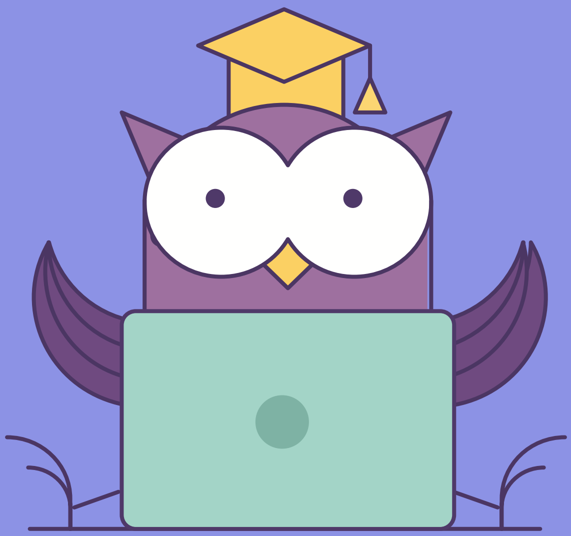
OTUS

ОНЛАЙН-ОБРАЗОВАНИЕ

Меня хорошо слышно && видно?



Напишите в чат, если есть проблемы!

Ставьте + если все хорошо

МНК

Метод наименьших квадратов.

Лоскутов Антон Алексеевич

Знакомство

OTUS

- Закончил механико-математический факультет МГУ им. Ломоносова
- Учился в Техносфере от Mail.Ru Group
- Являюсь ментором в Техносфере
- Работаю программистом-исследователем в Mail.Ru Group
- Веду лекции открытого курса mlcourse.ai

Правила вебинара

OTUS

Активно участвуем

Задаем вопросы в чат

Off-topic обсуждаем в Slack

Вопросы вижу в чате, могу ответить не сразу

Цели занятия

После занятия вы сможете:

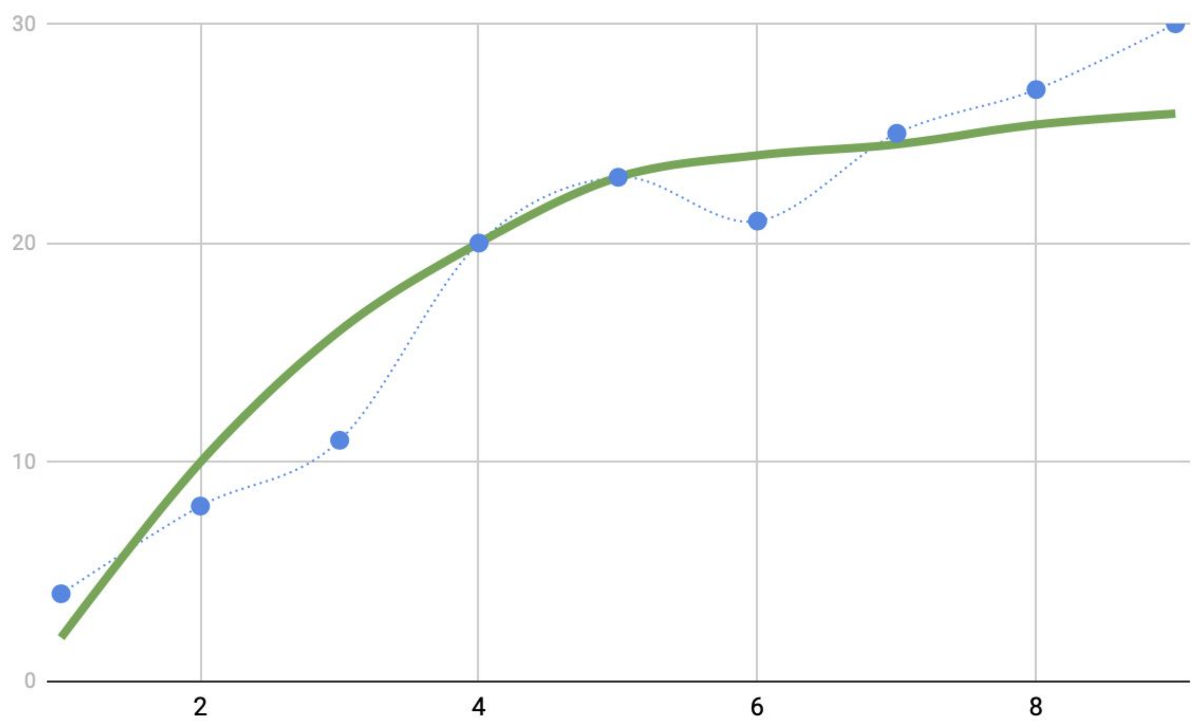
1. Применять методы оптимизации на практике.
2. Найти лучшую аппроксимацию аналитически.
3. Легко аппроксимировать почти любые функции.

Метод наименьших квадратов

OTUS

Пусть у нас есть некая зависимость одной переменной от другой

X	1	2	3	4	5	6	7	8	9
Y	4	8	11	20	23	21	25	27	30



Метод наименьших квадратов

Определение. Пусть $f(x)$ - аппроксимирующая функция для набора точек $(x_i; y_i)$. Тогда ошибками будет называть $e_i = y_i - f(x_i)$.

Задача. Давайте оценивать аппроксимирующие функции с помощью ошибок.

Проблема. Как именно с помощью ошибок можно оценивать?

Варианты:

- **Простая сумма:** $e(x) = e_1 + \ldots + e_n$ Слагаемые могут сократиться между собой
- **Сумма модулей:** $e(x) = |e_1| + \ldots + |e_n|$ Лучше подходит при ненормальном распределении ошибок
- **Сумма квадратов:** $e(x) = e_1^2 + \ldots + e_n^2$ Лучше подходит при нормальном и равномерном распределении ошибок
- **Сумма больших степеней:** $e(x) = e_1^{10} + \ldots + e_n^{10}$ Сложно вычислять и слишком сильно “наказываем” за большие ошибки

В прикладных задачах чаще встречается нормальное распределение

Метод наименьших квадратов

OTUS

Определение. Пусть задана такая зависимость: $y_t = f(x_t, b) + \varepsilon_t$, где ϵ_t - случайная ошибка модели и b - набор неизвестных параметров. Надо восстановить изначальную зависимость y от x. Для этого подберем параметры b наилучшим образом.

Определение. Введем функцию “ошибки”, с помощью которой будем оценивать параметры b

$$RSS(b) = e^T e = \sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n} (y_t - f(x_t, b))^2$$

Задача. Найти $\hat{b}_{OLS} = \arg\min_b RSS(b)$

Решение. Задачу можно решить с помощью методов оптимизации, а можно попытаться решить аналитически. Большинство задач можно решить аналитически, так что будем разбирать этот метод:

$$\sum_{t=1}^{n} (y_t - f(x_t, b)) \frac{\partial f(x_t, b)}{\partial b} = 0.$$

Метод наименьших квадратов (линейная модель)

Определение. Пусть задана линейная зависимость

$$y_t = \sum_{j=1}^{k} b_j x_{tj} + \varepsilon = x_t^T b + \varepsilon_t \quad <-> \quad y = Xb + \varepsilon.$$

Функциональное представление — Матричное представление

Определение. Функция ошибки в матричном представлении имеет вид

$$RSS = e^T e = (y - Xb)^T (y - Xb)$$

Если продифференцировать по вектору параметров b и приравняем производную к нулю, получаем

$$(X^T X)b = X^T y.$$

Метод наименьших квадратов (линейная модель)

OTUS

Попробуем сделать тоже самое, но в функциональном виде

$$((y_t - \sum_{j=1}^{k} b_j x_{tj} - \epsilon)^2)' = -2 \sum_{j=1}^{k} x_{tj}(y_t - \sum_{j=1}^{k} b_j x_{tj}) = 0$$

В расшифрованной матричной форме это будет выглядеть так

$$\begin{pmatrix} \sum x_{t1}^2 & \sum x_{t1}x_{t2} & \sum x_{t1}x_{t3} & \dots & \sum x_{t1}x_{tk} \\ \sum x_{t2}x_{t1} & \sum x_{t2}^2 & \sum x_{t2}x_{t3} & \dots & \sum x_{t2}x_{tk} \\ \sum x_{t3}x_{t1} & \sum x_{t3}x_{t2} & \sum x_{t3}^2 & \dots & \sum x_{t3}x_{tk} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{tk}x_{t1} & \sum x_{tk}x_{t2} & \sum x_{tk}x_{t3} & \dots & \sum x_{tk}^2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_k \end{pmatrix} = \begin{pmatrix} \sum x_{t1}y_t \\ \sum x_{t2}y_t \\ \sum x_{t3}y_t \\ \vdots \\ \sum x_{tk}y_t \end{pmatrix}$$

Получаем общую формулу для МНК-оценок линейной модели

$$\hat{b}_{OLS} = (X^T X)^{-1} X^T y = \left(\frac{1}{n} X^T X\right)^{-1} \frac{1}{n} X^T y$$

Пример метода наименьших квадратов

OTUS

1. Найдите МНК-оценку для парной линейной регрессии

$$y_t = a + bx_t + \varepsilon_t$$

Получаем систему уравнений

$$\begin{pmatrix} n & \sum_{t=1}^n x_t \\ \sum_{t=1}^n x_t & \sum_{t=1}^n x_t^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^n y_t \\ \sum_{t=1}^n x_t y_t \end{pmatrix}$$

Получаем решение для коэффициентов

$$\begin{cases} \hat{b} = \dfrac{n\sum_{t=1}^n x_t y_t - (\sum_{t=1}^n x_t)(\sum_{t=1}^n y_t)}{n\sum_{t=1}^n x_t^2 - (\sum_{t=1}^n x_t)^2} \\ \hat{a} = \dfrac{\sum_{t=1}^n y_t - \hat{b}\sum_{t=1}^n x_t}{n}. \end{cases}$$

Пример метода наименьших квадратов

1. Найдите МНК-оценку для парной линейной регрессии

$$y_t = a + bx_t + \varepsilon_t$$

2. Найдите МНК-оценку для линейной регрессии

$$y_t = bx_t + \epsilon_t$$

Получаем уравнение

$$\left(\sum x_t^2\right) b = \sum x_t y_t$$

Получаем решение

$$\hat{b} = \frac{\sum_{t=1}^{n} x_t y_t}{\sum_{t=1}^{n} x_t^2}$$

Больше про МНК с примерами и статистическими свойствами можно почитать тут:
Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика. страница 34 и далее.
http://math.isu.ru/ru/chairs/me/files/books/magnus.pdf

Есть вопросы или замечания?

Напишите в чат свои вопросы и замечания!

Ставьте + если все понятно

Антон Лоскутов

Slack: @LoskutovAnton

Спасибо за внимание!