УНИВЕРСИТЕТ ИТМО

Основы анализа данных.

Метод наименьших квадратов

Лекция 6

КМАИ

Секции

Суть метода наименьших квадратов

Метод наименьших квадратов для линейной функции

Метод наименьших квадратов для квадратичной функции

Матричный вид метода наименьших квадратов

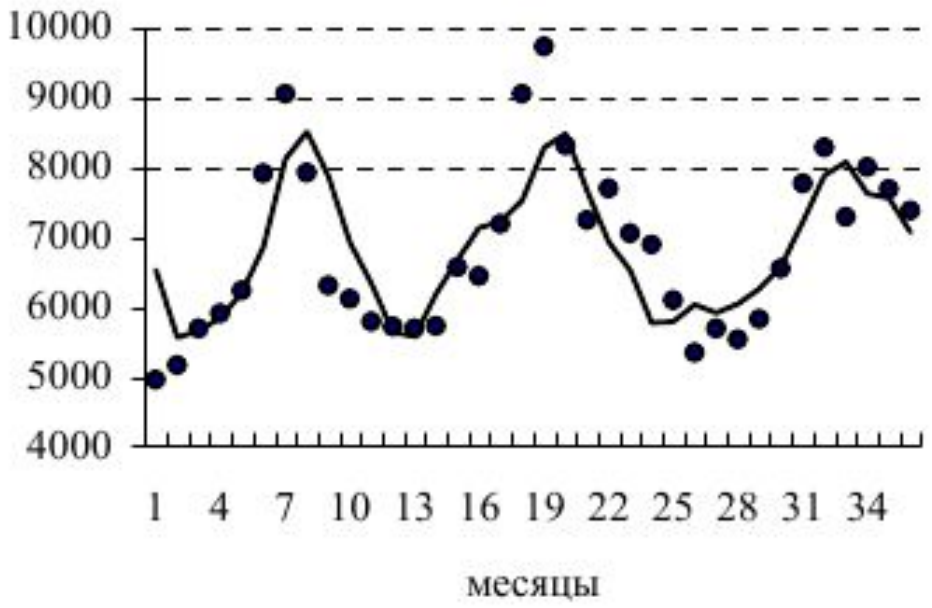
Фильтр Калмана

Метод наименьших квадратов

Наблюдаемая закономерность

$y_1, \dots, y_M$

$x_1, \dots, x_M$

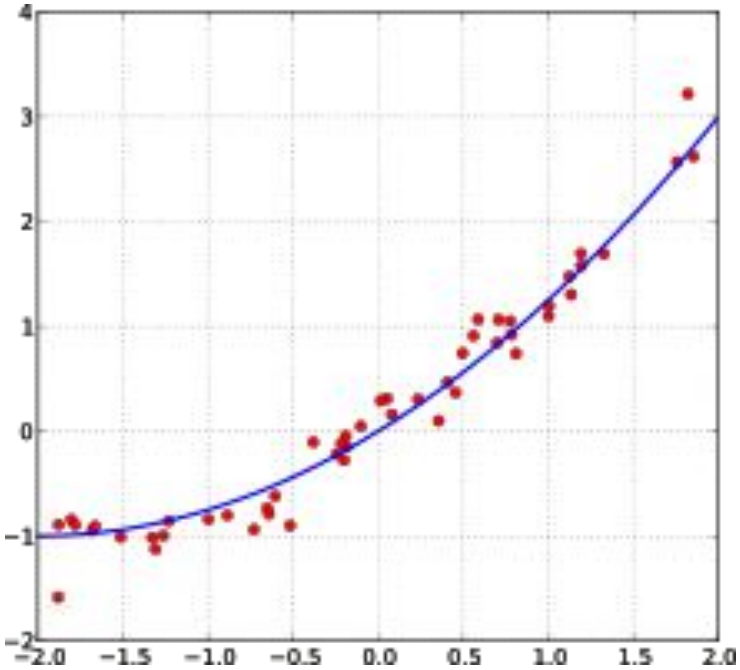


$$y_i = f_i(x, a) + \varepsilon_i$$

- Предполагаемая зависимость

Суть метода наименьших квадратов

Метод наименьших квадратов — математический метод, основанный на минимизации суммы квадратов отклонений некоторых функций от искомых переменных, применяемый для нахождения параметров функции аппроксимации.



$$y_i = f_i(x) + \varepsilon_i$$

$$\sum (y_i - f_i(x))^2 \rightarrow min$$

Функция неправдоподобия

Функция правдоподобия

Функция правдоподобия – функция максимальное значение которой соответствует наилучшему значению параметров интерполяции.

$$\widehat{a_{\text{МП}}} = \arg\max L(x_1, \dots, x_n | a)$$

Условие максимального правдоподобия

$$\frac{dl(x,a)}{da} = 0, \qquad l(x,a) = \ln(L(x_1, \dots, x_n | a))$$

***Теорема**: Не существует другого метода обработки результатов эксперимента, который дал бы лучшее приближение к истине, чем метод максимального правдоподобия.*

$$L = \sum (y_i - f_i(x, a))^2 \rightarrow min$$

$$y_i = f_i(x, a) + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma)$$

$$\frac{dL}{da} = \frac{y - f(x, a)}{\sigma^2} = 0$$

$$y - f(x, a) = 0$$

Секции

Суть метода наименьших квадратов

Метод наименьших квадратов для линейной функции

Метод наименьших квадратов для квадратичной функции

Матричный вид метода наименьших квадратов

Фильтр Калмана

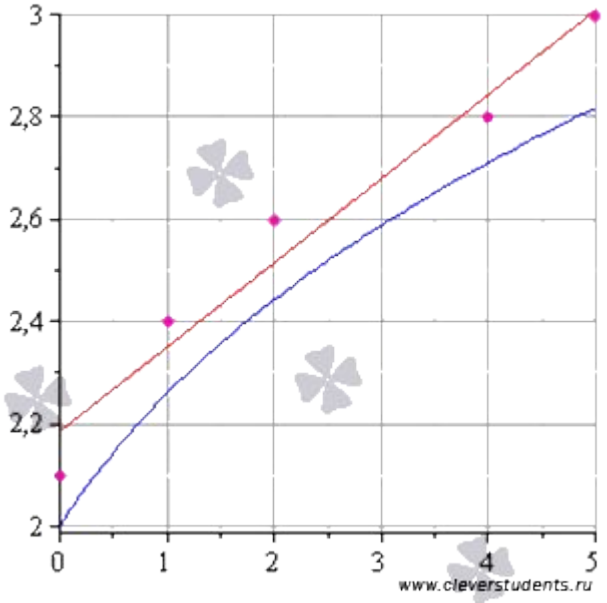
МНК линейной функции

Составляется и решается система из двух уравнений с двумя неизвестными.

$$y_i = f_i(x) + \varepsilon_i \longrightarrow y_i = ax_i + b + \varepsilon_i$$

Функция неправдоподобия:

$$L(a,b) = \sum (y_i - (ax_i + b))^2 \rightarrow min$$



Производная по параметрам:

$$\frac{dL(a,b)}{da} = 0 \quad \Rightarrow \quad -2\sum_{i=1}^{M}(y_i - (ax_i + b))x_i = 0$$

$$\frac{dL(a,b)}{db} = 0 \quad \Rightarrow \quad -2\sum_{i=1}^{M}(y_i - (ax_i + b)) = 0$$

$$a = \frac{n\sum_{i=1}^{M} x_i y_i - \sum_{i=1}^{M} x_i \sum_{i=1}^{M} y_i}{n\sum_{i=1}^{M} {x_i}^2 - (\sum_{i=1}^{M} x_i)^2}$$

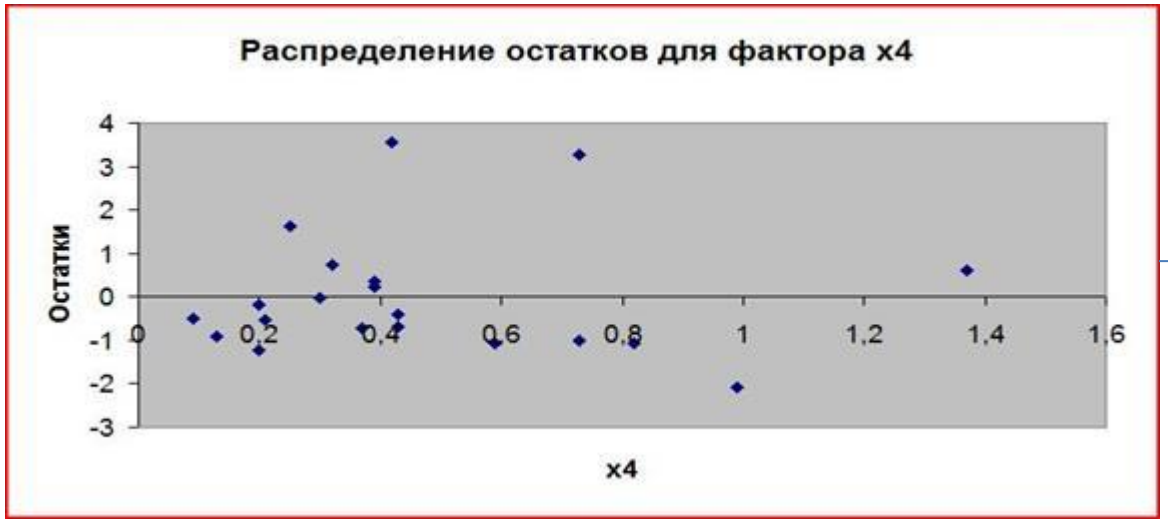
$$b = \frac{\sum_{i=1}^{M} y_i - a\sum_{i=1}^{M} x_i}{n}$$

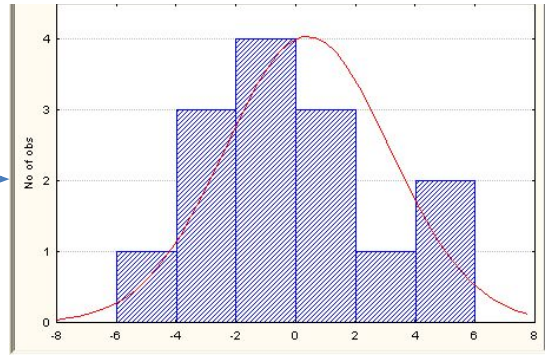
Оценки погрешности МНК

Погрешность МНК:

$$r = \sum (y_i - f_i(x_i)) = \sum (y_i - (ax_i + b))$$

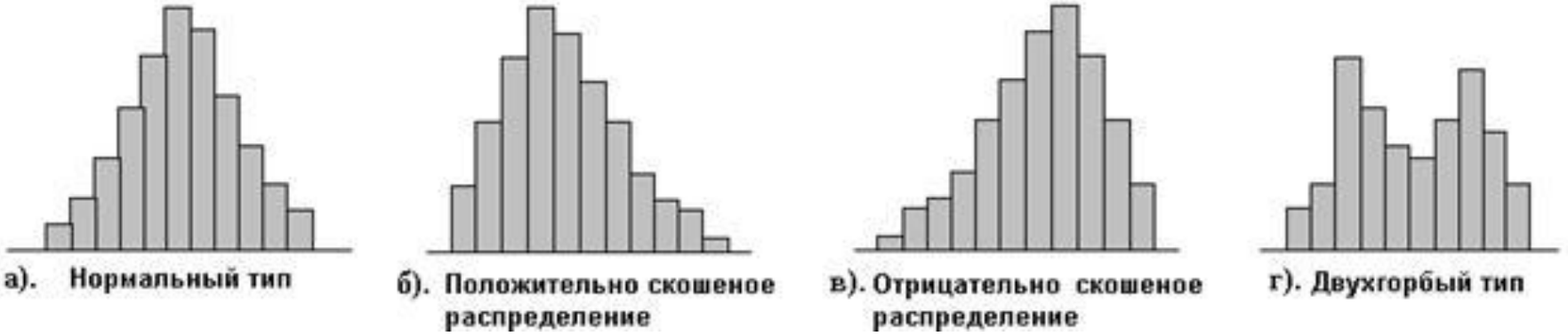
$$r^2 = \sum (y_i - f_i(x_i)^2 = \sum (y_i - (ax_i + b))^2 \rightarrow min$$





Проверка распределения остатков:

1. Визуально.



а). Нормальный тип

б). Положительно скошеное распределение

в). Отрицательно скошеное распределение

г). Двухгорбый тип

1. Сравнивая с табличным значением статистики хи-квадрат.

Секции

Суть метода наименьших квадратов

Метод наименьших квадратов для линейной функции

Метод наименьших квадратов для квадратичной функции

Матричный вид метода наименьших квадратов

Фильтр Калмана

МНК квадратичной функции

Квадратичная апроксимационная функция.

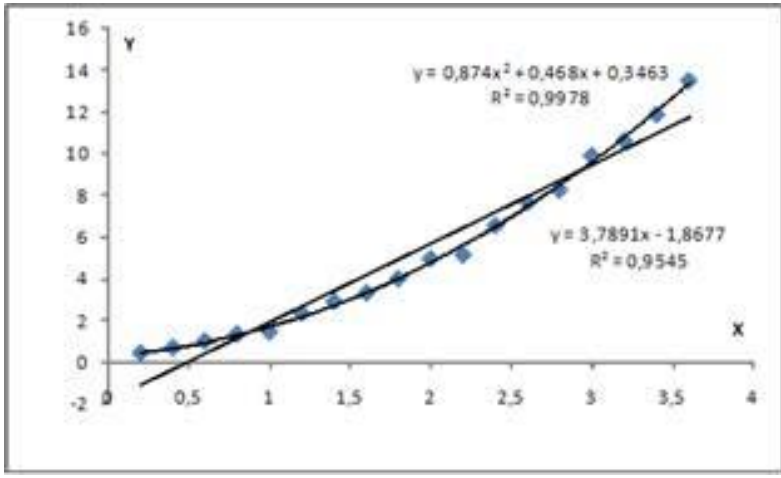
$$y_i = f_i(x) + \varepsilon_i$$

$$y_i = a_1 {x_i}^2 + a_2 x_i + a_3 + \varepsilon_i$$

Функция неправдоподобия:

$$L(a, b) =$$

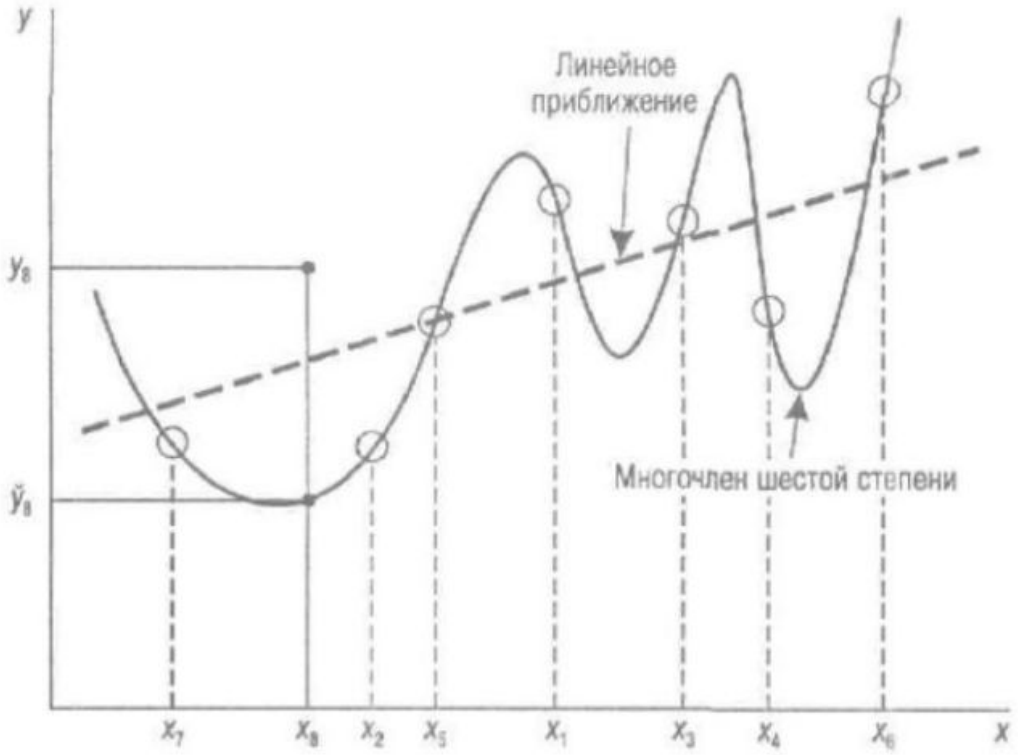
$$= \sum (y_i - (a_1 {x_i}^2 + a_2 x_i + a_3))^2$$



Проблема выбора степени полинома

1. Условие минимизации суммы квадратов отклонений в точке.
2. В других точках может наблюдаться эффект переобучения.
3. Вычислительная сложность.

Решение:
Увеличение количества измерений



Секции

Суть метода наименьших квадратов

Метод наименьших квадратов для линейной функции

Метод наименьших квадратов для квадратичной функции

Матричный вид метода наименьших квадратов

Фильтр Калмана

Матричный вид МНК

$$y_i = a_1 x_i^2 + a_2 x_i + a_3 + \varepsilon_i$$

← Квадратичная апроксимационная функция.

$$y_1 = a_1 x_1^2 + a_2 x_1 + a_3 + \varepsilon_i$$

$$y_2 = a_1 x_2^2 + a_2 x_2 + a_3 + \varepsilon_i$$

$$\dots$$

$$y_M = a_1 x_M^2 + a_2 x_M + a_3 + \varepsilon_i$$

$$Y = AX + U$$

Матричный вид МНК

$$Y = AX + U$$

$$U = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} \quad A = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}^T \quad X = \begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \vdots & \vdots & \vdots \\ x_M^2 & x_M & 1 \end{bmatrix}^T$$

Матричный вид МНК

$$Y = AX + U$$

$$U \to min$$

$$(Y - AX)^T (Y - AX) \to min$$

$$(X^T X)A = X^T Y \longrightarrow A = (X^T X)^{-1} X^T Y$$

Обобщенный МНК

$$(X^T X)A = X^T Y$$

Статистически независимые наблюдения.

Ковариация оценок наблюдений:

$$V(A) = \sigma^2 (X^T X)^{-1}$$

Ковариация ошибок наблюдений:

$$V(U) = \sigma^2 (U^T U)^{-1} = \sigma^2 I \longrightarrow \mathrm{W} = V(U)^{-1}$$

$$A = (X^T W X)^{-1} X^T W Y$$

Секции

Суть метода наименьших квадратов

Метод наименьших квадратов для линейной функции

Метод наименьших квадратов для квадратичной функции

Матричный вид метода наименьших квадратов

Фильтр Калмана

Матричный вид МНК

Фильтр Калмана — эффективный рекурсивный фильтр, оценивающий вектор состояния динамической системы, используя ряд неполных и зашумленных измерений.

Предположения фильтра Калмана:

- гауссовые априорные и апостериорные плотности вероятности вектора состояния на любой момент времени (в том числе начальный)
- гауссовые формирующие шумы
- гауссовые шумы наблюдений
- белые шумы наблюдений
- линейность модели наблюдений
- линейность модели формирующего процесса

Вопросы

1. Определение МНК и функция правдоподобия.
2. МНК линейной функции
3. МНК квадратичной функции
4. Матричный вид МНК