УДК 004:519.24:681.3.06

О. А. Павлов, М. М. Головченко,
В. В. Дрозд, В. С. Шаргородський

ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ МОДИФІКОВАНОГО МЕТОДУ УРАХУВАННЯ АРГУМЕНТІВ ДЛЯ ПОБУДОВИ БАГАТОВИМІРНИХ РЕГРЕСІЙ, ЗАДАНИХ НАДЛИШКОВИМ ОПИСОМ

Анотація: Розглядається задача побудови багатовимірної регресії, лінійної відносно невідомих коефіцієнтів, що задана надлишковим описом. Треба знайти структуру шуканої регресії, тобто виключити з її опису вхідні детерміновані змінні, значення яких не впливають чи практично не впливають на значення вихідної змінної, а також оцінити коефіцієнти, що залишилися. Раніше авторами для розв’язання цієї задачі було запропоновано метод, що є модифікацією загальновідомого методу групового урахування аргументів. Модифікація полягала в тому, що специфіка сформульованої задачі, а саме, представлення багатовимірної регресії надлишковим описом, дозволило суттєво спростити алгоритм знаходження множини часткових описів шуканої регресії, що статистично значимо містить правильну структуру. В якості регулярного критерію, що відбирає за перевірочною послідовністю даних з множини часткових описів той, що претендує на розв’язок, був використаний асиметричний критерій – сума квадратів відхилень значень вихідних даних від значень часткового опису багатовимірної регресії. В даній роботі пропонується підвищити ефективність модернізованого методу групового урахування аргументів за рахунок одночасного використання низки регулярних критеріїв алгоритмічною процедурою, що враховує специфіку сформульованої задачі. Наведені результати статистично значимого дослідження ефективності нової версії модифікованого методу групового урахування аргументів.

Ключові слова: регресійний аналіз, метод найменших квадратів, перевірочна послідовність, регулярний критерій, метод групового урахування аргументів, частковий опис.

1. Вступ

Побудова багатовимірної регресії (БР) за результатами активного експерименту, як показано, зокрема, в [1–15], є досі актуальною як в теоретичному, так і в практичному аспектах. Дуже важливим є частковий випадок, коли відомо, що БР є лінійною відносно невідомих коефіцієнтів, але невідомо, які детерміновані вхідні змінні суттєво, несуттєво впливають на величину вихідної змінної. Тобто, модель БР в загальному випадку є надлишковою. В [15] для розв’язання цієї задачі був запропонований модифікований метод групового урахування аргументів (ММГУА). В даній роботі пропонується підвищити ефективність ММГУА за рахунок комплексного використання регулярних критеріїв.

1.1. Формальна постановка задачі

БР, лінійна відносно невідомих коефіцієнтів, задається наступним надлишковим описом:

$$Y(\bar{x}) = \sum_{j=1}^{r} b_j \psi_j(\bar{x}) + E, \quad (1)$$

де $\bar{x} = (x_1, \dots, x_m)^{\mathrm{T}}$ – надлишковий вектор вхідних детермінованих змінних (деякі вхідні змінні, можливо, не впливають чи слабо впливають на значення вихідної змінної), E – випадкова величина, $ME = 0$, $DE = \sigma^2 < 0$, функція щільності випадкової величини E відома чи відома с точністю до числових значень її параметрів, $\psi_j(\bar{x}), j = \overline{1,r}$, – множина відомих базових скалярних функцій векторного аргументу, b_j – невідомі коефіцієнти. Необхідно знайти оцінки коефіцієнтів $b_j, j = \overline{1,r}$, за результатами активного чи пасивного експерименту

$$(\bar{x}_i \to y_i, i = \overline{1,n}), \quad (2)$$

де $y_i = \sum_{j=1}^{r} b_j \psi_j(\bar{x}_i) + \varepsilon_i, i = \overline{1,n}$; $\bar{x}_i = (x_{1i}, \dots, x_{mi})^{\mathrm{T}}$ – значення детермінованих вхідних змінних в i-му випробуванні, ε_i – реалізації випадкової величини E, n – кількість випробувань, що дозволяє використовувати критерій χ^2 для перевірки простої чи складної гіпотези.

1.2. Агрегована методологія використання ММГУА [15]

1.2.1. Знаходження оцінок $\hat{\boldsymbol{b}}_j, \boldsymbol{j} = \overline{\boldsymbol{1},\boldsymbol{r}}$, загальною процедурою методу найменших квадратів (МНК)

Нехай $\hat{b} = \left(\hat{b}_1, \dots, \hat{b}_r\right)^{\mathrm{T}}$, $y = (y_1, \dots, y_n)^{\mathrm{T}}$. Тоді

$$\hat{b} = (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}} y, \quad (3)$$

де

$$A = \begin{pmatrix} \psi_1(\bar{x}_1) & \psi_2(\bar{x}_1) & \dots & \psi_r(\bar{x}_1) \\ \vdots & \vdots & \vdots & \vdots \\ \psi_1(\bar{x}_n) & \psi_2(\bar{x}_n) & \dots & \psi_r(\bar{x}_n) \end{pmatrix}.$$

1.2.2. Алгоритм розбиття множини коефіцієнтів $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_r)$ на два класи $\boldsymbol{M}_1$, $\boldsymbol{M}_2$

Примітка 1. Алгоритм приводиться повністю, так як він використовується в основній частині цієї роботи.

Ранжування чисел $\hat{b}_j, j = \overline{1,r}$, по модулю їх значень: $\left|\hat{b}_{j_1}\right| \geq \left|\hat{b}_{j_2}\right| \geq \dots \geq \left|\hat{b}_{j_r}\right|$.

Перший крок. $b_{j_1} \in M_1$, $b_{j_r} \in M_2$. Якщо $\left|\hat{b}_{j_1} - \hat{b}_{j_2}\right| < \left|\hat{b}_{j_2}\right| - \left|\hat{b}_{j_r}\right|$, то $b_{j_2} \in M_1$, в протилежному випадку $b_{j_2}, b_{j_3}, \dots, b_{j_r} \in M_2$. Розбиття завершене.

 ISSN 1560-8956

l-й крок. $b_{j_l} \in M_1$. Якщо

$$\frac{1}{l}\sum_{i=1}^{l}\left|\hat{b}_{j_i}\right| - \left|\hat{b}_{j_{l+1}}\right| < \hat{b}_{j_{l+1}} - \left|\hat{b}_{j_r}\right|,$$

то $b_{j_{l+1}} \in M_1$, в протилежному випадку розбиття завершене, $M_1 = \left\{b_{j_1} \dots b_{j_l}\right\}$, $M_2 = \left\{b_{j_{l+1}} \dots b_{j_r}\right\}$.

За обмежену кількість кроків алгоритм остаточно завершує розбиття множини коефіцієнтів $b_j, j = \overline{1, r}$, на два класи M_1, M_2.

1.2.3. Формування множини часткових описів БР

Множину M_2 представляємо у вигляді об’єднання всіх різних підмножин M_2^j, $j = \overline{1, \sum_{t=1}^{|M_2|} C_{|M_2|}^t}$, $M_2 = \bigcup_j M_2^j$. Кожен частковий опис БР має вигляд

$$Y(\bar{x}) = \sum_{\forall b_l \in M_1} b_l \psi_l(\bar{x}) + \sum_{\forall b_l \in M_2^j} b_l \psi_l(\bar{x}) + E. \quad (4)$$

Дані експерименту $(\bar{x}_i \to y_i, i = \overline{1, n})$ розбиваємо на підпослідовність, за якою знаходяться МНК оцінки коефіцієнтів всіх часткових описів БР $(\bar{x}_i \to y_i, i = \overline{1, n_1})$ та перевірочну послідовність $(\bar{x}_i \to y_i, i = \overline{n_1 + 1, n})$, за якою, за допомогою регулярного критерію, знаходиться частковий опис БР, що претендує на правильну структуру БР.

1.2.4. Знаходження для кожного часткового опису БР значення регулярного критерію – залишкової суми квадратів (ЗСК)

$$\forall_{M_2^j} \text{ЗСК}\left(M_1, M_2^j\right) = \\ = \sum_{i=1}^{n-n_1}\left(y_{n_1+i} - \sum_{\forall b_l \in M_1} \hat{b}_l \psi_l\left(\bar{x}_{n_1+i}\right) - \sum_{\forall b_l \in M_2^j} \hat{b}_l \psi_l\left(\bar{x}_{n+i}\right)\right)^2 \quad (5)$$

Претендент на розв’язок – той, у якого (5) приймає найменше значення.

1.2.5. Остаточний розв’язок

Оцінки $b_j, j = \overline{1, r}$, знайденого часткового опису в п. 1.2.4 загальною процедурою МНК знаходяться по всій множині експериментальних даних $(\bar{x}_i \to y_i, i = \overline{1, n})$. Точність отриманого результату оцінюється за реалізацією критерію χ^2, побудованою за множиною оцінок реалізацій випадкової величини E:

$$\varepsilon_i = y_i - \sum_{\forall b_l \in M_1} \hat{b}_l \psi_l(\bar{x}_i) - \sum_{\forall b_l \in M_2^j(\text{opt})} \hat{b}_l \psi_l(\bar{x}_i), i = \overline{1, n}, \quad (6)$$

де $M_1, M_2^j(\text{opt})$ задає частковий опис, на якому досягається мінімум (5), а оцінки його коефіцієнтів знайдені МНК по всьому набору даних $(\bar{x}_i \to y_i, i = \overline{1, n})$. Якщо реалізація критерію χ^2, що перевіряє гіпотезу про розподіл випадкової величини E, яка знайдена по $\hat{\varepsilon}_i, i = \overline{1, n}$, попала в допустиму область, отриманий результат вважається допустимо достовірним.

Примітка 2. Вперше ММГУА описується для загальної моделі БР, лінійної відносно невідомих коефіцієнтів.

1.3. Постановка задачі дослідження

Як відомо, в алгоритмах класичного МГУА використовуються різні регулярні критерії, а саме, асиметричні критерії ЗСК, симетричний критерій $\text{ЗСК}_1 + \text{ЗСК}_2$, коли навчальну та перевірочну послідовності міняють місцями. Тому виникла пропозиція підвищити ефективність використання ММГУА, в постановці, викладеної в п. 1.2, з одночасним використанням регулярних критеріїв ЗСК_1, ЗСК_2, $\text{ЗСК}_1 + \text{ЗСК}_2$ та критерію χ^2 для перевірки гіпотези про розподіл випадкової величини E.

2. Розв'язання задачі дослідження

Перша проблема, яка виникає, – як розбити експериментальні дані на навчальну та перевірочну послідовності. Зазвичай використовують два варіанти – (50%, 50%) та (60%, 40%). Кожен з них має свої переваги та недоліки. В цьому дослідженні пропонується використати як один будь-який варіант розбиття, так і одночасно обидва варіанти.

2.1. Одноразове розбиття експериментальних даних

Будемо використовувати асиметричні критерії ЗСК_1, ЗСК_2 та симетричний критерій $\text{ЗСК}_1 + \text{ЗСК}_2$. Специфіка ММГУА, зокрема, використання алгоритму розбиття множини коефіцієнтів $\{b_1, \dots, b_r\}$ на два класи M_1, M_2 (п. 1.2.2) приводить до необхідності створення спеціальної процедури використання симетричного критерію. Дійсно, нехай $\left(\bar{x}_{n_1+i} \to y_{n_1+i}, i = \overline{n_1+1, n}\right)$ – це перша перевірочна послідовність, що породжує регулярний критерій ЗСК_1, а перевірочна послідовність $(\bar{x}_i \to y_i, i = \overline{1, n_1})$ – це друга перевірочна послідовність, що породжує регулярний критерій ЗСК_2. В першому випадку маємо розбиття множини коефіцієнтів $\{b_1, \dots, b_r\}$ на два класи M_1^1, M_2^1, а в другому випадку – M_1^2, M_2^2. Тоді для використання симетричного критерію розбиття множини коефіцієнтів $\{b_1, \dots, b_r\}$ на два класи задається наступним чином: $M_1 = M_1^1 \cap M_2^1$, $M_2 = \{b_1, \dots, b_r\}/M_1$. Для симетричного критерію множина всіх часткових описів БР має вигляд:

$$\forall_{M_2^j} \left\{Y(\bar{x}) = \sum\nolimits_{\forall b_l \in M_1} b_l \psi_l(\bar{x}) + \sum\nolimits_{\forall b_l \in M_2^j} b_l \psi_l(\bar{x}) + E\right\}. \tag{7}$$

В результаті послідовного використання регулярних критеріїв ЗСК_1, ЗСК_2, $\text{ЗСК}_1 + \text{ЗСК}_2$ можемо отримати три претенденти на розв'язок, які для скорочення позначимо

$$\left(M_1^1, M_2^{1,j_1}\right), \left(M_1^2, M_2^{2,j_2}\right), \left(M_1, M_2^{j_3}\right). \tag{8}$$

Примітка 3. Множини (8) однозначно задають структуру часткових описів БР, що є претендентами на розв’язок.

Нехай χ_1^2, χ_2^2, χ_3^2 – різні реалізації критерію χ^2, що перевіряє гіпотезу про розподіл випадкової величини E за множиною оцінок $\hat{\varepsilon}_i$ (6) для кожного з трьох часткових описів БР – претендентів на розв’язок.

Нехай χ_α^2 задає критичну область критерію

$$P(\chi^2 \geq \chi_\alpha^2) = 1 - \alpha. \tag{9}$$

Перший критерій розв’язку задачі регресії: шукана регресія задається частковим описом, на якому виконується

$$\min(\chi_1^2, \chi_2^2, \chi_3^2) < \chi_\alpha^2. \tag{10}$$

Обґрунтування критерію. Відомо, що чим більш невірна гіпотеза про розподіл випадкової величини, що перевіряється, тим сильніша тенденція к збільшенню значення реалізації критерію χ^2.

Другий критерій розв’язку задачі регресії:

а) з трьох претендентів на розв’язок вибирається той, який містить мінімальну кількість членів, якщо виконується

$$\max(\chi_1^2, \chi_2^2, \chi_3^2) - \min(\chi_1^2, \chi_2^2, \chi_3^2) < \chi_\alpha^2 - \max(\chi_1^2, \chi_2^2, \chi_3^2); \tag{11}$$

б) якщо лише дві реалізації критерію χ^2 задовольняють умові (10), то з двох претендентів на розв’язок вибирається той, який містить менше членів, при виконанні умови, аналогічній умові (11). В протилежному випадку залишається лише один претендент на розв’язок.

Примітка 4. Якщо претендентів на розв’язок два чи один (різні регулярні критерії вибирають один частковий опис БР), то розв’язок задачі регресії задається умовою (б). Якщо всі реалізації критерію належать до критичної області, то задача регресії на заданому об’ємі експериментальних даних $(\bar{x}_i \to y_i, i = \overline{1,n})$ ММГУА розв’язується некоректно.

2.2. Одночасне використання розбиття експериментальних даних на навчальну та перевірочну послідовності ((50%, 50%) та (60%, 40%))

В цьому випадку отримуємо в загальному випадку 6 реалізацій критерію χ^2 – $(\chi_1^2, \chi_2^2, \ldots, \chi_6^2)$. Нехай

$$\chi_i^2 < \chi_\alpha^2, i = \overline{1,6}. \tag{12}$$

Примітка 5. Якщо умова (12) виконується не для всіх реалізацій критерію χ^2, то для випадку трьох і менше реалізацій для знаходження розв’язку задачі регресії використовується алгоритм, приведений в п. 2.1. В протилежному випадку реалізується наступна алгоритмічна процедура.

Перший критерій знаходження розв’язку: розв’язок задачі регресії задається частковим описом, якому відповідає мінімальне значення критерію χ^2.

Другий критерій знаходження розв’язку задачі регресії: нехай для всіх реалізацій критерію χ^2 виконується умова, аналогічна умові (11). Тоді розв’язком задачі регресії є той частковий опис БР, що містить мінімум членів. В протилежному випадку (одна чи декілька реалізацій критерію χ^2 належать критичній області) пропонується використати алгоритм розбиття множини реалізацій критерію χ^2 на два класи M_1 і M_2 (див. п. 1.2.2), в якому на першому кроці $\max(\forall\chi_i^2) \in M_1$, $\min(\forall\chi_i^2) \in M_2$. Розв’язком задачі регресії є частковий опис БР з множини M_2, що містить мінімальну кількість членів.

Примітка 6. Статистично обґрунтовану рекомендацію використання першого чи другого критерію розв’язку задачі регресії можна отримати внаслідок аналізу результатів достатньої кількості статистичних імітаційних експериментів для різних розподілів випадкової величини E. На якісному рівні аналізу більш ефективним є другий критерій розв’язку задачі регресії.

3. Приклад реалізації статистичного імітаційного експерименту для нормального розподілу випадкової величини *E*

Для проведення даного дослідження пропонується інкрементально збільшувати значення середньоквадратичного відхилення при фіксованій кількості спостережень для кожного з запропонованих регулярних критеріїв і виділити критерій, який дає кращі середньозважені показники. Статистичне імітаційне моделювання було проведене для наступних параметрів: об’єм експериментальних даних ділиться на навчальну та перевірочну підпослідовності у відношенні (60%, 40%); кількість невідомих коефіцієнтів – 10; кількість випробувань n приймає значення 30, 40. χ_1^2 та χ_2^2 – це реалізації критерію χ^2 для перевірки гіпотези про розподіл випадкової величини E для претендентів на розв’язок, що породжені регулярним симетричним та асиметричним критеріями, відповідно. Коефіцієнти ідеальної моделі БР в кожному повторі імітаційного статистичного експерименту генеруються випадковим чином за рівномірним розподілом на відрізку [–10, 10]. Кількість нульових коефіцієнтів для кожної ідеальної моделі регресії дорівнює 3, їх індекси обираються випадковим чином. Міра порівнянь векторів $\hat{b}$ та b, де $\hat{b}$ – вектор оцінок коефіцієнтів моделі, b – вектор ідеальних значень моделі, має наступний вигляд: $d(\hat{b} - b) = \left\| \frac{\hat{b}}{\|\hat{b}\|} - \frac{b}{\|b\|} \right\|$, де $\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}$. Ця міра порівнянь була вибрана завдяки тому, що вона не залежить від довжини векторів, що порівнюються. Оцінки коефіцієнтів знайдені точно, якщо $d(\hat{b} - b)$ належить відрізку [0, 0.02]. Структура багатовимірної регресії заданої надлишковим описом має вигляд:

$$Y(\bar{x}) = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_1x_2 + b_6x_1x_4 + b_7x_2^2 + b_8x_3^2 + b_9x_4^3 + E. \qquad (13)$$

Результати проілюстровані в табл. 1–2. В кожній комірці приведених таблиць наведено: correct – відсоток правильних структур шуканої регресії, mean – середнє значення міри порівнянь для правильних структур, max – максимальне значення міри порівнянь для правильних структур. Кількість повторів імітаційного експерименту для кожної комірки – 1000.

Таблиця 1. **Результати моделювання для 10 повторів імітації основного експерименту, що складається з 30 випробувань**

σ^2	**Симетричний критерій ЗСК$_1$ + ЗСК$_2$**	**Несиметричний критерій ЗСК$_1$**	**Структура розв’язку задається $\min(\chi_1^2, \chi_2^2)$**
1	correct: 87.4 mean: 0.00135 max: 0.12569	correct: 81.0 mean: 0.00168 max: 0.33501	correct: 74.0 mean: 0.00152 max: 0.21569
2	correct: 86.7 mean: 0.00336 max: 0.20437	correct: 81.3 mean: 0.00384 max: 0.33501	correct: 72.8 mean: 0.00379 max: 0.25133
3	correct: 86.06667 mean: 0.0059 max: 0.32468	correct: 80.26667 mean: 0.00694 max: 0.33501	correct: 71.73333 mean: 0.00683 max: 0.33041
4	correct: 84.4 mean: 0.00938 max: 0.37185	correct: 78.9 mean: 0.01098 max: 0.58517	correct: 70.35 mean: 0.01094 max: 0.58515
5	correct: 83.76 mean: 0.01267 max: 0.40834	correct: 78.76 mean: 0.01443 max: 0.58517	correct: 70.76 mean: 0.01435 max: 0.58515
6	correct: 83.86667 mean: 0.01616 max: 0.43298	correct: 78.63333 mean: 0.01877 max: 0.58517	correct: 70.7 mean: 0.01881 max: 0.58515
7	correct: 83.65714 mean: 0.02027 max: 0.52842	correct: 78.4 mean: 0.02388 max: 0.58517	correct: 70.45714 mean: 0.02369 max: 0.58515
8	correct: 82.725 mean: 0.0247 max: 0.52842	correct: 77.775 mean: 0.02915 max: 0.58517	correct: 69.925 mean: 0.02892 max: 0.58515
9	correct: 82.31111 mean: 0.02961 max: 0.53409	correct: 77.02222 mean: 0.03545 max: 0.62139	correct: 69.48889 mean: 0.03494 max: 0.58515
10	correct: 82.12 mean: 0.03486 max: 0.70514	correct: 76.76 mean: 0.04123 max: 0.70514	correct: 69.38 mean: 0.04082 max: 0.71501

Таблиця 2. **Результати моделювання для 10 повторів імітації основного експерименту, що складається з 40 випробувань**

σ^2	**Симетричний критерій ЗСК$_1$ + ЗСК$_2$**	**Несиметричний критерій ЗСК$_1$**	**Структура розв’язку задається $\min(\chi_1^2, \chi_2^2)$**
1	correct: 92.2 mean: 0.001 max: 0.17242	correct: 89.2 mean: 0.00116 max: 0.22238	correct: 83.0 mean: 0.0012 max: 0.22238
2	correct: 93.4 mean: 0.00244 max: 0.17772	correct: 88.6 mean: 0.00299 max: 0.31679	correct: 84.4 mean: 0.00278 max: 0.22238

Закінчення таблиці 2

σ^2	Симетричний критерій ЗСК$_1$ + ЗСК$_2$	Несиметричний критерій ЗСК$_1$	Структура розв’язку задається $\min(\chi_1^2, \chi_2^2)$
3	correct: 92.06667 mean: 0.00454 max: 0.24839	correct: 87.66667 mean: 0.00529 max: 0.34136	correct: 83.33333 mean: 0.00503 max: 0.34136
4	correct: 91.45 mean: 0.00668 max: 0.24839	correct: 87.1 mean: 0.00771 max: 0.34136	correct: 82.65 mean: 0.00748 max: 0.34136
5	correct: 91.08 mean: 0.00933 max: 0.30096	correct: 86.28 mean: 0.01112 max: 0.47077	correct: 81.96 mean: 0.01059 max: 0.47077
6	correct: 90.76667 mean: 0.01195 max: 0.30096	correct: 86.1 mean: 0.01412 max: 0.47077	correct: 81.56667 mean: 0.01343 max: 0.47077
7	correct: 90.14286 mean: 0.01495 max: 0.31496	correct: 85.51429 mean: 0.01775 max: 0.48979	correct: 81.05714 mean: 0.01681 max: 0.48356
8	correct: 89.4 mean: 0.01831 max: 0.31496	correct: 84.825 mean: 0.02156 max: 0.48979	correct: 80.225 mean: 0.02057 max: 0.48356
9	correct: 88.6 mean: 0.02212 max: 0.40559	correct: 83.71111 mean: 0.02623 max: 0.51854	correct: 79.33333 mean: 0.02509 max: 0.48356
10	correct: 88.04 mean: 0.02587 max: 0.4489	correct: 82.96 mean: 0.03088 max: 0.56638	correct: 79.02 mean: 0.02934 max: 0.48505

Висновки

Розглянута можливість підвищення ефективності модифікованого методу групового урахування аргументів для оцінки багатовимірної регресії, заданої надлишковим описом, лінійної відносно невідомих коефіцієнтів, за рахунок комплексного використання регулярних критеріїв в сукупності з критерієм χ^2, що перевіряє гіпотезу про розподіл випадкової величини за оцінками її реалізацій.

Запропонована методологія та її алгоритмічна реалізація використання асиметричного та симетричного регулярних критеріїв для запропонованих варіантів розбиття експериментальних даних на навчальну та перевірочну послідовності, що використовує специфіку задачі регресії, заданої надлишковим описом.

Наведені результати статистичного імітаційного моделювання для випадкової величини, що має нормальний розподіл.

 ISSN 1560-8956

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Pavlov A.A., Holovchenko M.N., Drozd V.V. Efficiency substantiation for a synthetical method of constructing a multivariate polynomial regression given by a redundant representation. Bulletin of National Technical University “KhPI”. Series: System analysis, control and information technologies. 2023. Vol. 1. Iss. 9. P. 3–9. DOI: 10.20998/2079-0023.2023.01.01

2. Yu L. Using negative binomial regression analysis to predict software faults: a study of Apache Ant. *International Journal of Information Technology and Computer Science (IJITCS)*. 2012. Vol. 4. Iss. 8. P. 63–70. DOI: 10.5815/ijitcs.2012.08.08

3. Shahrel M.Z., Mutalib S., Abdul-Rahman S. PriceCop – price monitor and prediction using linear regression and LSVM-ABC methods for e-commerce platform. *International Journal of Information Engineering and Electronic Business (IJIEEB)*. 2021. Vol. 13. Iss. 1. P. 1–14. DOI: 10.5815/ijieeb.2021.01.01

4. Satter A., Ibtehaz N. A regression based sensor data prediction technique to analyze data trustworthiness in cyber-physical system. *International Journal of Information Engineering and Electronic Business (IJIEEB).* 2018. Vol. 10. Iss. 3. P. 15–22. DOI: 10.5815/ ijieeb.2018.03.03

5. Isabona J., Ojuh D.O. Machine learning based on kernel function controlled gaussian process regression method for in-depth extrapolative analysis of Covid-19 daily cases drift rates. *International Journal of Mathematical Sciences and Computing (IJMSC)*. 2021. Vol. 7. Iss. 2. P. 14–23. DOI: 10.5815/ ijmsc.2021.02.02

6. Sinha P. Multivariate polynomial regression in data mining: methodology, problems and solutions. *International Journal of Scientific & Engineering Research*. 2013. Vol. 4. Iss. 12. P. 962–965

7. Kalivas J.H. Interrelationships of multivariate regression methods using eigenvector basis sets. *Journal of Chemometrics*. 1999. Vol. 13. Iss. 2. P. 111–132. DOI: 10.1002/(SICI)1099-128X(199903/ 04)13:2<111::AID-CEM532>3.0.CO;2-N

8. Ortiz-Herrero L., Maguregui M. I., Bartolomé L. Multivariate (O)PLS regression methods in forensic dating. *TrAC Trends in Analytical Chemistry*. 2021. Vol. 141. P. 116278. DOI: 10.1016/ j.trac.2021.116278

9. Guo G., Niu G., Shi Q., Lin Q., Tian D., Duan Y. Multi-element quantitative analysis of soils by laser induced breakdown spectroscopy (LIBS) coupled with univariate and multivariate regression methods. *Analytical Methods*. 2019. Vol. 11. Iss. 23. P. 3006–3013. DOI: 10.1039/C9AY00890J

10. Настенко Е., Павлов В., Бойко Г., Носовец О. Многокритериальный алгоритм шаговой регрессии. *Біомедична інженерія і технологія*, 2020. № 3. С. 48–53. DOI: 10.20535/2617-8974.2020.3.195661

11. Babatunde G., Emmanuel A. A., Oluwaseun O. R., Bunmi O. B., Precious A. E. Impact of climatic change on agricultural product yield using *k*-means and multiple linear regressions. *International Journal of Education and Management Engineering (IJEME).* 2019. Vol. 9. No. 3. P. 16–26. DOI: 10.5815/ijeme.2019.03.02

12. Худсон Д. *Статистика для физиков: Лекции по теории вероятностей и элементарной статистике*. 1970. Мир, 296 с.

13. Pavlov A. A. Holovchenko M. N., Drozd V. V. Construction of a multivariate polynomial given by a redundant description in stochastic and deterministic formulations using an active experiment. Bulletin of National Technical University "KhPI". Series: System analysis, control and information technologies. 2022. Vol. 1. Iss. 7. P. 3–8. DOI: 10.20998/2079-0023.2022.01.01

14. Pavlov A., Holovchenko M., Mukha I., Lishchuk K., Drozd V. A Modified Method and an Architecture of a Software for a Multivariate Polynomial Regression Building Based on the Results of a Conditional Active Experiment. *Lecture Notes on Data Engineering and Communications Technologies.* 2023. Vol. 181. P. 207–222. DOI: 10.1007/978-3-031-36118-0_19

15. Pavlov A. A., Holovchenko M. N. Modified method of constructing a multivariate linear regression given by a redundant description. Bulletin of National Technical University "KhPI". Series: System analysis, control and information technologies. 2022. Vol. 2. Iss. 8. P. 3–8. DOI: 10.20998/2079-0023.2022.02.01

 ISSN 1560-8956