

*Пулеко І.В., к.т.н, доцент
Державний університет «Житомирська політехніка»,
Житомирський військовий інститут ім. С. П. Корольова
Шестак І.М., викладач,
Свистунович І.В., викладач
Житомирський військовий інститут ім. С. П. Корольова*

ПОЛІНОМІАЛЬНА РЕГРЕСІЯ ПРИ НЕГАУСІВСЬКОМУ РОЗПОДІЛІ ПОМИЛОК

Регресія відноситься до класичних методів машинного навчання з учителем і будується на оптимізаційних статистичних алгоритмах оцінювання. Основою регресійного аналізу є моделювання взаємозв'язку між деякою залежною змінною та однією або декількома незалежними змінними. Залежна змінна моделюється як адитивна суміш певного функціоналу від незалежних змінних і відповідних інформативних параметрів (коефіцієнтів) регресії та випадкової складової (регресійних помилок), що представляє варіації залежної змінної, які не пояснюються функцією від незалежних змінних. Для побудови моделі регресії застосовують поліноми, при цьому суть задачі буде зводитись до вибору виду і ступеня полінома та пошуку його оптимальних коефіцієнтів.

Кожна модель регресії в значній мірі спирається на деякі припущення, що повинні виконуватися, зокрема найбільш розповсюдженим є припущення про гаусівський (нормальний) розподіл помилок. Якщо ж припущення не виконуються, то методи отримання параметрів моделі стають не оптимальними, а помилки збільшуються. Тому, поліноміальна регресія з негаусівським розподілом помилок вимагає більш гнучкого підходу до оцінювання коефіцієнтів, оскільки класичний метод найменших квадратів, що базується на припущенні нормального розподілу помилок, вже не є оптимальним. Щоб розрахувати коефіцієнти поліноміальної регресії при негаусівському розподілі помилок, можна використовувати такі підходи:

Метод максимальної правдоподібності. Якщо відомий, закон розподілу помилок (наприклад, розподіл Пуассона, Лапласа або ін.), можна використовувати максимізацію функції правдоподібності, яка описує ймовірність отримання спостережуваних значень залежної змінної за умови моделі та обраного розподілу

Робастні методи регресії. Вони застосовуються в ситуаціях, коли помилки розподілені нерівномірно або присутні викиди, що можуть суттєво впливати на результати стандартної нелінійної регресії. Ці методи знижують вплив викидів і нелінійностей в даних.

Стохастичні поліноми Кунченка. Поліноми Кунченка застосовуються для моделювання та обробки негаусівських даних, що використовує нелінійні перетворення. Цей метод є особливо корисним, коли дані мають складну, нерегулярну структуру або відхиляються від нормального розподілу. Поліноми Кунченка можуть бути адаптовані до різних типів розподілів помилок, включаючи асиметричні, з «важкими хвостами» та мультимодальні. За рахунок врахування специфіки розподілу помилок, ці поліноми дозволяють отримати більш точні оцінки параметрів моделі та більш стійкі до впливу викидів.

Баєсові методи регресії. Використання Баєсового підходу дозволяє враховувати невизначеність щодо розподілу помилок і обчислювати ймовірнісні розподіли для коефіцієнтів. При цьому можна використовувати апріорну інформацію про розподіл коефіцієнтів і оновлювати їх на основі наявних даних та припущень про розподіл помилок.

Методи регресії на основі штучного інтелекту, такі як: генетичні алгоритми та нейронні мережі.

Якщо традиційні методи оптимізації не дають задовільних результатів через складність моделі або розподілу помилок, можна використовувати генетичні алгоритми, які шукають глобальний мінімум функції втрат у просторі параметрів.

У ситуаціях, коли розподіл помилок є невідомим або дані мають складну структуру, можна застосовувати штучні нейронні мережі. Вони можуть моделювати будь-яку складну нелінійну залежність між змінними, навіть за умови негаусівського розподілу помилок. Для цього необхідно визначити архітектуру нейронної мережі (кількість шарів, кількість нейронів, функції активації). Потім навчити мережу на основі набору даних, мінімізуючи відповідну функцію втрат, яка враховує розподіл помилок (наприклад, L1, Huber-функція або функція втрат для конкретного розподілу помилок).

Як загальний недолік генетичних алгоритмів та нейронних мереж слід відмітити необхідність в наявності великих за об'ємом репрезентативних та розмічених наборів даних і значний час, що затрачується на навчання.

Таким чином, для розрахунку коефіцієнтів поліноміальної регресії при негаусівському розподілі помилок варто використовувати альтернативні методи оптимізації, що краще підходять для таких умов. Це можуть бути методи максимальної правдоподібності, робастні методи, стохастичні поліноми Кунченка, Баєсові методи або алгоритми на основі штучного інтелекту. Вибір конкретного підходу залежить від характеру помилок і специфіки задачі.