МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ Н. Э. БАУМАНА

Факультет информатики и систем управления
Кафедра теоретической информатики и компьютерных технологий

Лабораторная работа №2
по курсу «Численные методы»

# «Аппроксимация функций методом наименьших квадратов»

**Выполнил:**
Студент группы ИУ9-62
Ситников К.А.

**Проверила:**
Домрачева А. Б.

Москва 2017

## Постановка задачи

Дана функция $f(x)$, определенная для набора точек $\{x_i\}_{i=1}^n$, и имеющая значения $\{y_i\}_{i=1}^n$ в этих точках:

$$y_i = f(x_i), i = 1 \ldots n$$

Необходимо аппроксимировать функцию $f(x)$ линейной функцией $g(x) = ax + b$ с помощью метода наименьших квадратов:

$$y_i \approx g(x_i), i = 1 \ldots n$$

## Теоретические сведения

Метод наименьших квадратов - метод, основанный на минимизации суммы квадратов отклонений некоторой функции от искомой переменной:

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

В случае аппроксимации линейной функцией соотношение (1) приобретает следующий вид:

$$\sum_{i=1}^{n} (y_i - ax_i - b)^2 \to min \tag{2}$$

Задача поиска экстремума заменяется эквивалентной задачей поиска экстремума исходной функции (равенства нулю производной):

$$\begin{cases} \frac{\partial f(a,b,x_i)}{\partial a} = 0 \\ \frac{\partial f(a,b,x_i)}{\partial b} = 0 \end{cases} \tag{3}$$

Подставим функцию из уравнения (2) в систему уравнений (3):

$$\begin{cases} -2x_i \sum_{i=1}^{n} (y_i - ax_i - b) = 0 \\ -2 \sum_{i=1}^{n} (y_i - ax_i - b) = 0 \end{cases}$$

$$\begin{cases} a \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \\ a \sum_{i=1}^{n} x_i + nb = \sum_{i=1}^{n} y_i \end{cases} \tag{4}$$

Оба уравнения в системе (4) линейны относительно $a, b$, поэтому оно имеет решение в виде:

$$b = \frac{D - aB}{n}$$

$$a = \frac{C - DB/n}{A - B^2/n}$$

где $A = \sum_{i=1}^{n} x_i^2, B = \sum_{i=1}^{n} x_i, C = \sum_{i=1}^{n} x_i y_i, D = \sum_{i=1}^{n} y_i$

## Реализация

Далее приведена простая реализация метода наименьших квадратов на языке Python. При задании дискретной области определения и области значения функции добавляется возмущение $\pm 10^{-1}$

```
import math
import random

i = 0
seq = [1, -1]
measurements = []

while i < 10:
    measurements.append((i +  random.choice(seq) * random.random()/10,
    i +  random.choice(seq) * random.random()/10))
    print(measurements[i])
    i += 1

A = 0.0
B = 0.0
C = 0.0
D = 0.0

for m in measurements:
    A += m[0] * m[0]
    B += m[0]
    C += m[0] * m[1]
    D += m[1]

a = (C - (B * D) / len(measurements)) / (A - (B * B) / len(measurements))
b = (D - B * a) / len(measurements)
print(a , b)
```

## Результаты

При тестировании программы в качестве входных данных использовались следующие данные:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| y | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

При задании дискретной области определения и области значения функции добавляется возмущение $\pm 10^{-1}$, имитирующее инструментальную погрешность (случай, когда точное значение в точках не может быть получено).

При подсчете коэффициентов $ab$ возникает вычислительная погрешность, связанная с ограничением разрядности чисел с плавающей точкой.

Значения вычисленные программой должны быть 1 и 0 для $ab$ соответственно, так как $f(x) = 1 * x + 0 = x$.

**Тест.** Входные данные после добавления возмущения:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| x | 0.01989 | 0.9774 | 1.9527 | 2.9707 | 3.9287 | 5.076 | 6.0244 | 6.9916 | 7.9273 | 9.0629 |
| y | 0.0818 | 0.9543 | 2.0223 | 2.9345 | 4.0969 | 4.9445 | 6.0230 | 7.0197 | 8.0582 | 8.9972 |

**Результат работы программы:**

$$y = 0.996725160550055x + 0.034753069120225175$$

$$a = 0.996725160550055$$

$$b = 0.034753069120225175$$

Инструментальная погрешность:

$$\Delta a = 0.996725160550055$$

$$\Delta b = 0.034753069120225175$$

## Вывод

В результате выполнения данной лабораторной работы был изучен метод наименьших квадратов,также он был реализован для случая линейной функции на языке Python. Было выяснено, что на итоговый результат влияют инструментальная и вычислительная погрешности.