Właściwości aproksymacji

Wstęp

W praktyce, w celu poprawy właściwości określonego procesu technologicznego pojawia się konieczność jego zamodelowania, czyli zbudowania opisu matematycznego zachodzących w procesie zjawisk. W obecnych czasach, w dobie powszechnego gromadzenia danych zwykle dysponujemy dużą ilością wartości opisujących zależność między wejściem a wyjściem procesu – dysponujemy parami {*x*, *y*} – szukamy zależności:

y = f(x)

Gdzie *x* opisuje dane wejściowe – np. temperaturę procesu technologicznego, z to inne parametry procesu, natomiast *y* to zarejestrowany parametr wyjściowy – np. moduł Yunga.

Zbudowanie powyższego modelu matematycznego często otwiera nam nowe możliwości np. w opisanym powyżej przykładzie pozwala na wyznaczenie temperatury procesu koniecznej do uzyskania określonego moduły Yunga, przy założeniu niezmienności innych parametrów (z) np. składu chemicznego materiału wejściowego.

Rozwiązaniem postawionego wyżej problemu, czyli znalezienia zależności y = f(x) jest aproksymacja. Jej idea polega na znalezieniu funkcji f(.), a najprostszym przykładem jest aproksymacja wielomianowa. Polega ona na znalezieniu współczynników wielomianu określonego stopnia np.

Aproksymacja liniowa $y \to f(x) = ax + b$ -> parametry to a,b

Funkcją kwadratową $y \to f(x) = a_2x^2 + a_1x + b$ -> parametry konieczne do otworzenia to a_2 a_1 b

Wielomianem n-tego rzędu $y \to f(x) = a_kx^k + a_{k-1}x^{k-1} + \ldots + a_2x^2 + a_1x + b$ gdzie parametry to: a_k ,a_{k-1}, … ,a_2 a_1 b

Zadania

1) Stwórz w matlabie funkcję obliczającą pierwiastek błędu średniokwadratowego (RMSE) dla zadanej pary liczb,

   ```
   function err = RMSE(y,d)
   ```

   gdzie y oraz d to dwie pary wektorów liczb. y – reprezentuje poprawną wartość którą powinniśmy uzyskać, natomiast d to wartość uzyskana z modelu. Błąd liczymy pomiędzy y1(i) a y2(i)
   RMSE oblicza się wg zależności:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - d_i)^2}$$

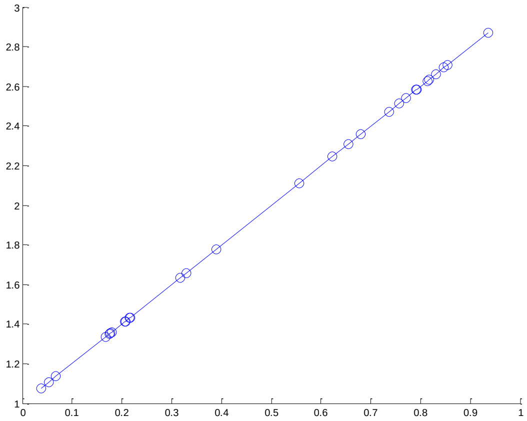
2) Stwórz skrypt który pozwoli na zbadanie właściwości aproksymacji. W pierwszym kroku poddana analizie zostanie prosta funkcja liniowa y = 2*x+1

```
clear; clc; %Wyczyszczenie pamięci
figure(1); clf; hold on; %Przygotowanie do generacji wykresu
x = rand(1,30); %generacja danych wejściowych - 30 próbek
x = sort(x); %Sortowanie danych od najmniejszych do największy - to ułatwi
             generację wykresów
y = 2*x+1; %przykładowa generacja danych wyjściowych.
```

Tak przygotowany skrypt zawiera przygotowany zbiór par danych x oraz y

Narysuj dane na wykresie korzystając z funkcji plot(), uwaga wykres ma przedstawiać jedynie zbiór punktów

3) Dokonaj aproksymacji punktów x,y wielomianem 1 rzędu (funkcja liniowa) korzystając z funkcji polyfit() – funkcja służy do aproksymacji wielomianowej i jako wyjście zwraca zbiór współczynników wielomianu.
Uzyskany wynik nanieś na wykres w postaci krzywej



Przeanalizuj uzyskane współczynniki wielomianu.
Czy uzyskane wsp. wielomianu różnią się względem wielomianu który posłużył do generacji danych y

4) Zwykle rejestrowane dane obarczone są szumem wynikającym zarówno z niedokładności przyrządów pomiarowych, metody pomiarowej jak i różnych innych czynników zakłócających. Dodaj szum do wcześniej wygenerowanych danych y, korzystając z funkcji rand(n,m) – funkcja generuje macierz o wymiarach n x m wartości losowych z rozkładu jednostajnego z przedziału [0,1]
`yy = y + a*(rand(1,length(y)) - 0.5);`
gdzie a to amplituda szumu. Ustaw ją na równą 1
Na wykresie przedstaw uzyskany zbiór punktów zgodnie z zależnością yy = f(x).
Rysunek dołącz do sprawozdania

5) Dokonaj aproksymacji punktów [x,yy] Wyznaczając współczynniki wielomianu 1 stopnia korzystając z funkcji polyfit().
Czy uzyskane wartości różnią się względem wartości uzyskanych w zadaniu 3. Jeśli tak to skąd się biorą różnice?

6) Wyznacz wartości d uzyskane na podstawie wyliczonego w poprzednim zadaniu wielomianu d = f(x) gdzie w miejsce f() wstaw odczytany wielomian.
   Wyznacz błąd RMSE(yy,d) – błąd dopasowania wielomianu do punktów.
   Ile wynosi jego wartość?
   Wyznacz błąd RMSE pomiędzy wynikami d, a oryginalnymi wynikami bez szumu RMSE(y,d). **Jaka jest wielkość tego błędu?**
   Porównaj obydwa błędy między sobą. Skąd się biorą różnice? Który jest mniejszy i dlaczego?
7) Stwórz nowy skrypt podobny do zadania 2, ale dane y wygeneruj nie z funkcji liniowej tylko kwadratowej. Samemu dobierz wsp. funkcji kwadratowej

8) Dokonaj aproksymacji wielomianem 1 stopnia jak w zadaniu 3. Stwórz wykres i umieść go w sprawozdaniu. Dla ułatwienia i wyznaczenia wartości d-f(x) możesz skorzystać z funkcji d = polyval(a,x), która na podstawie wsp wielomianu zapisanych z zmiennej a, wyznacza wartości d dla danych wartości x. **Policz i podaj błąd RMSE**
9) Powtórz zadanie 8 korzystając z wielomianu 2 stopnia. Odczytaj wsp. wielomianu i porównaj je z wsp. których użyłeś do wyznaczenia wartości y. **Czy wsp. wielomianów różnią się? Wyznacz i podaj błąd RMSE**
10) Powtórz zadanie 9 korzystając z wielomianu 10 stopnia. **Czy uzyskane wsp. wielomianu różnią się względem wyników uzyskanych w zad 9?**
    Policz i podaj błąd dopasowania RMSE.
11) Do danych y dodaj szum, podobnie jak w zadaniu 4.
12) Powtórz zadania 8 do 11 dokonując aproksymacji danych z szumem. Każdorazowo wyznacz błąd RMSE pomiędzy RMSE(yy,d) oraz RMSE(y,d). Pamiętaj że w pierwszym przypadku liczysz błąd dopasowania do danych które aproksymowałeś, natomiast w drugim przypadku jest to błąd do danych bez szumu – dopasowanie do ideału. **Zanotuj i skomentuj uzyskane wyniki. Zwróć uwagę co się dzieje z obydwoma błędami w zależności od rzędu wielomianu. Jak wpływa wyższy rząd wielomianu na dokładność dopasowania.**
13) Stwórz nowy jeden skrypt, który pozwoli na zbadanie wpływ liczby par {x,y} (liczby danych) na dokładność aproksymacji. W tym celu:
    a. Wygeneruj 30 danych losowych (x) jak w zadaniu 7
    b. Wyznacz wartości (y) dla zadanej przez siebie funkcji kwadratowej i danych (x)
    c. Stwórz zmienną (yy) dodając do (y) szum jak w zad 4.
    d. Dokonaj aproksymacji pary {yy,x} wielomianem 10 stopnia. I korzystając z funkcji polyval() wyznacz wartości d. Oblicz błędy dopasowania RMSE(yy,d) oraz RMSE(y,d)
    e. Wygeneruj nowy zestaw punktów losowych jak w podpunkcie a (wygeneruj 100 punktów) i zapisz do zmiennej xx
    f. Dokonaj złączenia danych x = [x xx];
    g. Powtórz podpunkty b do d
    h. Powtórz podpunkty e do g, ale teraz wygeneruj do zmiennej xx 1000 punktów.

    Jak wpływa liczba punktów na błędy RMSE? Który z błędów maleje i dlaczego