Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»

Факультет программной инженерии и компьютерной техники

Вычислительная математика

Лабораторная работа №1

Аппроксимация по методу наименьших модулей

Выполнила:

Гафурова Фарангиз Фуркатовна

Группа Р3220

Преподаватель:

Перл Ольга Вячеславовна

г. Санкт-Петербург

2025

Оглавление

Описание метода

Метод наименьших модулей (МНМ) является одним из методов аппроксимации функций. В отличие от метода наименьших квадратов, который минимизирует сумму квадратов отклонений, метод наименьших модулей минимизирует сумму абсолютных значений отклонений.

Пусть задан набор точек $(x_i, y_i), i = 1, \dots, n$. Мы ищем линейную функцию вида $y = ax + b$, которая аппроксимирует эти точки. Отклонение точки (x_i, y_i) от аппроксимирующей прямой $y = ax + b$ определяется как $r_i = |y_i - (ax_i + b)|$. Цель метода наименьших модулей – найти такие коэффициенты a и b, чтобы сумма $\sum_{i=1}^{n} |y_i - (ax_i + b)|$ была минимальна.

Процедура метода:

1. Упорядочить точки по значениям x.
2. Найти медианные значения x и y.
3. Вычислить медианный наклон k = median(yi - yj / xi - xj) i != j.
4. Определить свободный коэффициент b = ym - kxm.
5. Найти максимальное отклонение.

Геометрическая интерпретация:

С геометрической точки зрения, в декартовой плоскости координат, линейная функция $y = ax + b$ представляет собой прямую линию. Для данной точки $(x_i, y_i), r_i = |y_i - (ax_i + b)|$ представляет с собой вертикальное расстояние (без учета знака) от точки (x_i, y_i) до прямой $y = ax + b$. Цель метода заключается в нахождении такой прямой, при котором сумма абсолютных значений вертикальных расстояний всех точек до этой прямой будет минимальна.

Сравнение с методом наименьших квадратов:

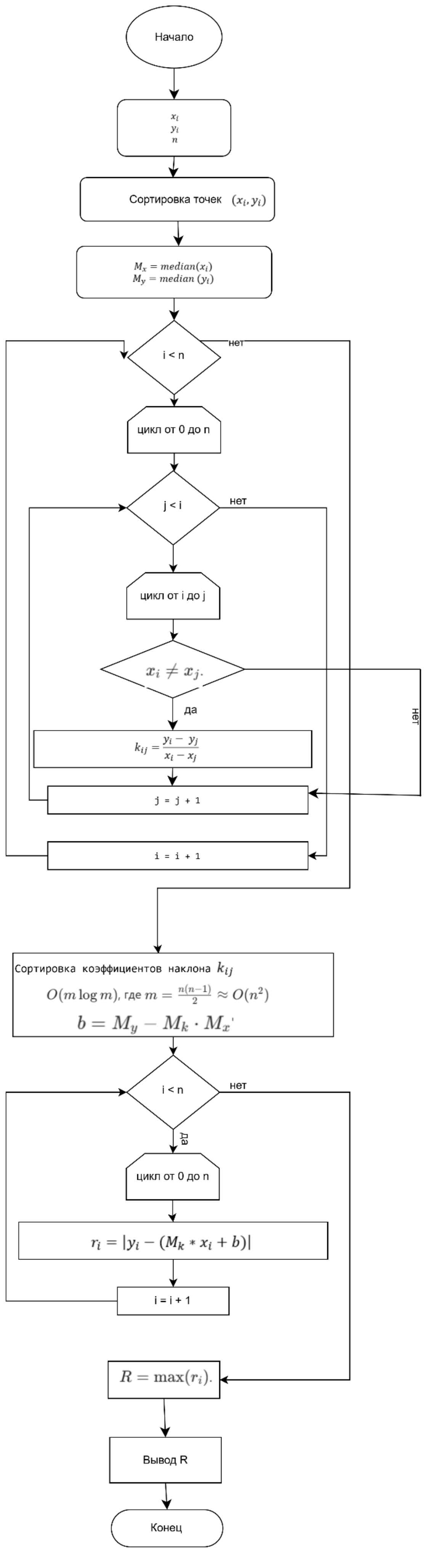
- Метод наименьших модулей:
  1. Минимизирует сумму абсолютных значений отклонений.
  2. Менее чувствителен к выбросам. Поскольку он учитывает только абсолютные значения ошибок, а не их квадраты, выбросы в данных не влияют настолько сильно на итоговую аппроксимирующую прямую.

(Например, если есть точка, которая находится далеко от остальных данных, метод наименьших модулей будет менее подвержен влиянию этой точки и лучше отобразит тенденцию большей части точек данных.)

3. Геометрическая интерпретация (описана выше)
4. Часто обеспечивает более устойчивую аппроксимацию, особенно когда данные имеют смещенное распределение или содержат много выбросов. Он лучше перехватывает общие характеристики данных и менее склонен к переобучению на локальных особенностях.

- Метод наименьших квадратов:
  1. Минимизирует сумму квадратов отклонений. То есть, для тех же данных и функции вида $y = ax + b$, отклонение точки (x_i, y_i), от прямой $y = ax + b$ считается как $r_i = \left(y_i - (ax_i + b)\right)^2$, и цель – найти a и b, при которых сумма $\sum_{i=1}^{n} \left(y_i - (ax_i + b)\right)^2$ достигает минимума.
  2. Более чувствителен к выбросам. Возведение ошибок в квадрат делает выбросы более влияющими на результат. Они могут привести к тому, что аппроксимирующая прямая сместится в сторону выброса.
  3. Геометрическая интерпретация: $r_i = \left(y_i - (ax + b)\right)^2$ представляет собой квадрат вертикального расстояния от точки (x_i, y_i), до прямой $y = ax + b$. Таким образом, метод наименьших квадратов минимизирует сумму квадратов этих расстояний.
  4. В случаях, когда данные равномерно распределены и не имеют заметных выбросов, может давать хорошую аппроксимацию. Однако, в присутствии выбросов он может быть склонен к переобучению и хуже отражать общую тенденцию данных.

Блок-схема



Начало
x_i
y_i
n
Сортировка точек (x_i, y_i)
$M_x = median(x_i)$
$M_y = median\ (y_i)$
i < n
нет
цикл от 0 до n
j < i
нет
цикл от i до j
$x_i \neq x_j.$
да
нет
$k_{ij} = \frac{y_i - y_j}{x_i - x_j}$
j = j + 1
i = i + 1
Сортировка коэффициентов наклона k_{ij}
$O(m \log m)$, где $m = \frac{n(n-1)}{2} \approx O(n^2)$
$b = M_y - M_k \cdot M_x$
i < n
нет
да
цикл от 0 до n
$r_i = |y_i - (M_k * x_i + b)|$
i = i + 1
$R = \max(r_i).$
Вывод R
Конец

Код численного метода (Python)

```
    #
    # Complete the 'approximate_linear_least_modules' function below.
    #
    # The function is expected to return a DOUBLE.
    # The function accepts following parameters:
    #  1. DOUBLE_ARRAY x_axis
    #  2. DOUBLE_ARRAY y_axis
    #

    def med(ar):
        return ar[len(ar) // 2]

    def koefNak(x_axis, y_axis, n_size):
        return [(y_axis[i] - y_axis[j]) / (x_axis[i] - x_axis[j])
            for i in range(n_size) for j in range(i) if x_axis[i] !=
x_axis[j]]


    def approximate_linear_least_modules(x_axis, y_axis):
        n_size = len(x_axis)

        points = sorted(zip(x_axis, y_axis))
        x_axis, y_axis = zip(*points)

        med_x = med(x_axis)
        med_y = med(y_axis)

        slopes = koefNak(x_axis, y_axis, n_size)

        slopes.sort()
        med_k = slopes[len(slopes) // 2]  # Медианный наклон
        med_b = med_y - med_k * med_x  # Смещение

        # Вычисляем отклонения
        res = [abs(y_axis[i] - (med_k * x_axis[i] + med_b)) for i in
range(n_size)]

        return max(res)
```

Примеры работы программы

Тест 1:

Значения x и y	Максимальное отклонение
$1\ 2\ 3\ 4\ 5\ 10\ -10\ 10\ -10\ 10$	20.0

Тест 2:

Значения x и y	Максимальное отклонение
1 2 3 4 5 -1 -0.5 1 1.5 2	0.66

Тест 3:

Значения x и y	Максимальное отклонение
1 2 3 4 5 10 1 1 1 1	9.0

Тест 4:

Значения x и y	Максимальное отклонение
-1 0 1 3 5 2 5 -1 6 3	6.5

Тест 5:

Значения x и y	Максимальное отклонение
0 1 -3 2 5 2 4 6 8 10	6.33

Тест 6:

Значения x и y	Максимальное отклонение
-1.5 -1.0 -0.5 0 1 3 6 9 12 15	3.0

Тест 7:

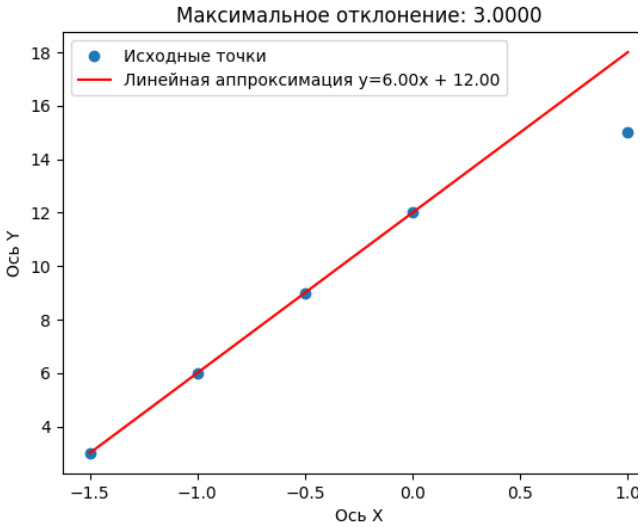
Значения x и y	Максимальное отклонение
0 1 6 7 9 1 2 3 4 5	1.1825

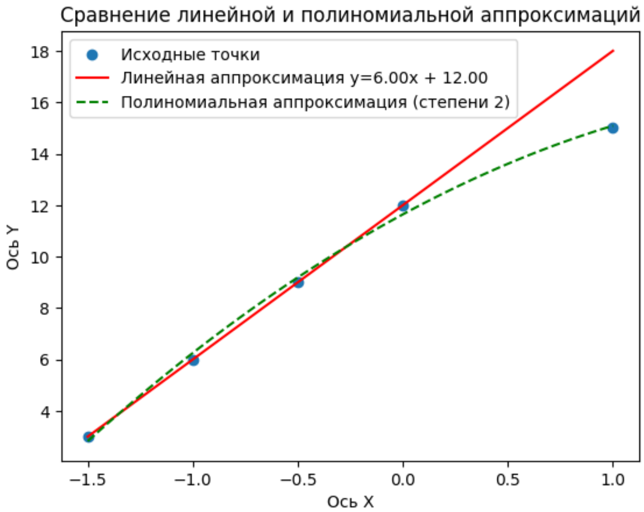
Тест 8:

Значения x и y	Максимальное отклонение
1 2 3 4 5 2.5 2.6 2.7 2.8 2.9	4.44

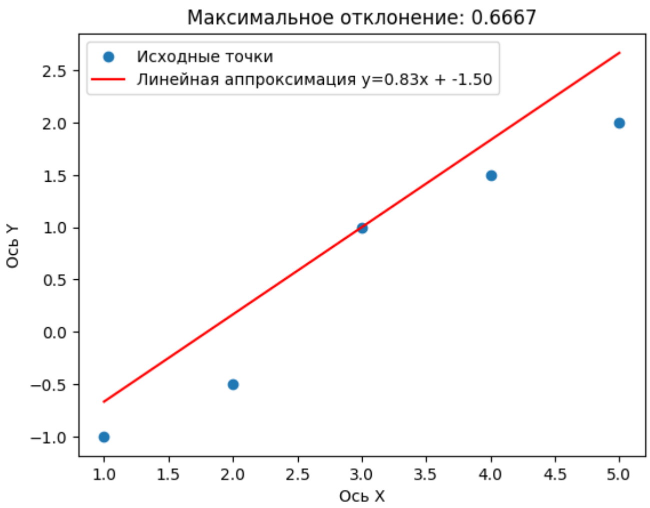
Графики:

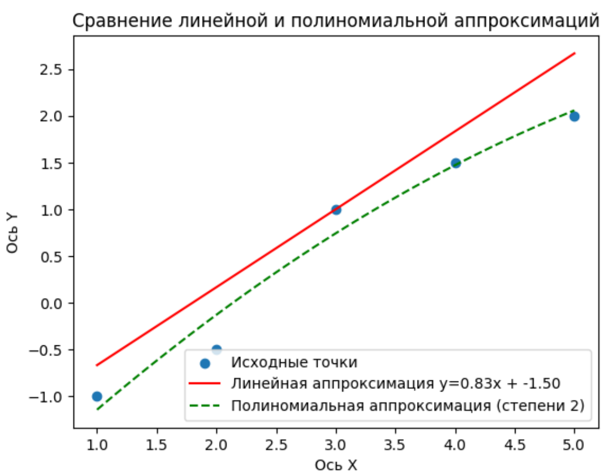
Значения x и y из Теста 6:





Значения x и y из Теста 2:





Выводы:

Метод наименьших модулей предназначен для аппроксимации наборов точек с целью минимизации суммы абсолютных отклонений от аппроксимирующей прямой $y = kx + b$. Для заданных точек $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ метод определяет коэффициенты k и b так, чтобы сумма $\sum_{i=1}^{n} |y_i - (kx_i + b)|$ была минимальной. Значение модуля наибольшего отклонения $\max|y_i - (kx_i + b)|$ служит мерой качества аппроксимации.

Сравнение с другими методами:

Метод наименьших квадратов: Метод наименьших квадратов минимизирует сумму квадратов отклонений $\sum_{i=1}^{n} \left(y_i - (kx_i + b)\right)^2$. Он более чувствителен к экстремальным значениям (выбросам), так как квадратичное увеличение отклонения больших значений сильнее влияет на итоговую функцию ошибки. В то время метод наименьших модулей, учитывающий абсолютные значения отклонений, более устойчив к выбросам.

Метод Ньютона: Метод Ньютона для аппроксимации функций обычно предполагает построение полинома Ньютона на основе заданных точек интерполяции. Полином Ньютона проходит через все заданные точки, в отличие от метода наименьших модулей, который аппроксимирует точки прямой. Метод Ньютона требует вычисления разделенных разностей, что может быть сложным для больших наборов точек. Если производные легко вычисляются, метод Ньютона может сходиться к точному значению быстрее, чем метод наименьших модулей. Однако, если функция имеет сложную структуру или производные вычисляются с трудом, метод наименьших модулей становится более предпочтительным.

Метод наименьших модулей применим, когда требуется получить аппроксимацию данных, имеющих относительно простую структуру, близкую к линейной, и когда важна устойчивость к выбросам. Также он полезен, когда данные имеют измеренные ошибки, распределенные равномерно без выраженного смещения. Однако, для сложных нелинейных зависимостей метод наименьших модулей может быть менее эффективен по сравнению с методами, специально разработанными для нелинейной аппроксимации.

В реализации, которая включает вычисление всех возможных наклонов между точками, временная сложность этого этапа составляет $O(n^2)$, где n - количество точек, из-за использования вложенного цикла для перебора всех пар точек. Затем сортировка этих наклонов имеет сложность $O(m \log m)$, где $m = \frac{n(n-1)}{2}$ – количество вычисленных наклонов. Существуют более эффективные алгоритмы, например, основанные на линейном программировании, с временной сложностью $O(n^3)$ в худшем случае.

Численная ошибка метода наименьших модулей возникает из-за округления при вычислениях, особенностей алгоритма и точностью представления чисел с плавающей запятой. Например, при вычислении коэффициентов наклона между точками могут возникать проблемы с точностью, особенно если координаты точек имеют очень большие или очень малые значения. Эти ошибки могут накапливаться при последующих операциях. Чем сложнее зависимость между данными, тем выше вероятность возникновения численной ошибки.

Метод наименьших модулей является полезным инструментом для аппроксимации данных в определенных случаях. Его преимущества включают устойчивость к выбросам и простоту применения для данных с линейной или близкой к линейной структуре. Однако, его ограничения заключаются в более высокой вычислительной сложности по сравнению с некоторыми методами и ограниченной применимости для сложных нелинейных зависимостей.