Основы теории идентификации
Лекция 2: Статический инструментарий теории идентификации

Метод наименьших квадратов

Арановский Станислав Владимирович

aranovskiysv@niuitmo.ru

УНИВЕРСИТЕТ ИТМО

Осень, 2014

Содержание

Содержание

Идентификация силы трения

Рассмотрим простую модель силы трения, Кулоново и вязкое трение:

$$F = F_c \mathrm{sign} v + F_v v,$$

где F_c – величина Кулонова трения, F_v – коэффициента вязкого трения, v – скорость движения тела.

Идентификация силы трения

Рассмотрим простую модель силы трения, Кулоново и вязкое трение:

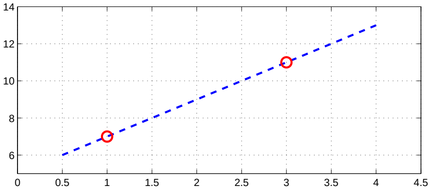
$$F = F_c \operatorname{sign} v + F_v v,$$

где F_c – величина Кулонова трения, F_v – коэффициента вязкого трения, v – скорость движения тела.

Пусть в ходе некоторого эксперимента мы измерили пары значений (v_1, F_1) и (v_2, F_2), $v_1, v_2 > 0$. Тогда мы можем легко подсчитать:

$$F_v = \frac{F_2 - F_1}{v_2 - v_1}$$

$$F_c = F_1 - F_v v_1 = F_2 - F_v v_2$$

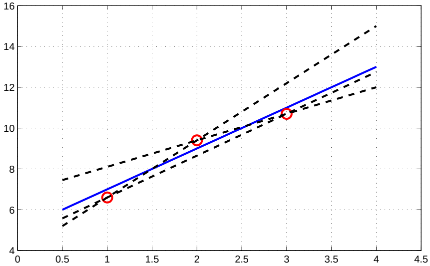


Идентификация силы трения

Рассмотрим простую модель силы трения, Кулоново и вязкое трение:

$$F = F_c \operatorname{sign} v + F_v v,$$

где F_c – величина Кулонова трения, F_v – коэффициента вязкого трения, v – скорость движения тела.
Но что делать, если мы измерили три точки?



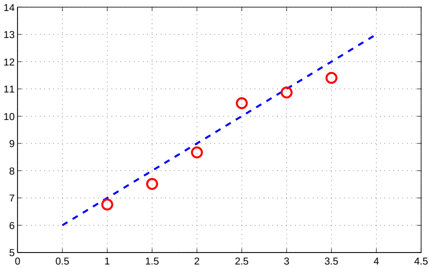
Идентификация силы трения

Рассмотрим простую модель силы трения, Кулоново и вязкое трение:

$$F = F_c \operatorname{sign} v + F_v v,$$

где F_c – величина Кулонова трения, F_v – коэффициента вязкого трения, v – скорость движения тела.

Но что делать, если мы измерили три точки? Много точек?



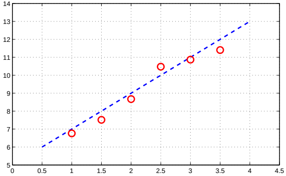
Идентификация силы трения

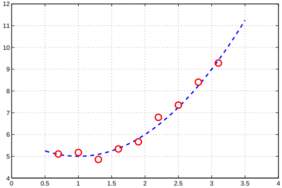
Рассмотрим простую модель силы трения, Кулоново и вязкое трение:

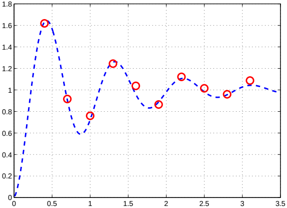
$$F = F_c \mathrm{sign} v + F_v v,$$

где F_c – величина Кулонова трения, F_v – коэффициента вязкого трения, v – скорость движения тела.

Как вообще найти наилучшую аппроксимацию?







Содержание

Постановка задачи

Регрессионная модель: $y = g(x_1, x_2, \ldots, x_m)$, где независимые переменные x_i – регрессоры.

Постановка задачи

Регрессионная модель: $y = g(x_1, x_2, \ldots, x_m)$, где независимые переменные x_i – регрессоры.
Линейная регрессия: $y = x_1\theta_1 + x_2\theta_2 + \ldots + x_m\theta_m = \sum_{i=1}^{m} x_i\theta_i$

Линейная регрессия

$$y = \boldsymbol{x}^\top \boldsymbol{\theta}$$

где $\boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{x} = \mathrm{col}\{x_i\}$ и $\boldsymbol{\theta} \in \mathbb{R}^m$,$\boldsymbol{\theta} = \mathrm{col}\{\theta_i\}$.

Постановка задачи

Регрессионная модель: $y = g(x_1, x_2, \ldots, x_m)$, где независимые переменные x_i – регрессоры.
Линейная регрессия: $y = x_1\theta_1 + x_2\theta_2 + \ldots + x_m\theta_m = \sum_{i=1}^{m} x_i\theta_i$

Линейная регрессия

$$y = \boldsymbol{x}^\top \boldsymbol{\theta}$$

где $\boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{x} = \mathrm{col}\{x_i\}$ и $\boldsymbol{\theta} \in \mathbb{R}^m, \boldsymbol{\theta} = \mathrm{col}\{\theta_i\}$.

Пример

$$F = F_c \mathrm{sign}\, v + F_v v = \begin{bmatrix} \mathrm{sign}\, v & v \end{bmatrix} \begin{bmatrix} F_c \\ F_v \end{bmatrix} = \boldsymbol{x}^\top \boldsymbol{\theta}$$

где $\boldsymbol{x} = \begin{bmatrix} \mathrm{sign}\, v \\ v \end{bmatrix}$ и $\boldsymbol{\theta} = \begin{bmatrix} F_c \\ F_v \end{bmatrix}$
Важно: используемая модель это только аппроксимация.

Постановка задачи, продолжение

Линейная регрессия

$$y = \boldsymbol{x}^\top \boldsymbol{\theta}$$

Пусть проведено N измерений $(y(k), \boldsymbol{x}(k))$, $k = 1, \dots, N$. Тогда можно записать $y(k) = \boldsymbol{x}^\top(k)\boldsymbol{\theta}$ или

$$Y = X\boldsymbol{\theta},$$

где $Y \in \mathbb{R}^N$, $Y = \mathrm{col}\{y(k)\}$ и $X \in \mathbb{R}^{N \times m}$, $X = \mathrm{col}\{\boldsymbol{x}^\top(k)\}$,

$$X = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_m(1) \\ x_1(2) & x_2(2) & \cdots & x_m(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(N) & x_2(N) & \cdots & x_m(N) \end{bmatrix}$$

Постановка задачи, продолжение

Линейная регрессия

$$y = \boldsymbol{x}^\top \boldsymbol{\theta} \Rightarrow Y = X\boldsymbol{\theta}$$

Задача: найти наилучшую аппроксимацию. Как оценить качество аппроксимации?

Постановка задачи, продолжение

Линейная регрессия

$$y = \boldsymbol{x}^\top \boldsymbol{\theta} \Rightarrow Y = X\boldsymbol{\theta}$$

Задача: найти наилучшую аппроксимацию. Как оценить качество аппроксимации?
Для $\hat{\boldsymbol{\theta}}$ найдем оценку $\hat{y}(k) = \boldsymbol{x}^\top(k)\hat{\boldsymbol{\theta}}$ и ошибку оценивания

$$e(k) = y(k) - \hat{y}(k) = y(k) - \boldsymbol{x}^\top(k)\hat{\boldsymbol{\theta}},$$
$$\boldsymbol{e} = \mathrm{col}\{e(k)\} \in \mathbb{R}^N.$$

Постановка задачи, продолжение

Линейная регрессия

$$y = \boldsymbol{x}^\top \boldsymbol{\theta} \Rightarrow Y = X\boldsymbol{\theta}$$

Задача: найти наилучшую аппроксимацию. Как оценить качество аппроксимации?

Для $\hat{\boldsymbol{\theta}}$ найдем оценку $\hat{y}(k) = \boldsymbol{x}^\top(k)\hat{\boldsymbol{\theta}}$ и ошибку оценивания

$$e(k) = y(k) - \hat{y}(k) = y(k) - \boldsymbol{x}^\top(k)\hat{\boldsymbol{\theta}},$$
$$\boldsymbol{e} = \mathrm{col}\{e(k)\} \in \mathbb{R}^N.$$

Сформируем критерий как сумму квадратов ошибки:

$$V(\hat{\boldsymbol{\theta}}) = \|\boldsymbol{e}\|_2^2 = \boldsymbol{e}^\top \boldsymbol{e} = \sum_{k=1}^{N} e^2(k)$$
$$= \sum_{k=1}^{N} (y(k) - \boldsymbol{x}^\top(k)\hat{\boldsymbol{\theta}})^2 = \left(Y - X\hat{\boldsymbol{\theta}}\right)^\top \left(Y - X\hat{\boldsymbol{\theta}}\right)$$

Постановка задачи, продолжение

Линейная регрессия

$$y = \boldsymbol{x}^\top \boldsymbol{\theta} \Rightarrow Y = X\boldsymbol{\theta}$$

Детерминированная задача НК

По заданным Y и X найти $\hat{\boldsymbol{\theta}}_{LS} = \arg\min_{\hat{\boldsymbol{\theta}}} V(\hat{\boldsymbol{\theta}})$ для

$$V(\hat{\boldsymbol{\theta}}) = \|\boldsymbol{e}\|_2^2 = \left(Y - X\hat{\boldsymbol{\theta}}\right)^\top \left(Y - X\hat{\boldsymbol{\theta}}\right)$$

Важно: в детерминированной постановке задачи не ставится вопрос идентификации некоторого *истинного* значения $\boldsymbol{\theta}$, не утверждается даже его существование. Вместо этого рассматривается поиск наилучшей линейной аппроксимации заданного набора данных. Такая постановка задачи встречается, например, в статистике.

Постановка задачи, продолжение

Линейная регрессия

$$y = \boldsymbol{x}^\top \boldsymbol{\theta} \Rightarrow Y = X\boldsymbol{\theta}$$

Детерминированная задача НК

По заданным Y и X найти $\hat{\boldsymbol{\theta}}_{LS} = \arg\min_{\hat{\boldsymbol{\theta}}} V(\hat{\boldsymbol{\theta}})$ для

$$V(\hat{\boldsymbol{\theta}}) = \|\boldsymbol{e}\|_2^2 = \left(Y - X\hat{\boldsymbol{\theta}}\right)^\top \left(Y - X\hat{\boldsymbol{\theta}}\right)$$

Существует альтернативный L_1 критерий

$$V(\hat{\boldsymbol{\theta}}) = \|\boldsymbol{e}\|_1 = \sum_{k=1}^{N} |e(k)|$$

МНК: Теорема

Детерминированная задача НК

По заданным Y и X найти $\hat{\boldsymbol{\theta}}_{LS} = \arg\min_{\hat{\boldsymbol{\theta}}} V(\hat{\boldsymbol{\theta}})$ для

$$V(\hat{\boldsymbol{\theta}}) = \|\boldsymbol{e}\|_2^2 = \left(Y - X\hat{\boldsymbol{\theta}}\right)^\top \left(Y - X\hat{\boldsymbol{\theta}}\right)$$

Теорема (Решение детерминированной задачи НК)

Минимум $V(\hat{\boldsymbol{\theta}}) = (Y - X\hat{\boldsymbol{\theta}})^\top (Y - X\hat{\boldsymbol{\theta}})$ *достигается при* $\hat{\boldsymbol{\theta}}$ *таких, что*

$$(X^\top X)\hat{\boldsymbol{\theta}} = X^\top Y.$$

Если $\Phi \in \mathbb{R}^{m\times m}$, $\Phi := X^\top X$ *имеет обратную, то*

- $\hat{\boldsymbol{\theta}}_{LS} := \left(X^\top X\right)^{-1} X^\top Y$,
- $V(\hat{\boldsymbol{\theta}})$ *имеет единственный минимум:* $\min_{\hat{\boldsymbol{\theta}}} V(\hat{\boldsymbol{\theta}}) = V(\hat{\boldsymbol{\theta}}_{LS}) = Y^\top Y - Y^\top X \left(X^\top X\right)^{-1} X^\top Y.$

МНК: Доказательство

$$V(\boldsymbol{\theta}) = (Y - X\boldsymbol{\theta})^\top (Y - X\boldsymbol{\theta})$$
$$= Y^\top Y - 2Y^\top X\boldsymbol{\theta} + \boldsymbol{\theta}^\top X^\top X\boldsymbol{\theta}$$

$$V(\boldsymbol{\theta}) \to \min \Rightarrow \nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}) = \mathbf{0}$$

Некоторые формулы:

$$\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}) = \mathrm{grad}_{\boldsymbol{\theta}}^\top V(\boldsymbol{\theta}) = \mathrm{col}\left\{\frac{\partial V(\boldsymbol{\theta})}{\partial \theta_i}\right\}$$

$$\nabla_{\boldsymbol{\theta}}(\boldsymbol{a}^\top \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}(\boldsymbol{\theta}^\top \boldsymbol{a}) = \boldsymbol{a}$$

$$\nabla_{\boldsymbol{\theta}}(\boldsymbol{\theta}^\top A\boldsymbol{\theta}) = (A + A^\top)\boldsymbol{\theta} = (\underbrace{\boldsymbol{\theta}^\top (A + A^\top)}_{\mathrm{grad}_{\boldsymbol{\theta}}})^\top$$

$X^\top X = (X^\top X)^\top \geq 0$ – симметрическая.

МНК: Доказательство, продолжение

$$\begin{aligned}
V(\boldsymbol{\theta}) &= Y^\top Y - 2Y^\top X\boldsymbol{\theta} + \boldsymbol{\theta}^\top X^\top X\boldsymbol{\theta} \\
\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}) &= -2X^\top Y + 2X^\top X\boldsymbol{\theta} \\
\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}) &= 0 \Leftrightarrow X^\top X\boldsymbol{\theta} = X^\top Y
\end{aligned}$$

Пусть $X^\top X$ имеет обратную. Тогда:

$$\boldsymbol{\theta} = \left(X^\top X\right)^{-1} X^\top Y =: \hat{\boldsymbol{\theta}}_{LS}$$

и

$$\begin{aligned}
V(\boldsymbol{\theta}) &= Y^\top Y - 2Y^\top X\boldsymbol{\theta} + \boldsymbol{\theta}^\top X^\top X\boldsymbol{\theta} \\
&= Y^\top Y - 2Y^\top X\left(X^\top X\right)^{-1} X^\top Y + Y^\top X\left(X^\top X\right)^{-1} X^\top Y \\
&= Y^\top Y - Y^\top X\left(X^\top X\right)^{-1} X^\top Y
\end{aligned}$$

МНК: Доказательство, продолжение

Проверим, что это действительно единственный минимум:

$$\begin{aligned} V(\hat{\theta}_{LS} + \tilde{\theta}) &= Y^\top Y - 2Y^\top X(\hat{\theta}_{LS} + \tilde{\theta}) + (\hat{\theta}_{LS} + \tilde{\theta})^\top X^\top X(\hat{\theta}_{LS} + \tilde{\theta}) \\ &= V(\hat{\theta}_{LS}) - 2Y^\top X\tilde{\theta} + 2\tilde{\theta}^\top X^\top X\hat{\theta}_{LS} + \tilde{\theta}^\top X^\top X\tilde{\theta} \\ &= V(\hat{\theta}_{LS}) + \tilde{\theta}^\top X^\top X\tilde{\theta} \geq V(\hat{\theta}_{LS}) \end{aligned}$$

МНК: Доказательство, продолжение

Проверим, что это действительно единственный минимум:

$$\begin{aligned} V(\hat{\theta}_{LS}+\tilde{\theta}) &= Y^\top Y - 2Y^\top X(\hat{\theta}_{LS}+\tilde{\theta}) + (\hat{\theta}_{LS}+\tilde{\theta})^\top X^\top X(\hat{\theta}_{LS}+\tilde{\theta}) \\ &= V(\hat{\theta}_{LS}) - 2Y^\top X\tilde{\theta} + 2\tilde{\theta}^\top X^\top X\hat{\theta}_{LS} + \tilde{\theta}^\top X^\top X\tilde{\theta} \\ &= V(\hat{\theta}_{LS}) + \tilde{\theta}^\top X^\top X\tilde{\theta} \geq V(\hat{\theta}_{LS}) \end{aligned}$$

Рассмотрим минимальное значение $V(\hat{\theta}) = V(\hat{\theta}_{LS})$:

$$V(\hat{\theta}_{LS}) = Y^\top Y - Y^\top X(X^\top X)^{-1}X^\top Y.$$

В каком случае оно нулевое?

МНК: Доказательство, продолжение

Проверим, что это действительно единственный минимум:

$$
\begin{aligned}
V(\hat{\theta}_{LS} + \tilde{\theta}) &= Y^\top Y - 2Y^\top X(\hat{\theta}_{LS} + \tilde{\theta}) + (\hat{\theta}_{LS} + \tilde{\theta})^\top X^\top X(\hat{\theta}_{LS} + \tilde{\theta}) \\
&= V(\hat{\theta}_{LS}) - 2Y^\top X\tilde{\theta} + 2\tilde{\theta}^\top X^\top X\hat{\theta}_{LS} + \tilde{\theta}^\top X^\top X\tilde{\theta} \\
&= V(\hat{\theta}_{LS}) + \tilde{\theta}^\top X^\top X\tilde{\theta} \geq V(\hat{\theta}_{LS})
\end{aligned}
$$

Рассмотрим минимальное значение $V(\hat{\theta}) = V(\hat{\theta}_{LS})$:

$$
V(\hat{\theta}_{LS}) = Y^\top Y - Y^\top X(X^\top X)^{-1}X^\top Y.
$$

В каком случае оно нулевое? При

$$
\mathcal{I}_{N\times N} - X(X^\top X)^{-1}X^\top = \mathbf{0}.
$$

МНК: Доказательство, продолжение

Проверим, что это действительно единственный минимум:

$$\begin{aligned} V(\hat{\theta}_{LS} + \tilde{\theta}) &= Y^\top Y - 2Y^\top X(\hat{\theta}_{LS} + \tilde{\theta}) + (\hat{\theta}_{LS} + \tilde{\theta})^\top X^\top X(\hat{\theta}_{LS} + \tilde{\theta}) \\ &= V(\hat{\theta}_{LS}) - 2Y^\top X\tilde{\theta} + 2\tilde{\theta}^\top X^\top X\hat{\theta}_{LS} + \tilde{\theta}^\top X^\top X\tilde{\theta} \\ &= V(\hat{\theta}_{LS}) + \tilde{\theta}^\top X^\top X\tilde{\theta} \geq V(\hat{\theta}_{LS}) \end{aligned}$$

Рассмотрим минимальное значение $V(\hat{\theta}) = V(\hat{\theta}_{LS})$:

$$V(\hat{\theta}_{LS}) = Y^\top Y - Y^\top X(X^\top X)^{-1}X^\top Y.$$

В каком случае оно нулевое? При

$$\mathcal{I}_{N\times N} - X(X^\top X)^{-1}X^\top = \mathbf{0}.$$

Частный случай: $N = m$, X – квадратная и $\det(X) \neq 0$.

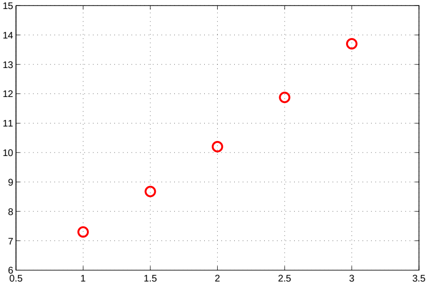
Содержание

Трение

Вернемся к задаче аппроксимации трения моделью

$$F = F_c \operatorname{sign} v + F_v v.$$

Пусть измерены точки (v_k, F_k): $(1,\ 7.3)$, $(1.5,\ 8.7)$, $(2,\ 10.2)$, $(2.5,\ 11.9)$, $(3,\ 13.7)$.



Трение

Вернемся к задаче аппроксимации трения моделью

$$F = F_c \operatorname{sign} v + F_v v.$$

Пусть измерены точки (v_k, F_k): $(1,\ 7.3)$, $(1.5,\ 8.7)$, $(2,\ 10.2)$, $(2.5,\ 11.9)$, $(3,\ 13.7)$.
Запишем:

$$\underbrace{\begin{bmatrix} 7.3 \\ 8.7 \\ 10.2 \\ 11.9 \\ 13.7 \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1.5 & 1 \\ 2 & 1 \\ 2.5 & 1 \\ 3 & 1 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} F_v \\ F_c \end{bmatrix}}_{\theta}$$

Трение

Вернемся к задаче аппроксимации трения моделью

$$F = F_c \operatorname{sign} v + F_v v.$$

Пусть измерены точки (v_k, F_k): $(1,\ 7.3)$, $(1.5,\ 8.7)$, $(2,\ 10.2)$, $(2.5,\ 11.9)$, $(3,\ 13.7)$.

$$\begin{bmatrix} 7.3 & 8.7 & 10.2 & 11.9 & 13.7 \end{bmatrix}^\top = \begin{bmatrix} 1 & 1.5 & 2 & 2.5 & 3 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}^\top \begin{bmatrix} F_v \\ F_c \end{bmatrix}$$

$$X^\top X = \begin{bmatrix} \sum_{k=1}^5 x_1(k)^2 & \sum_{k=1}^5 x_1(k)x_2(k) \\ \sum_{k=1}^5 x_1(k)x_2(k) & \sum_{k=1}^5 x_2(k)^2 \end{bmatrix} = \begin{bmatrix} 22.5 & 10 \\ 10 & 5 \end{bmatrix}$$

$$X^\top Y = \begin{bmatrix} \sum_{k=1}^5 x_1(k)y(k) \\ \sum_{k=1}^5 x_2(k)y(k) \end{bmatrix} = \begin{bmatrix} 111.5 \\ 51.75 \end{bmatrix}$$

$$\hat{\theta}_{LS} = \left(X^\top X\right)^{-1} X^\top Y = \begin{bmatrix} 3.2 \\ 3.95 \end{bmatrix}$$

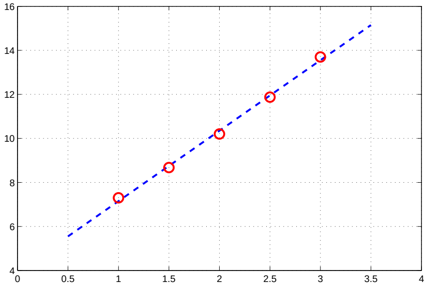
Трение

Вернемся к задаче аппроксимации трения моделью

$$F = F_c \operatorname{sign} v + F_v v.$$

Пусть измерены точки (v_k, F_k): $(1,\ 7.3)$, $(1.5,\ 8.7)$, $(2,\ 10.2)$, $(2.5,\ 11.9)$, $(3,\ 13.7)$.

$$\hat{\theta}_{LS} = \left(X^\top X\right)^{-1} X^\top Y = \begin{bmatrix} 3.2 \\ 3.95 \end{bmatrix}$$



Парабола

Рассмотрим задачу аппроксимации данных параболой
$y = ax^2 + bx + c$, или

$$y = \begin{bmatrix} x^2 & x & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix},$$

по $N = 5$ точкам: (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5).

Парабола

Рассмотрим задачу аппроксимации данных параболой
$y = ax^2 + bx + c$, или

$$y = \begin{bmatrix} x^2 & x & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix},$$

по $N = 5$ точкам: (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5). Тогда:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \underbrace{\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \\ x_4^2 & x_4 & 1 \\ x_5^2 & x_5 & 1 \end{bmatrix}}_{V} \begin{bmatrix} a \\ b \\ c \end{bmatrix},$$

Матрица Вандермонда

Матрица вида

$$V = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{m-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & x_N^2 & \cdots & x_N^{m-1} \end{bmatrix},$$

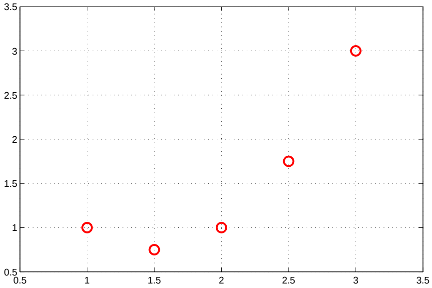
$V \in \mathbb{R}^{N\times m}$ называется матрицей Вандермонда. Для квадратной матрицы Вандермонда определитель

$$\det(V_{m\times m}) = \prod_{1\le i<j\le m} (x_j - x_i),$$

т.е. матрица невырождена, если все x_i различны.
Если $V \in \mathbb{R}^{N\times m}$, $N > m$, то $\det(V^\top V) \neq 0$ если $\operatorname{rank}(V) = m$, т.е. есть хотя бы m различных значений x_i.

Парабола, продолжение

Рассмотрим задачу аппроксимации данных параболой $y = ax^2 + bx + c$ по $N = 5$ точкам: $(1,\ 1)$, $(1.5,\ 0.75)$, $(2,\ 1)$, $(2.5,\ 1.75)$, $(3,\ 3)$.



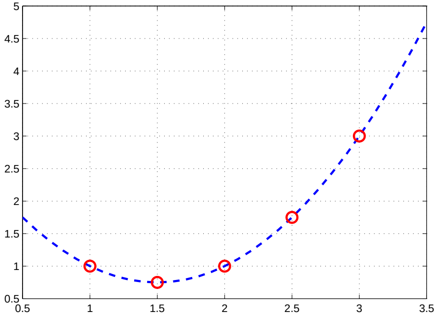
$$X^\top X = \begin{bmatrix} 142.12 & 55 & 22.5 \\ 55 & 22.5 & 10 \\ 22.5 & 10 & 5 \end{bmatrix} \qquad X^\top Y = \begin{bmatrix} 44.62 \\ 17.5 \\ 7.5 \end{bmatrix}$$

$$\hat{\theta}_{LS} = \left(X^\top X\right)^{-1} X^\top Y = \begin{bmatrix} 1 & -3 & 3 \end{bmatrix}^\top$$

Парабола, продолжение

Рассмотрим задачу аппроксимации данных параболой $y = ax^2 + bx + c$ по $N = 5$ точкам: $(1,\ 1)$, $(1.5,\ 0.75)$, $(2,\ 1)$, $(2.5,\ 1.75)$, $(3,\ 3)$.

$$\hat{\theta}_{LS} = \left(X^\top X\right)^{-1} X^\top Y = \begin{bmatrix} 1 & -3 & 3 \end{bmatrix}^\top$$



Содержание

Некоторые понятия из теории вероятности

Пусть $v \in \mathbb{R}^1$ – некоторая стационарная в широком смысле случайная величина.

- Плотность вероятности $\rho(v)$: $\int_{-\infty}^{\infty} \rho(v) dv = 1$.
- Математическое ожидание $\mathbb{E}\{v\} = \int_{-\infty}^{\infty} v\rho(v) dv$. Для N измерений $E_v(N) = \frac{1}{N}\sum_{k=1}^{N} v(k)$, $E_v(N) \to \mathbb{E}\{v\}$ при $N \to \infty$.
- Дисперсия $\mathbb{D}\{v\} = \mathbb{E}\{(v - \mathbb{E}\{v\})^2\} = \sigma_v^2$. Если $v \in \mathbb{R}^m$, то $\mathbb{D}\{v\} \in \mathbb{R}^{m\times m}$ – матрица дисперсий.

Некоторые понятия из теории вероятности, продолжение

Пусть $v \in \mathbb{R}^1$, $\mu \in \mathbb{R}^1$ – некоторые стационарные в широком смысле случайные величины.

- Ковариация: $\operatorname{cov}(v, \mu) = \mathbb{E}\{(v - \mathbb{E}\{v\})(\mu - \mathbb{E}\{\mu\})\}$. Аналог скалярного умножения. $\operatorname{cov}(v, v) = \mathbb{D}\{v\}$. Для двух векторов получается матрица ковариаций.
- Корреляция: $r_{v,\mu} = \operatorname{corr}(v, \mu) = \frac{\operatorname{cov}(v,\mu)}{\sigma_v \sigma_\mu}$. $\operatorname{corr}(v, v) = 1$.
- Функция автокорреляции: $R_v(\tau) = \frac{\mathbb{E}\{(v(k)-\mathbb{E}\{v\})(v(k+\tau)-\mathbb{E}\{v\})\}}{\sigma_v^2}$. $R_v(\tau) = R_v(-\tau)$. $R_v(0) = 1$.

Некоторые понятия из теории вероятности, продолжение

Величины v и μ называются взаимонезависимыми, если $\rho(v|\mu) = \rho(v)$. Для них справедливо:

- $\mathbb{E}\{v\mu\} = \mathbb{E}\{v\}\mathbb{E}\{\mu\}$
- $\mathrm{cov}(v, \mu) = 0$
- $\mathrm{corr}(v, \mu) = 0$

Если значения $v(k_1)$ и $v(k_2)$ являются независимыми для $k_1 \neq k_2$ (не автокореллированный процесс), то

- $\mathrm{cov}(V, V) = \mathbb{D}\{V\} = \sigma_v^2 \mathcal{I}$
- $R_v(\tau) = 0$ для $\tau \neq 0$

Постановка задачи

Дано

Будем полагать, что существует вектор истинных значений $\boldsymbol{\theta}_0$, а результаты измерений подвержены шуму:

$$y(k) = \boldsymbol{x}^\top(k)\boldsymbol{\theta}_0 + v(k),$$

где v – случайная величина.

Постановка задачи

Дано

Будем полагать, что существует вектор истинных значений $\boldsymbol{\theta}_0$, а результаты измерений подвержены шуму:

$$y(k) = \boldsymbol{x}^\top(k)\boldsymbol{\theta}_0 + v(k),$$

где v – случайная величина.

Надо

Требуется по набору доступных измерений Y, X сформировать такую оценку $\hat{\boldsymbol{\theta}}$, что $\mathbb{E}\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta}_0$. При этом желательно иметь малую дисперсию $\hat{\boldsymbol{\theta}}$.

Постановка задачи

Дано

Будем полагать, что существует вектор истинных значений $\boldsymbol{\theta}_0$, а результаты измерений подвержены шуму:

$$y(k) = \boldsymbol{x}^\top(k)\boldsymbol{\theta}_0 + v(k),$$

где v – случайная величина.

Надо

Требуется по набору доступных измерений Y, X сформировать такую оценку $\hat{\boldsymbol{\theta}}$, что $\mathbb{E}\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta}_0$. При этом желательно иметь малую дисперсию $\hat{\boldsymbol{\theta}}$.

Гипотезы

Нужные

H1 $\mathbb{E}\{v\} = 0$.

H2 v – взаимонезависимая со всеми x_i. Например, регрессор $\boldsymbol{x}$ –

Постановка задачи, продолжение

Дано: $y(k) = \boldsymbol{x}^\top(k)\boldsymbol{\theta}_0 + v(k)$.
Надо: $\hat{\boldsymbol{\theta}}$: $\mathbb{E}\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta}_0$.

Гипотезы

Нужные

H1 $\mathbb{E}\{v\} = 0$.

H2 v – взаимонезависимая со всеми x_i. Например, регрессор $\boldsymbol{x}$ – детерминированный.

H3 $\det(X^\top X) \neq 0$

Полезные

H4 $\mathbb{D}\{v\} = \sigma_v^2 = \mathrm{const}$.

H5 $v(k)$ – не автокорелллированный процесс, т.е. отдельные $v(i)$ и $v(j)$ взаимонезависимы при $i \neq j$, $\mathbb{D}\{V\} = \sigma_v^2 \mathcal{I}$.

МНК, Теорема 2

Решение стохастической задачи

Пусть линейная оценка формируется как $\hat{\boldsymbol{\theta}}_{LS} = (X^\top X)^{-1} X^\top Y$. Тогда если

- $\mathbb{E}\{v\} = 0$,
- пара $v, \boldsymbol{x}$ взаимонезависимы и выполняется $\mathbb{E}\{v x_i\} = \mathbb{E}\{v\}\mathbb{E}\{x_i\}$,

то оценка $\hat{\boldsymbol{\theta}}_{LS}$ – *несмещенная*, т.е. $\mathbb{E}\{\hat{\boldsymbol{\theta}}_{LS}\} = \boldsymbol{\theta}_0$.

Более того, если

- $\mathbb{D}\{v\} = \sigma_v^2 = \mathrm{const}$,
- $\mathbb{D}\{V\} = \sigma_v^2 \mathcal{I}$, т.е. $R_v(\tau) = 0$ для $\tau \neq 0$,

то $\hat{\boldsymbol{\theta}}_{LS}$ – *эффективная* оценка, т.е. наилучшая в классе линейных (BLUE - Best Linear Unbiased Estimation).

МНК, Теорема 2

Решение стохастической задачи

Пусть линейная оценка формируется как $\hat{\boldsymbol{\theta}}_{LS} = (X^\top X)^{-1} X^\top Y$. Тогда если

- $\mathbb{E}\{v\} = 0$,
- пара $v, \boldsymbol{x}$ взаимонезависимы и выполняется $\mathbb{E}\{v x_i\} = \mathbb{E}\{v\}\mathbb{E}\{x_i\}$,

то оценка $\hat{\boldsymbol{\theta}}_{LS}$ – *несмещенная*, т.е. $\mathbb{E}\{\hat{\boldsymbol{\theta}}_{LS}\} = \boldsymbol{\theta}_0$.
Более того, если

- $\mathbb{D}\{v\} = \sigma_v^2 = \mathrm{const}$,
- $\mathbb{D}\{V\} = \sigma_v^2 \mathcal{I}$, т.е. $R_v(\tau) = 0$ для $\tau \neq 0$,

то $\hat{\boldsymbol{\theta}}_{LS}$ – *эффективная* оценка, т.е. наилучшая в классе линейных (BLUE - Best Linear Unbiased Estimation).

Линейность

$Y = Y_1 + Y_2 \Rightarrow \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_1 + \hat{\boldsymbol{\theta}}_2$ и $Y = \alpha Y_1 \Rightarrow \hat{\boldsymbol{\theta}} = \alpha \hat{\boldsymbol{\theta}}_1$

Доказательство

Покажем несмещенность оценки. Действительно,

$$\mathbb{E}\{\hat{\boldsymbol{\theta}}_{LS}\} = \mathbb{E}\{(X^\top X)^{-1} X^\top Y\} = \mathbb{E}\{(X^\top X)^{-1} X^\top (X\boldsymbol{\theta}_0 + V)\}$$
$$= \boldsymbol{\theta}_0 + (X^\top X)^{-1}\mathbb{E}\{X^\top V\}.$$

Так как $\boldsymbol{x}$ и v взаимонезависимы и $\mathbb{E}\{v\} = 0$, то $\mathbb{E}\{X^\top V\} = 0$ и $\mathbb{E}\{\hat{\boldsymbol{\theta}}_{LS}\} = \boldsymbol{\theta}_0$.

Доказательство

Покажем несмещенность оценки. Действительно,

$$\begin{aligned}\mathbb{E}\{\hat{\boldsymbol{\theta}}_{LS}\} &= \mathbb{E}\{(X^\top X)^{-1}X^\top Y\} = \mathbb{E}\{(X^\top X)^{-1}X^\top(X\boldsymbol{\theta}_0 + V)\} \\ &= \boldsymbol{\theta}_0 + (X^\top X)^{-1}\mathbb{E}\{X^\top V\}.\end{aligned}$$

Так как $\boldsymbol{x}$ и v взаимонезависимы и $\mathbb{E}\{v\} = 0$, то $\mathbb{E}\{X^\top V\} = 0$ и $\mathbb{E}\{\hat{\boldsymbol{\theta}}_{LS}\} = \boldsymbol{\theta}_0$.
Рассмотрим дисперсию оценки.

$$\begin{aligned}\mathbb{D}\{\hat{\boldsymbol{\theta}}_{LS}\} &= \mathbb{E}\{(\hat{\boldsymbol{\theta}}_{LS} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_{LS} - \boldsymbol{\theta}_0)^\top\} = \mathbb{E}\{(X^\top X)^{-1}X^\top VV^\top X(X^\top X)^{-1}\} \\ &= (X^\top X)^{-1}X^\top\mathbb{E}\{VV^\top\}X(X^\top X)^{-1} = \boxed{\sigma_v^2(X^\top X)^{-1}}.\end{aligned}$$

Доказательство, продолжение

Покажем теперь, что эта дисперсия минимальна (в смысле квадратичной формы) среди всех несмещенных линейных оценок. Рассмотрим некоторую несмещенную линейную оценку $\boldsymbol{\theta} = MY$ с некоторой матрицей M.

Доказательство, продолжение

Покажем теперь, что эта дисперсия минимальна (в смысле квадратичной формы) среди всех несмещенных линейных оценок.
Рассмотрим некоторую несмещенную линейную оценку $\boldsymbol{\theta} = MY$ с некоторой матрицей M.
Так как $\mathbb{E}\{\boldsymbol{\theta}\} = \boldsymbol{\theta}_0$, то $\mathbb{E}\{MX\boldsymbol{\theta}_0\} + \mathbb{E}\{MV\} = \boldsymbol{\theta}_0$ и $MX = \mathcal{I}$. При этом $\mathbb{D}\{\boldsymbol{\theta}\} = \mathbb{E}\{MVV^\top M^\top\} = \sigma_v^2 MM^\top$.

Доказательство, продолжение

Покажем теперь, что эта дисперсия минимальна (в смысле квадратичной формы) среди всех несмещенных линейных оценок.
Рассмотрим некоторую несмещенную линейную оценку $\boldsymbol{\theta} = MY$ с некоторой матрицей M.
Так как $\mathbb{E}\{\boldsymbol{\theta}\} = \boldsymbol{\theta}_0$, то $\mathbb{E}\{MX\boldsymbol{\theta}_0\} + \mathbb{E}\{MV\} = \boldsymbol{\theta}_0$ и $MX = \mathcal{I}$. При этом $\mathbb{D}\{\boldsymbol{\theta}\} = \mathbb{E}\{MVV^\top M^\top\} = \sigma_v^2 MM^\top$.
Рассмотрим разность $Q := \mathbb{D}\{\boldsymbol{\theta}\} - \mathbb{D}\{\hat{\boldsymbol{\theta}}_{LS}\}$ и покажем, что $Q \geq 0$.

$$Q = \sigma_v^2 \left(MM^\top - \mathcal{I}(X^\top X)^{-1}\mathcal{I}^\top \right) = \sigma_v^2 \left(MM^\top - MX(X^\top X)^{-1}X^\top M^\top \right)$$
$$= \sigma_v^2 M \left(\mathcal{I} - P \right) M^\top,$$

где $P := X(X^\top X)^{-1}X^\top$, причем $PP^\top = P = P^\top$.

Доказательство, продолжение

Более того, $(\mathcal{I} - P)(\mathcal{I} - P)^\top = \mathcal{I} - P$. Тогда

$$Q = \sigma_v^2 M (\mathcal{I} - P)(\mathcal{I} - P)^\top M^\top = \sigma_v^2 S S^\top,$$

где $S \in \mathbb{R}^{m \times N}$, $S := M(\mathcal{I} - P)$.

Доказательство, продолжение

Более того, $(\mathcal{I} - P)(\mathcal{I} - P)^\top = \mathcal{I} - P$. Тогда

$$Q = \sigma_v^2 M (\mathcal{I} - P)(\mathcal{I} - P)^\top M^\top = \sigma_v^2 S S^\top,$$

где $S \in \mathbb{R}^{m \times N}$, $S := M(\mathcal{I} - P)$.
Тогда $Q \geq 0$ и

$$\boxed{\mathbb{D}\{\hat{\boldsymbol{\theta}}_{LS}\} \leq \mathbb{D}\{\boldsymbol{\theta}\}}$$

Содержание

Промежуточные итоги

Детерминированная задача

Для заданных X, Y требуется найти наилучшую аппроксимацию в классе линейных моделей $Y = X\boldsymbol{\theta}$ в смысле критерия наименьших квадратов

$$V(\boldsymbol{\theta}) = (Y - x\boldsymbol{\theta})^\top (Y - x\boldsymbol{\theta}).$$

Гипотеза: $\det(X^\top X) \neq 0$. *Решение:*

$$\boxed{\hat{\boldsymbol{\theta}}_{LS} = (X^\top X)^{-1} X^\top Y}$$

минимизирует указанный критерий, при этом

$$V(\hat{\boldsymbol{\theta}}_{LS}) = Y^\top Y - Y^\top X (X^\top X)^{-1} X^\top Y.$$

Промежуточные итоги

Стохастическая задча

Для модели

$$y(k) = \boldsymbol{x}^\top(k)\boldsymbol{\theta}_0 + v(k)$$

требуется сформировать несмещенную оценку вектора параметров $\boldsymbol{\theta}_0$.
Гипотезы: $\mathbb{E}\{v\} = 0$, $(\boldsymbol{x}^\top(k), v(k))$ не коррелируют, $\mathbb{D}\{V\} = \sigma_v^2\mathcal{I}$.
Решение:

$$\boxed{\hat{\boldsymbol{\theta}}_{LS} = (X^\top X)^{-1}X^\top Y}$$

является несмещенной оценкой, т.е. $\mathbb{E}\{\hat{\boldsymbol{\theta}}_{LS}\} = \boldsymbol{\theta}_0$. Если дополнительно $\mathbb{D}\{V\} = \sigma_v^2\mathcal{I}$, то

$$\boxed{\mathbb{D}\{\hat{\boldsymbol{\theta}}_{LS}\} = \sigma_v^2(X^\top X)^{-1}}$$

и оценка $\hat{\boldsymbol{\theta}}_{LS}$ является эффективной.