# Метод наименьших углов для логистической регрессии

*Скипор К. С., Стрижов В. В.*
skiporkonstantin@mail.ru, strijov@ccas.ru
Москва, МФТИ

Предлагается и исследуется алгоритм отбора признаков для решения задач восстановления логистической регрессии. Приводится математическое обоснование предложенного алгоритма. Алгоритм основан на методе наименьших углов для модели линейной регрессии с использованием дополнительной линеаризации функционала качества. Работа алгоритма проиллюстрирована задачей изучения факторов риска ишемических заболеваний сердца.

В работе рассматривается отыскание из множества признаков такого его подмножества, для которого их линейная комбинация наиболее точно описывает данные. В 1966 году Дрейпером был предложен ступенчатый алгоритм выбора признаков (Forward Stagewise) [1, 2, 3]. Позднее был предложен алгоритм Forward Selection [4], представляющий собой модифицированную версию Forward Stagewise. В 1970 году Хоэрл и Кеннард предложили метод гребневой регрессии (Ridge Regression) [5], в котором использовалась регуляризация — дополнительное ограничение [6] на задачу отбора признаков. Еще один метод регуляризации, Лассо (The Lasso), был предложен Тибширани в 1996 году [7]. В нем вводится ограничение на $L_1$-норму вектора параметров модели. В модели логистической регрессии этот метод также называется $L_1$-regularized Logistic Regression [2]. В 2002 году Эфрон, Хасти, Джонстон и Тибширани предложили метод наименьших углов (Least Angle Regression) [8]. Предложенный ими алгоритм LARS для линейных моделей заключается в последовательном добавлении признаков. На каждом шаге признак выбирается таким образом, что вектор регрессионных остатков равноуголен уже добавленным в модель признакам [9]. В 2004 году Мадиган и Ридгевэй предложили идею применения данного метода при использовании линеаризации для обобщенных линейных моделей, в частности, для модели логистической регрессии [10]. Реализация этой идеи лежит в основе данной работы.

## Постановка задачи отбора признаков

Дана выборка $D = \{(\boldsymbol{x}^i, y^i)\}_{i=1}^m$, в которой $i$-й объект описывается строкой из $n$ числовых признаков, $\boldsymbol{x}^i = (x_j^i)_{j=1}^n \in \mathbb{R}^n$ и метки класса $y^i \in \{0, 1\}$. Верхний индекс $i$ указывает порядковый номер объекта выборки, нижний индекс $j$ — порядковый номер признака. Векторы признаков $\boldsymbol{x}_j = (x_j^1, \ldots, x_j^m)^\intercal$ являются линейно независимыми свободными переменными, а вектор $\boldsymbol{y} = (y^1, \ldots, y^m)^\intercal$ является зависимой переменной. Предполагается, что $y^i$ имеет распределение Бернулли. Без ограничения общности будем считать, что признаки $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ стандартизованы

$$\|\boldsymbol{x}_j\|_1 = \sum_{i=1}^m x_j^i = 0, \quad \|\boldsymbol{x}_j\|_2^2 = \sum_{i=1}^m (x_j^i)^2 = 1. \tag{1}$$

Для удобства описания алгоритма обозначим матрицу признаков $X = (\boldsymbol{x}_1 \ldots \boldsymbol{x}_n)$ и вектор параметров $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)^\intercal$. Принята модель логистической регрессии, согласно которой

$$\boldsymbol{y} = \boldsymbol{\sigma}(X, \boldsymbol{\beta}) + \varepsilon, \tag{2}$$

где $\boldsymbol{\sigma}(X, \boldsymbol{\beta})$ — сигмоидная функция

$$\boldsymbol{\sigma}(X, \boldsymbol{\beta}) = \frac{1}{1 + \exp(-X\boldsymbol{\beta})}. \tag{3}$$

Обозначим функцию регрессии

$$\boldsymbol{\mu}(X, \boldsymbol{\beta}) = \sum_{j=1}^n \boldsymbol{x}_j \beta_j = X\boldsymbol{\beta}. \tag{4}$$

Критерием качества модели назначен функционал логарифма правдоподобия

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n \left( y^i \boldsymbol{x}^i \boldsymbol{\beta} - \ln(1 + \exp(\boldsymbol{x}^i \boldsymbol{\beta})) \right). \tag{5}$$

Требуется построить такой алгоритм последовательного добавления признаков, который на каждом шаге определяет набор признаков с множеством индексов $\mathcal{A}$ и соответствующий набору ненулевой вектор параметров $\boldsymbol{\beta}_{\mathcal{A}}$, такой, что $\boldsymbol{\beta}_{\mathcal{A}^c} = \mathbf{0}$, $\mathcal{A} \sqcup \mathcal{A}^c = \{1, \ldots, n\}$, который доставляет максимум приращению логарифма правдоподобия $\ell$ при условии, что скорость роста функционала $\ell$ по любому признаку из набора не меньше скорости роста по всем остальным признакам.

## Описание алгоритма

**Алгоритм LALR.** В настоящей работе предлагается новый алгоритм выбора признаков при восстановлении логистической регрессии — «Least Angle Logistic Regression (LALR)». Принята модель логистической регрессии (2), (3). Обозначим множество индексов $\mathcal{J} = \{1, \ldots, n\}$. Для некоторого

подмножества индексов $\mathcal{A} \subseteq \mathcal{J}$, назовем его *активным множеством*, определим матрицу *активных признаков*

$$X_{\mathcal{A}} = (\cdots\ s_j \boldsymbol{x}_j\ \cdots)_{j\in\mathcal{A}}, \tag{6}$$

где $s_j$, назовем его *знаком корреляции*, принимает значения $\pm 1$. Определим также матрицы разностей и сумм между активными признаками и некоторым фиксированным неактивным признаком $\boldsymbol{x}_d$, где $d \in \mathcal{A}^c$, в разбиении $\mathcal{A} \sqcup \mathcal{A}^c = \mathcal{J}$,

$$\begin{aligned} M_{d-} &= (\ \cdots\ \ s_j\boldsymbol{x}_j - s_d\boldsymbol{x}_d\ \ \cdots\ )_{j\in\mathcal{A}}, \\ M_{d+} &= (\ \cdots\ \ s_j\boldsymbol{x}_j + s_d\boldsymbol{x}_d\ \ \cdots\ )_{j\in\mathcal{A}}. \end{aligned} \tag{7}$$

Опишем алгоритм последовательного добавления признаков. Начальные значения положим $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\beta} = \mathbf{0}$, $\mathcal{A} = \varnothing$. Рассмотрим некоторый шаг алгоритма. Пусть $\boldsymbol{\mu}_{\mathcal{A}}$ есть текущее приближение функции регрессии на этом шаге. Тогда вектор текущих корреляций между признаками и вектором регрессионных остатков $\boldsymbol{y} - \boldsymbol{\sigma}(\boldsymbol{\mu}_{\mathcal{A}})$ имеет вид:

$$\boldsymbol{c} = X^{\intercal}\left(\boldsymbol{y} - \boldsymbol{\sigma}(\boldsymbol{\mu}_{\mathcal{A}})\right). \tag{8}$$

Примем знак корреляции

$$s_j = \operatorname{sign}(c_j), \qquad j \in \mathcal{J}. \tag{9}$$

Вычисляем матрицы $X_{\mathcal{A}}$, $M_{d-}$ и $M_{d-}$, согласно (6) и (7), для $d \in \mathcal{A}^c$. Обозначим $m \times m$-диагональную матрицу $W$ с элементами

$$W_{ii} = \sigma_i(\boldsymbol{\mu}_{\mathcal{A}})(1 - \sigma_i(\boldsymbol{\mu}_{\mathcal{A}})).$$

Также обозначим матрицы $A_{d-}$, $A_{d+}$ и векторы $\boldsymbol{b}_{d-}$, $\boldsymbol{b}_{d+}$

$$A_{d\pm} = M_{d\pm}^{\intercal} W X_{\mathcal{A}}, \tag{10}$$

$$\boldsymbol{b}_{d\pm} = M_{d\pm}^{\intercal}\left(\boldsymbol{y} - \boldsymbol{\sigma}(\boldsymbol{\mu}_{\mathcal{A}})\right), \tag{11}$$

для всех $d \in \mathcal{A}^c$.

Далее, используя введенные обозначения, вычисляем множество векторов $\Upsilon$,

$$\Upsilon = \{A_{d-}^{-1}\boldsymbol{b}_{d-}, A_{d+}^{-1}\boldsymbol{b}_{d+}\}_{d\in\mathcal{A}^c}, \tag{12}$$

Алгоритм обновляет текущее приближение функции регрессии $\boldsymbol{\mu}_{\mathcal{A}}$ (ниже приводится один вариант), скажем к

$$\boldsymbol{\mu}_{\mathcal{A}_+} = \boldsymbol{\mu}_{\mathcal{A}} + X_{\mathcal{A}}\boldsymbol{\gamma}_{\mathcal{A}}, \tag{13}$$

где оптимальный вектор параметров $\boldsymbol{\gamma}_{\mathcal{A}}$ определяется из условия

$$\boldsymbol{\gamma}_{\mathcal{A}} = \arg\min_{\boldsymbol{\gamma}\in\Upsilon}{}^{+}\left((\boldsymbol{s}_{\mathcal{A}} \circ \boldsymbol{c}_{\mathcal{A}})^{\intercal}\boldsymbol{\gamma}\right), \tag{14}$$

где $\min^+$ означает, что минимум берется только из положительных значений. Операция «$\circ$» — поэлементное (адамарово) умножение векторов.

Найденное решение $\boldsymbol{\gamma}_{\mathcal{A}}$ принадлежит множеству $\Upsilon$, поэтому для некоторого $d^* \in \mathcal{A}^c$ выполнено либо $\boldsymbol{\gamma}_{\mathcal{A}} = A_{d^*-}^{-1}\boldsymbol{b}_{d^*-}$, либо $\boldsymbol{\gamma}_{\mathcal{A}} = A_{d^*+}^{-1}\boldsymbol{b}_{d^*+}$. Так определяется оптимальный индекс $d^*$, соответствующий найденному решению $\boldsymbol{\gamma}_{\mathcal{A}}$. В случае, когда $\mathcal{A} = \varnothing$, что соответствует первому шагу, $d^*$ находится из условия максимума абсолютной корреляции:

$$d^* = \arg\max_{d\in\mathcal{J}} |c_d|.$$

Таким образом, определяется индекс $d^*$, соответствующий оптимальному признаку $\boldsymbol{x}_{d^*}$, и обновляется активное множество индексов, $\mathcal{A}_+ = \mathcal{A} \cup \{d^*\}$. Также обновляется вектор коэффициентов $\boldsymbol{\beta}$, используя (9), $\boldsymbol{\beta}_{\mathcal{A}} = \boldsymbol{\beta}_{\mathcal{A}} + \boldsymbol{s}_{\mathcal{A}} \circ \boldsymbol{\gamma}_{\mathcal{A}}$, нижний индекс $\boldsymbol{\beta}_{\mathcal{A}}$ указывает, что изменяются только компоненты, соответствующие активным признакам. Формула (14) дает приближенное значение вектора коэффициентов $\boldsymbol{\gamma}_{\mathcal{A}}$, поэтому алгоритм можно проитерировать для получения точного значения.

На последнем шаге, когда $\mathcal{A} = \mathcal{J}$, все дополнительные условия на скорость роста функционала $\ell$ выполнены автоматически, оптимальный вектор параметров находится из условия максимизации логарифма правдоподобия $\ell$, с помощью итерационного метода наименьших квадратов с перевзвешиванием элементов (IRLS) [11].

**Обоснование алгоритма.** Пусть имеется некоторое активное множество $\mathcal{A}$ и пусть к тому же известно текущее приближение функции регрессии $\boldsymbol{\mu}_{\mathcal{A}}$, (4). Запишем логарифм правдоподобия (5) через функцию регрессии $\boldsymbol{\mu}_{\mathcal{A}}$:

$$\ell(\boldsymbol{\mu}_{\mathcal{A}}) = \sum_{i=1}^{m}\left(y^i\mu_{\mathcal{A}}(\boldsymbol{x}^i) - \ln(1 + \exp(\mu_{\mathcal{A}}(\boldsymbol{x}^i)))\right). \tag{15}$$

Рассмотрим производную логарифма правдоподобия по некоторому вектору $\boldsymbol{x}_j$, обозначим ее $c_j$:

$$c_j = \frac{d}{d\gamma}\ell(\boldsymbol{\mu}_{\mathcal{A}} + \boldsymbol{x}_j\gamma)\Big|_{\gamma=0}, \tag{16}$$

откуда, пользуясь (3), в матричном виде получим (8):

$$\boldsymbol{c} = X^{\intercal}\left(\boldsymbol{y} - \boldsymbol{\sigma}(\boldsymbol{\mu}_{\mathcal{A}})\right).$$

**Замечание 1.** Вектор $\boldsymbol{c}$ есть вектор текущих корреляций векторов признаков и вектора остатков $\boldsymbol{y} - \boldsymbol{\sigma}(\boldsymbol{\mu}_{\mathcal{A}})$. Поэтому далее под вектором корреляций будем понимать вектор производных по направлению.

Обозначим знак корреляции, как это было сделано в (9). Таким образом, определяем матрицу активных признаков $X_{\mathcal{A}}$, согласно (6). Выразим новое приближение функции регрессии (13) через неизвестные коэффициенты $\boldsymbol{\gamma}$:

$$\boldsymbol{\mu}_{\mathcal{A}_+} = \boldsymbol{\mu}_{\mathcal{A}} + X_{\mathcal{A}}\boldsymbol{\gamma},$$

Основная цель состоит в поиске оптимального вектора коэффициентов $\boldsymbol{\gamma}$ и активного множества индексов $\mathcal{A}_+$.

Перейдем теперь к формальной интерпретации решаемой задачи. Под скоростью роста функционала по направлению понимается абсолютное значение производной функционала по направлению. Поэтому, решаемая задача заключается в максимизации приращения логарифма правдоподобия (15),

$$\ell(\boldsymbol{\mu}_{\mathcal{A}_+}) - \ell(\boldsymbol{\mu}_{\mathcal{A}}) \to \max_{\boldsymbol{\gamma}} \tag{17}$$

при условии, что абсолютная корреляция нового вектора остатков $\boldsymbol{y} - \boldsymbol{\sigma}(\boldsymbol{\mu}_{\mathcal{A}_+})$ на любой активный признак $\boldsymbol{x}_j$, $j \in \mathcal{A}$, не меньше абсолютной корреляции на любой неактивный признак $\boldsymbol{x}_d$, $d \in \mathcal{A}^c$, см. замечание 1. Запишем это условие через производную по направлению (16):

$$\left|\frac{d}{d\alpha}\ell(\boldsymbol{\mu}_{\mathcal{A}_+} + \boldsymbol{x}_j\alpha)\right|_{\alpha=0} \geqslant \left|\frac{d}{d\alpha}\ell(\boldsymbol{x}_{\mathcal{A}_+} + \boldsymbol{x}_d\alpha)\right|_{\alpha=0}, \tag{18}$$

для любых $j \in \mathcal{A}$ и $d \in \mathcal{A}^c$. Пользуясь обозначениями (10), (11) сформулируем лемму о линеаризации решаемой задачи.

**Лемма 1 (о линеаризованном виде).** *Задача (17), (18) при линеаризации эквивалентна задаче линейного программирования:*

$$(\boldsymbol{s}_{\mathcal{A}} \circ \boldsymbol{c}_{\mathcal{A}})^{\intercal}\boldsymbol{\gamma} \to \max_{\boldsymbol{\gamma}}$$
$$\begin{cases} A_{d-}\boldsymbol{\gamma} \leqslant \boldsymbol{b}_{d-}, \\ A_{d+}\boldsymbol{\gamma} \leqslant \boldsymbol{b}_{d+}, \\ \forall d \in \mathcal{A}^c. \end{cases} \tag{19}$$

Приведенная ниже теорема 4 позволяет существенно сократить количество опорных точек [12, 13], которые могут являться решением задачи (19). Для доказательства теоремы 4 сформулируем некоторые вспомогательные утверждения.

**Лемма 2.** *Пусть векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \boldsymbol{a}_{k+1} \in \mathbb{R}^n$ линейно независимы. Тогда векторы $(\boldsymbol{a}_1 + \boldsymbol{a}_{k+1}), \ldots, (\boldsymbol{a}_k + \boldsymbol{a}_{k+1})$, $\boldsymbol{a}_{k+1}$ также линейно независимы.*

**Определение 1.** *Точки $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k \in \mathbb{R}^n$ называются аффинно зависимыми, если существуют $\lambda_1, \ldots, \lambda_k$, одновременно ненулевые и такие, что*

$$\sum_{i=1}^{k} \lambda_i \boldsymbol{a}_i = 0, \qquad \sum_{i=1}^{k} \lambda_i = 0.$$

**Лемма 3.** *Пусть векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \boldsymbol{a}_{k+1} \in \mathbb{R}^n$ линейно независимы. Обозначим матрицы*

$$A = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k), \; A_+ = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \boldsymbol{a}_{k+1}),$$
$$C = (\boldsymbol{a}_1 - \boldsymbol{a}_{k+1}, \ldots, \boldsymbol{a}_k - \boldsymbol{a}_{k+1}).$$

*Матрица $A^{\intercal}C$ имеет полный ранг тогда и только тогда, когда столбцы матрицы $A^{\intercal}A_+$ аффинно независимы.*

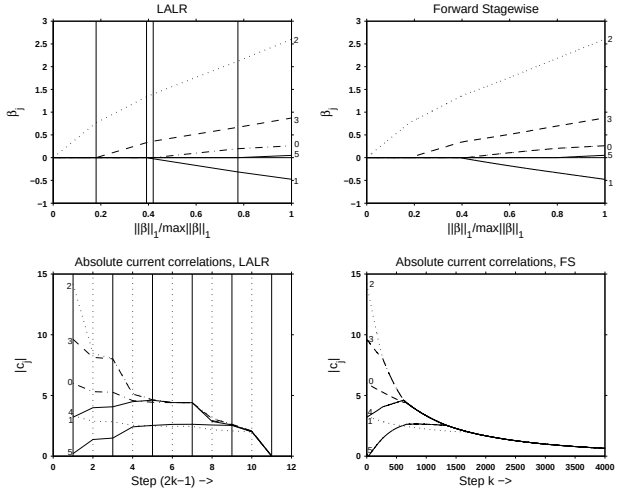


**Рис. 1.** Сравнение оценок коэффициентов для LALR и Forward Stagewise для модельных данных. Номера кривых соответствуют номерам признаков. Сплошные вертикальные линии обозначают шаги, а штриховые вертикальные — дополнительную итерацию для каждого шага.

Лемма 3 используется при доказательстве существования множества $\Upsilon$, (12). С помощью следующей теоремы формулируется утверждение о решении задачи линейного программирования (19).

**Теорема 4.** *Если ЗЛП (19) имеет решение $\boldsymbol{\gamma}^*$, то $\boldsymbol{\gamma}^* \in \Upsilon$, причем*

$$\boldsymbol{\gamma}^* = \arg\min_{\boldsymbol{\gamma} \in \Upsilon}{}^{+}\{(\boldsymbol{s}_{\mathcal{A}} \circ \boldsymbol{c}_{\mathcal{A}})^{\intercal}\boldsymbol{\gamma}\};$$

*где $\min^+$ означает, что минимум берется только из положительных значений.*

**Следствие 1.** *На каждом шаге алгоритма абсолютная корреляция текущего вектора остатков на любой активный признак при линеаризации одинакова и больше абсолютной корреляции на любой неактивный признак, т. е. справедливо*

$$\begin{cases} s_i c_i = s_j c_j, & i, j \in \mathcal{A}, \\ s_i c_i > s_d c_d, & i \in \mathcal{A}, d \in \mathcal{A}^c. \end{cases}$$

Следствие 1 представляет собой аналог основного свойства метода наименьших углов в линейных моделях: на каждом шаге вектор остатков лежит на биссекторе для добавленных признаков.

## Вычислительные эксперименты

**Модельные данные.** Сгенерируем $m = 50$ объектов с пятью независимыми, нормально распределенными признаками $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_5$, т. е. $\boldsymbol{x}^i = (x_1^i, \ldots, x_5^i) \sim \mathcal{N}_5(\mathbf{0}, \mathbf{I})$. Примем модель

$$\boldsymbol{y} = \frac{1}{1 + \exp(-(\beta_0 + \boldsymbol{x}_1\beta_1 + \boldsymbol{x}_2\beta_2 + \boldsymbol{x}_3\beta_3))} + \varepsilon.$$

**Таблица 1.** Результаты работы LALR.

| № | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0,1969 | 0,2606 | 9,2868 |
| 1 | 0 | 0 | −0,0250 | −0,3142 | −0,4733 | −21,4689 |
| 2 | 0,7769 | 1,3359 | 1,4005 | 2,1215 | 2,5999 | 91,6048 |
| 3 | 0 | 0,3313 | 0,3615 | 0,6677 | 0,8713 | 36,5161 |
| 4 | 0 | 0 | 0 | 0 | 0 | −8,2624 |
| 5 | 0 | 0 | 0 | 0 | 0,0513 | 1,6560 |

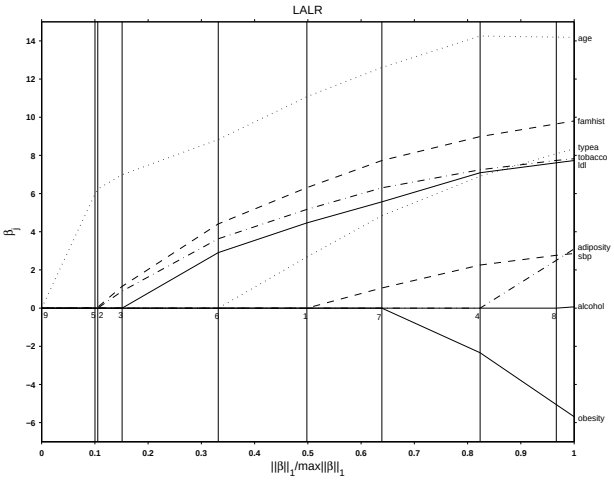


**Рис. 2.** Оценка коэффициентов алгоритма LALR для данных «South African Heart Disease». Вертикальные линии соответствуют шагам алгоритма. Данные стандартизованы, согласно (1).

В качестве параметров $\boldsymbol{\beta}$ возьмем, например, вектор $(\beta_0, \beta_1, \beta_2, \beta_3)^\intercal = (1, -2, 6, 3)^\intercal$. В нашей модели признаки $\boldsymbol{x}_4$ и $\boldsymbol{x}_5$ являются шумовыми. Результатом работы алгоритма является последовательность весов признаков, выбираемых на каждом шаге. В данном случае алгоритм сделает шесть шагов. В таблице 1 представлены результаты работы алгоритма. Первый столбец — номера признаков, первая строка — номер шага, а соответствующая ячейка таблицы — вес признака. Признаку с номером 0 соответствует константный признак. На рис. 1 показано сравнение оценок коэффициетов, полученных с помощью LALR и Forward Stagewise. По полученным результатам можно сделать вывод, что последовательность выбираемых признаков и их весов согласуется с исходной моделью.

**Данные «SAHD».** Проанализирована работа алгоритма на реальных данных «South African Heart Disease», см. [2]. Данные были впервые рассмотрены в [14]. Данные SAHD представляют собой сведения о физическом состоянии 462-х пациентов мужского пола белой расы возраста от 15 до 64 лет. Описание данных состоит из 9 признаков: $\boldsymbol{x}_1$ — sbp (sistolic blood pressure), $\boldsymbol{x}_2$ — tobacco, $\boldsymbol{x}_3$ — ldl (low-density lipoprotein), $\boldsymbol{x}_4$ — adiposity, $\boldsymbol{x}_5$ — famhist (family history), $\boldsymbol{x}_6$ — typea, $\boldsymbol{x}_7$ — obesity, $\boldsymbol{x}_8$ — alcohol, $\boldsymbol{x}_9$ — age; а также вектора меток класса chd: наличие «1», или отсутствие «0» инфаркта миокарда (MI) за время обследования. На рис. 2 представлены результаты работы алгоритма.

## Выводы

В данной работе предложен и исследован новый алгоритм LALR, решающий задачу отбора признаков в модели логистической регрессии, представляющий собой линеаризованный аналог алгоритма LARS. Проведена серия численных экспериментов на модельных и реальных данных «SAHD», результаты которых позволяют говорить о справедливости использования предложенного алгоритма.

## Литература

[1] *Draper N., Smith H.* APpied Regression Analysis. New York: Wiley, 1966. — 407 p.

[2] *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning. — New York: Springer, 2001.

[3] *Hastie T., Tibshirani R., Taylor J., Walther G.* Forward Stagewise Regression and the Monotone Lasso // Electronic Journal of Statistics, 2006. — Vol. 1.

[4] *Hocking R. R.* The Analysis and Selection of Variables in Linear Regression // Biometrics, 1976. — Vol. 32, № 1. — Pp. 1–49.

[5] *Hoerl A. E., Kennard R. W.* Ridge regression: biased estimation for nonorthogonal problems // Technometrics, 1970. — Vol. 12, № 1. — Pp. 55–67.

[6] *Ильин В. А.* О работах А. Н. Тихонова по методам решения некорректно поставленных задач // УМН, 1967. — Т. 22, № 2. — С. 168–175.

[7] *Tibshirani R.* Regression shrinkage and selection via the lasso. // Journal of the Royal Statistical Society, 1996. — Vol. 58, № 1. — Pp. 267–288.

[8] *Efron B., Hastie T., Johnstone I., Tibshirani R.* Least angle regression. // Annals of Statistics, 2004. — Vol. 32, № 2. — Pp. 407–499.

[9] *Lawson L., Hanson R.* Solving Least Squares Problems. — Englewood Cliffs: Prentice Hall, 1974.

[10] *Madigan D., Ridgeway G.* Discussion of least angle regression. // Annals of Statistics, 2004. — Vol. 32, № 2. — Pp. 465–469.

[11] *Rubin D. B.* Iteratively reweighted least squares. // Encyclopedia of statistical sciences, 1983. — Vol. 32. — Pp. 272–275.

[12] *Сухарев А. Г., Тимохов А. В.* Курс методов оптимизации. — Москва: ФИЗМАТЛИТ, 2005.

[13] *Измаилов А. Ф.* Численные методы оптимизации. — Москва: ФИЗМАТЛИТ, 2005.

[14] *Rousseauw J., du Plessis J., Ferreira J.* Coronary Risk Factor Screening in Three Rural Communities. // South African Medical Journal, 1983. — Vol. 64. — Pp. 430–436.