Szymon Ociepka *Data wykonania: 20.03.2024*

APROKSYMACJA

metoda najmniejszych kwadratów

Metody numeryczne

1. Wprowadzenie teoretyczne

Celem aproksymacji jest przybliżenie skomplikowanej funkcji inną, “prostszą”, która należy do określonej klasy funkcji oraz uzyskanie krzywej, która najlepiej pasuje do punktów doświadczalnych.

W teorii aproksymacji, celem jest identyfikacja tych krzywych. W przeciwieństwie do interpolacji, oczekujemy, że funkcja aproksymująca będzie jak najbliżej wszystkich punktów, czyli będzie jak najlepiej “dopasowana” do nich.

Aproksymujemy prostą: $y(x)=a_0+a_1 x$ (1)

Celem jest znalezienie parametrów dla naszej linii prostej tak, aby była **jak najbliżej** wszystkich punktów doświadczalnych (x_i, y_i). Osiąga się to poprzez minimalizację sumy:

$$S(a,b)=\sum_{i=1}^{n}\left[y_i-y\left(x_i\right)\right]^2=\sum_{i=1}^{n}\left[y_i-a_0-a_1 x_i\right]^2 \tag{2}$$

gdzie **n** to liczba punktów. Jest to funkcja dwóch zmiennych (a_0, a_1).

Poszukujemy wartości zmiennych, dla których nasza funkcja S jest minimalna. Funkcja wielu zmiennych osiąga minimum w punkcie, gdzie jej pochodne cząstkowe po wszystkich zmiennych są równe zero.

$$\frac{\partial s\left(a_0,a_1\right)}{\partial a_0}=0 \quad (3) \qquad \frac{\partial s\left(a_0,a_1\right)}{\partial a_1}=0 \tag{4}$$

$$a_1=\frac{n\sum_{i=1}^{n} x_i y_i-\sum_{i=1}^{n} x_i\sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2-\left(\sum_{i=1}^{n} x_i\right)^2} \tag{5}$$

$$a_0=\frac{\sum_{i=1}^{n} y_i\sum_{i=1}^{n} x_i^2-\sum_{i=1}^{n} x_i\sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2-\left(\sum_{i=1}^{n} x_i\right)^2} \tag{6}$$

Współczynnik korelacji to wskaźnik, który pokazuje, jak silna jest zależność między dwiema zmiennymi, x i y. Gdy jest bliski 1, oznacza to, że zmienne są silnie powiązane. Jeśli współczynnik wynosi dokładnie 1, oznacza to, że zmienne są całkowicie powiązane, co sugeruje, że istnieje prawdopodobnie pewna zależność funkcjonalna między nimi.

$$R=\frac{n\sum_{i=1}^{n}x_i y_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{\sqrt{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}\sqrt{n\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2}} \quad (7)$$

2. Opis implementacji numerycznej

Danymi wejściowymi do tej metody są punkty ze zmiennymi (x_1, x_2, … , x_n), które przyjmują odpowiednio wartości y_0, y_1, … , y_n.

```
cout << "\nWartosci x oraz y:" << endl;
for (int i = 0; i < n; i++) {
    tabx[i] = stod(tab[2 * i]);
    cout << "x" << i + 1 << " = " << tabx[i];
    taby[i] = stod(tab[2 * i + 1]);
    cout << "; y" << i + 1 << " = " << taby[i];
    cout << endl;
}
```

Rysunek 1: Fragment kodu (przypisywanie danych)

```
Wartosci x oraz y:
x1 = 2; y1 = 2
x2 = 4; y2 = 11
x3 = 6; y3 = 35
x4 = 8; y4 = 39
x5 = 10; y5 = 63
```

Rysunek 1.1: Fragment konsoli (wyświetlenie danych odczytanych z pliku)

Dodatkowo aby usprawnić testowanie programu, dane te wczytywane są z pliku dzięki użyciu biblioteki fstream.

```
ifstream plik("metoda3.txt");
string* tab = new string[liczbaLinii];
for (int i = 0; i < liczbaLinii; i++) {
    getline(plik, tab[i]);
}
plik.close();
```

Rysunek 2: Fragment kodu (odczytanie poszczególnych linii z pliku do elementów tablicy tab)

```
double a0(1), a1(1), r(1);
double sum(0); //suma iloczynów x[i] * y[i]
double sumx(0); //suma x-ów
double sumy(0); //suma y-ów
double sumx2(0); //suma kwadratów x[i]
double sumy2(0); ////suma kwadratów y[i]
```

Rysunek 3: Fragment kodu (zmienne lokalne potrzebne do obliczeń)

Znaczenie zmiennych:

$$\text{sum}=\sum_{i=1}^{n} x_i y_i \quad sumx=\sum_{i=1}^{n} x_i \quad sumy=\sum_{i=1}^{n} y_i$$

$$\text{sumx2}=\sum_{i=1}^{n} x_i^2 \quad \text{sumy2}=\sum_{i=1}^{n} y_i^2$$

```
for (int i = 0; i < n; i++) {
    sum += tabx[i] * taby[i];
    sumx += tabx[i];
    sumy += taby[i];
    sumx2 += tabx[i] * tabx[i];
    sumy2 += taby[i] * taby[i];
}
```

Rysunek 3.1: Fragment kodu (obliczanie a_1, a_0 oraz korelacji R)

W pętli **for** obliczane są odpowiednie sumy, które potem wykorzystane zostaną do obliczenia parametrów naszej prostej **a0**, **a1** oraz korelacji **R**.

3. Testy kodu numerycznego

Testy zostały przeprowadzone, aby sprawdzić działanie programu.

Test I: dla zbioru punktów {(2, 2), (4, 11), (6, 35), (8, 39), (10, 63)}; n = 5

```
Wartosci x oraz y:
x1 = 2; y1 = 2
x2 = 4; y2 = 11
x3 = 6; y3 = 35
x4 = 8; y4 = 39
x5 = 10; y5 = 63

a1 = 7.5
a0 = -15
Wspolczynnik korelacji jest rowny = 0.980581
```

Rysunek 4.1: Fragment konsoli (test dla 5 pkt. z przykładu z zajęć)

Test II: dla zbioru punktów {(2, 1), (5, 4), (7, 21), (11, 32)}; n = 4

```
Wartosci x oraz y:
x1 = 2; y1 = 1
x2 = 5; y2 = 4
x3 = 7; y3 = 21
x4 = 11; y4 = 32

a1 = 3.7076
a0 = -8.67251
Wspolczynnik korelacji jest rowny = 0.957485
```

Rysunek 4.2: Fragment konsoli (test dla 4 punktów)

Test III: dla zbioru punktów {(1, 34), (4, 11), (5, 1), (8, 102), (12, 323), (14, 20)}; n = 6

```
Wartosci x oraz y:
x1 = 1; y1 = 34
x2 = 4; y2 = 11
x3 = 5; y3 = 1
x4 = 8; y4 = 102
x5 = 12; y5 = 323
x6 = 14; y6 = 20

a1 = 11.7919
a0 = -4.64054
Wspolczynnik korelacji jest rowny = 0.474306
```

Rysunek 4.3: Fragment konsoli (test dla 6 punktów)

Test IV: dla zbioru punktów {(15, 34), (32, 11), (45, 162)}; n = 3

```
Wartosci x oraz y:
x1 = 15; y1 = 34
x2 = 32; y2 = 11
x3 = 45; y3 = 152

a1 = 3.66863
a0 = -46.838
Wspolczynnik korelacji jest rowny = 0.729611
```

Rysunek 4.4: Fragment konsoli (test dla 3 punktów)

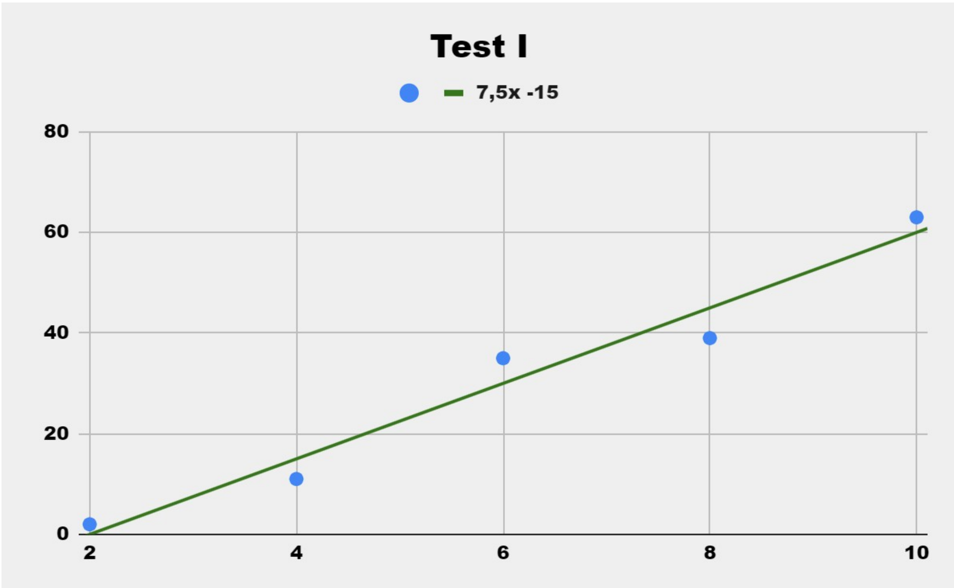
4. Analiza wyników

Poniższe tabele zawierają zestawienie wyników z testów przeprowadzonych w 3 części oraz wyników z kalkulatora online. Dodatkowo obok tabel znajdują się wykresy, na których widać zaznaczone punkty (kolor niebieski) oraz prostą powstałą z obliczonych parametrów (kolor zielony). Linia pozioma to zmienne x, linia pionowa to zmienne y.

TEST I

Szukana zmienna	Obliczona wartość zmiennej	Wartość z kalkulatora online
a_1	7.5	7.5
a_0	-15	-15
R	0.989581	0.9805807

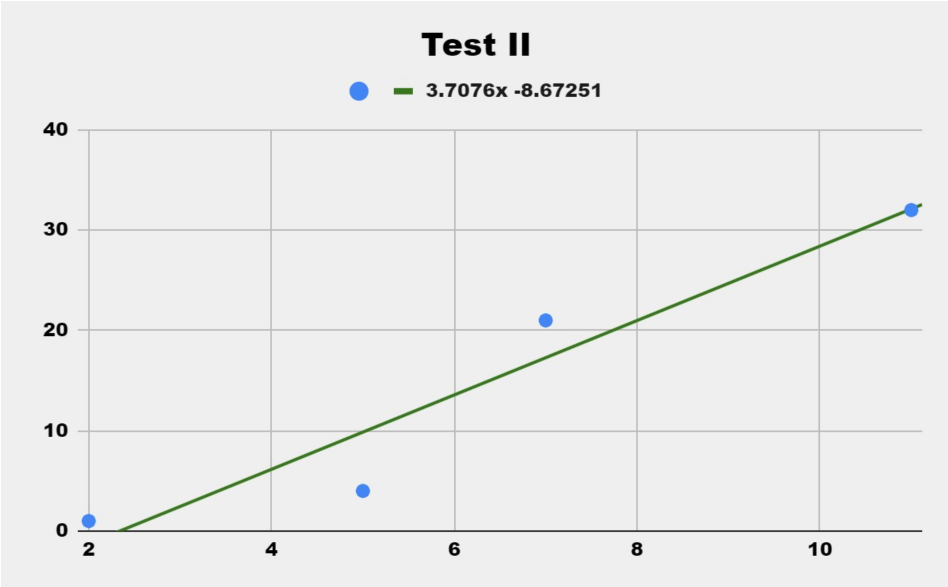
Widać, że wyniki są niemal identyczne, a korelacja jest prawie równa 1 co oznacza, że zmienne są dobrze skorelowane.



TEST II

Szukana zmienna	Obliczona wartość zmiennej	Wartość z kalkulatora online
a_1	3.7076	3.7076023
a_0	-8.67251	-8.6725146
R	0.957485	0.9574855

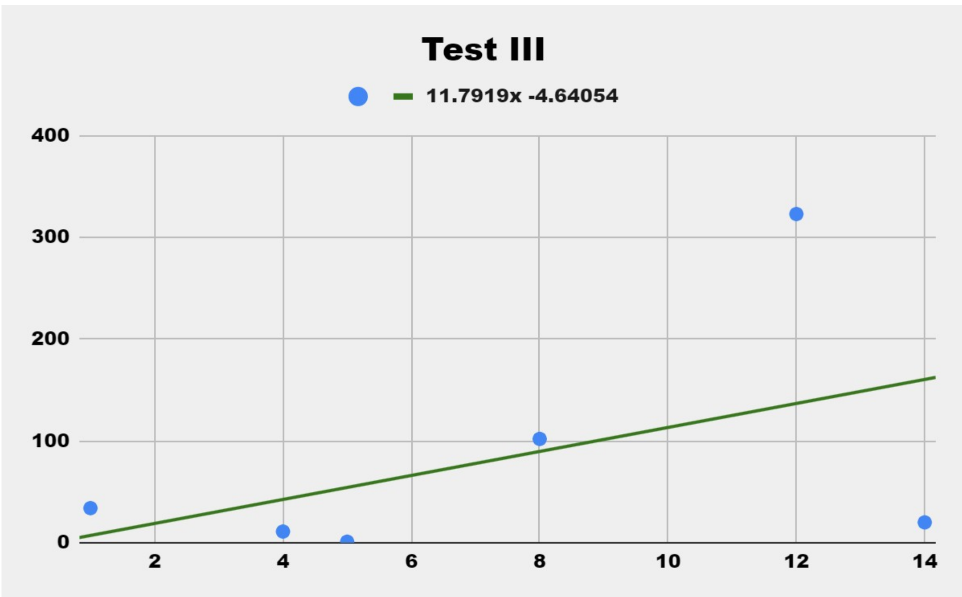
Podobnie jak w teście I widać, że wyniki są identyczne, a korelacja jest bliska wartości jeden co oznacza, że zmienne są dobrze skorelowane.



TEST III

Szukana zmienna	Obliczona wartość zmiennej	Wartość z kalkulatora online
a_1	11.7919	11.7918919
a_0	-4.64054	-4.6405405
R	0.474306	0.4743058

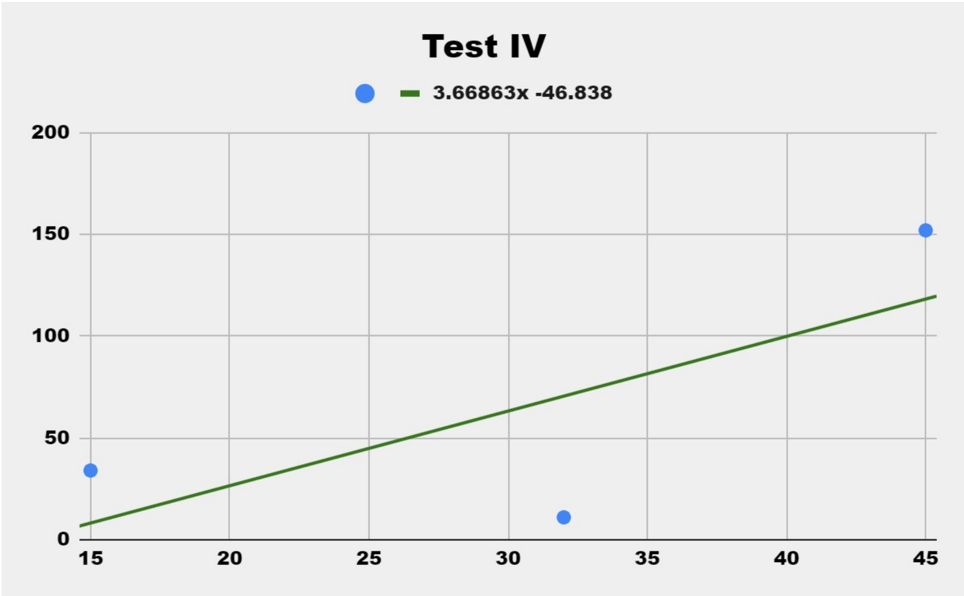
Wyniki są praktycznie identyczne, a korelacja jest nieco mniejsza od 0,5 co oznacza, że zmienne są słabo skorelowane. Widać to porównując 3 punkt (5, 1) do 5 punktu (12, 323).



TEST IV

Szukana zmienna	Obliczona wartość zmiennej	Wartość z kalkulatora online
a_1	3.66863	3.6686303
a_0	-46.838	-46.8379971
R	0.729611	0.7296114

Wyniki są praktycznie identyczne, a korelacja wynosi około 0,7 co oznacza, że zmienne są słabo skorelowane. Widać to porównując 2 punkt (32, 11) do 3 punktu (45, 152).



5. Wnioski

Metoda numeracyjna Aproksymacja - metoda najmniejszych kwadratów, jest potężnym narzędziem wykorzystywanym w analizie danych i statystyce. Jej głównym celem jest znalezienie funkcji, która najlepiej pasuje do zestawu danych poprzez minimalizację sumy kwadratów różnic między wartościami rzeczywistymi a wartościami przewidywanymi przez tę funkcję.

Dzięki metodzie aproksymacji można osiągnąć wiele korzyści. Umożliwia ona znalezienie matematycznej postaci, która najlepiej odzwierciedla dane, co może pomóc w zrozumieniu zależności między nimi. Pozwala na prognozowanie przyszłych wartości na podstawie już dostępnych danych, co jest niezwykle przydatne w analizie trendów i prognozowaniu przyszłych zdarzeń.

Patrząc na wyniki implementacji tej metody, napisany program wykazał się dużą dokładnością. Dzięki starannemu stworzeniu zmiennych oraz sczytaniu danych wejściowych z pliku, możliwe było uzyskanie wyników o wysokiej precyzji, co pozwoliło na wiarygodne wnioskowanie i prognozowanie na podstawie analizowanych danych. Metoda aproksymacji, dzięki swojej wszechstronności i skuteczności, stanowi nieocenione narzędzie w analizie danych i modelowaniu różnorodnych zjawisk.

6. Źródła

- prezentacja z zajęć „lab 4.pdf”
- arkusz kalkulacyjny Microsoft Excel
- goodcalculators.com/linear-regression-calculator/