

ИТМО

Отчет

По лабораторной работе №1

Аппроксимация по методу наименьших квадратов

Выполнил:
Рахматов Нематджон
Р3233

Преподаватель:
Перл Ольга
Вячеславовна

Санкт-Петербург
2024

Оглавление

Задача.....	2
Описание Метода.....	2
Блок-Схема Алгоритма.....	5
Код Программы.....	6
Примеры работ программы.....	7
Вывод.....	12

Задача

Дан набор точек, по которым необходимо построить линейную аппроксимацию по методу наименьших квадратов. Необходимо найти значение квадрата наибольшего отклонения среди заданных точек относительно полученной линейной аппроксимации. [Обратите внимание, что при защите лабораторной работы на занятии, желательно иметь несколько аппроксимирующих функций, кроме линейной.]

Формат входных данных:

x1 x2 x3 ...

y1 y2 y3 ...

где x1...xn - список значений аргумента для узлов интерполяции, y1...yn - список значений функции для соответствующего значения аргумента для узлов интерполяции. В тестах также вначале задаётся количество задаваемых точек, однако, в функцию этот параметр не передаётся.

Формат выходных значений: вещественное число, являющееся значением квадрата наибольшего отклонения исходных данных от полученной линейной аппроксимации.

Описание Метода

Метод наименьших квадратов (МНК) представляет собой численный метод, используемый для поиска оптимальных параметров линейной аппроксимации данных. Этот метод минимизирует сумму квадратов расстояний между наблюдаемыми значениями и значениями, предсказанными линейной функцией.

Постановка задачи:

Даны наборы точек (x, y), где x - аргумент, y - значение функции. Цель - найти линейную функцию вида $y = ax + b$, которая наилучшим образом соответствует этим точкам.

Минимизация суммы квадратов отклонений:

МНК стремится минимизировать сумму квадратов отклонений между наблюдаемыми значениями y и значениями, предсказанными аппроксимацией. Функционал ошибки определяется как:

$$E(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

где (x_i, y_i) - координаты i-ой точки, a - коэффициент наклона, b - свободный член.

Вычисление коэффициентов аппроксимации:

Коэффициенты a и b могут быть найдены аналитически с использованием следующих формул:

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

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

где n - количество точек.

Вычисление аппроксимированных значений:

Получив коэффициенты a и b, мы можем вычислить аппроксимированные значения y' для каждой точки:

$$y' = ax + b$$

Оценка точности аппроксимации:

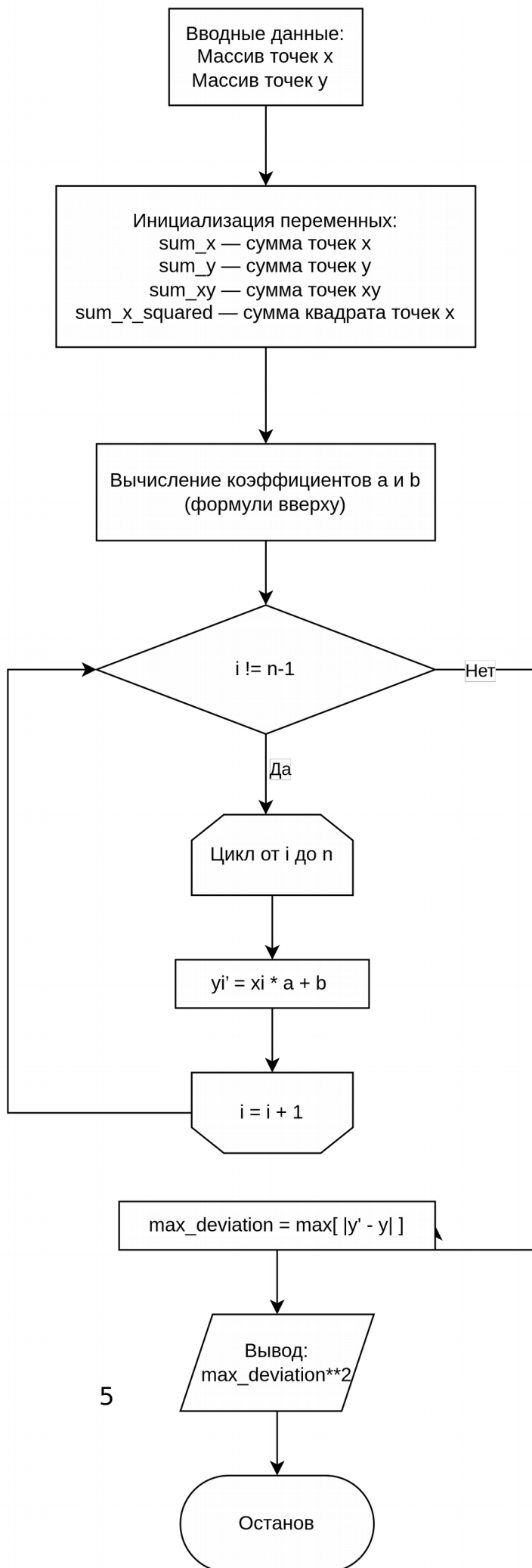
Максимальное отклонение между исходными значениями y и аппроксимированными значениями y' используется как мера точности аппроксимации:

$$\text{Макс. отклонение} = \max |y - y'|$$

Возвращается квадрат этого значения, поскольку часто интерес представляет не само отклонение, а его квадрат.

Метод наименьших квадратов широко применяется в различных областях, таких как статистика, регрессионный анализ и обработка экспериментальных данных. Важно отметить, что этот метод чувствителен к выбросам, поэтому в некоторых случаях может потребоваться использование других методов аппроксимации.

Блок-схема алгоритма



Код программы

```
# Main Function
def approximate_linear_least_squares(x_axis, y_axis):
    n = len(x_axis)
    sum_x = sum(x_axis)
    sum_y = sum(y_axis)
    sum_xy = sum(x * y for x, y in zip(x_axis, y_axis))
    sum_x_squared = sum(x ** 2 for x in x_axis)

    a = (n * sum_xy - sum_x * sum_y) / (n * sum_x_squared - sum_x ** 2)
    b = (sum_y - a * sum_x) / n

    y_approximated = [a * x + b for x in x_axis]

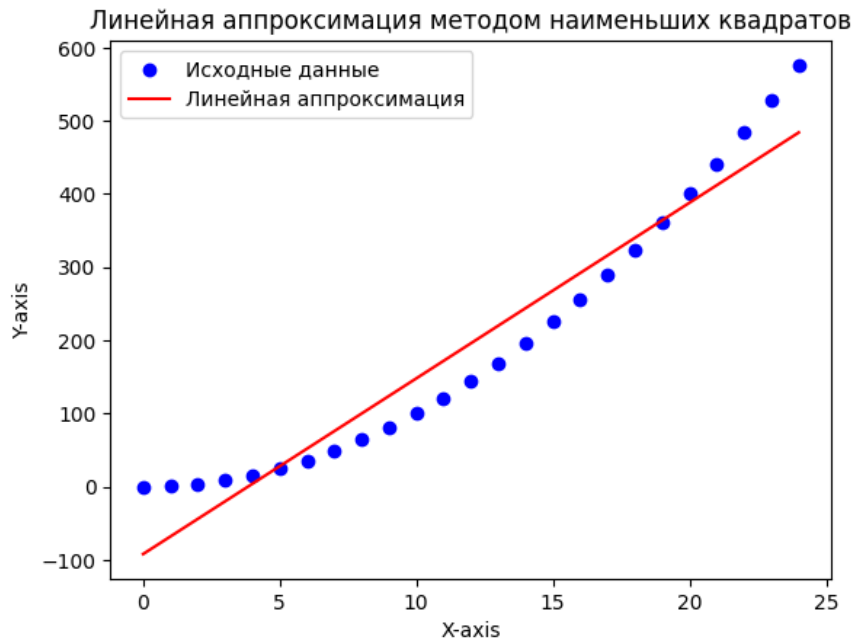
    max_deviation = max(abs(y - y_approx) for y, y_approx in zip(y_axis,
y_approximated))

    return max_deviation ** 2
```

Примеры работ:

1. $n = 25$, $x = [0:25)$, $y = [i*i \text{ for } i \text{ in range}(n)]$

Ответ:



$\text{max_deviation} = 8464$

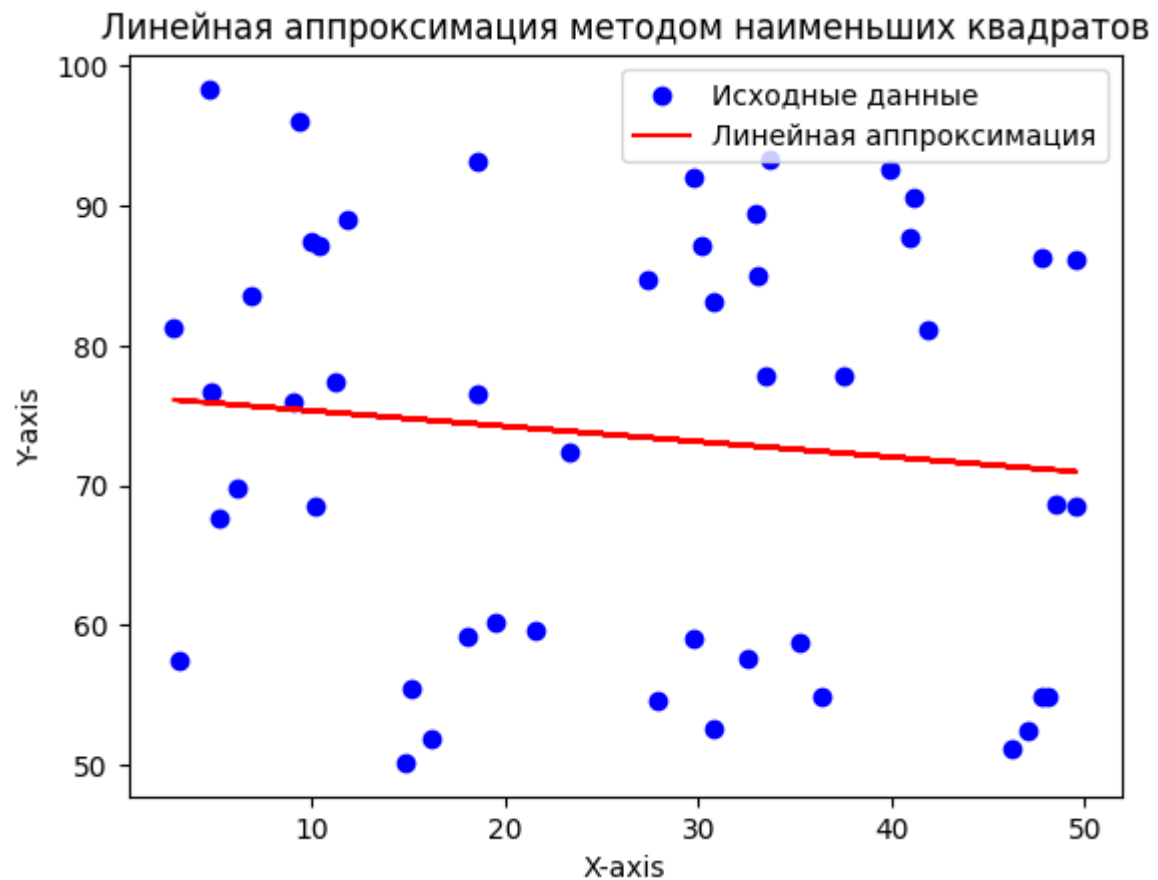
В этом примере видно что отклонение мало, 4 точки попали правильно. Но это аппроксимация поэтому мало точек. Этот метод больше всего предназначен для предикшенов и найти закономерности точек.

2. Рандомные точки.

$n = 50$, x = случайные значения в промежутке $[1, 50)$,

y = случайные значения в промежутке $[50, 100)$

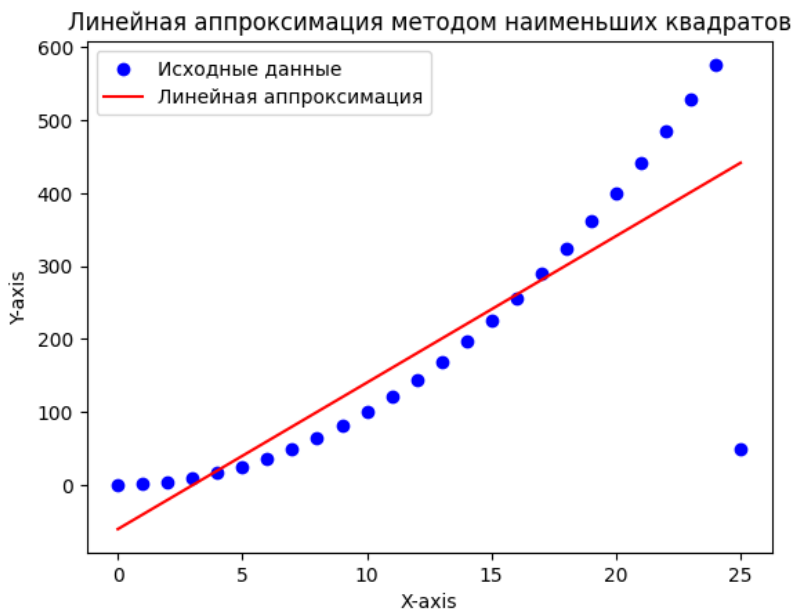
Ответ:



max_deviation = 606.97

В этот примере видно что линия в середине, это потому что между случайных точек нету не каких связи и поэтому оно догоняет модель ближе к середине.

3. Добавляем новую точку в дата сет и проверяем. Точку (25, 50)
Ответ:



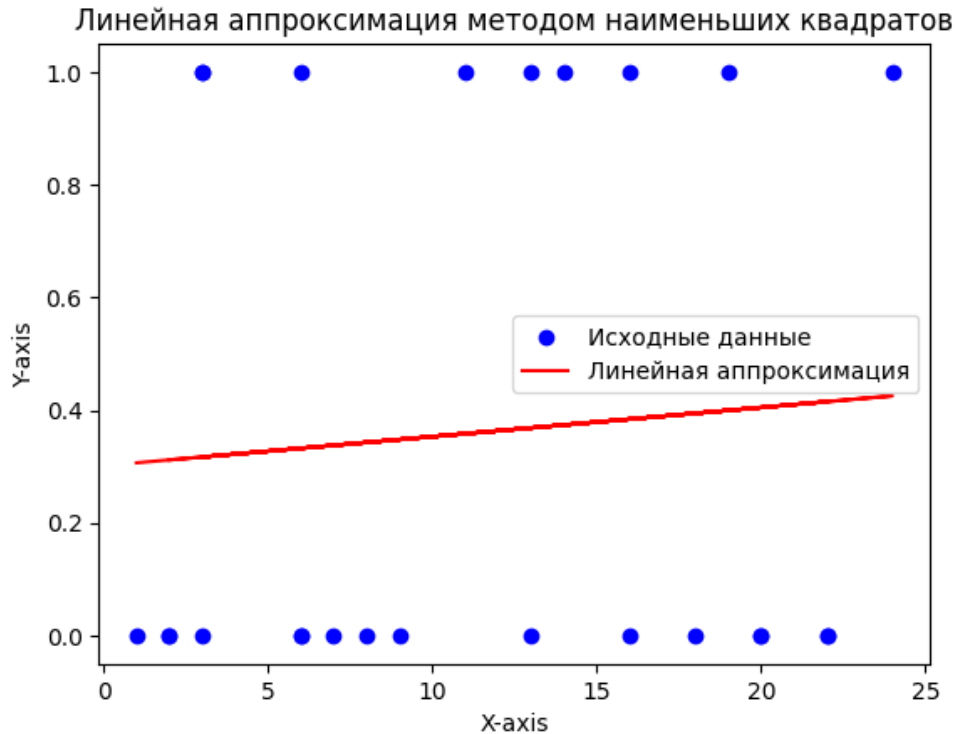
$\text{max_deviation} = 153235.445$

Тут можно видеть что при добавление новой точки которая не имеет связи с другими точками результат увеличивался. Это изза того что это точка рандомная и не имеет связи с прежнемы точками.

4. Если y будет всегда $[0,1]$

Ответ:

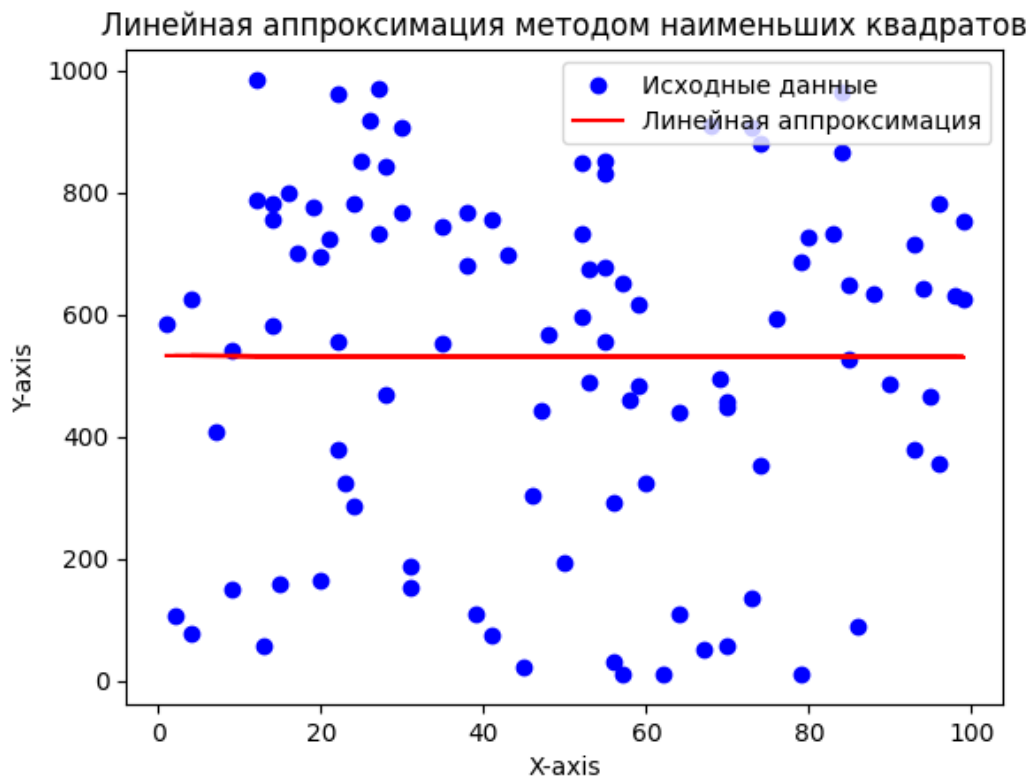
$\max_deviation = 0.466$



Видим что ошибок модел старается близится где там много точек. К примеру в начале оно близко к 0 потому что в этом промежутке много 0. И патом растет и приближается к 1. Это говорит нам что если будет всегда 1 то и модел будет всегда вывести 1. Если на половину то в середине. Ещё это нам говорит что чем много данных есть тем и аппроксимация точнее будет.

5. $n = 100$, x = случайные значения в промежутке $[0, 100)$,
 y = случайные значения в промежутке $[0, 1000)$

Ответ:



$\text{max_deviation} = 270690.063$

В этом примере плотно разбросаны точки и видим что линия прямая и находится немножко вверх от середине. Все точки случайные поэтому прямая линия. Но направлена где точек много.

Вывод

В результате выполнения лабораторной работы по аппроксимации методом наименьших квадратов были получены следующие выводы. Метод успешно применен к различным наборам данных, представляющим собой точки в пространстве. Результаты подтверждают, что алгоритм корректно вычисляет коэффициенты линейной аппроксимации, предоставляя адекватные значения на выходе. При сравнении с другими методами аппроксимации, такими как полиномиальная аппроксимация или метод наименьших модулей, становится ясным, что метод наименьших квадратов более чувствителен к выбросам. Это может привести к искажению результатов в случае наличия аномалий в данных. Метод наименьших квадратов хорошо подходит для аппроксимации линейных зависимостей между переменными. Он легко реализуем и обеспечивает точные результаты в большинстве случаев. Однако, следует учитывать, что при наличии нелинейных зависимостей использование метода может привести к недостоверным результатам. Также стоит отметить, что метод чувствителен к выбросам. Численная ошибка метода зависит от структуры данных и наличия возможных выбросов. В случае отсутствия выбросов метод предоставляет точные результаты. Однако, при наличии аномалий или выбросов, ошибка может значительно увеличиться. Алгоритмическая сложность метода наименьших квадратов зависит от количества точек и выполняется за линейное время. Это делает метод эффективным для обработки больших объемов данных. Однако, вычисление обратной матрицы при использовании аналитической формулы для коэффициентов может увеличить вычислительную сложность в некоторых случаях. Таким образом, метод наименьших квадратов представляет собой мощный инструмент для аппроксимации линейных зависимостей, но требует внимательного обращения при обработке выбросов и осторожности при анализе результатов.