

А. Г. А р с и т о в (Ульяновск, УлГУ). **Проблема мультиколлинеарности в методе наименьших квадратов.**

Введение. Очень часто возникает задача определения параметров математической модели по известным входным данным и по полученным откликам системы.

Постановка задачи. Будем рассматривать линейную регрессионную модель $z = Ax + v$, где A — матрица эксперимента, а z — отклик системы. Задача заключается в поиске такого $\hat{x}$, который удовлетворяет критерию МНК:

$$J(x) = \|z - Ax\|^2 \rightarrow \min.$$

Проблема мультиколлинеарности. Часто в прикладных задачах регрессионного анализа возникает проблема плохой численной обусловленности матрицы данных, а именно, ее мультиколлинеарности. Мультиколлинеарность матрицы A означает, что входные данные имеют повторы. Тогда нормальная система уравнений для данной модели $A^T A x = A^T z$, имеет вырожденную матрицу $A^T A$. Естественно, попытки отыскания решения с помощью рассмотрения нормальной системы уравнений заранее обречены на неудачу.

Рекуррентные и нерекуррентные алгоритмы. Поскольку проблема мультиколлинеарности относится к категории некорректных задач, ее можно решать известными нерекуррентными классическими методами [1]. Однако наряду с этими традиционными подходами существует менее известная в среде специалистов по регрессионному анализу возможность применения эффективных алгоритмов дискретной фильтрации [2].

Преимущества этих алгоритмов состоят в следующем: они не требуют ни решения систем, ни обращения матриц и гарантируют вычисление решения, численно устойчивы и быстро сходятся к решению. Также они не требуют хранения и обработки всей матрицы A .

Ковариационный алгоритм с матрицей ковариации P_i при начальном значении $P_0 = \varepsilon^{-1}I$ с $0 < \varepsilon \ll 1$ в форме Калмана имеет вид

$$\alpha_i = a_i^T P_{i-1} a_i + 1, \quad K_i = P_{i-1} a_i / \alpha_i,$$

$$P_i = P_{i-1} - K_i a_i^T P_{i-1}, \quad \hat{x}_i = \hat{x}_{i-1} + K_i (\zeta_i - a_i^T \hat{x}_{i-1}).$$

Пример. Рассмотрим пример в условиях «жесткой» мультиколлинеарности:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ \dots & \dots \\ 1 & 1 \end{bmatrix}, \quad z = \begin{bmatrix} 1 \\ 2 \\ \dots \\ i \end{bmatrix}, \quad P_0 = \begin{bmatrix} 1/\varepsilon & 0 \\ 0 & 1/\varepsilon \end{bmatrix}.$$

Применяя алгоритм Калмана, получим следующий результат:

$$\hat{x}_i = \frac{i(i+1)}{2(2i+\varepsilon)} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \xrightarrow{\varepsilon \rightarrow 0} \frac{i+1}{4} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Нормальная система будет иметь вид:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \hat{x} = \frac{i+1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Эта система не дает нормального псевдорешения, тогда как алгоритм Калмана сходится к нему при $\varepsilon \rightarrow 0$.

Заключение. Перспективность работы с рекуррентными алгоритмами связана с дальнейшей их модификацией и детальным изучением численной устойчивости в условиях мультиколлинеарности.

СПИСОК ЛИТЕРАТУРЫ

1. *Radhakrishna Rao C., Toutenburg H.* Linear Models: Least Squares and Alternatives. Springer, 1999.
2. *Park P., Kailath T.* New Square-Root Algorithms for Kalman Filtering. — IEEE Trans. On Automatic Control, 1995.