

А.Н. Тырсин, К.К. Костин

ОЦЕНИВАНИЕ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ КАК ЭКСТРЕМАЛЬНАЯ ЗАДАЧА

Работа выполнена при финансовой поддержке РФФИ № 16-06-00048.

Рассматриваются особенности оценивания логистической регрессии как экстремальной задачи, реализующей критерий максимального правдоподобия. Показано, что эта задача имеет существенные отличия по сравнению с поиском локального экстремума. Описан устойчивый алгоритм вычисления коэффициентов разделяющего уравнения логистической регрессии. Приведены результаты статистического моделирования сравнительного анализа оценок логистической регрессии предложенного алгоритма с известными алгоритмами.

Ключевые слова: логистическая регрессия; метод максимального правдоподобия; алгоритм спуска; случайный поиск; модель; целевая функция.

Логистическая регрессия является одним из распространенных методов классификации данных в разных областях [1. С. 181–242; 2. С. 283–369; 3. С. 152–214; 4. С. 321–329].

Основная цель логистической регрессии состоит в разделении множества исходных значений линейной границей (разделяющей прямой, плоскостью или гиперплоскостью) на две соответствующих двум заданным классам области. Логистическая регрессия прогнозирует вероятность некоторого события, находящуюся в пределах от 0 до 1 [4. С. 322–324]. Следует указать, что число классов L может быть больше двух, в этом случае они имеют мультиномиальную регрессию, которую можно построить с помощью $L - 1$ независимых логистических регрессий. Поэтому будем рассматривать задачу классификации с двумя классами.

В настоящее время построение логистической регрессии обычно осуществляют в виде оптимизационной задачи, реализующей критерий максимального правдоподобия [5. С. 124]. Ряд авторов [6. С. 31–32; 7] предлагают оценивать коэффициенты разделяющего уравнения методом Ньютона–Рафсона. Однако исследования на модельных данных этого и других алгоритмов оценивания логистической регрессии, реализованных в статистических пакетах STATISTICA и SPSS, показали неустойчивость оценок. Этот недостаток отмечается также в [7].

Цель статьи – исследование особенностей процедуры оценивания логистической регрессии как экстремальной задачи, а также разработка устойчивого алгоритма вычисления коэффициентов разделяющего уравнения.

1. Построение логистической регрессии по критерию максимального правдоподобия

Пусть имеется выборка прецедентов (обучающая выборка):

$$(\mathbf{x}_i, y_i), i = 1, 2, \dots, n, \quad (1)$$

где $\mathbf{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \dots \\ x_{im} \end{pmatrix} = \begin{pmatrix} 1 \\ x_{i2} \\ \dots \\ x_{im} \end{pmatrix}$ – вектор значений i -го объекта, $\forall i x_{i1} = 1$; $\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \dots \\ \mathbf{x}_n^T \end{pmatrix} = \begin{pmatrix} 1 & x_{12} & \dots & x_{1m} \\ 1 & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n2} & \dots & x_{nm} \end{pmatrix}$;

$y_i \in \{-1; 1\}$ – бинарная переменная, указывающая на принадлежность i -го объекта соответствующему классу, например, для определенности первому классу при $y_i = -1$ и второму – при $y_i = 1$; $M = m - 1$ – число признаков у каждого объекта; n – количество наблюдений.

Классификацию осуществляют с помощью функции $h(\mathbf{x}) = \frac{1}{1 + \exp\{-\mathbf{b}^T \mathbf{x}\}}$, принимающей значения в интервале (0; 1). Пороговым значением является $h(\mathbf{x}) = 0,5$. В [6. С. 29–32] показано, что максимизация логарифма правдоподобия эквивалентна минимизации функционала

$$Q(\mathbf{b}) = \sum_{i=1}^n \ln(1 + e^{-y_i \mathbf{b}^T \mathbf{x}_i}) \rightarrow \min_{\mathbf{b} \in \mathbb{R}^m}, \quad (2)$$

где $\mathbf{b}^T = (b_1 \ b_2 \ \dots \ b_m)$ – искомый вектор коэффициентов разделяющей линейной границы, описываемой в общем случае уравнением гиперплоскости

$$b_1 + \sum_{j=2}^m b_j x_j = 0.$$

Аналитически задача минимизации (2) не решается. Поэтому для оценивания вектора коэффициентов \mathbf{b} используют итерационные алгоритмы спуска для решения экстремальных задач. В [6. С. 32–34] описан итерационный алгоритм Ньютона–Рафсона решения задачи (2). Он состоит в следующем.

В качестве нулевого приближения можно взять решение задачи классификации методом многомерной линейной регрессии

$$\mathbf{b}^{(0)} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Y}). \quad (3)$$

Затем начинается итерационный процесс, на k -м шаге которого уточняется вектор коэффициентов $\mathbf{b}^{(k)}$:

$$\mathbf{b}^{(k)} = \mathbf{b}^{(k-1)} - h_k (Q''(\mathbf{b}^{(k-1)}))^{-1} Q'(\mathbf{b}^{(k-1)}),$$

где $Q'(\mathbf{b}^{(k)})$ – вектор первых производных (градиент) функционала $Q(\mathbf{b})$ в точке $\mathbf{b}^{(k)}$; $Q''(\mathbf{b}^{(k)})$ – матрица вторых производных (гессиан) функционала $Q(\mathbf{b})$ в точке $\mathbf{b}^{(k)}$; h_k – величина шага, который можно положить равным 1, но его подбор на каждом шаге способен увеличить скорость сходимости. Метод Ньютона–Рафсона описан в ряде учебников по методам оптимизации, например в [8. С. 223–226].

Статистическое моделирование методом Монте–Карло [9] показало неустойчивость в некоторых случаях работы этого алгоритма, а также других алгоритмов, реализованных в статистических пакетах. Отметим основные причины этого.

В итерационных алгоритмах поиска минимума требуется задать начальное приближение вектора $\mathbf{b}^{(0)}$. Плохое начальное приближение, значительно отличающееся от точки локального минимума, может значительно ухудшить условия сходимости итерационного алгоритма. В результате статистического моделирования выяснилось, что использование в качестве начального приближения $\mathbf{b}^{(0)}$ решения задачи классификации в виде линейной регрессии часто дает неудовлетворительный результат. Покажем это на примере.

Пример 1. Сгенерируем выборку прецедентов (\mathbf{x}_i, y_i) , $\mathbf{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \end{pmatrix} = \begin{pmatrix} 1 \\ x_{i2} \\ x_{i3} \end{pmatrix}$, $i = 1, 2, \dots, n$, следующим образом.

Объем выборки $n = 100$, число признаков $M = 2$, первая половина выборки ($i = 1, 2, \dots, 50$) относится к первому классу ($y_i = -1$), а вторая – ко второму классу ($y_i = 1$). Пусть в каждом классе признаки являются взаимно независимыми гауссовскими случайными величинами. Это означает, что первый и второй классы представляют собой случайные выборки из двумерных гауссовых случайных векторов $\mathbf{X}^{(1)} = (X_2^{(1)}, X_3^{(1)})$ и $\mathbf{X}^{(2)} = (X_2^{(2)}, X_3^{(2)})$ соответственно. Зададим у всех компонент одно и то же среднее квадратическое отклонение $\sigma_x = 1,5$ и математические ожидания: $M[X_2^{(1)}] = 10$, $M[X_3^{(1)}] = 16$, $M[X_2^{(2)}] = 4$, $M[X_3^{(2)}] = 4$. На рисунке 1 приведен результат классификации с помощью метода наименьших квадратов на основе (3). Разделяющая прямая не только не классифицирует выборку прецедентов на два класса, но и значительно удалена от всех объектов. Хотя классы в этом примере легкоразделимы. Такое неудачное начальное приближение значительно затрудняет работу итерационных алгоритмов решения задачи (2).

Вторая причина неустойчивости решений задачи (2) состоит в том, что итерационные алгоритмы спуска рассчитаны на поиск локального минимума. А задача (2) при корректной классификации всех прецедентов имеет нижнюю грань на бесконечности, равную нулю. Причем качество построения разделяющей границы при формальном спуске уже не будет связано со значением целевой функции $Q(\mathbf{b})$. Поэтому требуется корректировка стратегии спуска.

Третья причина состоит в том, что при достаточно малых значениях целевой функции $Q(\mathbf{b})$ некоторые величины $-y_i \mathbf{b}^T \mathbf{x}_i$ могут принимать очень большие значения, что приводит к неустранимым вычислительным погрешностям и даже к останову работы алгоритма из-за переполнения памяти.

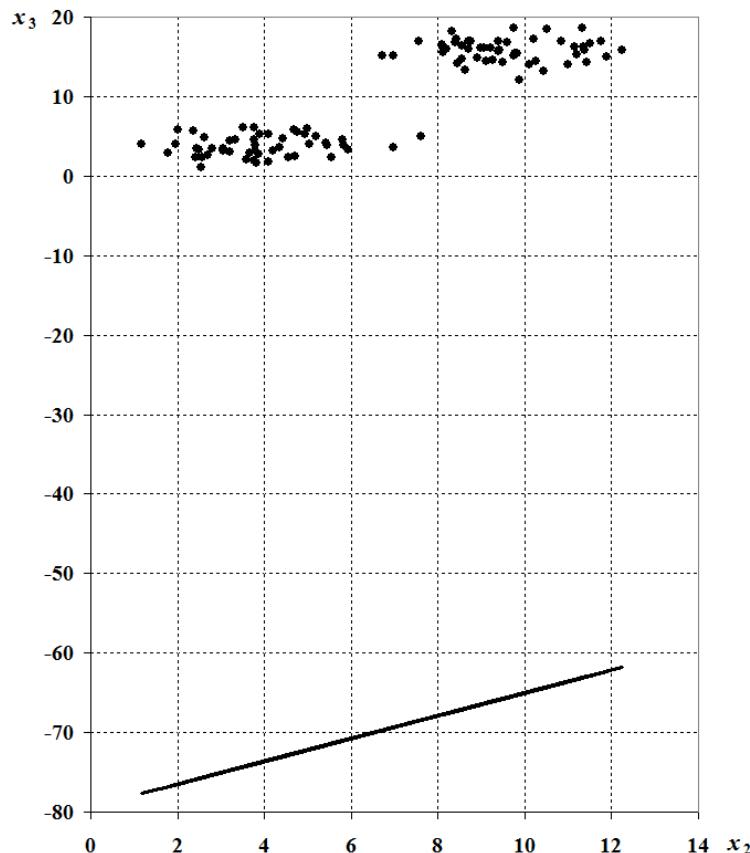


Рис. 1. Пример построения разделяющей границы с помощью линейной регрессии

Таким образом, приведенные недостатки непосредственного использования итерационных алгоритмов спуска требуют исследования особенностей задачи (2).

2. Исследование целевой функции задачи оценивания коэффициентов разделяющей границы по критерию максимального правдоподобия

Исследуем свойства целевой функции $Q(\mathbf{b})$. Рассмотрим вначале произвольное i -е слагаемое целевой функции $Q(\mathbf{b})$, т.е. функцию

$$z_i(\mathbf{b}) = \ln(1 + e^{-y_i \mathbf{b}^T \mathbf{x}_i}) = \ln\left[1 + \exp\left\{-y_i \sum_{k=1}^m b_k x_{ik}\right\}\right].$$

Найдем ее частные производные:

$$\frac{\partial z_i}{\partial b_k} = -\frac{y_i x_{ik} e^{-y_i \mathbf{b}^T \mathbf{x}_i}}{1 + e^{-y_i \mathbf{b}^T \mathbf{x}_i}}, \quad k = 1, 2, \dots, m.$$

Так как все частные производные непрерывно дифференцируемы и $\frac{\partial z_i}{\partial b_1} \neq 0$, то функция $z_i(\mathbf{b})$ не имеет стационарных точек, а значит, у нее нет экстремумов.

Далее найдем вторые производные $\frac{\partial^2 z_i}{\partial b_k \partial b_j}$. Они равны

$$\frac{\partial^2 z_i}{\partial b_k \partial b_j} = y_i^2 x_{ik} x_{ij} U_i(\mathbf{b}), \quad k, j = 1, 2, \dots, m,$$

где $U_i(\mathbf{b}) = \frac{e^{-y_i \mathbf{b}^T \mathbf{x}_i}}{(1 + e^{-y_i \mathbf{b}^T \mathbf{x}_i})^2} > 0$.

Поскольку все вторые производные всюду непрерывны, то $\frac{\partial^2 z_i}{\partial b_k \partial b_j} = \frac{\partial^2 z_i}{\partial b_j \partial b_k}$. Исследуем функцию $z_i(\mathbf{b})$ на выпуклость. Ее гессиан равен

$$H(z_i) = (q_{kj}(\mathbf{b})), \quad q_{kj}(\mathbf{b}) = \frac{\partial^2 z_i}{\partial b_k \partial b_j} = y_i^2 x_{ik} x_{ij} U_i(\mathbf{b}), \quad k, j = 1, 2, \dots, m.$$

Воспользуемся критерием Сильвестра [10. С. 507]. Определим главные миноры A_1, A_2, \dots, A_m гессиана $H(z_i)$. Поскольку $x_{i1} = 1, y_i^2 = 1$, то $A_1 = y_i^2 x_{i1}^2 U_i(\mathbf{b}) = U_i(\mathbf{b}) > 0$. Можно показать, что все остальные главные миноры

$$A_k = (U_i(\mathbf{b}))^k \cdot \begin{vmatrix} 1 & x_{i1}x_{i2} & \dots & x_{i1}x_{ik} \\ x_{i2}x_{i1} & x_{i2}^2 & \dots & x_{i2}x_{ik} \\ \dots & \dots & \dots & \dots \\ x_{ik}x_{i1} & x_{ik}x_{i2} & \dots & x_{ik}^2 \end{vmatrix} = 0, \quad k = 2, 3, \dots, m.$$

Получили, что квадратичная форма относительно переменных h_1, h_2, \dots, h_m

$$\Phi(h_1, h_2, \dots, h_m) = \sum_{k=1}^m \sum_{j=1}^m q_{kj}(\mathbf{b}) h_k h_j$$

является неотрицательно определенной. Это означает выпуклость функции $z_i(\mathbf{b})$ на \mathbf{R}^m .

Перейдем теперь к рассмотрению функции $Q(\mathbf{b})$. Она является суммой n собственных выпуклых функций $z_i(\mathbf{b})$, следовательно, $Q(\mathbf{b})$ также выпукла [11. С. 49].

Найдем частные производные функции $Q(\mathbf{b})$:

$$\frac{\partial Q}{\partial b_k} = - \sum_{i=1}^n \frac{y_i x_{ik} e^{-y_i \mathbf{b}^T \mathbf{x}_i}}{1 + e^{-y_i \mathbf{b}^T \mathbf{x}_i}}, \quad k = 1, 2, \dots, m.$$

Все частные производные всюду на \mathbf{R}^m непрерывно дифференцируемы, поэтому экстремумы функции $Q(\mathbf{b})$ могут быть только в стационарных точках, в которых градиент равен нулю, т.е.

$$\sum_{i=1}^n \frac{y_i x_{ik} e^{-y_i \mathbf{b}^T \mathbf{x}_i}}{1 + e^{-y_i \mathbf{b}^T \mathbf{x}_i}} = 0, \quad k = 1, 2, \dots, m. \quad (4)$$

Система нелинейных уравнений (4) имеет достаточно сложный вид, установить ее несовместность или найти решения аналитически не представляется возможным. Однако удалось подобрать примеры, когда система (4) имеет решения. Это означает, что целевая функция $Q(\mathbf{b})$ может иметь стационарные точки, а значит, ввиду ее выпуклости может иметь локальные минимумы, не являющиеся глобальным минимумом.

Найдем вторые производные $\frac{\partial^2 Q}{\partial b_k \partial b_j}$. С учетом того, что $\forall i \quad y_i^2 = 1$, они равны

$$\frac{\partial^2 z_i}{\partial b_k \partial b_j} = \sum_{i=1}^n x_{ik} x_{ij} U_i(\mathbf{b}), \quad k, j = 1, 2, \dots, m.$$

Так как все вторые производные всюду непрерывны, то $\frac{\partial^2 Q}{\partial b_k \partial b_j} = \frac{\partial^2 Q}{\partial b_j \partial b_k}$. Гессиан функции $Q(\mathbf{b})$ равен

$$H(Q) = (Q_{kj}(\mathbf{b})), \quad Q_{kj}(\mathbf{b}) = \frac{\partial^2 z_i}{\partial b_k \partial b_j} = \sum_{i=1}^n x_{ik} x_{ij} U_i(\mathbf{b}), \quad k, j = 1, 2, \dots, m.$$

Опять воспользуемся критерием Сильвестра и определим главные миноры A_1, A_2, \dots, A_m гессиана $H(Q)$. Поскольку $x_{i1} = 1$, то $A_1 = \sum_{i=1}^n x_{i1}^2 U_i(\mathbf{b}) = \sum_{i=1}^n U_i(\mathbf{b}) > 0$ как сумма строго положительных чисел.

Остальные главные миноры с учетом выпуклости $Q(\mathbf{b})$ равны

$$A_k = \begin{vmatrix} \sum_{i=1}^n U_i(\mathbf{b}) & \sum_{i=1}^n x_{i2} U_i(\mathbf{b}) & \dots & \sum_{i=1}^n x_{ik} U_i(\mathbf{b}) \\ \sum_{i=1}^n x_{i2} U_i(\mathbf{b}) & \sum_{i=1}^n x_{i2}^2 U_i(\mathbf{b}) & \dots & \sum_{i=1}^n x_{i2} x_{ik} U_i(\mathbf{b}) \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^n x_{ik} U_i(\mathbf{b}) & \sum_{i=1}^n x_{ik} x_{i2} U_i(\mathbf{b}) & \dots & \sum_{i=1}^n x_{ik}^2 U_i(\mathbf{b}) \end{vmatrix} \geq 0, \quad k = 2, 3, \dots, m.$$

Рассмотрим частные случаи.

Пусть $m = 2, n = 2$. После преобразований получим $A_2 = (x_{12} - x_{22})^2 U_1(\mathbf{b}) U_2(\mathbf{b}) \geq 0$.

Пусть $m = 2, n = 3$. После преобразований получим $A_2 = \sum_{i=1}^{n-1} \sum_{j=i+1}^n (x_{i2} - x_{j2})^2 U_i(\mathbf{b}) U_j(\mathbf{b}) \geq 0$.

Таким образом, квадратичная форма в случае нескольких наблюдений может быть положительно определенной или неотрицательно определенной. Это означает, что функция $Q(\mathbf{b})$:

- является выпуклой на \mathbf{R}^m и не имеет максимумов;
- может иметь стационарные точки, в которых могут быть локальные минимумы.

Поскольку равенство нулю главных миноров будет только в некоторых точках, например в вырожденных случаях равенства всех измерений, то функция $Q(\mathbf{b})$ может оказаться в окрестности стационарной точки строго выпуклой. Это означает, что в такой стационарной точке у функции $Q(\mathbf{b})$ будет локальный минимум. Из этого можно сделать выводы.

Использование численных методов спуска первого или второго порядков при наличии у целевой функции стационарных точек может привести к неправильному построению логистической регрессии по двум причинам. Во-первых, можно в качестве решения получить локальный минимум, который будет грубой оценкой коэффициентов уравнения логистической регрессии. Во-вторых, спуск может стремиться к нижней грани целевой функции, а вектор коэффициентов логистической регрессии будет оставаться грубым решением.

3. Устойчивый алгоритм вычисления коэффициентов разделяющего уравнения логистической регрессии

Имеем выборку прецедентов (1), в которой первые N_1 объектов относятся к первому классу ($y_i = -1$), а остальные – ко второму классу ($y_i = 1$).

С учетом вышеизложенного приведем описание алгоритма. Он состоит в следующем.

1. Найдем середину $\tilde{\mathbf{x}}^{(0)}$ отрезка, соединяющего центры тяжести первого и второго классов $\tilde{\mathbf{x}}^{(1)} = \frac{1}{N_1} \sum_{i=1}^{N_1} \mathbf{x}_i, \quad \tilde{\mathbf{x}}^{(2)} = \frac{1}{n - N_1} \sum_{i=N_1+1}^n \mathbf{x}_i$.

2. Осуществляем параллельный сдвиг начала координат в точку $\tilde{\mathbf{x}}^{(0)}$.

3. Через точку $\tilde{\mathbf{x}}^{(0)}$ строим гиперплоскость $S: s_1 + s_2 x_2 + \dots + s_m x_m = 0$, ортогональную вектору $\mathbf{d} = \tilde{\mathbf{x}}^{(2)} - \tilde{\mathbf{x}}^{(1)}$.

4. Получаем начальную оценку $\mathbf{b}^{(0)}$ вектора разделяющей гиперплоскости как $\mathbf{b}^{(0)} = \mathbf{s} / |\mathbf{s}|$, где \mathbf{s} – вектор коэффициентов гиперплоскости S . При этом длина $|\mathbf{b}^{(0)}| = 1$. Зададим параметр $\beta_0 = 1$.

5. С помощью итерационного алгоритма ищем с заданной точностью ε решение задачи (2). На любом k -м шаге алгоритма, многократно генерируя случайный вектор $\mathbf{p}^{(l)}$, решаем задачу

$$\begin{cases} Q(\mathbf{b}^{(k,l)}) \rightarrow \min_{\mathbf{b}^{(k)} \in \mathbb{R}^m}, \\ \mathbf{b}^{(k,l)} = \alpha(\mathbf{b}^{(k,l-1)} + h_k \mathbf{p}^{(l)}), \\ |\mathbf{b}^{(k,l)}| = \beta_{k-1}, \mathbf{b}^{(k,1)} = \mathbf{b}^{(k-1)}. \end{cases}$$

Если требуемая точность решения будет достигнута ($Q(\mathbf{b}^{(k,l)}) < \varepsilon$), то в качестве решения задачи используем последнюю оценку $\mathbf{b}^{(k,l)}$. В противном случае фиксируем последнюю оценку $\mathbf{b}^{(k)} = \mathbf{b}^{(k,l)}$, увеличиваем длину векторов оценок, т.е. $\beta_k = \gamma \cdot \beta_{k-1}$, где $\gamma > 1$ – коэффициент растяжения и возвращающийся к п. 5 ($k := k + 1$).

Следует отметить, что слишком малое значение параметра точности ε может привести к большим длинам оценок векторов $\mathbf{b}^{(k,l)}$ и, как указано выше, к потере точности алгоритма.

4. Вычислительные эксперименты

Проведем с помощью метода статистических испытаний Монте-Карло несколько экспериментов для исследования эффективности предложенного алгоритма по сравнению с алгоритмом «Quasi-Newton», реализованным в статистическом пакете STATISTICA. Он показал наилучшие результаты оценивания среди алгоритмов логистической регрессии в STATISTICA.

Число опытов $L = 200$. Выборка прецедентов имеет тот же вид, что и в примере 1, различия заключаются в объемах выборок первого N_1 и второго N_2 классов ($N_1 + N_2 = n$) и математических ожиданиях $M[X_2^{(1)}]$, $M[X_3^{(1)}]$, $M[X_2^{(2)}]$, $M[X_3^{(2)}]$. Поскольку с точностью до постоянного множителя векторов коэффициентов \mathbf{b} может быть бесконечное множество, то для удобства сравнения результаты оценивания будем задавать в нормированном виде с единичной длиной.

Пример 2. Рассмотрим случай одинаковых объемов выборок в классах. Зададим $N_1 = N_2 = 20$. Результаты оценивания коэффициентов логистической регрессии по предложенному алгоритму и алгоритму «Quasi-Newton» в пакете STATISTICA приведены в табл. 1.

Т а б л и ц а 1

Результаты оценивания логистической регрессии по предложенному алгоритму и алгоритму «Quasi-Newton» в пакете STATISTICA для случая $N_1 = N_2 = 20$

Матожидания признаков	Параметр	Предложенный алгоритм			Алгоритм «Quasi-Newton»		
		b_1	b_2	b_3	b_1	b_2	b_3
$M[X_2^{(1)}] = 6$,	β_j	0,00000	-0,89443	-0,44721	0,00000	-0,89443	-0,44721
	\bar{b}_j	-0,00925	-0,88467	-0,46612	-0,13761	-0,78647	-0,28092
	$s(b_j)$	0,26070	0,10410	0,13367	0,47889	0,21434	0,29080
	Δb_j	-0,00925	0,00975	-0,01891	-0,13761	0,10795	0,16629
$M[X_2^{(1)}] = 16$,	β_j	0,99723	-0,06648	-0,03324	0,99723	-0,06648	-0,03324
	\bar{b}_j	0,99721	-0,06691	-0,03297	0,99701	-0,04299	-0,05718
	$s(b_j)$	0,00047	0,00958	0,00923	0,00063	0,03115	0,03347
	Δb_j	-0,00002	-0,00042	0,00027	-0,00023	0,02350	-0,02394

Окончание табл. 1

Матожидания признаков	Параметр	Предложенный алгоритм			Алгоритм «Quasi-Newton»		
		b_1	b_2	b_3	b_1	b_2	b_3
$M[X_2^{(1)}] = 26$, $M[X_3^{(1)}] = 23$, $M[X_2^{(2)}] = 14$, $M[X_3^{(2)}] = 17$	β_j	0,99931	-0,03331	-0,01666	0,99931	-0,03331	-0,01666
	\bar{b}_j	0,99930	-0,03345	-0,01648	0,99917	-0,03187	-0,01801
	$s(b_j)$	0,00009	0,00428	0,00429	0,00060	0,01393	0,01250
	Δb_j	0,00000	-0,00014	0,00018	-0,00014	0,00144	-0,00135

Пример 3. Рассмотрим случай разных объемов выборок в классах. Зададим $N_1 = 20$, $N_2 = 40$. Результаты оценивания коэффициентов логистической регрессии по предложенному алгоритму и алгоритму «Quasi-Newton» в пакете STATISTICA приведены в табл. 2.

Таблица 2

Результаты оценивания логистической регрессии по предложенному алгоритму и алгоритму «Quasi-Newton» в пакете STATISTICA для случая $N_1 = 20$, $N_2 = 40$

Матожидания признаков	Параметр	Предложенный алгоритм			Алгоритм «Quasi-Newton»		
		b_1	b_2	b_3	b_1	b_2	b_3
$M[X_2^{(1)}] = 6$, $M[X_3^{(1)}] = 3$, $M[X_2^{(2)}] = -6$, $M[X_3^{(2)}] = -3$	β_j	0,00000	-0,89443	-0,44721	0,00000	-0,89443	-0,44721
	\bar{b}_j	0,09740	-0,89644	-0,43234	-0,10119	-0,73443	-0,18157
	$s(b_j)$	0,25587	0,08658	0,14076	0,55009	0,27014	0,41928
	Δb_j	0,09740	-0,00201	0,01488	-0,10119	0,15999	0,26565
$M[X_2^{(1)}] = 16$, $M[X_3^{(1)}] = 13$, $M[X_2^{(2)}] = 4$, $M[X_3^{(2)}] = 7$	β_j	0,99723	-0,06648	-0,03324	0,99723	-0,06648	-0,03324
	\bar{b}_j	0,99724	-0,06699	-0,03194	0,99702	-0,03436	-0,06328
	$s(b_j)$	0,00035	0,00852	0,00881	0,00068	0,03844	0,03698
	Δb_j	0,00001	-0,00051	0,00130	-0,00022	0,03213	-0,03004
$M[X_2^{(1)}] = 26$, $M[X_3^{(1)}] = 23$, $M[X_2^{(2)}] = 14$, $M[X_3^{(2)}] = 17$	β_j	0,99931	-0,03331	-0,01666	0,99931	-0,03331	-0,01666
	\bar{b}_j	0,99931	-0,03303	-0,01676	0,99894	-0,02170	-0,02813
	$s(b_j)$	0,00010	0,00499	0,00493	0,00090	0,02441	0,02436
	Δb_j	0,00001	0,00028	-0,00011	-0,00037	0,01161	-0,01148

В таблицах 1 и 2 приняты следующие обозначения: β_j – теоретические значения коэффициентов логистической регрессии; \bar{b}_j – средние значения оценок коэффициентов логистической регрессии; $s(b_j)$ – средние квадратические значения ошибок оценок коэффициентов логистической регрессии; Δb_j – средние значения ошибок оценок коэффициентов логистической регрессии.

Анализ таблиц показывает, что предложенный алгоритм значительно выигрывает по сравнению с известным.

Заключение

Исследования показали вычислительную неустойчивость алгоритмов оценивания логистической регрессии, реализующих критерий максимального правдоподобия. С целью повышения точности оценок предложен новый итерационный алгоритм. Он включает последовательность задач минимизации нулевого порядка на основе случайного поиска. На каждом k -м шаге длина вектора коэффициентов $\mathbf{b}^{(k)}$ является фиксированной. Ее постепенно увеличивают до достижения целевой функцией требуемого значения. Результаты статистического моделирования показали более высокую точность оценок логистической регрессии предложенного алгоритма по сравнению с известными.

ЛИТЕРАТУРА

1. Azen R., Walker C.M. Categorical Data Analysis for the Behavioral and Social Sciences. Routledge, 2011. 283 p.
2. Lachin J.M. Biostatistical Methods: the Assessment of Relative Risks. 2nd edition. Wiley, 2011. 644 p.
3. Shoukri M.M., Pause C.A. Statistical Methods for Health Sciences. 2nd edition. CRC Press, 1999. 390 p.
4. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. 6-е изд., перераб. и доп. М. : Дело, 2004. 576 с.
5. Мятлев В.Д., Панченко Л.А., Ризниченко Г.Ю., Терехин А.Т. Теория вероятностей и математическая статистика. Математические модели. М. : Академия, 2009. 320 с.
6. Воронцов К.В. Лекции по алгоритмам восстановления регрессии. 2007. 37 с. URL: <http://www.ccas.ru/voron/download/Regression.pdf> (дата обращения: 15.03.2016)
7. Васильев Н.П., Егоров А.А. Опыт расчета параметров логистической регрессии методом Ньютона–Рафсона для оценки зи-мостойкости растений // Математическая биология и биоинформатика. 2011. Т. 6, № 2. С. 190–199. URL: http://www.matbio.org/article_pdf.php?id=82 (дата обращения: 15.03.2016)
8. Пантелеев А.В., Легова Т.А. Методы оптимизации в примерах и задачах. 2-е изд., испр. М. : Высшая школа, 2005. 544 с.
9. Бусленко Н.П., Шрейдер Ю.А. Метод статистических испытаний (Монте-Карло) и его реализация на цифровых вычисли-тельных машинах. М. : ФИЗМАТЛИТ, 1961. 226 с.
10. Ильин В.А., Садовничий В.А., Сендов Бл.Х. Математический анализ. Начальный курс. 2-е изд., перераб. М. : Изд-во МГУ, 1985. 662 с.
11. Рокафеллар Р. Выпуклый анализ : пер. с англ. М. : Мир, 1973. 472 с.

Тырсин Александр Николаевич, д-р техн. наук. E-mail: at2001@yandex.ru

Уральский федеральный университет имени первого Президента России Б.Н. Ельцина (г. Екатеринбург)

Костин Кирилл Константинович. E-mail: lemwwwar@gmail.com

Южно-Уральский государственный университет (национальный исследовательский университет) (г. Челябинск)

Поступила в редакцию 13 марта 2017 г.

Tyrsin Alexander N., Kostin Kirill K. (Ural Federal University named after the first President of Russia B.N. Yeltsin, South Ural State University (National Research University), Russian Federation).

Consideration of estimation of logistic regression as an optimization problem.

Keywords: logistic regression; the maximum likelihood method; the algorithm of descent; random search; model; objective function.

DOI: 10.17223/19988605/40/6

The logistic regression is one of the common methods of data classification in various spheres. The main goal of logistic regression is a separation of multiple input values with a linear boundary (a dividing line, plane or hyperplane) on two classes corresponding to two given region. Logistic regression predicts the probability of some events that are in the range from 0 to 1. Should indicate that the L classes can be more than two, in this case we have a multinomial regression. It can be built using $L - 1$ independent logistic regressions.

The research of an estimation algorithm logistic regression on model data implemented in statistical packages, showed instability of the estimates.

The goal of this article is research of the features of the procedure of estimation of logistic regression as an extreme problem, as well as the development of a stable algorithm of computing the coefficients of the separating equation.

Nowadays the logistic regression is usually built as an optimization problem $Q(\mathbf{b}) = \sum_{i=1}^n \ln(1 + e^{-y_i \mathbf{b}^T \mathbf{x}_i}) \rightarrow \min_{\mathbf{b} \in \mathbb{R}^n}$ that implements the criterion of maximum likelihood. In this article the features of estimation of logistic regression are considered as an optimization problem. There is shown that this problem has significant differences compared to the search of a local extremum.

There are investigated causes of the instability of estimates of the logistic regression. It is shown that using $\mathbf{b}^{(0)}$ as initial approximation for the solve of the classification task as a linear regression often yields unsatisfactory results. The second reason for the instability of estimates of maximum likelihood is that the iterative descent algorithms are designed to find a local minimum. The target function with the correct classification of all precedents has zero as the lower bound at infinity. Moreover the quality of the building of the dividing line in the formal descent wouldn't be associated with the value of the objective function $Q(\mathbf{b})$. So adjust of the strategy of the descent is needed. The third reason is that the some values, for example, $y_i \mathbf{b}^T \mathbf{x}_i$ can be very large when the value of objective function $Q(\mathbf{b})$ is small. It leads to fatal errors for computing and even can stop the operation of the algorithm due to memory overflow.

The stable algorithm for calculating the coefficients of the separating equation of the logistic regression is described. As initial approximation a hyperplane is used. The hyperplane is orthogonal to the segment connecting the centers of gravity of classes, and is passed through the middle of this segment. The iterative algorithm involves a sequence of problems of minimizing zero-order based on a random search. At each k -th step the length of the vector of coefficients $\mathbf{b}^{(k)}$ is fixed. It is gradually increased until the objective function achieves the required value.

The results of statistical modeling of comparative analysis of estimation of logistic regression with the existing algorithm are given.

REFERENCES

1. Azen, R. & Walker, C.M. (2011) *Categorical Data Analysis for the Behavioral and Social Sciences*. Routledge.
- 2 Lachin, J.M. (2011) *Biostatistical Methods: The Assessment of Relative Risks*. 2nd ed. Wiley.
3. Shoukri, M.M. & Pause, C.A. (1999) *Statistical Methods for Health Sciences*. 2nd ed. CRC Press.
4. Magnus, Ja.R., Katyshev, P.K. & Pereseckij, A.A. (2004) *Ekonometrika. Nachal'nyy kurs* [Econometrics. The Starting Course]. 6th ed. Moscow: Delo. (In Russian).
5. Myatlev, V.D., Panchenko, L.A., Riznichenko, G.Yu. & Terekhin, A.T. (2009) *Teoriya veroyatnostey i matematicheskaya statistika. Matematicheskie modeli* [Probability theory and mathematical statistics. Mathematical models]. Moscow: Akademiya.
6. Vorontsov, K.V. (2007) *Lektsii po algoritmam vosstanovleniya regressii* [Lectures about the recovery algorithms regression]. [Online] Available from: <http://www.ccas.ru/voron/download/Regression.pdf>.
7. Vasiliev, N.P. & Egorov A.A. (2011) Experience of logistic regression parameters calculation by Newton-Rafson method to estimation resistance to cold of plants. *Matematicheskaya biologiya i bioinformatika – Mathematical Biology and Bioinformatics*. 6(2). pp. 190–199. (In Russian).
8. Panteleev, A.V. & Letova, T.A. (2005) *Metody optimizatsii v primerakh i zadachakh* [Optimization methods in examples and problems]. 2nd ed. Moscow: Vysshaya shkola.
9. Buslenko, N.P. & Shreider Yu.A. (1961) *Metod statisticheskikh ispytaniy (Monte-Karlo) i ego realizatsiya na tsifrovyykh vychislitel'nykh mashinakh* [The method of statistical trials (Monte-Carlo) and its implementation on digital computers]. Moscow: FIZ-MATLIT.
10. Ilyin, V.A., Sadovnichiy, V.A. & Sendov, Bl.H. (1985) *Matematicheskiy analiz. Nachal'nyy kurs* [Mathematical analysis. The starting Course]. 2nd ed. Moscow: Moscow State University.
11. Rockafellar, R.T. (1970) *Vypuklyy analiz* [Convex Analysis]. Translated from English by A. Ioffe, V. Tikhomirov. Princeton: Princeton University Press.