направления (восходящего или нисходящего) [8].

- 3) Индикатор PPO (Percentage Price Oscillator, Процентный Ценовой Осциллятор) — это технический индикатор, используемый в анализе финансовых рынков для оценки импульса (моментума) цены актива [9].
- 4) Индикатор TRIX (Triple Exponential Average) — это технический индикатор, используемый в анализе финансовых рынков для определения трендов и потенциальных точек разворота цены [10].
- 5) Индикатор OBV (On-Balance Volume, Балансовый объём) — это технический индикатор, который связывает объём торгов с движением цены актива, представляющий собой кумулятивную скользящую среднюю объёма торгов, взятого со знаком плюс в случае растущего рынка и со знаком минус в случае падающего [11].
- 6) Индикатор ATR (Average True Range, средний истинный диапазон) — это технический индикатор, который является измерителем волатильности и говорит, что он отражает диапазон возможного изменения цен актива на выбранном временном интервале [12].

Обучение моделей

На основе полученных данных было решено построить 2 модели машинного обучения: классическую линейную регрессию и модель Ridge, от модели Lasso было решено отказаться, так как влияние L1-регуляризации может обнулить коэффициенты признаков, что при малом их количестве может привести к серьёзному недобучению [13].

Для применения моделей было проведено масштабирование данных с помощью метода StandardScaler, чтобы избежать серьёзного разрыва в коэффициентах регрессии [14].

Таблица 2.
Результаты работы модели

Название модели	MAPE	Итоговая формула (коэффициенты с округлением до 4 знаков после запятой)
LinearRegression	0.054	$price = 1.8733 * rsi + 0.3765 * adx + 0.7439 * ppo + 2.3939 * trix + 0.1111 * OBV + 6.2349 * ATR$
Ridge	0.036	$price = 0.6144 * rsi + 3.7986 * adx + 0.3844 * ppo + 8.6111 * trix + 10.2500 * OBV + 0.7481 * ATR$

Оценка качества модели была проведена при помощи метрики mean absolute percentage error, так как значения стоимостей бумаг могут сильно отличаться между собой.

Модель Ridge имеет параметр alpha, который означает штраф за сложность модели [15], для данного признака была проведена калибровка значений от 0.1 до 1, с помощью кросс-валидации на 5 наборах данных, лучший результат параметра alpha = 0.73 со значением MAPE = 0.036. Итоги обучения моделей будут указаны в таблице 2.

Заключение

В данной работе был проведен анализ использования линейной регрессии в качестве метода предсказания цены акций — были отобраны признаки, с помощью сингулярного разложения, а также обучены следующие модели: базовая линейная регрессия и модель Ridge, которая использует L2-регуляризацию с коэффициентом alpha = 0.73.

Модель Ridge показала себя лучше на 1.4 %, а её общая ошибка получилась 2.4 %, что позволяет рассматривать модель как рекомендательную систему в помощь эксперту.

ЛИТЕРАТУРА

1. Майкл Льюис. Flash Boys: Высокочастотная революция на Уолл-стрит = Flash Boys A Wall Street Revolt. — М.: Альпина Паблишер, 2015. — 348 с.
2. В.В. Стрижов. Функция ошибки в задачах восстановления регрессии // Заводская лаборатория. Диагностика материалов. — 2012. — №79. — С. 65–73.
3. Рудзейт О.Ю., Зайнетдинов А.Р., Недяк А.В., Рагулин П.Г. Прогнозирование цены акции с помощью метода регрессионного анализа // «Отходы и ресурсы». — 2020. — №4. — С. 23–32.
4. Павлов В.А., Солобуто А.В. Предобработка данных в задаче прогнозирования динамики ценных бумаг // Современные и информационные технологии в социальной сфере. — Чебоксары: Волжский филиал МАДИ, 2024. — С. 42–45.
5. Хорн Р., Джонсон Ч. Матричный анализ. — М.: Мир, 1989. — 656 с.
6. William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery. 2.6 Singular Value Decomposition // Numerical Recipes in C. — 2nd edition. — Cambridge: Cambridge University Press.
7. Донна Л. МакКормик, Джейфри Оуэн Кац Энциклопедия торговых стратегий. — М.: Альпина Паблишер, 2012. — 394 с.

8. Стивен Б. Акелис *Technical Analysis from A to Z*. — 2 изд. McGraw Hill, 2013. — 400 с.
9. Колби Роберт. Энциклопедия технических индикаторов рынка. — 2-е изд. — М.: «Альпина Бизнес Букс», 2004. — 837 с.
10. J. Welles Wilder, Jr. *New Concepts in Technical Trading Systems* — М.: ПРИЗМА-15. pp 40–41
11. Лебо Ч., Лукас Д.В Компьютерный анализ фьючерсных рынков. — 6 изд. — М.: Альпина Паблишер, 2011. — 304 с.
12. Кауфман Перри. Системы и методы биржевой торговли. — 1-е изд. — М.: Альпина PRO, 2024. — 1248 с.
13. F. Bach. Bolasso: model consistent lasso estimation through the bootstrap // ICML, 2008.
14. StandardScaler sklearn documentation // [Электронный ресурс] URL: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html> (дата обращения: 03.05.2025).
15. Норман Дрейпер, Гарри Смит. Прикладной регрессионный анализ. Множественная регрессия. — 3 изд. — М.: Диалектика, 2016. — 912 с.

© Солобуто Алексей Викторович (29352138@s.mfua.ru); Павлов Валерий Анатольевич (29359332@s.mfua.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»