Лабораторная работа № 1. МНК

Цель работы.

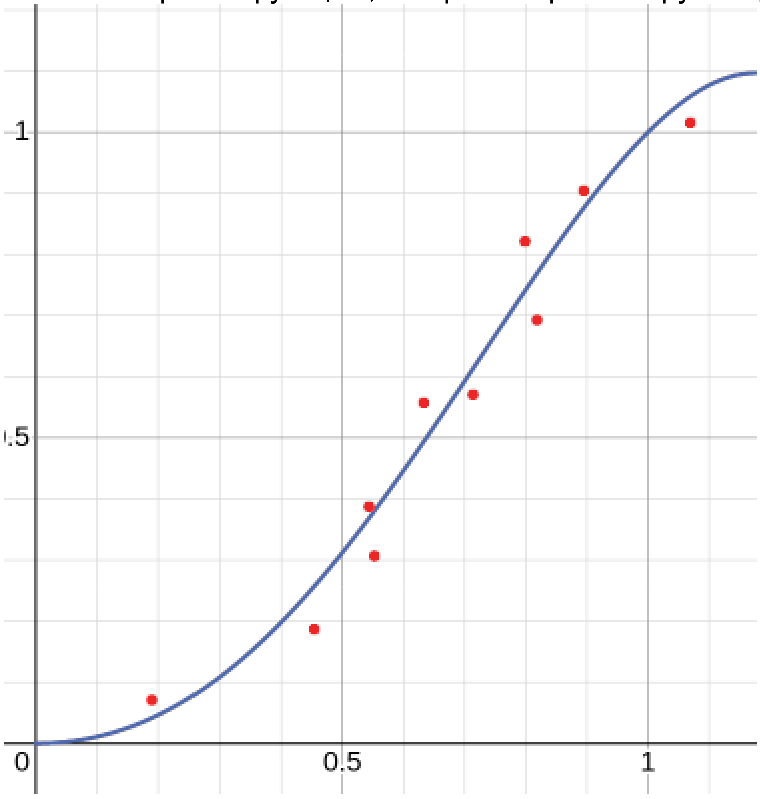
Изучить метод наименьших квадратов и реализовать его.

Общие положения.

Для выполнения работы вам необходимо знать метод наименьших квадратов. В работе требуется написать программу на языке программирования Python. При написании программы нельзя использовать встроенную в numpy функцию numpy.linalg.lstsq. Проверка работы происходит через автоматизированную систему moodle.

Решаемая задача:

Дан некоторый набор (x_i, y_i) точек на плоскости.
Хотим построить функцию, которая аппроксимирует заданный набор точек.



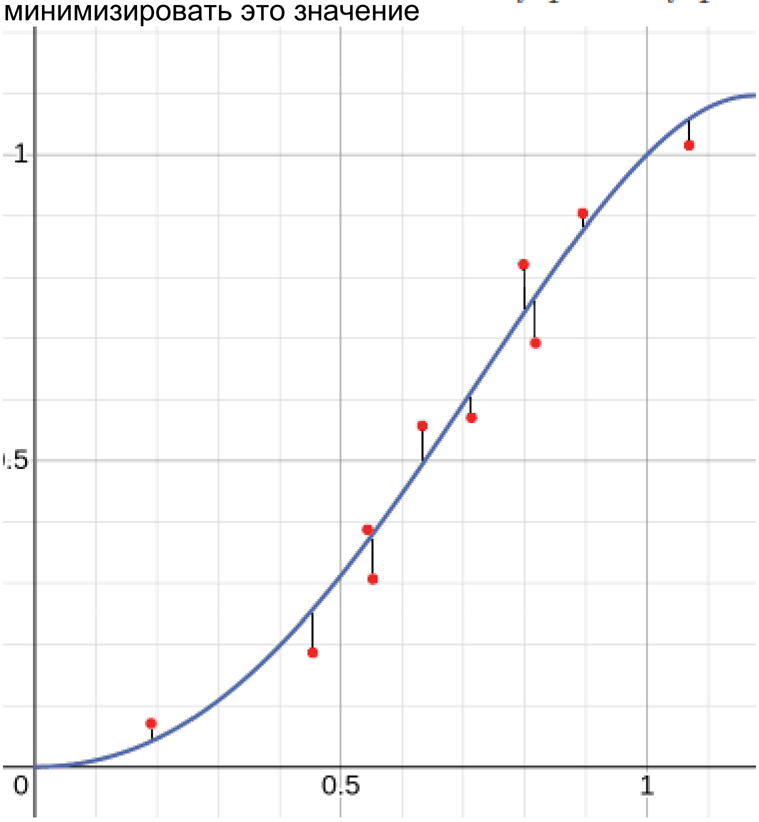
МНК идея:
Ищем прямую, у которой сумма квадратов расстояний до заданных точек наименьшая.

МНК описание общего решения:
Пусть аппроксимирующая функция будет представлена в виде:
$y = f(x, b_0, b_1, \dots, b_n)$, где f – известная функция(многочлен), а $b_0, b_1, \dots, b_n$ – неизвестные параметры функции.
Для решения задачи необходимо найти такие значения параметров функции, чтобы значения исследуемой функции в точках x_i примерно совпадали со значениями аппроксимирующей функции.
Пусть $\varepsilon_i = f(x_i, b_0, b_1, \dots, b_n) - y_i, i = 1, 2, \dots, m$ – ошибка (отклонение) значений исходной и исследуемой функций в точке x_i.

Сумма квадратов ε_i будет величиной, показывающей меру отклонения f от известных значений y_i, обозначим её как S. Чтобы найти аппроксимирующую функцию, надо минимизировать это значение

$$S = \sum_{i=1}^{m} \varepsilon_i^2 = \sum_{i=1}^{m} (f(x_i) - y_i)^2 \to min$$

.



Для этого потребуется найти его производную и приравнять к нулю.

Величина S зависит от переменных функции $b_0, b_1, \ldots, b_n$, они независимы, поэтому для оценки производной мы можем оценить частные производные по этим переменным.

$$\frac{\partial S}{\partial b_0} = 0; \frac{\partial S}{\partial b_1} = 0; \ \ldots; \ \frac{\partial S}{\partial b_n} = 0$$

Из полученных соотношений можно составить систему уравнений для определения значений $b_0, b_1, \ldots, b_n$.

МНК пример матричной реализации с линейным случаем:

Пусть даны (x_i, y_i) - координаты точек из заданного набора

$y = ax + b$ – прямая, которую мы хотим найти

нам нужно минимизировать сумму квадратов расстояний, n - количество точек

$$\sum_{i=1}^{n} (f(x_i) - y_i)^2 \to min$$

в идеале у нас должна быть верна система уравнений

$$\begin{cases} ax_1 + b = y_1 \\ ax_2 + b = y_2 \\ ax_3 + b = y_3 \\ \dots \\ ax_n + b = y_n \end{cases}$$

запишем систему в матричном виде

$$A = \begin{bmatrix} x_1 \; 1 \\ x_2 \; 1 \\ \dots \\ x_n \; 1 \end{bmatrix}$$

$$x = \begin{bmatrix} a \\ b \end{bmatrix}$$

$$b = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}$$

тогда $Ax = b$.

но на самом деле $Ax = b + e$, e - ошибка(погрешность)

таким образом нам нужно минимизировать погрешности, а точнее их квадраты

т.к e - вектор, то $< e, e > \to min$ при скалярном произведении получается число, которое нужно минимизировать

$$e = Ax - b$$
$$e^T e = (Ax - b)^T (Ax - b)$$

нужно оценить производную

$$\frac{d\varepsilon^T \varepsilon}{dx} = 0$$

(найти те x, которые зануляют эту производную)
преобразуем

$$\varepsilon^T \varepsilon = (Ax - b)^T (Ax - b) = \left((Ax)^T - b^T\right)(Ax - b) = \left(x^T A^T - b^T\right)(Ax - b) =$$
$$= x^T A^T Ax - x^T A^T b - b^T Ax + b^T b =$$
$$= x^T A^T Ax - 2x^T A^T b + b^T b$$

$$\boxed{\begin{gathered} (A*x)^T = x^T*A^T \\ b^T*(A*x) = (A*x)^T*b = x^T*A^T*b \end{gathered}}$$

применяем матричное дифференцирование

(производная произведения в первом слагаемом):

$$\frac{d\varepsilon^T\varepsilon}{dx}=\left(A^TA+\left(A^TA\right)^T\right)x-2A^Tb=\left(A^TA+A^TA\right)x-2A^Tb=2A^TAx-2A^Tb$$

$$\frac{d\varepsilon^T\varepsilon}{dx}=0$$

$$2A^TAx-2A^Tb=0$$

$$A^TAx=A^Tb$$

$$x=\left(A^TA\right)^{-1}A^Tb$$

И вот это решение X и есть решение по МНК.

Порядок выполнения.

Дан набор точек, который получен зашумлением некоторой прямой $y = ax + b$ с исходными коэффициентами.
Используя метод наименьших квадратов, вам необходимо реализовать функцию solution, которая будет возвращать исходные коэффициенты прямой.

Контрольные вопросы и задачи.

1. Методом наименьших квадратов посчитайте расстояние между функцией $y = x^2 + 1$ и набором точек

i	1	2	3	4	5	6	7	8	9	10
xi	-0.5	0.81	0.99	0.6	5	2.9	1.4	3.2	0.01	2.13
yi	-3.67	2.99	1.83	3.95	2.11	4.15	0.23	2.22	1.01	4.56

2. Назовите основное отличие интерполяции точек от их аппроксимации.
3. Для чего нужен resampling?

Дополнительные материалы, информационные источники.

Информация об интерполяции, аппроксимации и методе наименьших квадратов:
https://portal.tpu.ru/SHARED/m/MBB/uchebnaya_rabota/Model/Tab/Interp_app.pdf