Взвешенный метод наименьших квадратов

Дмитрий Алексеевич Павлов
инж.-иссл. Лаборатории Чебышёва СПбГУ
d.a.pavlov@spbu.ru

СПбГУ, МКН
2021

Линейная регрессия

Модель

Простая (парная) регрессия $y = \beta_0 + \beta_1 x + \varepsilon$

Множественная регрессия $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$

y – зависимая переменная (наблюдаемая)

x_i – влияющие переменные (наблюдаемые)

β_j – параметры модели (ненаблюдаемые)

ε – случайные ошибки (шум)

Наблюдения

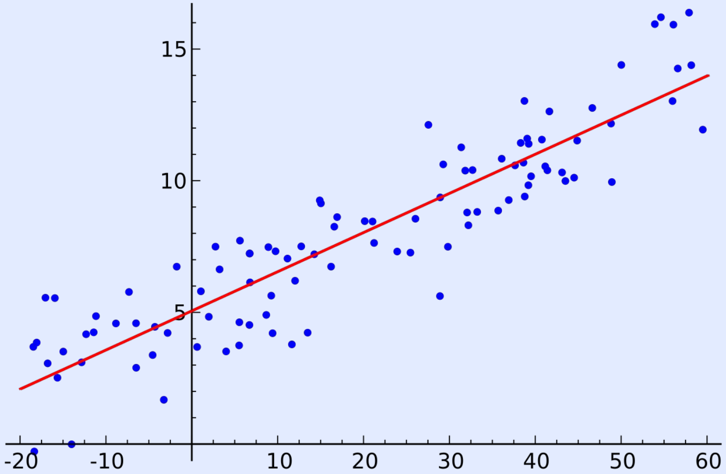
$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

Линейная регрессия (продолжение)

Дано

$\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$ – наблюдения

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}$$



Найти $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k):$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \hat{\beta}_0 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \hat{\beta}_1 \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} + \cdots + \hat{\beta}_k \begin{pmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{nk} \end{pmatrix} + \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

Метод наименьших квадратов

$$\sum_{i=1}^{n} r_i^2 = (\mathbf{y} - X\hat{\boldsymbol{\beta}})^T(\mathbf{y} - X\hat{\boldsymbol{\beta}}) \to \min$$

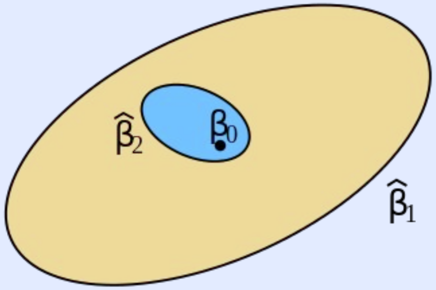
$$\frac{\partial}{\partial\hat{\boldsymbol{\beta}}}\sum_{i=1}^{n} r_i^2 = \frac{\partial}{\partial\hat{\boldsymbol{\beta}}}\hat{\boldsymbol{\beta}}^T X^T X\hat{\boldsymbol{\beta}} - 2\frac{\partial}{\partial\hat{\boldsymbol{\beta}}}\hat{\boldsymbol{\beta}}^T X^T\mathbf{y} = 2(X^T X\hat{\boldsymbol{\beta}} - X^T\mathbf{y})$$

$$\frac{\partial}{\partial\hat{\boldsymbol{\beta}}}\sum_{i=1}^{n} r_i^2 = 0 \Rightarrow \hat{\boldsymbol{\beta}} = (X^T X)^{-1}X^T y$$ – линейная (по *y*)

$$\mathrm{E}[\varepsilon_i | X] = 0 \;\Rightarrow\; \mathrm{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$$ – несмещённая

$$\lim_{n\to\infty} \mathbb{P}[|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}| < \delta] = 0, \;\forall\delta$$ – состоятельная

оценка



Теорема Гаусса-Маркова

Если модель данных *правильно специфицирована* и

$$\mathrm{rank}(X) = k \qquad \mathrm{Var}[\varepsilon_i] = \sigma^2 < \infty$$

$$\mathrm{E}[\varepsilon_i | X] = 0 \qquad \mathrm{Cov}[\varepsilon_i, \varepsilon_j] = 0, \ \forall i \neq j$$

то

$$\sum r_i^2 = \min$$

$$\mathrm{Var}[\hat{\boldsymbol{\beta}}] = \sigma^2 (X^T X)^{-1}$$

$\hat{\boldsymbol{\beta}}$ является лучшей из линейных несмещённых оценок $\boldsymbol{\beta}$

(BLUE – Best Linear Unbiased Estimator)

$$\tilde{\boldsymbol{\beta}} = ((X^T X)^{-1} + D)\mathbf{y} \ \Rightarrow \ \mathrm{Var}[\tilde{\boldsymbol{\beta}}] = \mathrm{Var}[\hat{\boldsymbol{\beta}}] + \sigma^2 D D^T$$

Другие оценки

- Взвешенный метод наименьших квадратов (учитывает различные $\mathrm{Var}[\varepsilon_i]$, см. далее)

Остальное см. в учебниках:

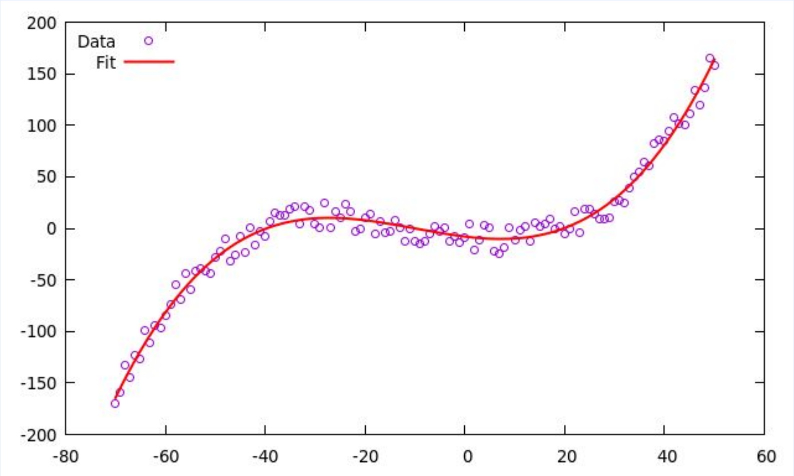
- Обобщённый метод наименьших квадратов (учитывает корреляции между ε_i)
- Метод регуляризации Тихонова (учитывает априорное распределение β)
- Метод максимального правдоподобия (эквивалентен МНК при условиях теоремы Гаусса-Маркова и нормальном распределении ε_i)
- Метод наименьших полных квадратов (учитывает ошибки в зависимых переменных)
- Метод наименьших модулей (устойчив к выбросам в данных)

Полиномиальная регрессия

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k + \varepsilon$$

- Модель осталась линейной!
- Корреляции между x^i не нарушают условий теоремы Гаусса-Маркова.

```
set key left top

system "seq -70 50 | awk -e '\
{ print $1,
   1e-3 * $1**3 + 0.03 * $1**2 - 0.6 * $1 - 8 \
   + 30 * (rand() - 0.5) }' > ./data.txt"

f(x) = a * x**3  + b * x**2 + c * x + d

fit f(x) "data.txt" using 1:2 via a, b, c, d

plot 'data.txt' using 1:2 with points pt 6 t "Data",\
      f(x) t "Fit" lc rgb "red" lw 2

pause -1
```




```
Final set of parameters            Asymptotic Standard Error
=======================            ==========================
a               = 0.00100756       +/- 2.422e-05     (2.404%)
b               = 0.0301535        +/- 0.001039      (3.446%)
c               = -0.591481        +/- 0.05355       (9.054%)
d               = -7.91562         +/- 1.289         (16.28%)
```

- Обобщается на произвольное количество переменных.
- Вместо мономов могут быть ортогональные многочлены: Чебышёва, Лежандра, …

Взвешивание

$\mathrm{Var}[\varepsilon_i] = \sigma_i^2$ Если дисперсии ошибок различаются, МНК уже не является лучшей оценкой (хотя продолжает быть несмещённой).

$$\frac{y_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{x_{i1}}{\sigma_i} + \cdots + \beta_k \frac{x_{ik}}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i} \qquad \mathrm{Var}\left[\frac{\varepsilon_i}{\sigma_i}\right] = 1 = \mathrm{const}$$

$$\sum_{i=1}^{n} \frac{r_i^2}{\sigma_i^2} = (\sqrt{W}\mathbf{y} - \sqrt{W}X\hat{\boldsymbol{\beta}})^T(\sqrt{W}\mathbf{y} - \sqrt{W}X\hat{\boldsymbol{\beta}}) \to \min \qquad \sqrt{W} = \begin{bmatrix} \frac{1}{\sigma_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_n} \end{bmatrix}$$

$$\mathbf{y}' = \sqrt{W}\mathbf{y},\ X' = \sqrt{W}X,\ \mathbf{r}' = \sqrt{W}\mathbf{r}$$

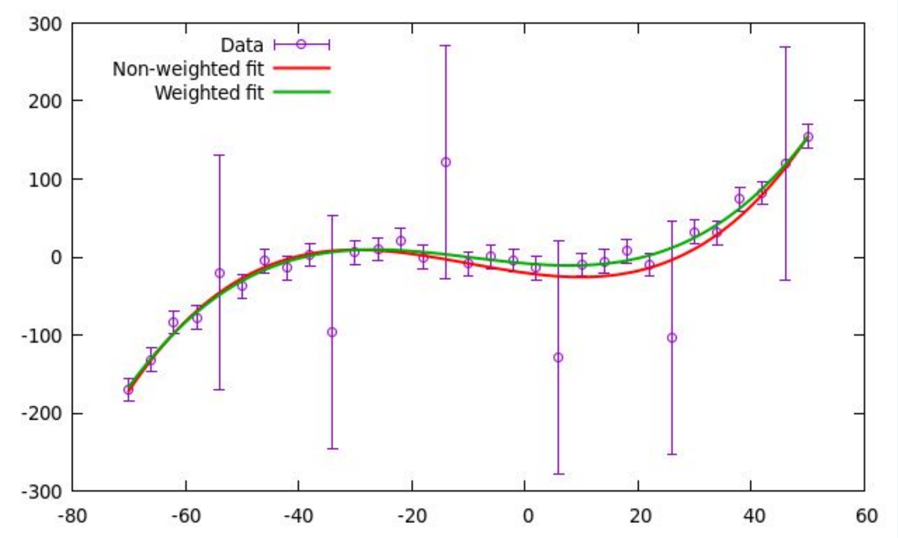
В такой постановке вновь применима теорема Гаусса-Маркова.

$\hat{\boldsymbol{\beta}} = (X^T W X)^{-1} X^T W y$ — лучшая из линейных несмещённых оценок

$$\mathrm{Var}[\hat{\boldsymbol{\beta}}] = (X^T W X)^{-1}$$

Взвешивание: пример

```
set key left top

system "seq -70 4 50 | awk -e '                        \
{ if (NR % 5 == 0) sigma = 150; else sigma = 15;        \
  print $1, 1e-3 * $1**3 + 0.03 * $1**2 - 0.6 * $1 - 8 \
  + 2 * sigma * (rand() - 0.5), sigma}' > ./data.txt"

f(x) = a * x**3  + b * x**2 + c * x + d
f1(x) = a1 * x**3  + b1 * x**2 + c1 * x + d1

fit f(x)  'data.txt' using 1:2           via a,  b,  c,  d
fit f1(x) 'data.txt' using 1:2:3 yerrors via a1, b1, c1, d1

plot 'data.txt' using 1:2:3 with yerrorbars pt 6 t "Data",\
      f(x)  t "Non-weighted fit" lc rgb "red" lw 2,      \
      f1(x) t "Weighted fit"     lc rgb "#00AA00" lw 2
pause -1
```



Невзвешенный МНК

```
Final set of parameters            Asymptotic Standard Error
=======================            ==========================
a               = 0.00112155       +/- 0.0002194     (19.57%)
b               = 0.0334969        +/- 0.009513      (28.4%)
c               = -0.990355        +/- 0.5097        (51.46%)
d               = -20.2313         +/- 12.47         (61.62%)
```

Взвешенный МНК

```
Final set of parameters            Asymptotic Standard Error
=======================            ==========================
a1              = 0.000986942      +/- 4.749e-05     (4.812%)
b1              = 0.0278733        +/- 0.002147      (7.702%)
c1              = -0.623062        +/- 0.1102        (17.69%)
d1              = -8.09056         +/- 2.745         (33.93%)
```

Оценка дисперсии и ошибка единицы веса

Пусть дисперсия в МНК неизвестна: $\mathrm{Var}[\varepsilon_i] = \sigma^2 = ?$

$\hat{\sigma}^2 = \dfrac{\sum r_i^2}{n-k}$ – несмещённая оценка дисперсии

Ошибка единицы веса во взвешенном МНК:

$$\mu = \frac{\sum \frac{r_i^2}{\sigma_i^2}}{n-k}$$

(reduced chi-square, regression standard error, variance of unit weight)

$\mu \gg 1$ – плохое соответствие между моделью и данными

$\mu \ll 1$ – «перетренированная» модель или слишком пессимистические априорные ошибки (σ_i)

μ около 1 – полный порядок

Разложение Холецкого

Как вычислить $(X^\mathrm{T}X)^{-1}$ или $(X^\mathrm{T}WX)^{-1}$?

Кто сказал «методом Гаусса»? Никогда не применяйте метод Гаусса.

$A = X^\mathrm{T}WX$ — симметричная положительно определённая матрица.

$$A = LL^T,\ L = \begin{bmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ l_{(k+1)1} & \dots & l_{(k+1)k} & l_{(k+1)(k+1)} \end{bmatrix}$$

$$C \equiv A^{-1} \quad AC = E \Leftrightarrow LB = E,\ L^T C = B$$

Так как матрица L — нижняя треугольная, то обе системы решаются последовательным исключением неизвестных.

Осталось найти L.

Разложение Холецкого (продолжение)

$$
\begin{aligned}
l_{11} &= \sqrt{a_{11}} \\
l_{j1} &= a_{j1}/l_{11}, \quad j = 2..(k+1) \\
l_{ii} &= \sqrt{a_{ii} - \sum_{p=1}^{i-1} l_{ip}^2}, \quad i = 2..(k+1) \\
l_{ji} &= \left(a_{ji} - \sum_{p=1}^{i-1} l_{ip} l_{jp}\right)/l_{ii}, \quad i = 2..k,\ j = (i+1)..(k+1)
\end{aligned}
$$

Так как *A* положительно определенная, то элементы под корнем не могут быть отрицательными.

Выделять память под *L* необязательно: можно замещать элементы *A*.

Выделять память под *X* тем более необязательно: *A* нужно строить, последовательно обрабатывая строки $x_{i1} .. x_{ik}$.

Ссылки

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC392707/pdf/pnas00019-0030.pdf

http://web.hku.hk/~pingyu/6005/LN/LN3_Least%20Squares%20Estimation-Finite-Sample%20Properties.pdf

http://web.hku.hk/~pingyu/6005/LN/LN5_Least%20Squares%20Estimation-%20Large-Sample%20Properties.pdf

https://en.wikipedia.org/wiki/Ordinary_least_squares

https://stats.stackexchange.com/a/149111

https://en.wikipedia.org/wiki/Proofs_involving_ordinary_least_squares

https://en.wikipedia.org/wiki/Weighted_least_squares

https://en.wikipedia.org/wiki/Cholesky_decomposition