SPRAWOZDANIE
Z ÓSMEGO ZADANIA NUMERYCZNEGO
MACIEJ WÓJCIK

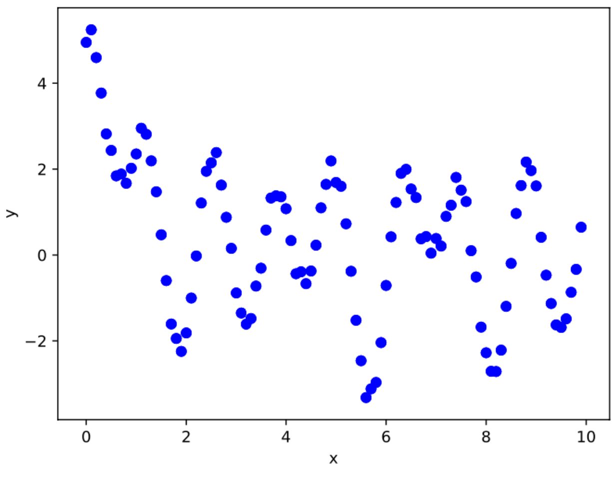
Spis treści:

1. Polecenie zadania

(Zadanie numeryczne NUM8) Zadany jest zbiór punktów zilustrowany poniżej (plik dostępny jest do pobrania na platformie Pegaz), dwie liczby w każdym wierszu to współrzędne x i y). Punkty te modelujemy za pomocą funkcji $F(x) = a \cdot \sin(2x) + b \cdot \sin(3x) + c \cdot \cos(5x) + d \cdot \exp(-x)$.

(a) Znajdź wartości współczynników a-d które najlepiej opisują te dane w sensie metody najmniejszych kwadratów. Rezultat przedstaw graficznie. Rozwiązując to zadanie nie można korzystać z procedur bibliotecznych służących do aproksymacji. Poza tym, użycie procedur z zakresu algebry liniowej jest dozwolone.

(b) Zaproponuj inną funkcję $G(x)$ (która zależy od kilku parametrów) i wygeneruj zbiór punktów w postaci $(x, G(x) + \delta y)$, gdzie δy to losowe zaburzenia. Powtórz dopasowanie z pkt. (a) dla swoich danych i sprawdź, czy udało się odtworzyć wartości ustalonych wcześniej parametrów. Poeksperymentuj zmieniając ilość wygenerowanych punktów i wielkość zaburzeń.



Rysunek 1: Dane do zadania numerycznego NUM8.

Podpunkt a polega na znalezieniu wartości współczynników, które najlepiej odwzorowują podane dane.

W podpunkcie b należy zaproponować inną funkcję, a następnie wygenerować jej zbiór punktów, gdzie do y dodajemy losowe zaburzenie.

2. Instrukcja uruchomienia

Aby uruchomić program należy skorzystać z polecenia:

make runA

make runB

Lub

python3 NUM8.py exercise_a

pyton3 NUM8.py exercise_b

3. Metoda najmniejszych kwadratów

Korzystając z metody najmniejszych kwadratów możemy z przybliżeniem obliczyć współczynniki zadanej funkcji, tak aby ta najlepiej pasowała do zadanych punktów.

Metoda ta polega na dopasowaniu linii, która będzie leżała jak najbliżej wszystkich punktów, tak aby suma kwadratów odległości punktów od linii była jak najmniejsza.

4. Aproksymacja punktowa

Mamy N par punktów $\{(x_i, y_i)\}_{i=1}^{N}$, gdzie x_i jest dokładną wartością argumentu, a y_i jest znaną wartością funkcji.

Każdej zmierzonej (i obarczonej błędem) wartości y_1 odpowiada wartość teoretyczna y_z, jaką zmienna y „powinna" przyjmować dla danej wartości zmiennej x.

Przyjmujemy, że nasza wartość teoretyczna jest kombinacją liniową pewnych znanych funkcji:

$$y_z = a_1 * f_1(x_i) + a_2 * f_2(x_i) + \ldots + a_s * f_s(x_i)$$

Zespół wszystkich wartości teoretycznych możemy przedstawić jako:

$$y_z = Ap$$

Stosując powyższe do przykładu z podpunktu a otrzymujemy:

$$y_z = \begin{bmatrix} sin(2x_1) & sin(3x_1) & cos(5x_1) & exp(-x_1) \\ sin(2x_2) & sin(3x_2) & cos(5x_2) & exp(-x_2) \\ \dots & \dots & \dots & \dots \\ sin(2x_n) & sin(3x_n) & cos(5x_n) & exp(-xn) \end{bmatrix} \begin{bmatrix} a_1 \\ b_2 \\ a_3 \\ a_4 \end{bmatrix}$$

Naszym zadaniem jest policzenie przybliżonego wektora $p = [a_1, a_2, a_3, a_4]^T$ w celu obliczenia współczynników.

Skorzystamy z: $$Ap = y$$

Będzie to nad określony układ równań, który poza wyjątkowymi sytuacjami, nie ma ścisłego rozwiązania. W celu rozwiązania skorzystamy z metody SVD, która dostarcza przybliżonego rozwiązania takich układów, optymalnego w sensie najmniejszych kwadratów.

Równanie przemnożymy z lewej strony o A^T otrzymując:

$$A^T A p = A^T y$$

Następnie przy użyciu SVD przekształcamy obliczając odwrotność $A^T A$ i otrzymujemy:

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

Dzięki tym przekształceniom możemy obliczyć wektor p, który zawiera przybliżone wartości naszych poszukiwanych współczynników.

W podpunkcie b zaproponowałem funkcję:

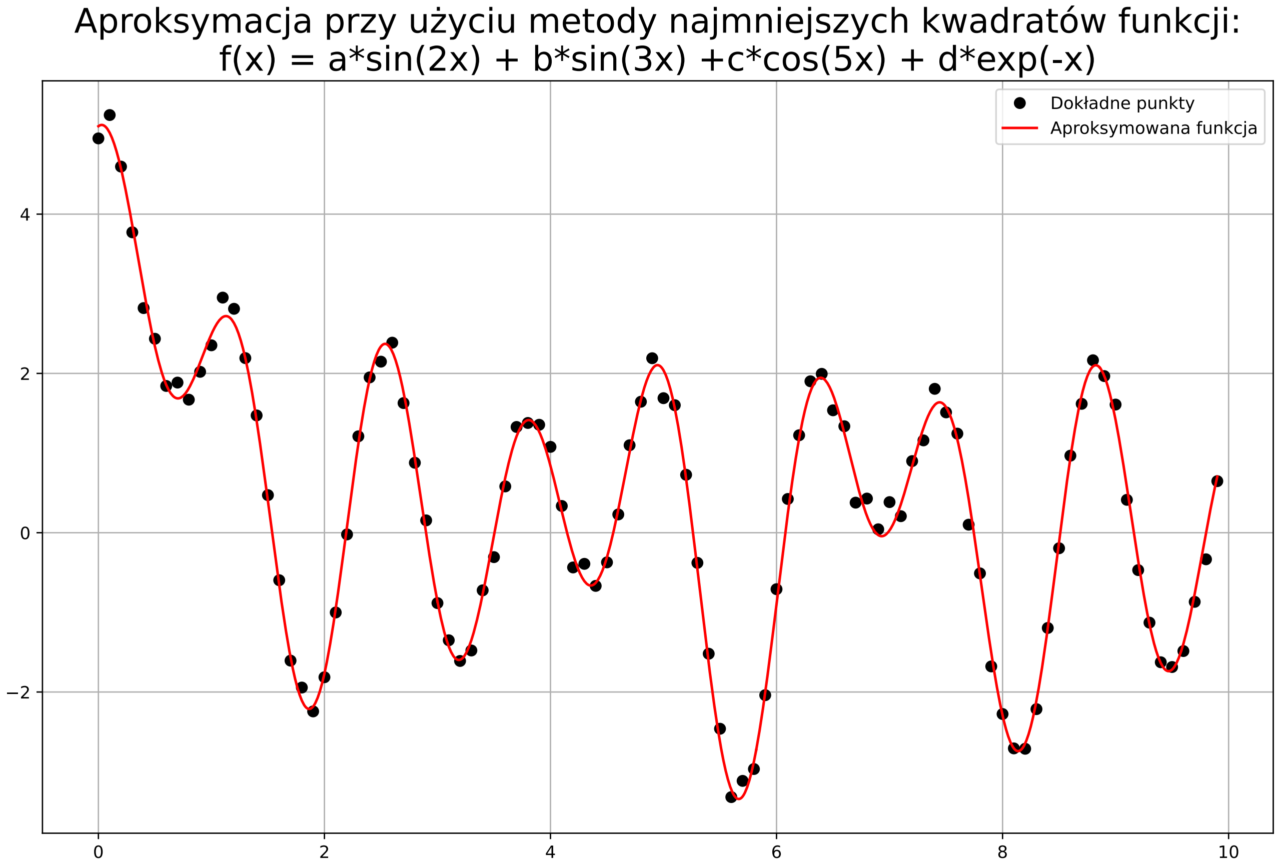
$$g(x) = a * 2x \ + \ b * cos(3x) \ + \ c * cos(5x)$$

I ustaliłem współczynniki: $2, \ 3, \ 1$

Do każdej wartości y dodaję losową liczbę z przedziału $< -1, \ 1 >$ w celu zaburzenia.

5. Wyniki

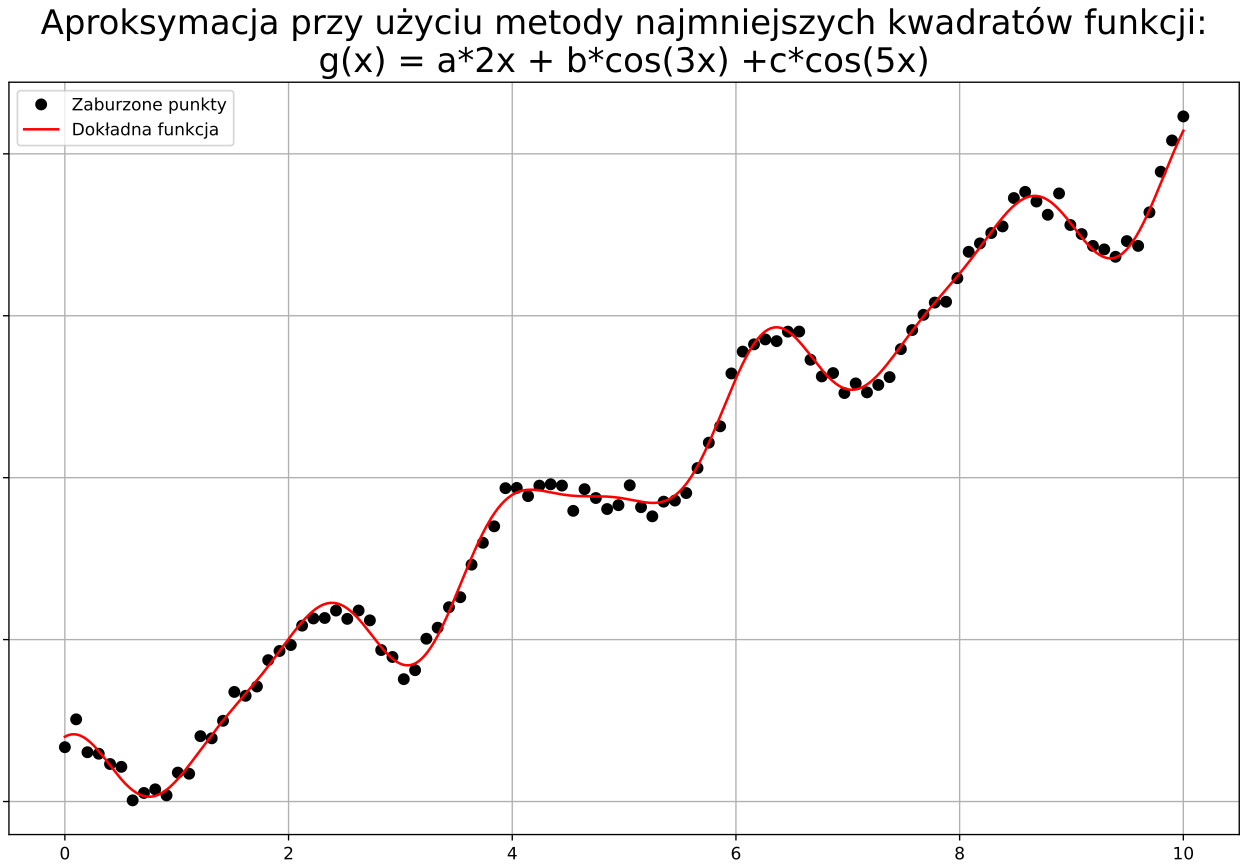
podpunkt a)



Znalezione współczynniki:

$$a = 0.66767124$$
$$b = 1.07293276$$
$$c = 1.69693956$$
$$d = 3.40636597$$

podpunkt b)



Dokładne współczynniki:

$$a = 2, \quad b = 3, \quad c = 1$$

Znalezione współczynniki*:

$$a = 2.00001165$$
$$b = 3.00358837$$
$$c = 1.00642184$$

*(współczynnik były obliczane dla większej ilości punktów niż są przedstawione na wykresie w celu uzyskania lepszej precyzji. Na wykresie liczba punktów jest mniejsza w celach lepszej widoczności wykresu)

6. Wnioski

W zależności od tego, jaka jest gęstość punktów na danym przedziale x, uzyskujemy inną dokładność przybliżenia obliczanych współczynników.

W podpunkcie b, dla 10000 punktów na przedziale od 0 do 10, przybliżone wartości współczynników wynoszą:

$$a = 2.00001165$$
$$b = 3.00358837$$
$$c = 1.00642184$$

Dla porównania, w tym samym podpunkcie, dla 100 punktów na przedziale od 0 do 10, te wartości wynoszą:

$$a = 1.9970751$$
$$b = 3.08662584$$
$$c = 1.13102974$$

Zauważyłem też, że gdy zwiększymy przedział losowania liczby do zaburzeń, a nie zmienimy ilości punktów, to również wtedy otrzymujemy gorszą aproksymację.

Oznacza to, że jeśli używamy większego zakresu liczb do zaburzania, to potrzebujemy też większej ilości punktów w celu uzyskania lepszego wyniku przybliżenia.

Na powyższych przykładach widać, że im gęściej rozłożone są punkty, tym dokładniej jesteśmy w stanie obliczyć te wartości.