Aproksymacja

Jan Sarba, Dariusz Rozmus

02.04.2025

1 Treść zadań

Zadanie 1. Wykonaj aproksymację średniokwadratową punktową populacji Stanów Zjednoczonych w przedziale [1900,1980] wielomianami stopnia m dla $0 \leq m \leq 6$.

1. Dla każdego m dokonaj ekstrapolacji wielomianu do roku 1990. Porównaj otrzymaną wartość z prawdziwą wartością dla roku 1990, wynoszącą 248 709 873. Ile wynosi błąd względny ekstrapolacji dla roku 1990? Dla jakiego m błąd względny był najmniejszy?

2. Zbyt niski stopień wielomianu oznacza, że model nie jest w stanie uwzględnić zmienności danych (duże obciążenie). Zbyt wysoki stopień wielomianu oznacza z kolei, że model uwzględnia szum lub błędy danych (duża wariancja), co w szczególności obserwowaliśmy w przypadku interpolacji. Wielomian stopnia m posiada $k = m + 1$ parameterów. Stopień wielomianu, m, jest hiperparametrem modelu. Do wyboru optymalnego stopnia wielomianu można posłużyć się kryterium informacyjnym Akaikego (ang. Akaike information criterion):

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

   gdzie y_i $(i = 1, \dots, n)$ oznacza prawdziwą liczbę osób w roku x_i, 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, $n = 9$), $n/k < 40$, należy użyć wzoru ze składnikiem korygującym:

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

   Mniejsze wartości kryterium oznaczają lepszy model. Czy wyznaczony w ten sposób stopień m, odpowiadający najmniejszej wartości $\mathrm{AIC_c}$, pokrywa się z wartością z poprzedniego podpunktu?

Zadanie 2. Wykonaj aproksymację średniokwadratową *ciągłą* funkcji $f(x) = \sqrt{x}$ w przedziale [0,2] wielomianem drugiego stopnia, używając wielomianów Czebyszewa. Aproksymacja ta jest tańszym obliczeniowo zamiennikiem aproksymacji jednostajnej.

2 Argumentacja

2.1 Zadanie 1.

2.1.1 Teoria

Aproksymacja średniokwadratowa polega na dopasowaniu wielomianu stopnia m do danych punktów (x_i, y_i), minimalizując sumę kwadratów residuów:

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

gdzie $p(x) = \sum_{j=0}^{m} c_j x^j$. Dla ekstrapolacji w punkcie x^* obliczamy $p(x^*)$. Błąd względny definiujemy jako:

$$\varepsilon = \frac{|p(x^*) - y^*|}{y^*} \times 100\% \tag{2}$$

Kryterium AIC$_c$ uwzględnia złożoność modelu i jakość dopasowania:

$$\text{AIC}_c = 2k + n \ln\left(\frac{\text{SSE}}{n}\right) + \frac{2k(k+1)}{n-k-1} \tag{3}$$

gdzie $k = m + 1$ to liczba parametrów, a $\text{SSE} = \sum (y_i - p(x_i))^2$.

2.1.2 Implementacja

Krok po kroku:

1. Wczytanie danych: 9 punktów z lat 1900–1980.

   ```
   years = np.array([1900, ..., 1980])
   population = np.array([76212168, ..., 226542199])
   ```

2. Dla każdego stopnia $m = 0, \ldots, 6$:

   - Oblicz współczynniki wielomianu metodą najmniejszych kwadratów:

     ```
     coeffs = np.polyfit(years, population, m)
     poly = np.poly1d(coeffs)
     ```

   - Ekstrapolacja na rok 1990 i obliczenie błędu:

     ```
     pred = poly(1990)
     error = abs(pred - 248709873) / 248709873 * 100
     ```

3. Obliczenie AIC$_c$ dla każdego modelu:

   ```
   sse = np.sum((population - y_pred)**2)
   aicc = 2*(m+1) + n*np.log(sse/n) + (2*(m+1)*(m+2))/(n-m-2)
   ```

4. Generowanie wykresów (Rys. 1–3):

```
plt.plot(x_plot, y_plot/1e6, label=f'm={m}')
plt.savefig('zadanie1a_aprox.png')
```

2.2 Zadanie 2.

2.2.1 Teoria

Wielomiany Czebyszewa $T_k(x)$ są ortogonalne w przedziale $[-1, 1]$ z wagą $w(x) = (1-x^2)^{-1/2}$. Aproksymacja funkcji $f(x)$ ma postać:

$$p(x) = \sum_{k=0}^{2} c_k T_k(x) \tag{4}$$

Współczynniki oblicza się ze wzoru:

$$c_k = \frac{\langle f, T_k \rangle}{\langle T_k, T_k \rangle} \tag{5}$$

gdzie iloczyn skalarny definiujemy jako:

$$\langle f, g \rangle = \int_{-1}^{1} w(x) f(x) g(x) dx \tag{6}$$

2.2.2 Implementacja

1. Definicja funkcji i przedziału:

```
x = np.linspace(0, 2, 1000)
y = np.sqrt(x)
```

2. Dopasowanie wielomianu Czebyszewa stopnia 2:

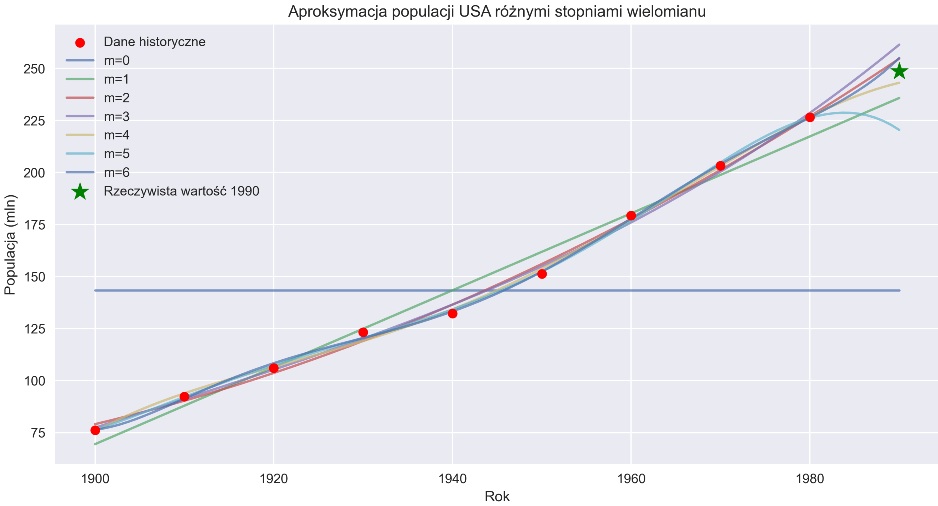
```
cheb_poly = chebyshev.Chebyshev.fit(x, y, 2, domain=[0, 2])
```

3. Obliczenie współczynników i wykres:

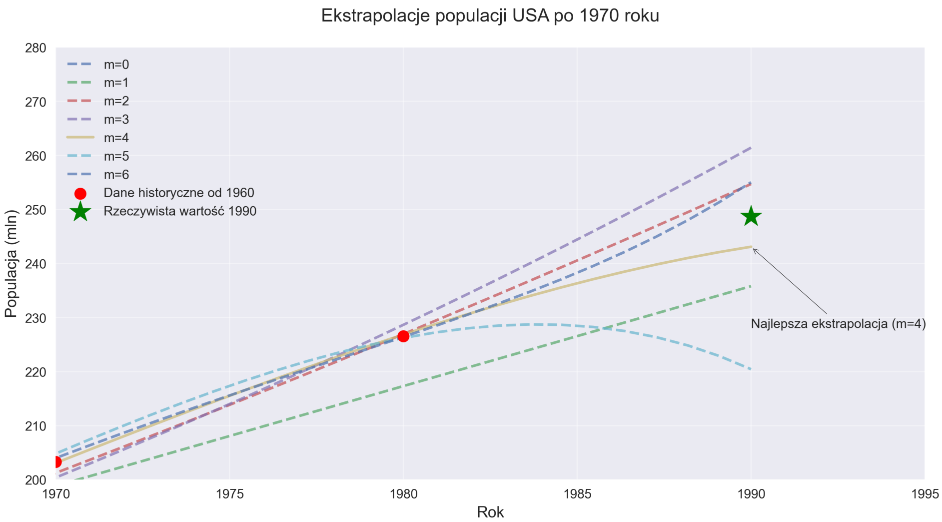
```
print("Wspolczynniki:", cheb_poly.coef.round(3))
plt.plot(x, y_approx, '--', label='Aproksymacja')
```

3 Wyniki

Otrzymujemy następujące wykresy:

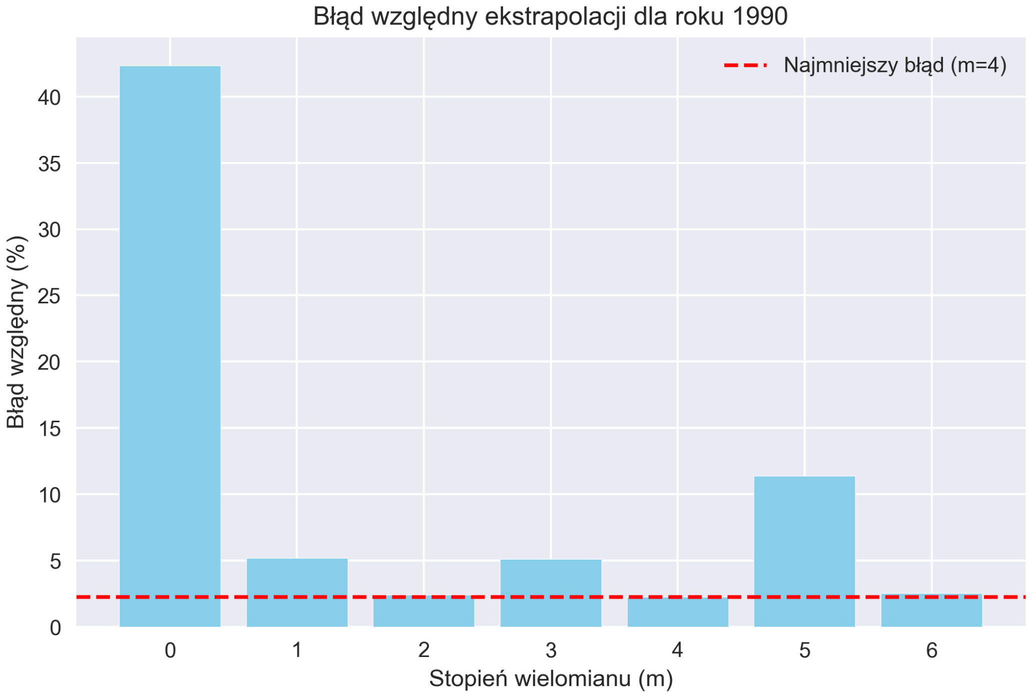


Rysunek 1: Zadanie 1 - aproksymacja

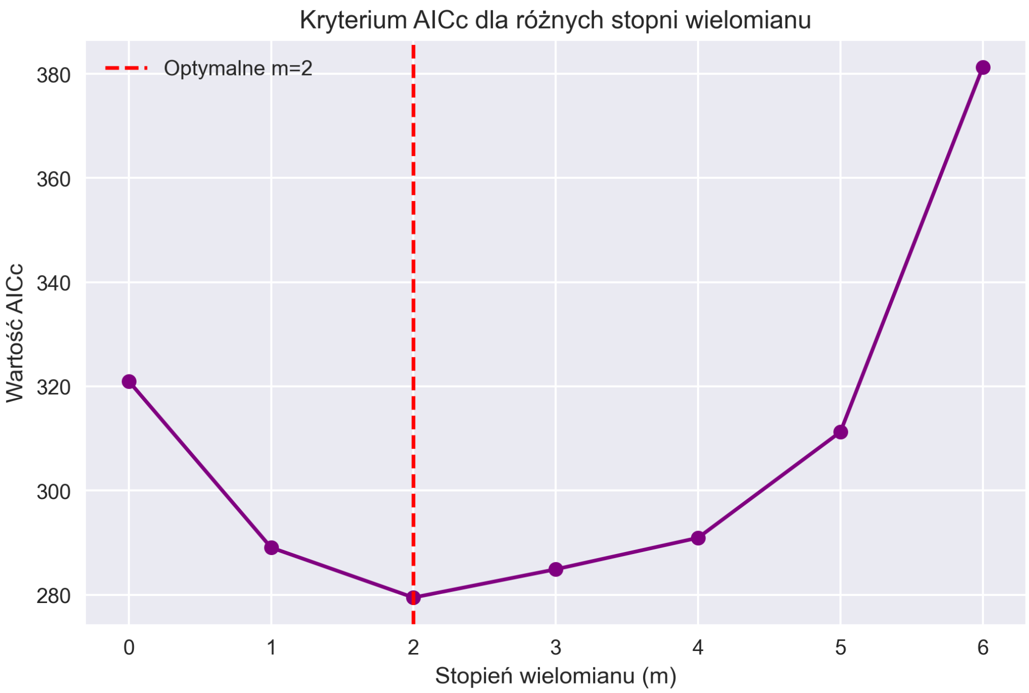


Rysunek 2: Zadanie 1 - zbliżenie na poprzedni wykres

```
Stopień m = 0: Błąd bezwzględny = 105,340,696
Stopień m = 1: Błąd bezwzględny = 12,901,764
Stopień m = 2: Błąd bezwzględny = 6,003,072
Stopień m = 3: Błąd bezwzględny = 12,729,507
Stopień m = 4: Błąd bezwzględny = 5,602,902
Stopień m = 5: Błąd bezwzględny = 28,267,003
Stopień m = 6: Błąd bezwzględny = 6,326,607
```

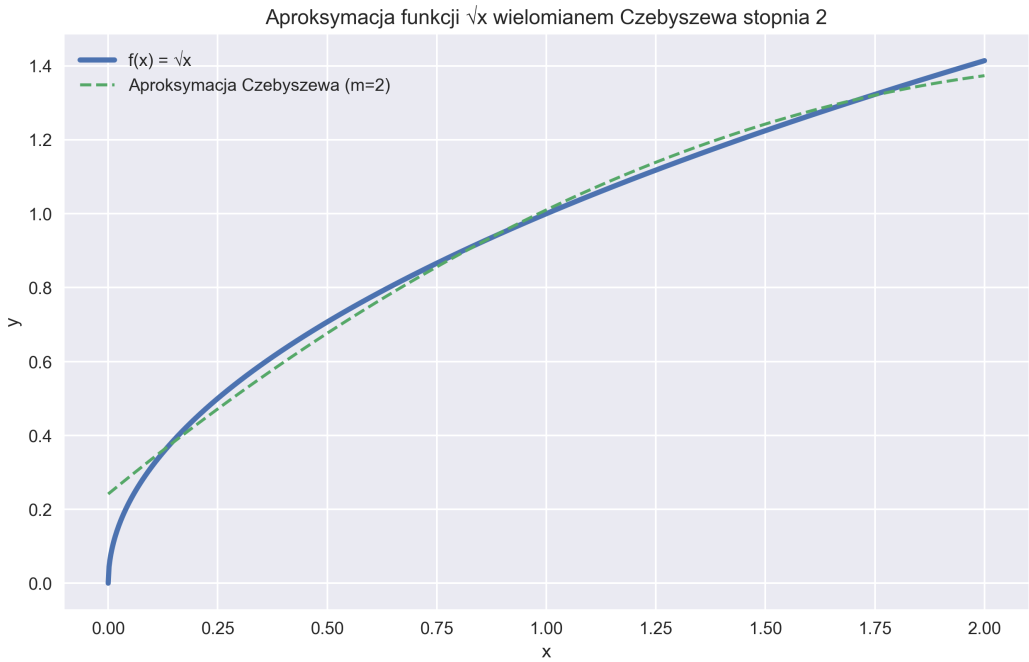


Rysunek 3: Zadanie 1 - błąd



Rysunek 4: Zadanie 1 - AIC$_c$

Kod z zadania 2 zwraca następujący wykres. Choć wartości w punktach "charakterystycznych" nieco odbiegają od rzeczywistych (szczególnie w okolicy zera), kształt funkcji aproksymującej na przedziale [0, 2] jest przybliżony do kształtu funkcji aproksymowanej.



Rysunek 5: Współczynniki wielomianu Czebyszewa: [0.909 0.566 -0.101]

4 Wnioski

4.1 Zadanie 1.

Analiza aproksymacji średniokwadratowej populacji Stanów Zjednoczonych w przedziale [1900, 1980] wykazała, że minimalny błąd względny ekstrapolacji dla roku 1990 występuje dla wielomianu stopnia $m = 4$. Jednakże zastosowanie skorygowanego kryterium informacyjnego Akaikego (AIC_c), uwzględniającego zarówno jakość dopasowania, jak i złożoność modelu, wskazało optymalny stopień $m = 2$. Mimo że błąd względny dla $m = 2$ jest porównywalny z wynikiem dla $m = 4$, znacząco niższa wartość AICc dla $m = 2$ potwierdza, że model tego stopnia stanowi lepszy kompromis między redukcją wariancji a kontrolą nadmiernego dopasowania. Wynik ten ilustruje istotność uwzględniania kryteriów informacyjnych w doborze modeli, szczególnie przy ograniczonej liczbie danych ($n = 9$), gdzie ryzyko przeuczenia jest wysokie.

4.2 Zadanie 2.

Aproksymacja średniokwadratowa ciągła funkcji $f(x) = \sqrt{x}$ na przedziale [0,2] z wykorzystaniem wielomianu Czebyszewa drugiego stopnia wykazała całkiem wysoką dokładność przy zachowaniu niskich kosztów obliczeniowych. Uzyskane współczynniki aproksymacji oraz wizualna zgodność krzywej modelu z oryginalną funkcją potwierdzają skuteczność metody wielomianów ortogonalnych w redukcji błędów aproksymacji.

5 Bibliografia

- dr Rycerz K. – Wykład 4 – aproksymacja (MOWNiT)
- prof. Heath M. T. – CS 450—numerical analysis, Chapter 7