МЕТОД НАИМЕНЬШИХ УГЛОВ ДЛЯ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ

К. С. Скипор

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

Ключевые слова: выбор признаков, линейное программирование, логистическая регрессия, машинное обучение, метод наименьших углов

1. Введение

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 [1].

EXAMPLE

См. также [2, 3, 4, 5, 6, 7].

2. Постановка задачи отбора признаков

Дана выборка $D = \{(\mathbf{x}^i, y^i)\}_{i=1}^m$, в которой i-й объект описывается строкой из n числовых признаков, $\mathbf{x}^i = (x_j^i)_{j=1}^n \in \mathbb{R}^n$ и метки класса $y^i \in \{0, 1\}$. Верхний индекс i указывает порядковый номер объекта выборки, нижний индекс j — порядковый номер признака. Векторы признаков $\mathbf{x}_j = (x_j^1, \ldots, x_j^i, \ldots, x_j^m)^T$ являются линейно независимыми свободными переменными, а вектор $\mathbf{y} = (y^1, \ldots, y^i, \ldots, y^m)^T$ является зависимой переменной.

EXAMPLE

Требуется построить такой алгоритм последовательного добавления признаков, что на каждом шаге:

- определяются набор *активных признаков* с *активным множеством* индексов $\mathcal{A}$ и соответствующий набору ненулевой вектор параметров $\boldsymbol{\beta}_{\mathcal{A}}$, такой что $\boldsymbol{\beta}_{\mathcal{A}^c} = \mathbf{0}$, $\mathcal{A} \sqcup \mathcal{A}^c = \{1, \ldots, n\}$;
- набор *активных признаков* и вектор параметров $\boldsymbol{\beta}_{\mathcal{A}}$ доставляют максимум приращению логарифма правдоподобия ℓ;

Date: 14 февраля 2011 г.

Научный руководитель В. В. Стрижов.

- скорость роста функционала ℓ по любому активному признаку не меньше скорости роста по любому неактивному признаку.

3. Описание алгоритма

Будем считать, что принята линейная модель

$$\mathbf{y} = \boldsymbol{\mu}(X, \boldsymbol{\beta}) + \varepsilon,$$

где функция регрессии $\boldsymbol{\mu}(X, \boldsymbol{\beta})$, представляющая собой приближение вектора y, имеет вид

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

Критерием качества назначена среднеквадратичная ошибка

$$S(X, \boldsymbol{\beta}) = \|\mathbf{y} - \boldsymbol{\mu}(X, \boldsymbol{\beta})\|^2.$$

Замечание 1. *Как и в случае LARS, вектор EXAMPLE.*

Перейдем теперь к формальной постановке решаемой задачи.
EXAMPLE

Лемма 1. *Задача EXAMPLE.*

Для доказательства теоремы 1 сформулируем некоторые вспомогательные утверждения.

Лемма 2. *Пусть векторы* $\mathbf{a}_1, \ldots, \mathbf{v}a_k, \mathbf{v}a_{k+1} \in \mathbb{R}^n$ *линейно независимы. EXAMPLE*

Для использования леммы 2 определим понятие афинной зависимости векторов

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

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

EXAMPLE
С помощью следующей теоремы формулируется утверждение о решении задачи линейного программирования (1).

Теорема 1. *Если ЗЛП (1) имеет решение* $mathbf g^*$*, то*

$$\mathbf{g}^* \in \Upsilon, \tag{2}$$

причем

$$EXAMPLE \tag{3}$$

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

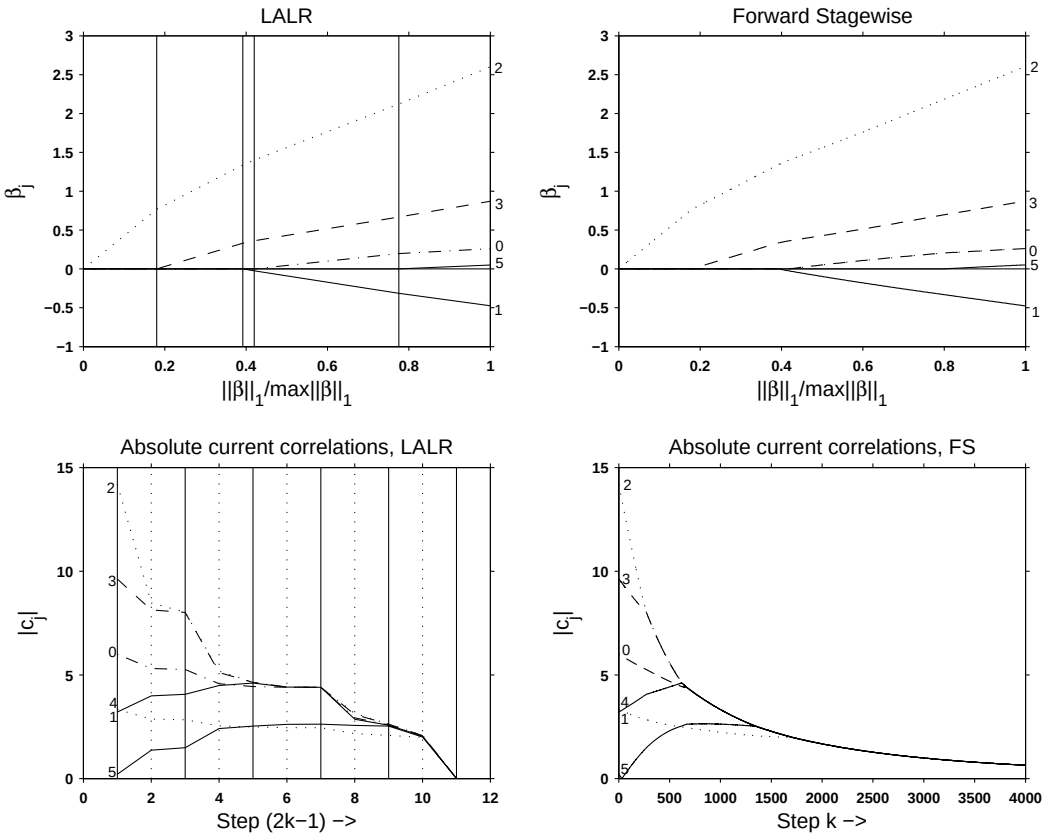

Рис. 1. Сравнение оценок коэффициентов для LALR и Forward Stagewise для модельных данных. Номера кривых соответствуют номерам признаков. Сплошные вертикальные линии обозначают шаги, а штриховые вертикальные — дополнительную итерацию для каждого шага.

4. Вычислительный эксперимент

Сравним предложенный алгоритм с описанным в Forward Stagewise. На каждом шаге алгоритм выбирает признак $\mathbf{x}_{j^*}$, имеющий наибольшую корреляцию c_{j^*} с текущим вектором остатков $\mathbf{y} - \mathbf{s}(\mathbf{m})$ и делает небольшое смещение γ текущего приближения в направлении выбранного признака $\mathbf{x}_{j^*}$,

$$j^* = \arg\max |c_j| \quad \text{и} \quad \boldsymbol{\mu} \to \boldsymbol{\mu} + \gamma \operatorname{sign}(c_{j^*})\mathbf{x}_{j^*}.$$

Чем меньше абсолютная величина смещения γ, тем точнее получается оценка параметров $\boldsymbol{\beta}$.

Сгенерируем $m = 50$ объектов с пятью независимыми, нормально распределенными признаками $\mathbf{x}_1, \cdots, \mathbf{x}_5$, т.е $\mathbf{x}_i = (x_{i1}, \cdots, x_{i5}) \sim \mathcal{N}_5(\mathbf{0}, \mathbf{I})$. Примем модель

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

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

в качестве параметров $\boldsymbol{\beta}$ возьмем, например, вектор $(\beta_0, \beta_1, \beta_2, \beta_3)^T = (1, -2, 6, 3)^T$. В нашей модели признаки $\mathbf{x}_4$ и $\mathbf{x}_5$ 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.

EXAMPLE

5. Заключение

В данной работе предложен и исследован новый алгоритм LALR, решающий задачу отбора признаков в модели логистической регрессии.

EXAMPLE

Список литературы

[1] Forward stagewise regression and the monotone lasso / T. Hastie, J. Taylor, R. Tibshirani, G. Walther // *Electronic Journal of Statistics.* — 2007. — Vol. 1, no. 1. — Pp. 1–29.

[2] *Тихонов А. Н.* Решение некорректно поставленных задач и метод регуляризации // *ДАН.* — 1963. — Т. 151, № 3. — С. 501–504.

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

[4] *Friedman J. H.* Greedy function approximation: A gradient boosting machine // *Annals of Statistics.* — 2000. — Vol. 29. — Pp. 1189–1232.

[5] *Friedman J., Hastie T., Tibshirani R.* Additive logistic regression: a statistical view of boosting // *Annals of Statistics.* — 1998. — Vol. 28. — P. 2000.

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

[7] *Draper N., Smith H.* Appied Regression Analysis. — New York: Wiley, 1966. — 709 pp.

Московский физико-технический институт, ФУПМ, каф. «Интеллектуальные системы»