УДК 519.862.6
DOI: 10.22213/2410-9304-2023-3-40-47

Алгоритм интерпретации линейных и неэлементарных линейных регрессионных моделей в условиях мультиколлинеарности

М. П. Базилевский, кандидат технических наук, доцент,
Иркутский государственный университет путей сообщения, Иркутск, Россия

При построении моделей машинного обучения все больше исследователей приходят к пониманию того, что помимо точности модели важна и ее интерпретируемость, означающая степень понятности человеку. В настоящее время в науке формируется новая область – интерпретируемое машинное обучение. Статья посвящена исследованию вопросов интерпретации линейных регрессионных моделей. Традиционно линейные регрессии принято интерпретировать в условиях слабой корреляции объясняющих переменных. В таком случае по известному шаблону можно объяснить влияние каждой входной переменной на выходную переменную. Часто же объясняющие переменные сильно коррелируют между собой. В такой ситуации рекомендуют исключать сильно коррелирующие переменные, что приводит, во-первых, к снижению качества модели, во-вторых, к потере целостности исследования и интерпретации изучаемого процесса или явления. В данной работе предложен алгоритм интерпретации линейной регрессии, построенной при любой степени корреляции объясняющих переменных. При слабой корреляции всех объясняющих переменных алгоритм дает традиционную интерпретацию линейной регрессии, а при сильной корреляции, как следует из доказанной в статье теоремы, интерпретация регрессии с несколькими переменными сводится к интерпретации уравнения с одной переменной, но при этом не теряется информация о связях между парами объясняющих переменных. Алгоритм может применяться не только для обычных линейных регрессий, но и для неэлементарных, содержащих помимо объясняющих переменных их пары, преобразованные с помощью бинарных операций min и max. Работа алгоритма продемонстрирована на конкретных примерах.

Ключевые слова: машинное обучение, интерпретируемость, линейная регрессия, неэлементарная линейная регрессия, мультиколлинеарность, метод наименьших квадратов.

Введение

В настоящее время машинному обучению [1–3] уделяется значительное внимание как инструменту, позволяющему эффективно справляться с задачами прогнозирования сложных явлений и процессов. Однако, как отмечено в [4], использование лишь одной метрики, такой как точность модели, недостаточно для полного описания большинства реальных задач. Поэтому растет понимание важности интерпретации построенной модели, что способствует выявлению и устранению ее «уязвимых» мест и, в конечном счете, повышает доверие к модели у экспертов из данной предметной области. В результате в науке сегодня формируется новая область машинного обучения – интерпретируемое машинное обучение [5, 6]. К сожалению, четкого определения интерпретируемости модели пока не существует.

В монографии К. Молнара утверждается, что линейные регрессии являются широко используемыми интерпретируемыми моделями машинного обучения. Модель множественной линейной регрессии имеет вид

$$y_i = \alpha_0 + \sum_{j=1}^{l} \alpha_j x_{ij} + \varepsilon_i, \ i = \overline{1,n}, \qquad (1)$$

где n – объем выборки; l – число объясняющих (независимых) переменных; y_i, $i = \overline{1,n}$, x_{ij}, $i = \overline{1,n}$, $j = \overline{1,l}$ – значения объясняемой (зависимой) переменной y и объясняющих переменных x_1, x_2, ..., x_l; α_0, α_1, ..., α_l – неизвестные параметры; ε_i, $i = \overline{1,n}$ – ошибки аппроксимации.

Существует широкий спектр методов оценивания неизвестных параметров линейной регрессии (1). Описание многих из них можно найти в монографии [7]. С вычислительной точки зрения самым эффективным из них считается метод наименьших квадратов (МНК). Обозначим оцененную с помощью МНК модель (1) в виде

$$\tilde{y} = \tilde{\alpha}_0 + \sum_{j=1}^{l} \tilde{\alpha}_j x_j, \qquad (2)$$

где $\tilde{y}$ – расчетные (прогнозные) значения переменной y; $\tilde{\alpha}_0$, $\tilde{\alpha}_1$, ..., $\tilde{\alpha}_l$ – оценки неизвестных параметров.

Интерпретация весовых коэффициентов в уравнении (2) традиционно происходит по следующему текстовому шаблону: изменение переменной x_j на 1 единицу (при неизменных значениях остальных объясняющих переменных) приводит к изменению прогноза $\tilde{y}$ на $\tilde{\alpha}_j$ единиц.

Однако такая интерпретация справедлива только тогда, когда объясняющие переменные не коррелируют между собой, т. е. при отсутствии мультиколлинеарности [8, 9]. В противном случае изменение переменной x_j может приводить к изменению других объясняющих переменных, поэтому уравнение (2) превращается в некую смесь линейных зависимостей и его интерпретация затрудняется.

В работе [10] авторы утверждают, что необходимым и достаточным условием однозначной интерпретации линейной регрессии является диагональный вид корреляционной матрицы объясняющих переменных. С этим нельзя не согласиться. В той же работе авторы в зависимости от степени коррелируемости факторов выделяют следующие уровни интерпретируемости линейной регрессии: точная, приближенная, грубая, неверная и невозможная. По мнению автора данной статьи, классифицировать интерпретацию линейной регрессии как невозможную при сильной корреляции объясняющих переменных не совсем верно.

Цель данной работы состоит в разработке алгоритма интерпретации некоторых видов линейных по параметрам регрессионных моделей в условиях корреляции объясняющих переменных.

Алгоритм интерпретации

Предлагается следующий алгоритм интерпретации влияния j^*-й переменной на $\tilde{y}$ в полученном с помощью МНК уравнении линейной регрессии (2).

1. Определяются коэффициенты парной корреляции между выбранной переменной x_{j^*} и всеми другими объясняющими переменными.

2. Для заданного уровня значимости α для каждого найденного выборочного коэффициента корреляции проверяется гипотеза [11] о равенстве генерального коэффициента корреляции нулю. В результате формируется два индексных множества: M_1, содержащее номера объясняющих переменных значимо коррелирующих с x_{j^*}, и M_2, включающее номера объясняющих переменных незначимо коррелирующих с x_{j^*}.

3. С помощью МНК оцениваются парные линейные регрессии, в которых зависимыми переменными выступают переменные с номерами из множества M_1, а независимой переменной – x_{j^*}. Обозначим коэффициенты при переменных этих регрессий k_1, k_2 и т. д.

4. В уравнении (2) переменные с номерами из множества M_1 заменяются на полученные на предыдущем шаге зависимости и приводятся подобные. Обозначим буквой w полученный коэффициент при переменной x_{j^*}.

5. Осуществляется интерпретация по следующему текстовому шаблону: изменение переменной x_{j^*} на 1 единицу приводит к изменению переменных с номерами из множества M_1 на k_1, k_2 и т. д. единиц соответственно, что в совокупности (при неизменных значениях объясняющих переменных с номерами из множества M_2) приводит к изменению прогноза $\tilde{y}$ на w единиц.

Если множество M_1 пусто, т. е. все объясняющие переменные вообще не коррелируют между собой либо коррелируют слабо, то интерпретация уравнения (2) по представленному алгоритму эквивалентна рассмотренной во введении традиционной в регрессионном анализе интерпретации линейной регрессии.

Ответ на вопрос, как интерпретируется линейная регрессия по предложенному алгоритму при сильной корреляции всех объясняющих переменных, дает следующая теорема.

Теорема. Пусть с помощью МНК оценены парные линейные регрессии зависимости объясняющих переменных с номерами $\{1,2,3,...,l\}\setminus\{j^*\}$ от переменной x_{j^*}. Тогда при замене в уравнении (2) всех объясняющих переменных с номерами $\{1,2,3,...,l\}\setminus\{j^*\}$ найденными соотношениями будет получено уравнение оцененной с помощью МНК парной линейной регрессии зависимости y от x_{j^*}.

Доказательство. Проведем стандартизацию (нормирование) всех переменных по формулам [12]:

$$y^* = \frac{y-\overline{y}}{\sigma_y}\,;\; x_j^* = \frac{x_j-\overline{x_j}}{\sigma_{x_j}},\; j=\overline{1,l}\,,$$

где σ_y, σ_{x_j}, $j=\overline{1,l}$ – стандартные отклонения переменных.

Тогда модель множественной линейной регрессии (1) в стандартизованном масштабе принимает вид:

$$y_i^* = \sum_{j=1}^{l} \beta_j x_{ij}^* + \xi_i,\ i = \overline{1,n}, \quad (3)$$

где β_j, $j = \overline{1,l}$ – стандартизованные коэффициенты регрессии.

МНК-оценки модели (3) являются результатом решения системы линейных алгебраических уравнений:

$$R_{xx} \cdot \beta = R_{yx}, \quad (4)$$

где R_{xx} – матрица парных коэффициентов корреляции между объясняющими переменными; R_{yx} – вектор парных коэффициентов корреляции между переменной y и объясняющими переменными; β – вектор стандартизованных коэффициентов регрессии.

Обозначим оцененную с помощью МНК модель (3) в виде:

$$\tilde{y}^* = \sum_{j=1}^{l} \tilde{\beta}_j x_j^*, \quad (5)$$

где $\tilde{\beta}_1, ..., \tilde{\beta}_l$ – оценки неизвестных параметров.

Пусть с помощью МНК оценены парные линейные регрессии с зависимыми переменными $x_2^*, x_3^*, ..., x_l^*$ и независимой переменной x_1^*:

$$\tilde{x}_2^* = \tilde{\beta}_{21} x_1^*,\ \tilde{x}_3^* = \tilde{\beta}_{31} x_1^*,\ ...,\ \tilde{x}_l^* = \tilde{\beta}_{l1} x_1^*, \quad (6)$$

в которых, как следует из (4), оценки равны соответствующим коэффициентам парной корреляции, т. е. $\tilde{\beta}_{21} = r_{x_1x_2}$, $\tilde{\beta}_{31} = r_{x_1x_3}$, ..., $\tilde{\beta}_{l1} = r_{x_1x_l}$.

Подставим найденные зависимости (6) в уравнение (5):

$$\tilde{y}^* = \tilde{\beta}_1 x_1^* + \tilde{\beta}_2 r_{x_1x_2} x_1^* + \tilde{\beta}_3 r_{x_1x_3} x_1^* + ... + \tilde{\beta}_l r_{x_1x_l} x_1^* =$$
$$= \left(\tilde{\beta}_1 + \tilde{\beta}_2 r_{x_1x_2} + \tilde{\beta}_3 r_{x_1x_3} + ... + \tilde{\beta}_l r_{x_1x_l}\right) x_1^*.$$

Можно заметить, что в этом выражении в скобках содержится левая часть первого уравнения системы (4), поэтому $\tilde{\beta}_1 + \tilde{\beta}_2 r_{x_1x_2} + \tilde{\beta}_3 r_{x_1x_3} + ... + \tilde{\beta}_l r_{x_1x_l} = r_{yx_1}$.

Отсюда следует, что $\tilde{y}^* = r_{yx_1} x_1^*$, т. е. уравнение (5) стало уравнением оцененной с помощью МНК в стандартизованном виде парной линейной регрессии зависимости y^* от x_1^*. Аналогичные рассуждения справедливы для любой выбранной переменной.

Теорема доказана.

Из теоремы следует, что если множество M_2 пусто, т. е. все объясняющие переменные коррелируют сильно между собой, то интерпретация влияния любой объясняющей переменной x_{j^*} на y по представленному алгоритму сводится к интерпретации коэффициента парной линейной регрессии y от x_{j^*}. В таком случае интерпретацию нужно проводить по следующему текстовому шаблону: изменение переменной x_{j^*} на 1 единицу приводит к изменению других переменных из множества M_1 на k_1, k_2 и т. д. единиц соответственно, что в совокупности приводит к изменению прогноза $\tilde{y}$ на w единиц.

Стоить заметить, что проверка значимости коэффициента парной корреляции эквивалентна проверке значимости коэффициента парной линейной регрессии между теми же переменными. В обоих случаях проверка осуществляется с помощью t-критерия Стьюдента. Поэтому на 2-м шаге представленного алгоритма вместо проверки значимости коэффициентов парной корреляции можно сразу переходить к оцениванию соответствующих парных линейных регрессий, в которых проверять значимость коэффициентов при переменных по t-критерию.

Отметим также, что предложенный алгоритм интерпретации можно применять и для неэлементарных линейных регрессий (НЛР) [13–15] с бинарными операциями min и max:

$$y_i = \alpha_0 + \sum_{j=1}^{l} \alpha_j x_{ij} + \sum_{j=1}^{p} \alpha_j^{\min} \min\{x_{i,\mu_{j1}}, k_j^{\min} x_{i,\mu_{j2}}\} +$$
$$+ \sum_{j=1}^{p} \alpha_j^{\max} \max\{x_{i,\mu_{j1}}, k_j^{\max} x_{i,\mu_{j2}}\} + \varepsilon_i,\ i = \overline{1,n},$$

где $\alpha_j^{\min}$, $\alpha_j^{\max}$, $k_j^{\min}$, $k_j^{\max}$, $j = \overline{1,p}$ – неизвестные параметры; $p = C_l^2$ – число всех возможных комбинаций пар входных переменных; μ_{j1}, μ_{j2}, $j = \overline{1,p}$ – элементы первого и второго столбца матрицы M размера $p \times 2$, содержащей по строкам в лексикографическом порядке индексы всех возможных комбинаций пар входных переменных.

Для этого оцененную НЛР сначала необходимо представить в виде кусочно-заданной функции, а потом на каждом из интервалов, составляющих область ее определения, применить предложенный алгоритм.

Термин «неэлементарная» появился потому, что в качестве регрессоров в НЛР выступают неэлементарные функции. Неэлементарные линейные регрессии являются частным случаем квазилинейных регрессий.

Примеры интерпретации регрессионных моделей

Для демонстрации работы предложенного алгоритма интерпретации линейных регрессий были использованы данные из статьи [16] по следующим переменным: y – отправление грузов железнодорожным транспортом общего пользования в Иркутской области (млн руб.); x_2 – процент трудоспособного населения от общей численности; x_3 – численность рабочей силы (тыс. чел.); x_5 – численность пенсионеров (тыс. чел.); x_8 – число собственных легковых автомобилей на 1000 человек населения (шт.); x_{18} – число предприятий и организаций; x_{20} – кредиторская задолженность организаций (млн руб.); x_{22} – производство электроэнергии (млрд кВт.ч); x_{58} – тарифы на грузовые перевозки (железнодорожный транспорт), усл. ед. С помощью предложенного алгоритма для различных моделей интерпретировалось влияние переменной x_2 на y.

Пример 1. Имитировалась ситуация, когда линейная регрессия построена в условиях сильной корреляции объясняющих переменных. Для этого была оценена зависимость переменной y от x_2, x_3, x_5 и x_8:

$$\begin{aligned}\tilde{y} = -481,309 + 4,29223x_2 + \\ +0,03786x_3 + 0,3385x_5 - 0,0379x_8.\end{aligned} \quad (7)$$

Коэффициент детерминации модели (7) R^2 составляет 0,902. Наименьшая по модулю величина среди парных коэффициентов корреляции объясняющих переменных составляет 0,746 между x_3 и x_8.

Для интерпретации были построены парные линейные регрессии:

$$\tilde{x}_3 = 463,232 + \underset{(6,242)}{13,3279}\,x_2,\ R^2 = 0,672, \quad (8)$$

$$\tilde{x}_5 = 1326,2 - \underset{(-9,092)}{10,2552}\,x_2,\ R^2 = 0,813, \quad (9)$$

$$\tilde{x}_8 = 1025,34 - \underset{(-6,577)}{13,6747}\,x_2,\ R^2 = 0,695, \quad (10)$$

где в скобках под коэффициентами указаны наблюдаемые значения t-критерия Стьюдента, подтверждающие их значимость на уровне $\alpha = 0,01$.

Подстановка зависимостей (8)–(10) в уравнение (7), как и следовало ожидать, преобразовала его в уравнение оцененной с помощью МНК парной линейной регрессии:

$$\tilde{y} = -53,7045 + 1,8437x_2.$$

Тогда справедлива следующая интерпретация: с увеличением трудоспособного населения x_2 на 1 % примерно на 13328 человек увеличивается численность рабочей силы x_3, примерно на 10255 человек снижается численность пенсионеров x_5, примерно на 13,67 штук снижается число собственных легковых автомобилей x_8, что в совокупности приводит к увеличению отправления грузов y примерно на 1,844 млн руб.

Пример 2. Имитировалась ситуация, когда линейная регрессия построена в условиях различной корреляции объясняющих переменных. Для этого была оценена зависимость переменной y от x_2, x_3, x_5, x_8, x_{18}, x_{20}, x_{22} и x_{58}:

$$\begin{aligned}\tilde{y} = -443,324 + 3,42019x_2 + 0,04439x_3 + \\ +0,3252x_5 - 0,05637x_8 + 0,0001369x_{18} - \\ -4,446 \cdot 10^{-6}x_{20} + 0,229x_{22} - 0,000466x_{58}.\end{aligned} \quad (11)$$

Для (11) $R^2 = 0,906$.

Для интерпретации были построены парные линейные регрессии зависимости объясняющих переменных от x_2. Со значимыми коэффициентами оказались регрессии (8)–(10), а также

$$\tilde{x}_{20} = 1,1677 \cdot 10^6 - \underset{(-2,984)}{16691,6}\,x_2,\ R^2 = 0,319, \quad (12)$$

$$\tilde{x}_{58} = 24892,3 - \underset{(-8,118)}{369,698}\,x_2,\ R^2 = 0,776. \quad (13)$$

Подстановка зависимостей (8)–(10), (12), (13) в уравнение (11) позволила получить следующее равенство:

$$\begin{aligned}\tilde{y} = -66,085 + 1,6944x_2 + \\ +0,0001369x_{18} + 0,22904x_{22}.\end{aligned} \quad (14)$$

Тогда справедлива следующая интерпретация: с увеличением трудоспособного населения x_2 на 1 % примерно на 13328 человек увеличивается численность рабочей силы x_3, примерно на 10255 человек снижается численность пенсионеров x_5, примерно на 13,67 штук снижается число собственных легковых автомобилей x_8, примерно на 16691,6 млн руб. снижается кредиторская задолженность x_{20}, примерно на 369,7 усл. ед. снижаются тарифы на грузовые перевозки x_{58}, что в совокупности (при неизменных значениях переменных x_{18} и x_{22}) приводит к увеличению отправления грузов y примерно на 1,694 млн руб.

Заметим, что уравнение (14) разнится с уравнением оцененной с помощью МНК линейной регрессии зависимости y от x_2, x_{18}, x_{22}.

Как видно, в примерах 1 и 2 влияние переменной x_2 на y интерпретировано по-разному: в примере 1 с ростом x_2 на 1 % переменная y увеличивается примерно на 1,844 млн руб., а в примере 2 – на 1,694 млн руб.

Это естественно, поскольку интерпретация велась по двум разным моделям, построенным по данным с не абсолютно полным отсутствием корреляции между объясняющими переменными.

Ориентироваться в такой ситуации следует на модель с наибольшим значением коэффициента детерминации R^2.

Пример 3. Интерпретировалась построенная в статье [17] НЛР:

$$\begin{aligned}\tilde{y} = &-24,5274 + 1,1895\min\{x_2, 0,000933x_{18}\} - \\ &-0,0196\min\{x_5, 0,006754x_{20}\} - \\ &-0,0323\min\{x_8, 0,11725x_{58}\} + \\ &+0,0254\max\{x_3, 23,079x_{22}\}. \end{aligned} \qquad (15)$$

Для (15) $R^2 = 0,946$.

В той же работе НЛР (15) представлена в виде кусочно-заданной функции. Подстановка в нее зависимостей (8)–(10), (12), (13) привела к следующему выражению:

$$\tilde{y} = \begin{cases} -259,153 + 3,913x_2 + 0,0011x_{18} \text{ при } \frac{x_2}{x_{18}} \ge 0,000933, \frac{x_5}{x_{20}} \ge 0,00675, \frac{x_8}{x_{58}} \ge 0,117, \frac{x_3}{x_{22}} \ge 23,08; \\ -270,919 + 3,575x_2 + 0,0011x_{18} + 0,5857x_{22} \text{ при } \frac{x_2}{x_{18}} \ge 0,000933, \frac{x_5}{x_{20}} \ge 0,00675, \frac{x_8}{x_{58}} \ge 0,117, \frac{x_3}{x_{22}} < 23,08; \\ -197,681 + 2,95x_2 + 0,0011x_{18} \text{ при } \frac{x_2}{x_{18}} \ge 0,000933, \frac{x_5}{x_{20}} \ge 0,00675, \frac{x_8}{x_{58}} < 0,117, \frac{x_3}{x_{22}} \ge 23,08; \\ -209,447 + 2,612x_2 + 0,0011x_{18} + 0,5857x_{22} \text{ при } \frac{x_2}{x_{18}} \ge 0,000933, \frac{x_5}{x_{20}} \ge 0,00675, \frac{x_8}{x_{58}} < 0,117, \frac{x_3}{x_{22}} < 23,08; \\ -133,346 + 1,944x_2 + 0,0011x_{18} \text{ при } \frac{x_2}{x_{18}} \ge 0,000933, \frac{x_5}{x_{20}} < 0,00675, \frac{x_8}{x_{58}} \ge 0,117, \frac{x_3}{x_{22}} \ge 23,08; \\ -145,112 + 1,606x_2 + 0,0011x_{18} + 0,5857x_{22} \text{ при } \frac{x_2}{x_{18}} \ge 0,000933, \frac{x_5}{x_{20}} < 0,00675, \frac{x_8}{x_{58}} \ge 0,117, \frac{x_3}{x_{22}} < 23,08; \\ -71,873 + 0,981x_2 + 0,0011x_{18} \text{ при } \frac{x_2}{x_{18}} \ge 0,000933, \frac{x_5}{x_{20}} < 0,00675, \frac{x_8}{x_{58}} < 0,117, \frac{x_3}{x_{22}} \ge 23,08; \\ -83,64 + 0,643x_2 + 0,0011x_{18} + 0,5857x_{22} \text{ при } \frac{x_2}{x_{18}} \ge 0,000933, \frac{x_5}{x_{20}} < 0,00675, \frac{x_8}{x_{58}} < 0,117, \frac{x_3}{x_{22}} < 23,08; \\ -259,153 + 5,103x_2 \text{ при } \frac{x_2}{x_{18}} < 0,000933, \frac{x_5}{x_{20}} \ge 0,00675, \frac{x_8}{x_{58}} \ge 0,117, \frac{x_3}{x_{22}} \ge 23,08; \\ -270,919 + 4,764x_2 + 0,5857x_{22} \text{ при } \frac{x_2}{x_{18}} < 0,000933, \frac{x_5}{x_{20}} \ge 0,00675, \frac{x_8}{x_{58}} \ge 0,117, \frac{x_3}{x_{22}} < 23,08; \\ -197,681 + 4,14x_2 \text{ при } \frac{x_2}{x_{18}} < 0,000933, \frac{x_5}{x_{20}} \ge 0,00675, \frac{x_8}{x_{58}} < 0,117, \frac{x_3}{x_{22}} \ge 23,08; \\ -209,447 + 3,801x_2 + 0,5857x_{22} \text{ при } \frac{x_2}{x_{18}} < 0,000933, \frac{x_5}{x_{20}} \ge 0,00675, \frac{x_8}{x_{58}} < 0,117, \frac{x_3}{x_{22}} < 23,08; \\ -133,346 + 3,134x_2 \text{ при } \frac{x_2}{x_{18}} < 0,000933, \frac{x_5}{x_{20}} < 0,00675, \frac{x_8}{x_{58}} \ge 0,117, \frac{x_3}{x_{22}} \ge 23,08; \\ -145,112 + 2,795x_2 + 0,5857x_{22} \text{ при } \frac{x_2}{x_{18}} < 0,000933, \frac{x_5}{x_{20}} < 0,00675, \frac{x_8}{x_{58}} \ge 0,117, \frac{x_3}{x_{22}} < 23,08; \\ -71,873 + 2,171x_2 \text{ при } \frac{x_2}{x_{18}} < 0,000933, \frac{x_5}{x_{20}} < 0,00675, \frac{x_8}{x_{58}} < 0,117, \frac{x_3}{x_{22}} \ge 23,08; \\ -83,639 + 1,832x_2 + 0,5857x_{22} \text{ при } \frac{x_2}{x_{18}} < 0,000933, \frac{x_5}{x_{20}} < 0,00675, \frac{x_8}{x_{58}} < 0,117, \frac{x_3}{x_{22}} < 23,08. \end{cases}$$

Таким образом, в зависимости от интервала меняется степень влияния переменной x_2 на y. Свое наименьшее значение (0,643) коэффициент при переменной x_2 достигает при условиях

$\frac{x_2}{x_{18}} \geq 0,000933$, $\frac{x_5}{x_{20}} < 0,00675$, $\frac{x_8}{x_{58}} < 0,117$, $\frac{x_3}{x_{22}} < 23,08$, а наибольшее (5,103) при условиях $\frac{x_2}{x_{18}} < 0,000933$, $\frac{x_5}{x_{20}} \geq 0,00675$, $\frac{x_8}{x_{58}} \geq 0,117$, $\frac{x_3}{x_{22}} \geq 23,08$. В последнем случае справедлива следующая интерпретация: при условиях $\frac{x_2}{x_{18}} < 0,000933$, $\frac{x_5}{x_{20}} \geq 0,00675$, $\frac{x_8}{x_{58}} \geq 0,117$, $\frac{x_3}{x_{22}} \geq 23,08$ с увеличением трудоспособного населения x_2 на 1 % примерно на 13328 человек увеличивается численность рабочей силы x_3, примерно на 10255 человек снижается численность пенсионеров x_5, примерно на 13,67 штук снижается число собственных легковых автомобилей x_8, примерно на 16691,6 млн руб. снижается кредиторская задолженность x_{20}, примерно на 369,7 усл. ед. снижаются тарифы на грузовые перевозки x_{58}, что в совокупности (при неизменных значениях переменных x_{18} и x_{22}) приводит к увеличению отправления грузов y примерно на 5,103 млн руб.

Заключение

В работе предложен алгоритм интерпретации влияния j^*-й переменной на $\tilde{y}$ в полученном с помощью МНК уравнении линейной регрессии. Алгоритм может работать в условиях произвольной степени корреляции объясняющих переменных. При слабой корреляции объясняющих переменных из алгоритма вытекает традиционная схема интерпретации весовых коэффициентов уравнения линейной регрессии. Доказано, что при сильной корреляции всех объясняющих переменных интерпретация множественной линейной регрессии предложенным алгоритмом эквивалентна интерпретации коэффициентов соответствующих парных регрессий. Работа алгоритма успешно продемонстрирована не только на примере линейных регрессий, но и на примере неэлементарных линейных регрессий.

Библиографические ссылки

1. A guide to machine learning for biologists / J. G. Greener, S. M. Kandathil, L. Moffat, D. T. Jones // Nature Reviews Molecular Cell Biology. 2022. Vol. 23, no. 1. Pp. 40-55.

2. *Janiesch C., Zschech P., Heinrich K.* Machine learning and deep learning // Electronic Markets. 2021. Vol. 31, no. 3. Pp. 685-695.

3. *Sarker I. H.* Machine learning: Algorithms, real-world applications and research directions // SN computer science. 2021. Vol. 2, no. 3. P. 160.

4. *Doshi-Velez F., Kim B.* Towards a rigorous science of interpretable machine learning // arXiv preprint arXiv:1702.08608. 2017.

5. *Molnar C.* Interpretable machine learning. Lulu.com, 2020.

6. *Du M., Liu N., Hu X.* Techniques for interpretable machine learning // Communications of the ACM. 2019. Vol. 63, no. 1. Pp. 68-77.

7. *Носков С. И.* Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных. Иркутск : Облинформпечать, 1996. 321 с.

8. *Shrestha N.* Detecting multicollinearity in regression analysis // American Journal of Applied Mathematics and Statistics. 2020. Vol. 8, no. 2. Pp. 39-42.

9. *Kim J. H.* Multicollinearity and misleading statistical results // Korean journal of anesthesiology. 2019. Vol. 72, no. 6. Pp. 558-569.

10. *Горбач А. Н., Цейтлин Н. А.* Покупательское поведение: анализ спонтанных последовательностей и регрессионных моделей в маркетинговых исследованиях. Киев : Освiта УкраЇны, 2011. 220 с.

11. *Гефан Г. Д.* Эконометрика. Иркутск : ИрГУПС, 2005. 84 с.

12. *Фёрстер Э., Рёнц Б.* Методы корреляционного и регрессионного анализа. М. : Финансы и статистика, 1983. 303 с.

13. *Базилевский М. П.* Метод построения неэлементарных линейных регрессий на основе аппарата математического программирования // Проблемы управления. 2022. № 4. С. 3–14.

14. *Базилевский М. П.* Оценивание линейно-неэлементарных регрессионных моделей с помощью метода наименьших квадратов // Моделирование, оптимизация и информационные технологии. 2020. Т. 8, № 4 (31).

15. *Базилевский М. П.* Отбор информативных операций при построении линейно-неэлементарных регрессионных моделей // International Journal of Open Information Technologies. 2021. Т. 9, № 5. С. 30–35.

16. *Базилевский М. П.* Метод построения неэлементарных линейных регрессий на основе аппарата математического программирования // Проблемы управления. 2022. № 4. С. 3–14.

17. Там же.

References

1. Greener J.D., Kandathil S.M., Moffat L., Jones D.T. [A guide to machine learning for biologists]. *Nature Reviews Molecular Cell Biology*, 2022, vol. 23, no. 1, pp. 40-55.

2. Janiesch C., Zschech P., Heinrich K. [Machine learning and deep learning]. *Electronic Markets*, 2021, vol. 31, no. 3, pp. 685–695.

3. Sarker I.H. [Machine learning: Algorithms, real-world applications and research directions]. *SN computer science*, 2021, vol. 2, no. 3. pp. 160.

4. Doshi-Velez F., Kim B. [Towards a rigorous science of interpretable machine learning]. *arXiv preprint arXiv:1702.08608*, 2017.

5. Molnar C. *Interpretable machine learning*. Lulu. com, 2020.

6. Du M., Liu N., Hu X. [Techniques for interpretable machine learning]. *Communications of the ACM*, 2019, vol. 63, no. 1, pp. 68–77.

7. Noskov S.I. *Tekhnologiya modelirovaniya ob"ektov s nestabil'nym funktsionirovaniem i neopredelennost'yu v dannykh* [Technology for modeling objects with unstable operation and uncertainty in data]. Irkutsk, Oblinformpechat' Publ., 1996, 321 p. (in Russ.).

8. Shrestha N. [Detecting multicollinearity in regression analysis]. *American Journal of Applied Mathematics and Statistic*s, 2020, vol. 8, no. 2, pp. 39-42.

9. Kim J.H. [Multicollinearity and misleading statistical results]. *Korean journal of anesthesiology*, 2019, vol. 72, no. 6, pp. 558-569.

10. Gorbach A.N., Tseitlin N.A. *Pokupatel'skoe povedenie: analiz spontannykh posledovatel'nostei i regressionnykh modelei v marketingovykh issledovaniyakh* [Buying Behavior: An Analysis of Spontaneous Sequences and Regression Models in Marketing Research]. Kiev, Osvita UkraÏny Publ., 2011, 220 p. (in Russ.).

11. Gefan G.D. *Ekonometrika* [Econometrics]. Irkutsk, IrGUPS Publ., 2005, 84 p. (in Russ.).

12. Ferster E., Rents B. *Metody korrelyatsionnogo i regressionnogo analiza* [Methods of correlation and regression analysis]. Moscow, Finansy i statistika Publ., 1983, 303 p. (in Russ.).

13. Bazilevskii M.P. *Metod postroeniya neelementarnykh lineinykh regressii na osnove apparata matematicheskogo programmirovaniya* [A method for constructing nonelementary linear regressions based on mathematical programming]. *Control Sciences*, 2022, no. 4, pp. 3–14. (in Russ.).

14. Bazilevskii M.P. Otsenivanie lineino-neelementarnykh regressionnykh modelei s pomoshch'yu metoda naimen'shikh kvadratov [Estimation linear non-elementary regression models using ordinary least squares]. *Modeling, optimization and information technology*, 2020, vol. 8, no. 4 (31). (in Russ.).

15. Bazilevskii M.P. Otbor informativnykh operatsii pri postroenii lineino-neelementarnykh regressionnykh modelei [Selection of informative operations in the construction of linear non-elementary regression models]. *International Journal of Open Information Technologies*, 2021, vol. 9, no. 5, pp. 30–35. (in Russ.).

16. Bazilevskii M.P. *Metod postroeniya neelementarnykh lineinykh regressii na osnove apparata matematicheskogo programmirovaniya* [A method for constructing nonelementary linear regressions based on mathematical programming]. *Control Sciences*, 2022, no. 4, pp. 3–14. (in Russ.).

17. Ibid.

* * *

Interpretation Algorithm of Linear and Non-elementary Linear Regression Models Under Multicollinearity Conditions

M. P. Bazilevskiy, PhD in Engineering, Associate Professor, Department of Mathematics, Irkutsk State Transport University, Irkutsk, Russia

When constructing machine learning models, more and more researchers come to understanding that in addition to the accuracy of the model, its interpretability is also important, that means the degree of understandability to a person. Currently, a new field is being formed in science - interpretable machine learning. This article is devoted to the study of linear regression models interpretation questions. Traditionally, linear regressions are interpreted in terms of weak correlation of explanatory variables. In this case, according to a well-known pattern, it is possible to explain the influence of each input variable on the output variable. Often the explanatory variables are highly correlated with each other. In such a situation, it is recommended to exclude strongly correlated variables, that leads, firstly, to a decrease in the model quality, and secondly, to integrity loss of the study and the interpretation of a process or phenomenon under study. In this paper, we propose an algorithm for interpreting a linear regression constructed for any degree of explanatory variablecorrelation. The algorithm gives the traditional interpretation of linear regression with weak correlation of all explanatory variables, while in case of strong correlation, as it follows from the theorem proved in the article, regression interpretation with several variables is reduced to interpreting an equation with one variable, without losing information about the relationships between pairs of explanatory variables. The algorithm can be used not only for simple linear regressions, but also for non-elementary ones, containing, in addition to explanatory variables, their pairs transformed by means of binary operations min and max. The operation of the algorithm is demonstrated on specific examples.

Keywords: machine learning, interpretability, linear regression, non-elementary linear regression, multicollinearity, least squares.

Получено: 04.05.23