УДК 004.4

М. А. Карпенко[1]

Реалізація методу лінійної регресії для прогнозування фінансових показників

Вінницький національний технічний університет

Анотація

Розглянуто особливості реалізації методу лінійної регресії для прогнозування фінансових показників в інтерактивній програмній системі моніторингу та аналізу фінансових інструментів.

Ключові слова: *фінансові інструменти, регресійний аналіз, методи прогнозування.*

Abstract

The features of the implementation of the linear regression method for forecasting financial indicators in an interactive software system for monitoring and analyzing financial instruments are considered.

Keywords: *financial instruments, regression analysis, forecasting methods.*

Прогнозування фінансових показників є важливим завданням для багатьох компаній і банківських установ. Завдяки цьому можна ухвалювати рішення щодо інвестицій, планування бюджетів, управління ризиками та загальної стратегії компанії. Одним з найбільш ефективних інструментів для прогнозування є методи машинного навчання (МН). Вони дозволяють моделювати складні взаємозв'язки між даними і знаходити закономірності, які можуть бути неочевидними при використанні традиційних статистичних методів [1].

Лінійна регресія — це один із фундаментальних методів статистичного аналізу та машинного навчання, який застосовується для моделювання взаємозв'язку між незалежними змінними та залежною змінною. Основне завдання лінійної регресії полягає в прогнозуванні значень залежної змінної на основі набору незалежних змінних. Такий підхід знайшов широке застосування в економіці, фінансах, маркетингу та інших сферах, де необхідно передбачити майбутні показники на основі історичних даних.

Модель лінійної регресії намагається знайти найкраще представлення залежності між незалежними змінними $x_1,\ x_2,\ \dots, x_n$ і залежною змінною y. Рівняння лінійної регресії виглядає так:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \qquad (1)$$

де y – прогнозоване значення;

$x_1,\ x_2,\ \dots, x_n$ – незалежні змінні;

$\beta_1,\ \beta_2,\ \dots, \beta_n$ – коефіцієнти регресії;

ϵ – залишок (помилка прогнозу).

Ціль: Знайти значення коефіцієнтів $\beta_1,\ \beta_2,\ \dots, \beta_n$, що мінімізують суму квадратів залишків між реальними і прогнозованими значеннями.

Метод найменших квадратів мінімізує функцію втрат:

$$SSE = \sum_{i=1}^{m}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{m}(y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_2 x_{i2}))^2 \quad (2)$$

де SSE – сума квадратів залишків (Sum of Squared Errors);

y_i – фактичне значення;

$\hat{y}_i$ – прогнозоване значення для кожного і-го спостереження.

Коефіцієнти β можна знайти, розв'язавши систему рівнянь, яка мінімізує значення SSE.

Коефіцієнти β для багатовимірної лінійної регресії можна обчислити за допомогою матричного підходу. Нехай X — матриця розміром m × (n + 1), де кожен рядок — це вектор спостережень для змінних $x_1,\ x_2,\ \ldots, x_n$, а y — вектор фактичних значень y. Тоді формула для обчислення коефіцієнтів виглядає так:

$$\beta = (X^T X)^{-1} X^T y \quad (3)$$

де X^T – транспонована матриця X;

$(X^T X)^{-1}$ – обернена матриця до $X^T X$;

$X^T y$ – добуток транспонованої матриці X на вектор y.

В таблиці 1 представленні тестові дані для аналізу.

Таблиця 1 – Тестові дані для аналізу методом лінійної регресії

ВВП (в %)	ІСЦ (індекс)	Чистий прибуток (млн грн)
3.5	1.8	15
4.2	2.1	18
3.8	1.9	16
3.6	2.0	14

Позначимо.

1. x_1 – ВВП.
2. x_2 – ІСЦ.
3. y – чистий прибуток.

Матриця ознак (X):

$$X = \begin{bmatrix} 1 & 3.5 & 1.8 \\ 1 & 4.2 & 2.1 \\ 1 & 3.8 & 1.9 \\ 1 & 3.6 & 2.0 \end{bmatrix} \quad (4)$$

Вектор залежної змінної (y):

$$y = \begin{bmatrix} 15 \\ 18 \\ 16 \\ 14 \end{bmatrix} \quad (5)$$

Тоді коефіцієнти β можна знайти за формулою (2.3).

Після побудови моделі важливо оцінити її точність. Один з основних показників точності — коефіцієнт детермінації R^2, що показує частку поясненої варіації залежної змінної:

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y}_i)^2} \quad (6)$$

де $\sum_{i=1}^{m}(y_i - \hat{y}_i)^2$ – сума квадратів залишків;

$\sum_{i=1}^{m}(y_i - \bar{y}_i)^2$ – загальна сума квадратів;

$\bar{y}$ – середнє значення фактичних значень.

Якщо $R^2 \approx 1$, це свідчить про високу точність моделі.

Після побудови моделі можна прогнозувати нові значення. Наприклад, для значень $x_1 = 3\%$ і $x_2 = 2$ прогнозоване значення у обчислюється за формулою:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad (7)$$

Припустимо, ми отримали значення коефіцієнтів: $\beta_0 = 10$, $\beta_1 = 0.8$, $\beta_2 = 0.5$. Тоді:

$$y = 10 + 0.8 \times 3 + 0.5 \times 2 = 10 + 2.4 + 1 = 13.4 \text{ млн грн} \quad (8)$$

Застосування лінійної регресії для прогнозування фінансових показників дозволяє отримати швидкі прогнози і надає можливість оцінювати вплив кожного фактора на цільовий показник. Така модель є базовою і може бути доповнена нелінійними компонентами або розширена за допомогою інших методів машинного навчання для досягнення кращої точності.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Іваненко С.М., Петров О.В. Основи статистичного аналізу: Лінійна регресія та її застосування, 2023. - Наукова думка – 280 с.

Карпенко Максим Анатолійович, студент групи 2ПІ-23м, факультет інформаційних технологій та комп’ютерної інженерії, Вінницький національний технічний університет, Україна

Maksym Karpenko, student of group 2PI-23m, Faculty for Information Technologies and Computer Engineering, Vinnytsia National Technical University, Ukraine, diskmax008@gmail.com.