

МАТЕМАТИЧНИЙ АПАРАТ ЗАБЕЗПЕЧЕННЯ ОБРОБКИ СТАТИСТИЧНИХ ДАНИХ

Щехорський А.,
*кандидат фізико-математичних наук,
доцент комп'ютерних наук
та інформаційних технологій,
Житомирський державний університет
імені Івана Франка,
м. Житомир, Україна*

Математичних методів обробки монотонних одновимірних масивів статистичних даних із заданим кроком, як відомо, існує доволі багато. В моделях регресії найбільш поширений метод найменших квадратів. Побудова моделей регресії за методом найменших квадратів вагомо пов'язана з вибором специфікації моделі. Вивчення моделей регресії з лінійною специфікацією є остаточним. У випадку нелінійної специфікації, навіть зведенням нелінійної моделі до лінійної, остаточним його назвати не можна. Проблема полягає у виборі функцій зведеної лінійної моделі.

У випадку нелінійної специфікації її вибір намагаються представити як лінійну комбінацію $\sum \beta_i f_i(x)$ кортежу (списку) відомих простих функцій $(1, f_1(x), \dots, f_n(x))$, наприклад, кортежу степеневих функцій $(1, x, \dots, x^n)$. Завдяки такому вибору однофакторна нелінійна модель регресії $Y = \beta X + U$ (Y – результуюча змінна, X – регулярна змінна, U – змінна збурення, β – коефіцієнти) зводиться до лінійної моделі регресії стосовно нових “регуляторних” змінних $1, x, \dots, x^n$, якими є степеневі функції.

Адекватність специфікації моделі статистичним даним залежить від такого показника як коефіцієнт детермінації. Його

значення повинно бути не меншим 0,95. Існують приклади, в яких жодна апроксимація поліноміальним кортежем результуючого показника не може задовольняти нормативне значення коефіцієнта детермінації не меншим 0,95. Виявляється, в таких прикладах, поліноміальна апроксимація “дробовим кортежем” $(1, x^{\alpha_1}, \dots, x^{\alpha_n})$ (α_i - дійсні додатні числа) фіксованого розміру таке нормативне значення може забезпечити, завдяки дійсним значенням показників α_i . Можна навести приклади статистичних даних результуючого показника, апроксимація якого, дробовим кортежем заданого фіксованого розміру теж стає не можливою, не зважаючи на дійсні значення степенів α_i . Апроксимація стає можливою, коли змінювати дробові кортежі як за розміром так і за варіацією дробових степенів.

Таким чином, для досягнення нормативних значень, згаданої апроксимації, потрібно змінювати не тільки розмір дробового кортежу але і значення показників степенів. Для збереження статистичних оцінок параметрів моделі (наявності числв степенів вільності) такі зміни неминуче вимагатимуть зміни розміру бази вхідних статистичних даних. Практично змінити вхідну базу даних неможливо за фіксованого її розміру, а значить і фіксованого розміру регуляторної бази даних, яка безпосередньо завантажується в код алгоритму модельного процесу. Єдиний вихід це знаходження способу розширення бази даних до забезпечення необхідної норми коефіцієнта детермінації моделі. Такий алгоритм розширення бази даних знайдено. Взагалі, знайдений алгоритм апроксимації результуючого показника моделі регресії дробовими кортежами змінних розмірів можливим розширенням бази даних з максимальним значенням степеня кортежу рівним 6. На рис.1 приведений приклад апроксимації результуючого показника дробовим кортежем $f(x) = (1, x, x^{3,6}, x^{4,6}, x^5)$ за методом найменших квадратів.

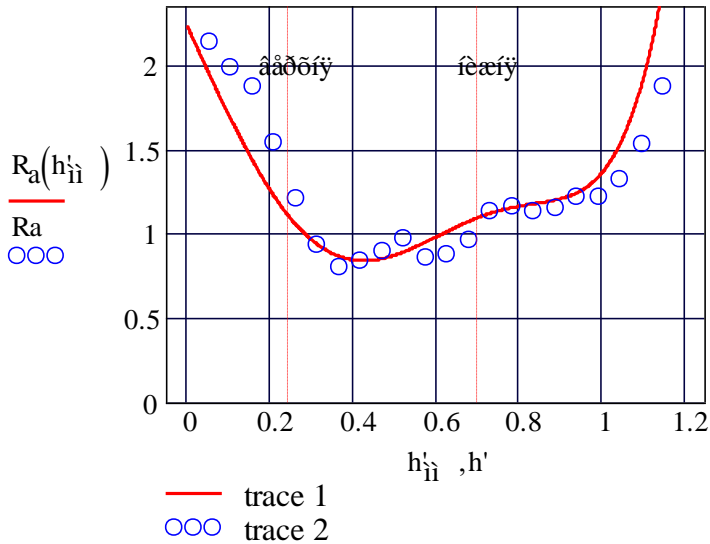


Рис.1. Приклад апроксимації результуючого показника дробовим кортежем $f(x) = (1, x, x^{3.6}, x^{4.6}, x^5)$ за методом найменших квадратів

Приведемо прикладний аспект апроксимації, який полягає в статистичній обробці визначення показників якості (наприклад шорсткості) оброблених поверхонь деталей з порошковими газотермічними покриттями [1]. Обробка даних здійснювалась в середовищі MathcadPro. В кодах Mathcad і інших мовах програмування кортежі повинні бути фіксованих розмірів, кортеж фактично виступає в ролі шаблону. Вимога до зміни розміру кортежа в коді моделі може бути реалізована на алгоритмічній мові PYTHON. Крім того, апроксимація в [1]. відбувалась за шаблонами шляхом підбору степеневих функцій.

Особливість запропонованого кортежного методу побудови нелінійних моделей регресії степеневими функціями полягає: 1) в його застосуванні для обробки любых одновимірних, одно крокових монотонних масивів даних; 2) апроксимації дробовими кортежами дозволяє отримувати її результати, слідуючи крок в крок, за статистичними даними, що дає

можливість здійснювати точні внутрішні прогнози результуючого показника. 3) в його застосуванні для обробки багатовимірних одно крокових монотонних масивів багатофакторних моделях регресії.

References

1. Schehorsky A., Kravchenko M., Kozakov O., Polonsky L. (2014). Статистичний метод визначення розподілу показників якості поверхонь, оброблених у товщі газотермічних покриттів на різних глибинах [Statistical method of finding quality indicators distribution of surfaces treated in thickness of gas-thermal coatings in various], *Вісник ТНТУ імені Івана Пулюя* Vol. 4. URL: <http://library.tntu.edu.ua/vysnyk/visnyk-tntu/>