Politechnika Krakowska
im. Tadeusza Kościuszki w Krakowie

Wydział Fizyki, Matematyki i Informatyki
Fizyka Techniczna

Juliusz Chojenka, Paula Świerska, Wojciech Tejchman

Dopasowanie funkcji do danych

Kraków 2016

1 Opis zadania

Celem projektu było przetestowanie różnych metod dopasowania. Wykorzystano trzy metody: Lagrange'a, najmniejszych kwadratów, Cubic Spline. Metody te przetestowano na danych wygenerowanych dla funkcji kwadratowej oraz dla danych pobranych z internetu ilości małżeństw, rozwodów i separacji w Polsce.

2 Metoda Lagrange'a

Dane podane były w postaci punktów: $(x_0, y_0), ..., (x_n, y_n)$

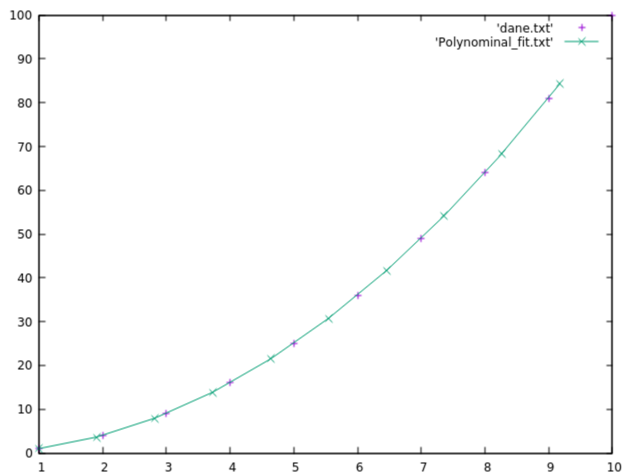
Dopasowywany wielomian był w postaci:

$$g(x) = \sum_{i=0}^{n} y_i \lambda_i(x),$$

$$\text{gdzie: } \lambda_i(x) = \prod_{i \neq j} \frac{x - x_j}{x_i - x_j}$$

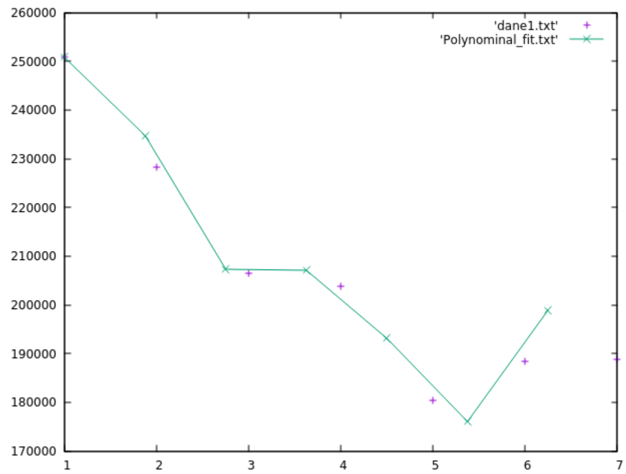
2.1 Wyniki

Dopasowanie do funkcji kwadratowej:



Rysunek 1: Fioletowe punkty są zadanymi danymi w postaci: $y = x^2$, niebieska linia przedstawia funkcję dopasowaną do danych metodą Lagrange'a.

Dopasowanie do ilości małżeństw, rozwodów i separacji w Polsce:



Rysunek 2: Fioletowe punkty są zadanymi danymi, gdzie na osi x znajdują się kolejne lata (od 2009 roku), natomiast na osi y znajduje się ilość małżeństw, rozwodów i separacji w Polsce. Niebieska linia przedstawia funkcję dopasowaną do danych metodą Lagrange'a.

3 Metoda najmniejszych kwadratów

Żądamy minimalizacji χ^2, która mierzy odchylenie zadanej zależności funkcyjnej od ptk. doświadczalnych. Poniżej znajduje się wyprowadzenie wzorów dla funkcji liniowej $f(x) = ax + b$, wzory dla funkcji eksponencjalnej wyprowadzane są analogicznie.

$$\chi^2(a,b) = \sum_{i=1}^{n} \frac{(y_i - ax_i - b)^2}{\sigma_i^2}$$

gdzie σ_i to odchylenie standardowe

$$S = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}$$

$$S_x = \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}$$

$$S_y = \sum_{i=1}^{n} \frac{y_i}{\sigma_i^2}$$

$$S_{xx} = \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2}$$

$$S_{xy} = \sum_{i=1}^{n} \frac{y_i x_i}{\sigma_i^2}$$

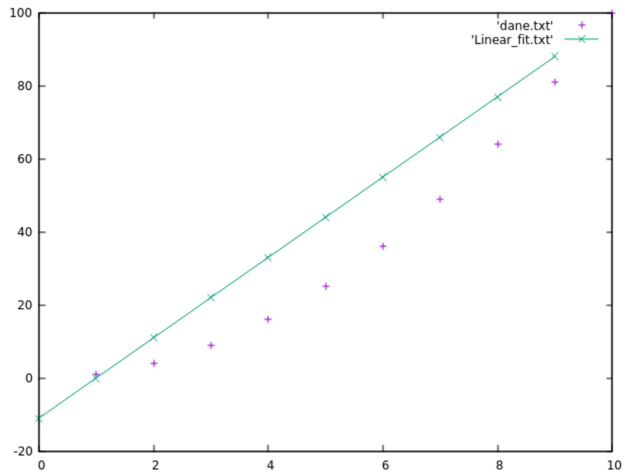
$$a = \frac{SS_{xy} - S_x S_y}{\Delta}$$

$$b = \frac{S_{xx} S_y - S_x S_{xy}}{\Delta}$$

$$\Delta = SS_{xx} - S_x^2$$

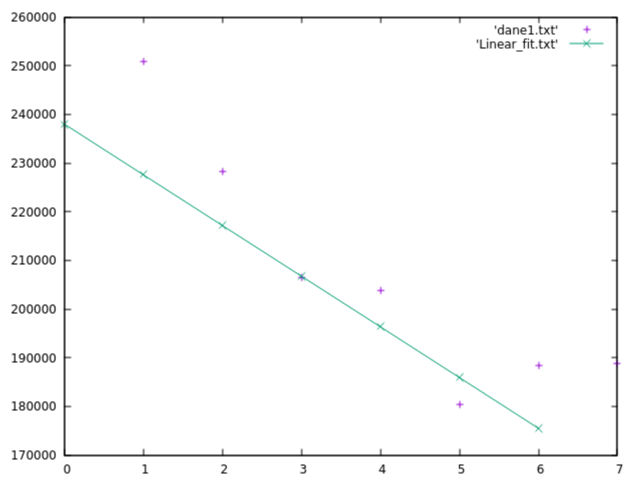
3.1 Wyniki

Dopasowanie do funkcji kwadratowej, za pomocą funkcji liniowej:



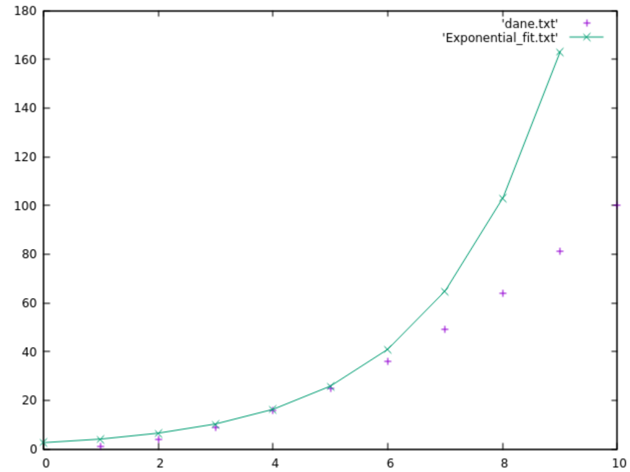
Rysunek 3: Fioletowe punkty są zadanymi danymi w postaci: $y = x^2$, niebieska linia przedstawia funkcję dopasowaną do danych metodą najmniejszych kwadratów.

Dopasowanie do ilości małżeństw, rozwodów i separacji w Polsce:



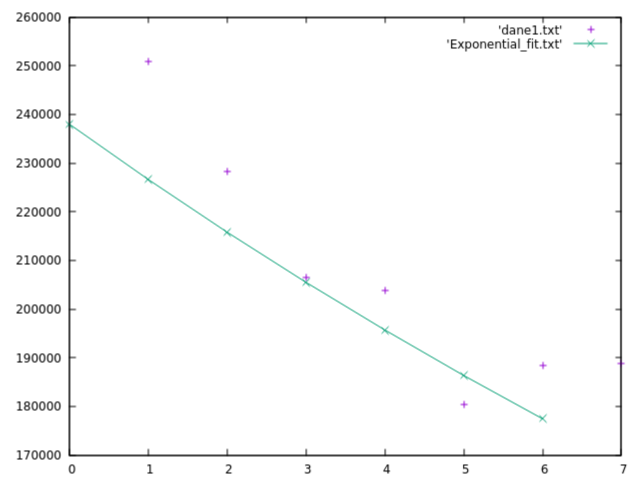
Rysunek 4: Fioletowe punkty są zadanymi danymi, gdzie na osi x znajdują się kolejne lata (od 2009 roku), natomiast na osi y znajduje się ilość małżeństw, rozwodów i separacji w Polsce. Niebieska linia przedstawia funkcję dopasowaną do danych metodą najmniejszych kwadratów.

Dopasowanie do funkcji kwadratowej, pomocą funkcji exp.:



Rysunek 5: Fioletowe punkty są zadanymi danymi w postaci: $y = x^2$, niebieska linia przedstawia funkcję dopasowaną do danych metodą najmniejszych kwadratów.

Dopasowanie do ilości małżeństw, rozwodów i separacji w Polsce, za pomocą funkcji exp.:



Rysunek 6: Fioletowe punkty są zadanymi danymi, gdzie na osi x znajdują się kolejne lata (od 2009 roku), natomiast na osi y znajduje się ilość małżeństw, rozwodów i separacji w Polsce. Niebieska linia przedstawia funkcję dopasowaną do danych metodą najmniejszych kwadratów.

4 Metoda: Cubic Spline

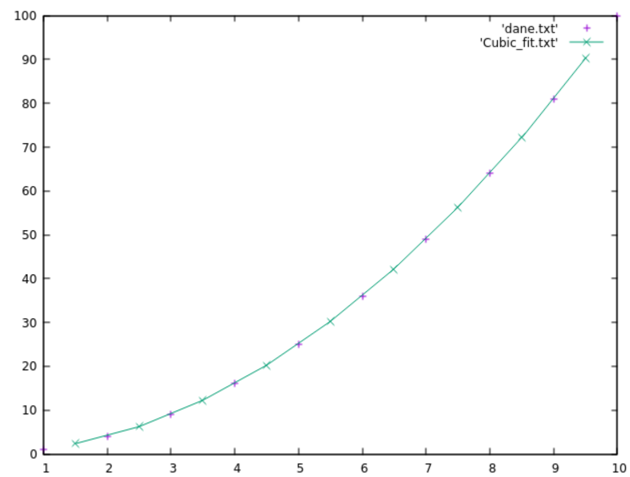
Dane w postaci punktów: $(x_0, y_0), ..., (x_n, y_n) \quad x_0 < ... < x_n$
$g(x) \simeq g_i(x) \quad x \in [x_i, x_{i+1}]$ $g_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + y_i$

Muszą być spełnione zależności:

$$g_i(x_{i+1}) = y_{i+1}$$
$$g'_{i-1}(x_i) = g_i(x_i)$$
$$g''_{i-1}(x_i) = g''_i(x_i)$$

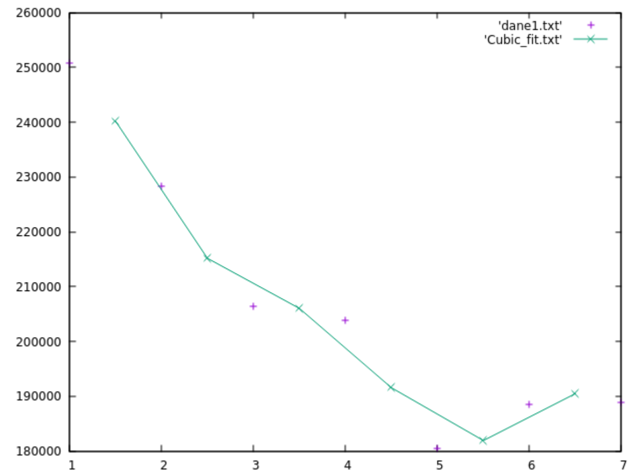
4.1 Wyniki

Dopasowanie do funkcji kwadratowej:



Rysunek 7: Fioletowe punkty są zadanymi danymi w postaci: $y = x^2$, niebieska linia przedstawia funkcję dopasowaną do danych metodą Cubic Spline.

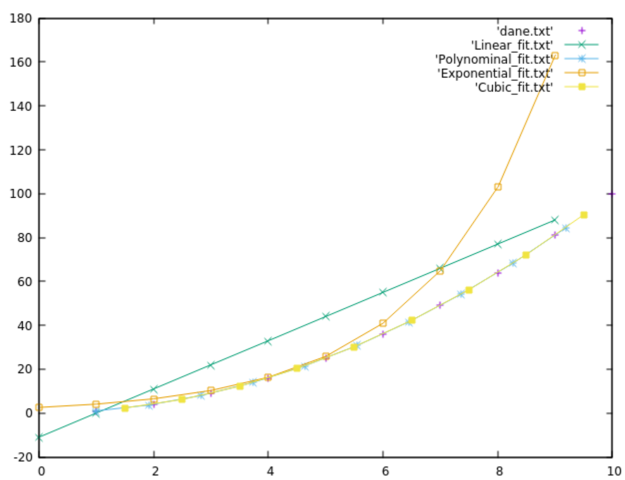
Dopasowanie do ilości małżeństw, rozwodów i separacji w Polsce:



Rysunek 8: Fioletowe punkty są zadanymi danymi, gdzie na osi x znajdują się kolejne lata (od 2009 roku), natomiast na osi y znajduje się ilość małżeństw, rozwodów i separacji w Polsce. Niebieska linia przedstawia funkcję dopasowaną do danych metodą Cubic Spline.

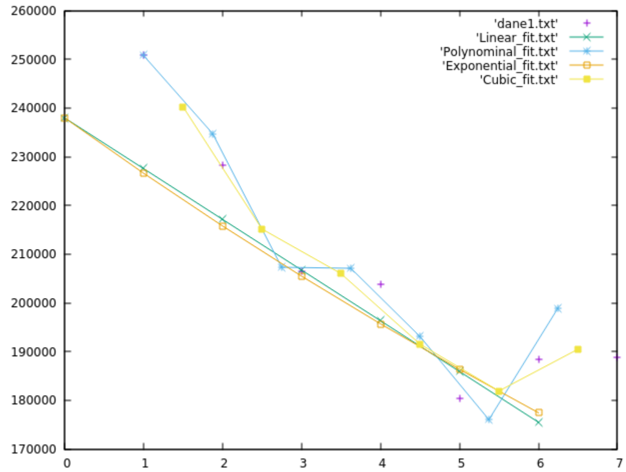
5 Porównanie wszystkich metod

Dopasowanie do funkcji kwadratowej:



Rysunek 9: Fioletowe punkty są zadanymi danymi w postaci: $y = x^2$. Kolory zielony, niebieski, pomarańczowy i żółty reprezentują dopasowania odpowiednimi metodami: funkcji liniowej metodą najmniejszych kwadratów, Lagrange'a, funkcji exp. metodą najmniejszych kwadratów oraz Cubic Spline.

Dopasowanie do ilości małżeństw, rozwodów i separacji w Polsce



Rysunek 10: Fioletowe punkty są zadanymi danymi, gdzie na osi x znajdują się kolejne lata (od 2009 roku), natomiast na osi y znajduje się ilość małżeństw, rozwodów i separacji w Polsce. Kolory zielony, niebieski, pomarańczowy i żółty reprezentują dopasowania odpowiednimi metodami: funkcji liniowej metodą najmniejszych kwadratów, Lagrange'a, funkcji exp. metodą najmniejszych kwadratów oraz Cubic Spline.

6 Wnioski

- Metoda Lagrange'a uzyskała bardzo dobre wyniki dla dopasowania do funkcji kwadratowej. Do danych nie uzyskanych żadną funkcją (ilości małżeństw, rozwodów i separacji w Polsce), metoda ta nie uzyskała dobrych wyników i nie można wysnuć lepszych wniosków.

- Metoda najmniejszych kwadratów dla dopasowania funkcji liniowej do funkcji kwadratowej, nie daje zaskakujących wyników i wniosków. Natomiast z dopasowania do bardziej losowych danych pozwala na wywnioskowanie czy dane mają tendencję rosnącą czy malejącą. Dzięki temu możemy być pewni, że na przestrzeni sześciu lat ilości małżeństw, rozwodów i separacji w Polsce spadła.

- Metoda najmniejszych kwadratów dla dopasowania funkcji exp. do funkcji kwadratowej daje lepsze wyniki niż dopasowanie funkcji liniowej tą metodą, jednak wyniki dalej nie są rewelacyjne. Przy dopasowaniu do ilości małżeństw, rozwodów i separacji w Polsce daje podobne wnioski jak dopasowanie funkcją liniową tą metodą.

- Metoda Cubic Spline daje bardzo dobre wyniki dla funkcji kwadratowej oraz wydaje się, że daje najbliższe wartości dla danych "losowych" niestety nie można z nich wysnuć lepszych wniosków.