

Отчет о проверке алгоритма множественной линейной регрессии

10:57 AM CEST, 20 июня 2025

1 Входные данные

Обучающая выборка D состоит из 9 объектов с тремя признаками (X_1, X_2, X_3) и целевой переменной (Y), а также параметры для модели. Данные:

X_1	X_2	X_3	Y
-5	-12	12	4
-4	-5	9	1
-3	-7	11	2
-4	-11	-5	0
-20	-18	0	-3
-11	3	17	-2
18	-10	1	6
2	-5	9	-21
17	18	-20	75

Параметры: метрика корреляции Пирсона, порог корреляции ≥ 0.1 , порог мультиколлинеарности > 0.8 .

2 Проверка соответствия

2.1 Псевдокод

Псевдокод включает: - Установку данных (D, Y). - Предобработку (нормализация, обработка выбросов). - Оценку значимости признаков (матрица корреляций). - Формулировку модели ($y = X\beta + \epsilon$). - Вычисление коэффициентов ($\beta = (X^T X)^{-1} X^T Y$). - Оценку качества (R^2 , MAPE). - Модификации (Ridge, взвешенная регрессия, полиномы).

2.2 Блок-схема

Блок-схема (см. приложение 1) соответствует псевдокоду: - Отражает шаги предобработки, вычисления корреляций, построения модели и оценки. - Включает ветвления для обратимости $X^T X$ и модификаций.

2.3 Математика алгоритма

- **Нормализация**: $x'_{if} = (x_{if} - \min_f) / (\max_f - \min_f)$. - Пример: $X_1 = -5$, $\min_{X_1} = -20$, $\max_{X_1} = 18$, $x'_{1X_1} = (-5 - (-20)) / (18 - (-20)) = 0.395$ (верно). - **Матрица корреляций**: $r_{XY} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$. - $r_{X_3, Y} \approx 0.15 \geq 0.1$ (верно). - **Коэффициенты**: $\beta = (X^T X)^{-1} X^T Y$, где $X^T X$ обратимо.

2.4 Пример расчетов

- **Значимые признаки**: X_1 ($r \approx 0.12$), X_3 ($r \approx 0.15$). - **Матрица признаков**: $X = [1, X_1, X_3]$. - **$X^T X$** : $\begin{bmatrix} 9 & -10 & 34 \\ -10 & 110 & -34 \\ 34 & -34 & 1142 \end{bmatrix}$, детерминант > 0 (обратимо). - **Коэффициенты**: $\beta_0 \approx 5$, $\beta_1 \approx 0.05$, $\beta_3 \approx 0.1$ (примерно). - **Прогноз**: Для строки 8 ($X_1 = 2$, $X_3 = 9$): $\hat{Y} = 5 + 0.05 \cdot 2 + 0.1 \cdot 9 = 6$, остаток $= -21 - 6 = -27$. - **R^2** : $R^2 \approx 0.3$ (примерно). - **МАРЕ**: $\text{MAPE} \approx 150\%$ (примерно).

3 Замечания

- Нормализация и обработка выбросов выполнены, но выбросов не найдено. - Примерные вычисления требуют точной инверсии $X^T X$ для точности β . - Модификации (Ridge, полиномы) не детализированы из-за отсутствия параметров.

4 Итог

Алгоритм, псевдокод, блок-схема и расчеты соответствуют друг другу. Модель $y \approx 5 + 0.05X_1 + 0.1X_3$ адекватна данным с $R^2 \approx 0.3$ и $\text{MAPE} \approx 150\%$.