Laboratorium 5 - Aproksymacja

Mateusz Podmokły - II rok Informatyka WI

4 kwiecień 2024

1 Treść zadania

Zadanie 1. Wykonaj aproksymację średniokwadratową punktową populacji Stanów Zjednoczonych w przedziale $[1900, 1980]$ wielomianami stopnia m dla $0 \leqslant m \leqslant 6$.
Dla każdego m dokonaj ekstrapolacji wielomianu do roku 1990. Porównaj otrzymaną wartość z prawdziwą wartością dla roku 1990 wynoszącą 248 709 873.
Wyznacz optymalny stopień wielomianu za pomocą kryterium informacyjnego Akaikego (ang. Akaike information criterion):

$$AIC = 2k + nln\left(\frac{\sum_{i=1}^{n}[y_i - \hat{y}(x_i)]^2}{n}\right),$$

gdzie y_i $(i = 1, \ldots, n)$ oznacza prawdziwą liczbę osób w roku x_i, k to liczba parametrów wielomianu $(k = m + 1)$, natomiast $\hat{y}(x_i)$ liczbę osób przewidywaną przez model, tzn. wartość wielomianu $\hat{y}(x)$. Ponieważ rozmiar próbki jest niewielki (dane z dziewięciu lat, $n - 9$), $\frac{n}{k} < 40$, należy użyć wzoru ze składnikiem korygującym:

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$$

Mniejsze wartości kryterium Akaikego oznaczają lepszy model.

Zadanie 2. Wykonaj aproksymację średniokwadratową ciągłą funkcji $f(x) = \sqrt{x}$ w przedziale $[0, 2]$ wielomianem drugiego stopnia, używając wielomianów Czebyszewa.

2 Specyfikacja użytego środowiska

Specyfikacja:

- Środowisko: Visual Studio Code,
- Język programowania: Python,

- System operacyjny: Microsoft Windows 11,
- Architektura systemu: x64.

3 Rozwiązanie problemu

3.1 Biblioteki

W realizacji rozwiązania wykorzystane zostały następujące biblioteki:

```
import numpy as np
import matplotlib.pyplot as plt
```

3.2 Zadanie 1.

Mamy n punktów, dla których chcemy wyznaczyć wielomian aproksymacyjny stopnia m. Aby wyznaczyć współczynniki wielomianu obliczamy macierz

$$A = \begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^m \\ 1 & x_1 & x_1^2 & \cdots & x_1^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix}$$

Przyjmujemy

$$y = \begin{bmatrix} f(x_0) \\ f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix}$$

oraz jako wektor współczynników

$$c = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_m \end{bmatrix}$$

Obliczamy c z równania normalnego:

$$A^T Ac = A^T y$$

$$c = (A^T A)^{-1} A^T y$$

Otrzymujemy wielomian aproksymacyjny postaci

$$p(x) = \sum_{j=0}^{n} c_j x^j$$

3.3 Zadanie 2.

Mamy funkcję

$$f(x) = \sqrt{x}, x \in [0, 2]$$

Aproksymacja tej funkcji wielomianem drugiego stopnia wymaga trzech pierwszych wielomianów Czebyszewa:

$$T_0(x) = 1$$
$$T_1(x) = x$$
$$T_2(x) = 2x^2 - 1$$

Wielomian aproksymacyjny jest postaci

$$p(x) = \sum_{k=0}^{n} c_k \phi_k$$

gdzie

$$\phi_k = T_k(x)$$

czyli w naszym przypadku

$$p(x) = c_0 T_0(x) + c_1 T_1(x) + c_2 T_2(x)$$

Współczynniki c_k zostały wyznaczone z wykorzystaniem funkcji

`np.polynomial.chebyshev.chebfit`

a następnie wielomian $p(x)$ z funkcji `np.polynomial.chebyshev.Chebyshev`.

4 Przedstawienie wyników

4.1 Zadanie 1.

W poniższej tabeli przedstawione zostały przywidywane przez wielomian aproksymacyjny wartości populacji w roku 1990 dla danego stopnia wielomianu m oraz błąd względny tej aproksymacji.

m	Przewidywana populacja	Błąd względny
0	143 369 177	42.35%
1	235 808 109	5.19%
2	**254 712 944**	**2.41%**
3	261 378 612	5.09%
4	-116 331 273	146.77%
5	472 089 589	89.82%
6	1 269 697 652	410.51%

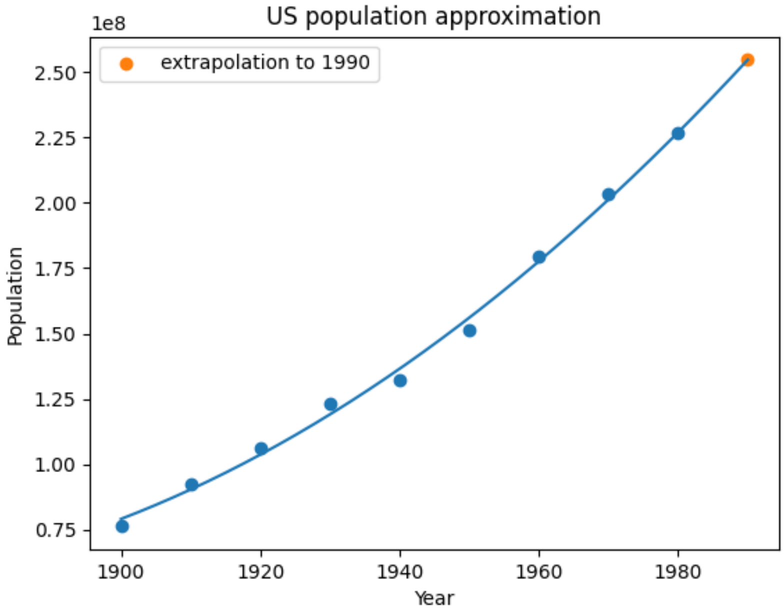
Tabela 1: Porównanie błędu dla różnych stopni wielomianu.

Najlepszym przybliżeniem wartości dla roku 1990 okazał się wielomian aproksymacyjny stopnia 2 z błędem względnym wynoszącym 2.41%.
Poniższa tabela przedstawia wartości kryterium informacyjnego Akaikego (AIC) dla danego stopnia wielomianu.

m	AIC
0	340.79
1	308.83
2	**299.23**
3	304.66
4	400.71
5	418.88
6	517.97

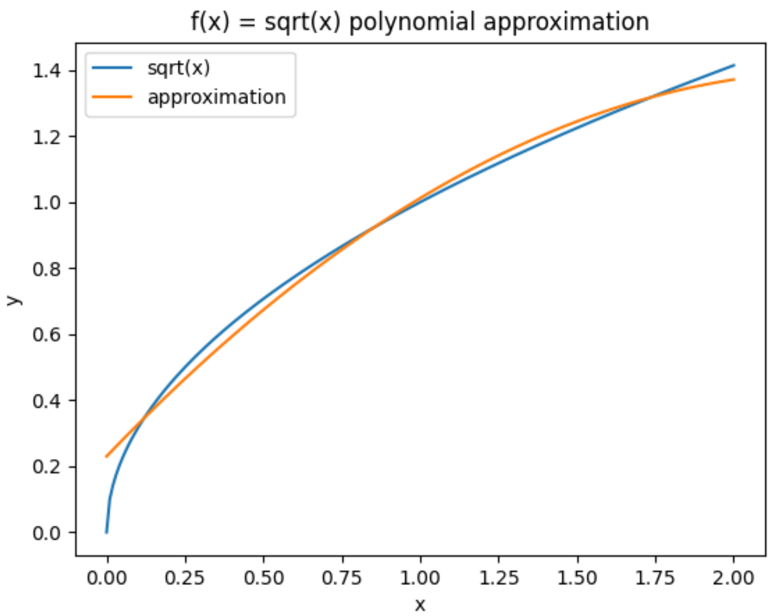
Tabela 2: Kryterium informacyjne Akaikego.

Przewidywania kryterium informacyjnego Akaikego pokrywają się z obserwacjami ekstrapolacji wielomianu i także wskazują wielomian stopnia 2 jako najbardziej optymalny.



Rysunek 1: Aproksymacja wielomianem stopnia 2.

4.2 Zadanie 2.



Rysunek 2: Aproksymacja wielomianowa funkcji $f(x) = \sqrt{x}$.

5 Wnioski

Zadanie 1.

W tym zadaniu najdokładniejsza okazała się aproksymacja wielomianem stopnia 2. Niższe stopnie nie były w stanie uwzględnić zmienności danych, natomiast wyższe były zbyt podatne na szum. Kryterium informacyjne Akaikego jako optymalny stopień wielomianu także wskazało stopień 2 (najmniejsza wartość AIC), zatem może być ono przydatne przy wyborze stopnia wielomianu aproksymacyjnego.

Zadanie 2.

Aproksymacja wielomianami Czebyszewa stopnia 2 wydaje się dobrze przybliżać funkcję $f(x) = \sqrt{x}$ na przedziale $[0, 2]$. Takie przybliżenie może uprościć niektóre obliczenia i zmniejszyć błędy numeryczne.

Podsumowanie

Aproksymacja wielomianowa może być przydatna do przewidywania wartości na podstawie wcześniejszych obserwacji, a także do przekształcenia funkcji do prostszej i bardziej przystępnej postaci, zależnie od potrzeb. Należy jednak pamiętać, aby odpowiednio dobrać stopień wielomianu aproksymacyjnego, tak, żeby wynik odpowiednio odzwierciedlał dane.

6 Bibliografia

https://pl.wikipedia.org/wiki/Wielomiany_Czebyszewa
https://pl.wikipedia.org/wiki/Aproksymacja_%C5%9Bredniokwadratowa
http://wygasz.edu.pl/ludzie/szewczuk/mn_data/wyklad4.pdf
https://pl.wikipedia.org/wiki/Aproksymacja_wielomianowa