Akademia Górniczo-Hutnicza
im. Stanisława Staszica w Krakowie

AGH University of Science
and Technology

Metody numeryczne

Janusz Miller

Katedra Informatyki Stosowanej
AGH University of Science and Technology

9.11.2021

AGH

www.agh.edu.pl

Część VI

Aproksymacja

Terminologia

- Łac. *approximare* — przybliżać. Zazwyczaj byty (np. rozwiązania, funkcje) **skomplikowane** aproksymuje się bytami (rozwiązaniami, funkcjami) **prostszymi**.
- U nas: **Aproksymacja** – model relacji (zależności) zmiennej zależnej od określonych zmiennych niezależnych.
- W statystyce mówi się o **regresji** jako metodzie badania związku między zmienną **objaśnianą** a zmiennymi **objaśniającymi**.

Kiedy w praktyce pojawia się potrzeba aproksymacji

Chcemy zbadać ewentualny związek między zmiennymi – wielkościami np. fizycznymi, biologicznymi, ekonomicznymi, społecznymi itp.

Uwaga: Nie określamy między nimi relacji przyczyna – skutek.

Co należy zrobić?

Kiedy w praktyce pojawia się potrzeba aproksymacji

Etapy badania związku

- Zebrać dane ilościowe, np. pomiary badanych wielkości.
- Rozstrzygnąć, czy istnieje między nimi statystycznie istotna zależność – analiza korelacji.
- Jeżeli istnieje, to zbudować model zależności
  - jakościowo – strukturę,
  - ilościowo – wyznaczyć wartości parametrów (współczynników modelu).
- Przeprowadzić testy zgodności pomiarów z hipotezą (modelem), np. χ^2.

Kiedy w praktyce pojawia się potrzeba aproksymacji

Miejsce aproksymacji

Ten wykład obejmuje tylko niektóre techniki obliczeniowe stosowane w przypadku aproksymacji rozumianej jako budowanie modelu na podstawie zebranych już i sprawdzonych danych.

Pozostałe elementy – akwizycja danych, analiza statystyczna – na innych przedmiotach.

Aproksymacja, interpolacja, ekstrapolacja

- W szerszym znaczeniu aproksymacji – interpolacja funkcji też jest aproksymacją tej funkcji.
  Aproksymacja jest pojęciem szerszym niż interpolacja.
- Określenie „***curve fitting***” – dopasowywanie albo dobieranie krzywej — stosuje się w tym szerszym znaczeniu aproksymacji (tzn. obejmuje też interpolację).
  W dalszej części będziemy używać terminu aproksymacja w węższym znaczeniu.
- Ekstrapolacja – tworzenie modelu zależności dla obszaru poza zakresem danych.

Aproksymacja funkcji vs. aproksymacja dyskretna

Aproksymacja funkcji – (a. integralna)

- dana funkcja aproksymowana $f(x)$,
- szukana funkcja $P_n(x; a_0, a_1, \ldots, a_n)$, która minimalizuje

$$e_2(a_0, a_1, \ldots, a_n) = \int_a^b (f(x) - P_n(x))^2 \, dx, \qquad (1)$$

Aproksymacja dyskretna – (a. punktowa)

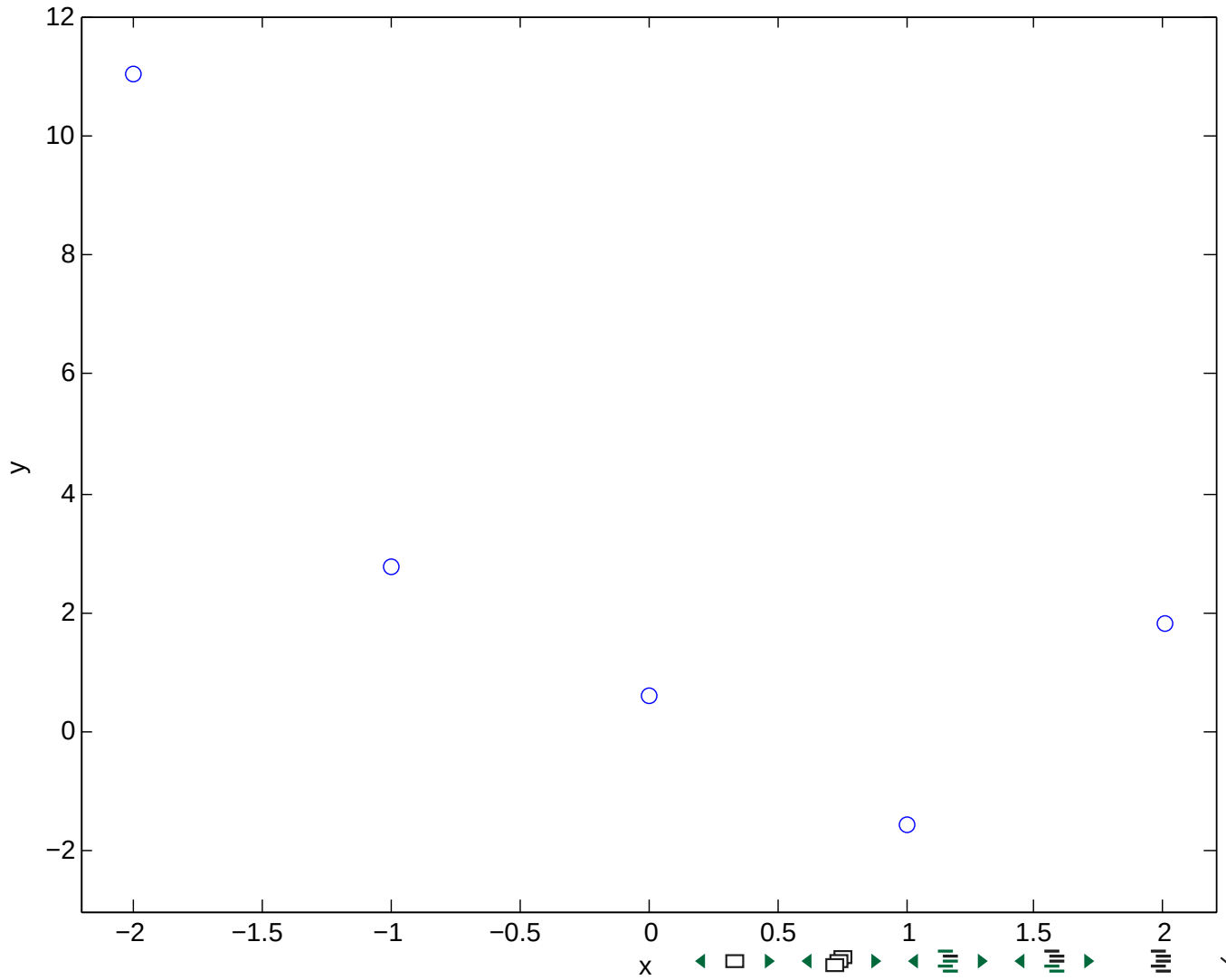
- dany zbiór punktów (danych aproksymowanych) $\{(x_0, y_0), \ldots, (x_m, y_m)\}$,
- szukana funkcja $P_n(x; a_0, a_1, \ldots, a_n)$, która minimalizuje [1]

$$E_2(a_0, a_1, \ldots, a_n) = \sum_{i=0}^{m} (y_i - P_n(x_i))^2 . \qquad (2)$$

[1] Wyrażenia (1) i (2) są przykładowymi wskaźnikami jakości aproksymacji.

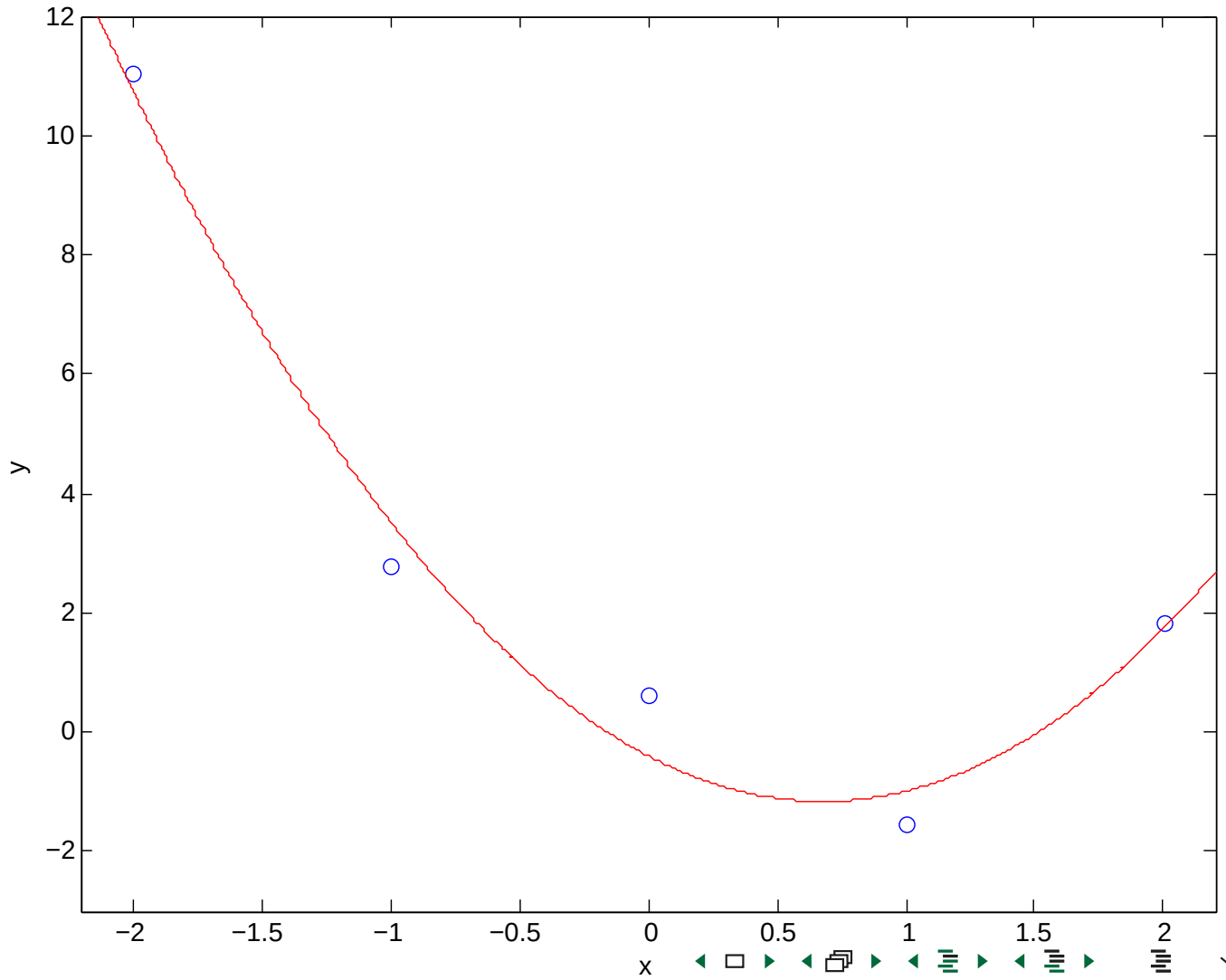
Przykład aproksymacji dyskretnej

– dane aproksymowane ... i funkcja aproksymująca



Przykład aproksymacji dyskretnej

– dane aproksymowane ... i funkcja aproksymująca



Aproksymacja funkcji vs. aproksymacja dyskretna c.d.

Konstruowanie kwadratur, znaczenie błędu danych

Kwadratury Newtona–Cotesa, Gaussa, (Eulera – Maclaurina) i wiele innych były konstruowane wg schematu:

1. Obliczyć wartości funkcji podcałkowej (i ew. jej pochodnej) w ściśle określonych punktach.
2. Wyznaczyć wielomian **interpolacyjny** Lagrange'a (ew. Hermite'a).
3. Obliczyć całkę oznaczoną z otrzymanego wielomianu.

Podsumowując: Funkcję podcałkową „przybliżano" funkcją **interpolującą**. Dlaczego interpolującą? Należy zauważyć, że:

- Funkcja „przybliżająca" nie minimalizuje wskaźnika całkowego (1).
- Uzasadnione jest wyznaczanie funkcji **interpolującej, a nie aproksymującej**, gdy dane wyznaczane są z pełną dokładnością, tj. nie ma **błędu** danych.

Kiedy interpolacja, a kiedy aproksymacja

- Wybór interpolacji (zamiast aproksymacji) jest uzasadniony, gdy mamy pełną i dokładną (jakościową i ilościową) wiedzę o zależności, tzn.
  1. wiemy co od czego zależy,
  2. ilościowo możemy tę zależność dokładnie wyznaczyć.

  Np. Zależność wyrażoną wzorem algebraicznym.
- W kontraście do przykładu kwadratur – „klasyczna" aproksymacja dyskretna zakłada
  1. istnienie ukrytych, nieuwzględnionych zależności[2],
  2. istnienie „błędu" czy „niepewności" danych aproksymowanych.

[2] Odkrywanie tych zależności jest jednym z zadań sztucznej inteligencji

W praktyce: Skąd „niepewność”

– danych i rodzaju zależności

Zazwyczaj nie dysponujemy wiedzą o wszystkich czynnikach wpływających na zmienną objaśnianą.

W szczególności – czasem wiemy o zależności od jakiegoś czynnika, ale:

- mamy za mało pomiarów tego czynnika,
- nie mamy dostępu do danych,
- nie został zmierzony,
- nie da się go zmierzyć.

W praktyce: Skąd „niepewność”

Przypadek idealny

Ideałem byłoby uwzględnienie **wszystkich czynników** determinujących zmienną zależną.
Wszystkie inne zależności można byłby traktować jako losowe.
Jeżeli

- składniki losowe oscylują wokół zera,
- ich wariancje są stałe,
- nieskorelowane w czasie,
- nieskorelowane ze zmiennymi objaśniającymi,
- mają rozkład normalny.

to możemy składniki zakłócające pojmować jako generowane przez proces białego szumu.

Rozkład normalny

Funkcja gęstości prawdopodobieństwa dla rozkładu normalnego

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right).$$

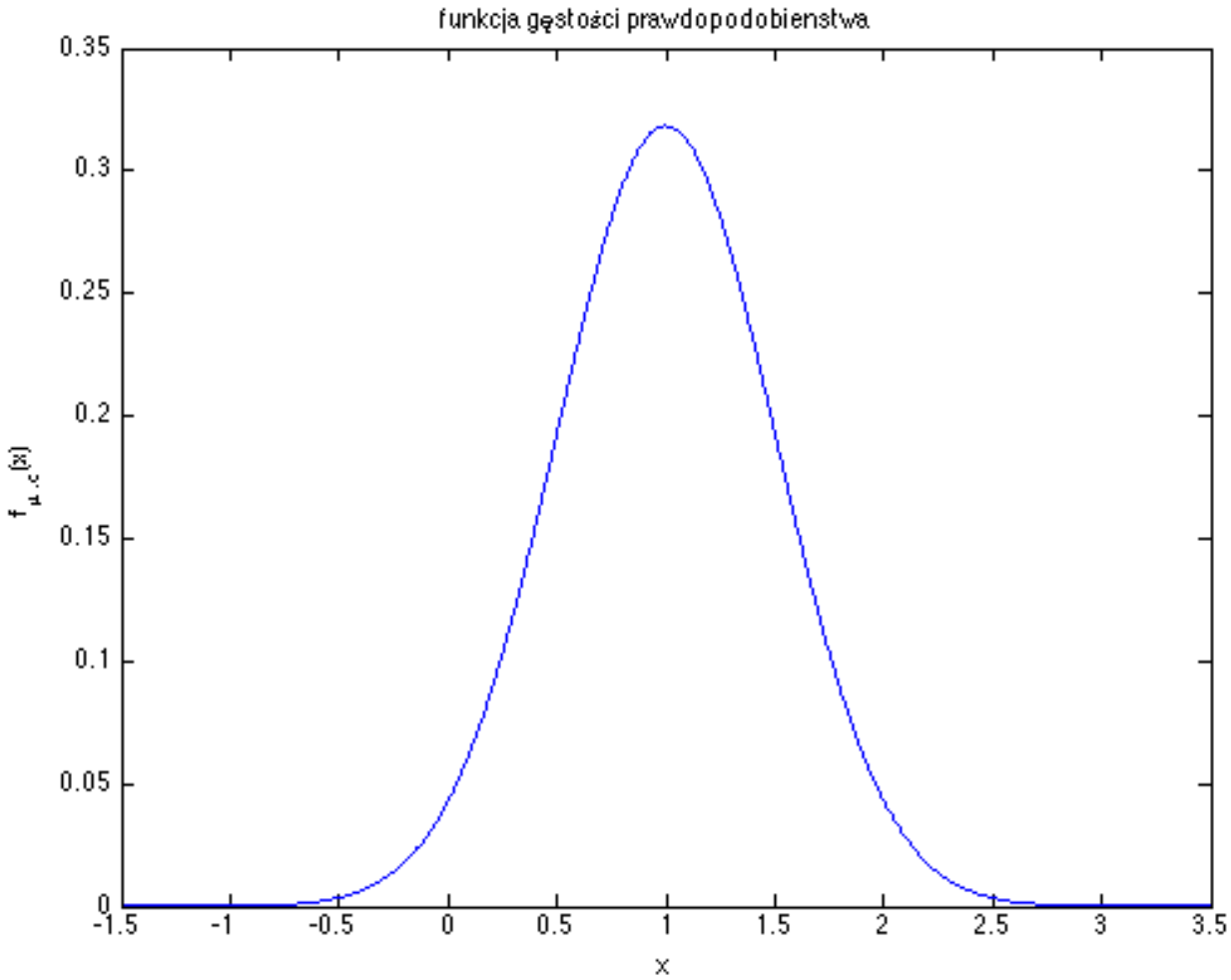
μ – wartość średnia,
σ – odchylenie standardowe.

Centralne twierdzenie graniczne:

Jeśli jakaś wielkość jest sumą lub średnią **bardzo wielu drobnych losowych czynników**, to niezależnie od rozkładu każdego z tych czynników jej **rozkład będzie zbliżony do normalnego**.

Rozkład normalny

dla $\mu = 1, \sigma = 0.5$:



Zamiast ideału – praktyka

Aby jakaś zależność wyłoniła się z szumu „zakłóceń” – konieczna jest duża liczba próbek (pomiarów, danych) – każdej z uwzględnianych w modelu zmiennej objaśniającej (niezależnej).

Problem:
Które zmienne objaśniające możemy pominąć w modelu?

Do rozstrzygnięcia, czy istnieje statystycznie istotna relacja (zależność) między zmienną objaśniającą a objaśnianą, służą testy odpowiednich hipotez.

Przykłady

Przykłady:

- Model zapotrzebowania na energię elektryczną regionu w ciągu dnia.
- Długość życia mieszkańca danego regionu.
- Codzienny popyt na artykuły w sklepie.
- Liczebność populacji np. śledzi w Bałtyku.
- Kursy akcji na giełdzie.

Znaczenie poprawnej akwizycji danych

Zbieranie i obróbka danych – najbardziej pracochłonna, wciąż nie do końca zautomatyzowana:

- Obróbka statystyczna - sprawdzanie hipotez statystycznej zależności między zmiennymi, ich korelacji własnej i wzajemnej.
- Decyzje w sprawie np.:
  - liczby zmiennych objaśniających,
  - klasy funkcji aproksymującej,
  - wskaźnika jakości aproksymacji,
  - współczynnika wygładzania,
  - progu błędów grubych.

Klasyfikacja aproksymacji dyskretnej

(w węższym znaczeniu)

- Wg kryterium postaci funkcji aproksymującej:
  - Aproksymacja parametryczna
    (w tym wielomianowa, trygonometryczna, wymierna itp.)
  - Aproksymacja nieparametryczna
  - Aproksymacja funkcjami sklejanymi – ma cechy aproksymacji parametrycznej i nieparametrycznej.
- Wg kryterium wymiaru przestrzeni danych:
  - 2D – *curve fitting*,
  - 3D – *surface fitting*,
  - itd.

Metody obliczeniowe dla aproksymacji - plan

1. Aproksymacja parametryczna
   a. jako minimalizacja kwadratu błędu przybliżenia,
   b. z perspektywy probabilistycznej – regresja,
   c. problem doboru liczby parametrów i wygładzanie,
   d. inne przypadki aproksymacji dyskretnej.
2. Aproksymacja nieparametryczna – przykłady.

1. Aproksymacja parametryczna

a. Jako minimalizacja błędu przybliżenia

Trzeba podjąć dwie decyzje:

- Wybór normy błędu aproksymacji
- Wybór klasy funkcji aproksymującej

1. Aproksymacja parametryczna

a. Jako minimalizacja błędu przybliżenia

- Jeżeli jako normę błędu aproksymacji przyjmiemy **normą średniokwadratową**, to będzie to **aproksymacja średniokwadratowa**.
- Jeżeli ponadto funkcja aproksymująca będzie **liniowo zależna od jej współczynników**, to zadanie aproksymacji średniokwadratowej można sprowadzić do **zagadnienia liniowego** z nadokreślonym układem równań.

1. Aproksymacja parametryczna

a. Jako minimalizacja błędu przybliżenia

Przykład aproksymacji wielomianem stopnia n

- Ustalenia początkowe:

  Dane aproksymowane – zbiór punktów: $\{(x_k, y_k)\}, \quad k = 0, 1, \ldots, m$

  Funkcja aproksymująca: wielomian stopnia $n, (n < m)$: $f_n(x) = a_0 + a_1 x + \ldots + a_n x^n$

  Norma aproksymacji (błąd aproksymacji): $E_2(a_0, a_1, \ldots, a_n) = \sum_{k=0}^{m} [y_k - f_n(x_k)]^2$

- Szukane:

  Współczynniki funkcji aproksymującej $\{a_0, a_1, \ldots, a_n\}$, dla których norma błędu aproksymacji jest najmniejsza.

1. Aproksymacja parametryczna

a. Jako minimalizacja błędu przybliżenia

Zapisanie warunków zgodności danych i funkcji aproksymującej prowadzi do układu równań dla $k = 0, 1, \ldots, m$:

$$y_k - (a_0 + a_1 x_k + \ldots + a_n x_k^n) = 0,$$

a w postaci macierzowej $\mathbf{Xa} = \mathbf{y}$, czyli

$$\begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ 1 & x_3 & x_3^2 & \cdots & x_3^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}.$$

W przypadku typowym dla aproksymacji $n < m$ jest to **nadokreślony** układ równań liniowych, z reguły – **sprzeczny**.

1. Aproksymacja parametryczna

a. Jako minimalizacja błędu przybliżenia

Dyskretna aproksymacja średniokwadratowa a rozwiązywanie nadokreślonego układu równań liniowych

Plan na następne 5 minut:

1. Jak w postaci macierzowej zapisać warunek konieczny minimalizacji błędu aproksymacji średniokwadratowej.
2. Sprowadzenie problemu aproksymacji średniokwadratowej do rozwiązania układu równań liniowych z kwadratową macierzą współczynników.

1. Aproksymacja parametryczna

a. Jako minimalizacja błędu przybliżenia

1. Istnienie minimum — dowód pomijamy.
2. Warunek konieczny:
   Zerowanie się wszystkich pochodnych cząstkowych $\frac{E(a)}{da} = 0$ czyli $\frac{\partial E(a_i)}{\partial a_i} = 0$ dla $i = 0, 1, \ldots, n$.
3. Jeżeli norma jest klasy co najmniej C^1, (a $||\ ||_2$ jest), to otrzymujemy układ równań normalnych

$$\frac{\partial}{\partial a_i} \sum_{k=0}^{m} [y_k - f_n(x_k)]^2 =$$

$$= \frac{\partial}{\partial a_i} \sum_{k=0}^{m} [y_k - (a_0 + a_1 x_k + \ldots + a_n x_k^n)]^2$$

$$= 2 \sum_{k=0}^{m} [y_k - (a_0 + a_1 x_k + \ldots + a_n x_k^n)] \left(-x_k^i\right) = 0$$

1. Aproksymacja parametryczna

a. Jako minimalizacja błędu przybliżenia

Układ równań normalnych w postaci macierzowej

Po przekształceniu ostatniej postaci otrzymujemy układ równań dla $i = 0, 1, \ldots, n$

$$a_0 \sum_{k=0}^{m} x_k^i + a_1 \sum_{k=0}^{m} x_k^{i+1} + \cdots + a_n \sum_{k=0}^{m} x_k^{i+n} = \sum_{k=0}^{m} x_k^i y_k,$$

który w postaci macierzowej można zapisać w postaci

$$\mathbf{X}^\mathsf{T}\mathbf{X}\mathbf{a} = \mathbf{X}^\mathsf{T}\mathbf{y},$$

1. Aproksymacja parametryczna

a. Jako minimalizacja błędu przybliżenia – przykład

Dla n=2, m=3:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ x_0 & x_1 & x_2 & x_3 \\ x_0^2 & x_1^2 & x_2^2 & x_3^2 \end{bmatrix} \begin{bmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ x_0 & x_1 & x_2 & x_3 \\ x_0^2 & x_1^2 & x_2^2 & x_3^2 \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$\mathbf{X}^{\dagger} = \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}$$

1. Aproksymacja parametryczna

Dygresja o numerycznym rozwiązywaniu układu równań liniowych

Jak znaleźć rozwiązanie układu równań

$$\mathbf{X}^\mathsf{T}\mathbf{X}\mathbf{a} = \mathbf{X}^\mathsf{T}\mathbf{y}\,:$$

- analitycznie: $\mathbf{a} = \mathbf{X}^\dagger\mathbf{y}$, gdzie $\mathbf{X}^\dagger = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\,\mathbf{X}^\mathsf{T}$ – macierz pseudoodwotna Moore'a-Penrose'a,
- metodą eliminacyjną Gaussa,
- najlepiej: zauważyć, że $\mathbf{X}^\mathsf{T}\mathbf{X}$ jest macierzą symetryczną $\longrightarrow$ metodą Choleskiego (przez rozkład $\mathbf{L}\mathbf{L}^\mathsf{T}$),
- MATLAB: używając operatora $\backslash$ „dzielenia przez macierz" $\mathbf{a} = (\mathbf{X}^\mathsf{T}\mathbf{X})\backslash(\mathbf{X}^\mathsf{T}\mathbf{y})$.

1. Aproksymacja parametryczna

Inna funkcja aproksymująca, ale nadal liniowo zależna od współczynników (parametrów)

Co w ww. algorytmie trzeba zmienić w przypadku, gdy zamiast wielomianu wybierzemy inną funkcję liniowo zależną od jej parametrów $f(x; a_0, a_1, \ldots, a_n)$?

- W wielomianie funkcjami bazowymi były funkcje potęgowe.
- Do macierzy $\mathbf{X}$ należy wpisać wartości nowych funkcji bazowych obliczonych dla kolejnych odciętych punktów danych.

1. Aproksymacja parametryczna

Podsumowanie aproksymacji średniokwadratowej

Zagadnienie aproksymacji dyskretnej możemy sprowadzić do rozwiązania **układu równań liniowych** z kwadratową macierzą współczynników **jeżeli**:

1. Zależność funkcji aproksymującej od jej parametrów jest liniowa, np.:
   - wielomian,
   - szereg funkcji trygonometrycznych $f(x) = a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx)$,
   - rzadziej - szereg funkcji wykładniczych $f(x) = \sum_{k=0}^{n} a_k e^{kx}$,
   - itp.
2. Jakość (błąd) aproksymacji mierzymy normą $\| \ \|_2$.

Ogólnie - pochodna błędu aproksymacji względem parametrów funkcji aproksymującej musi **liniowo** zależeć od tych parametrów.

1. Aproksymacja parametryczna

W kontraście – Aproksymacja (regresja) nieliniowa

1. Funkcja aproksymująca o nieliniowej zależności od parametrów:
   - Padé approximant of order [m/n]
     $$R(x) = \frac{\sum_{j=0}^{m} a_j x^j}{1+\sum_{k=1}^{n} b_k x^k} = \frac{a_0+a_1x+a_2x^2+\cdots+a_mx^m}{1+b_1x+b_2x^2+\cdots+b_nx^n},$$
   - $f(x) = \sum_{k=0}^{n} a_k \exp(b_k x)$,
   - funkcje będące liniową kombinacją RBF (radial basis functions), przykład radialnej funkcji bazowej: $\Phi(r) = \exp\left(-\varepsilon r^2\right)$, $r = \|\mathrm{x} - \mathrm{x}_i\|$.
2. Inne stosowane normy (terminologia stosowana w statystyce):
   - Mean Absolute Error (MAE):
     $E_1(a_0, a_1, \ldots, a_n) = \frac{1}{m+1} \sum_{k=0}^{m} |y_k - f_n(x_k)|$,
   - minimax problem:
     $E_\infty(a_0, a_1, \ldots, a_n) = \max_{k=0,1,\ldots,m} \{|y_k - f_n(x_k)|\}$

1. Aproksymacja dyskretna

Dygresja o danych uczących i testowych

W praktyce, dane dla aproksymacji dzieli się na dwa podzbiory:

- **zbiór treningowy (uczący)** – te dane służą do wyznaczenia „próbnego" modelu (funkcji aproksymującej),
- **zbiór testowy** – dla weryfikacji, sprawdzenia własności modelu (innych niż minimalizowany błąd aproksymacji), np. gładkości modelu.

Przy tworzeniu podzbiorów należy unikać regularności.

1. Aproksymacja parametryczna

b. Z perspektywy probabilistycznej – regresja[3]

- **Regresja** – metoda estymowania wartości oczekiwanej zmiennej zależnej y przy znanych wartościach innej zmiennej (lub zmiennych) niezależnych x.
- Celem regresji jest przewidywanie wartości zmiennej zależnej dla zadanej, nowej wartości zmiennej niezależnej (na podstawie wcześniej poznanego zbioru treningowego par wartości zmiennej niezależnej i zależnej).
- Treningowe wartości zmiennych zależnych są obarczone błędem o pewnym rozkładzie prawdopodobieństwa.
- Jeżeli rozkład błędów jest rozkładem normalnym (Gaussa) to **suma kwadratów błędów jest funkcją największej wiarygodności** (*reliability*).

[3] Ch.M.Bishop: Pattern Recognition and Machine Learning. Springer 2006. Legalnie dostępne w internecie – podrozdział 1.2.5 Curve fitting re-visited.

1. Aproksymacja parametryczna

c. Wygładzanie funkcji aproksymującej i dobór stopnia wielomianu aproksymującego

Zwiększanie liczby parametrów funkcji aproksymującej monotonicznie **zmniejsza błąd** i prowadzi do interpolacji.

Inna interpretacja błędu w interpolacji, inna w aproksymacji.

- W interpolacji nie powinno być niezgodności funkcji interpolującej i danych. Błąd interpolacji to niezgodność poza punktami danych.
- W aproksymacji dane, z założenia, są obarczone błędem. Zadaniem aproksymacji jest „odseperowanie" badanej zależności od wpływu zakłóceń.
- Błąd aproksymacji nie musi mieć znaczenia pejoratywnego.

1. Aproksymacja parametryczna

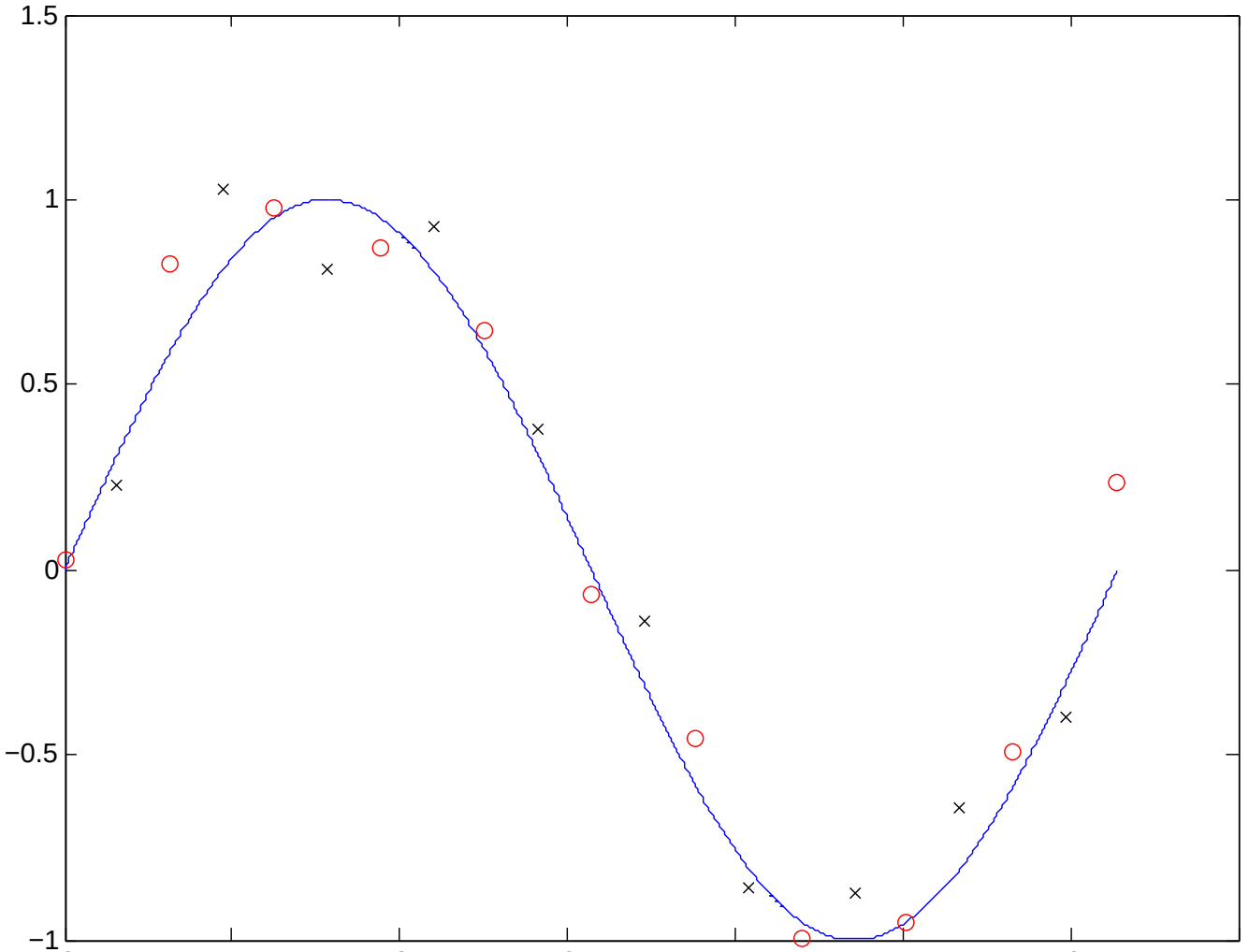
c. Przykład problemu doboru stopnia wielomianu aproksymującego

Przykład prób aproksymacji wielomianami stopnia coraz wyższego

- Tworzymy abstrakcyjne dane dla aproksymacji.
- Zakładamy, że zmienna objaśniana zależy od zmiennej objaśniającej wg funkcji sinus.
- Dane dla aproksymacji zostały zmierzone z błędem o równomiernym rozkładzie losowym.
- Dane zostały podzielone na równoliczne podzbiory: treningowy (zaznaczony kółkami) i testowy (zaznaczony krzyżykami).
- Na kolejnych wykresach są: dane oraz wielomiany aproksymujące stopni: 3, 5, 8, i 10.

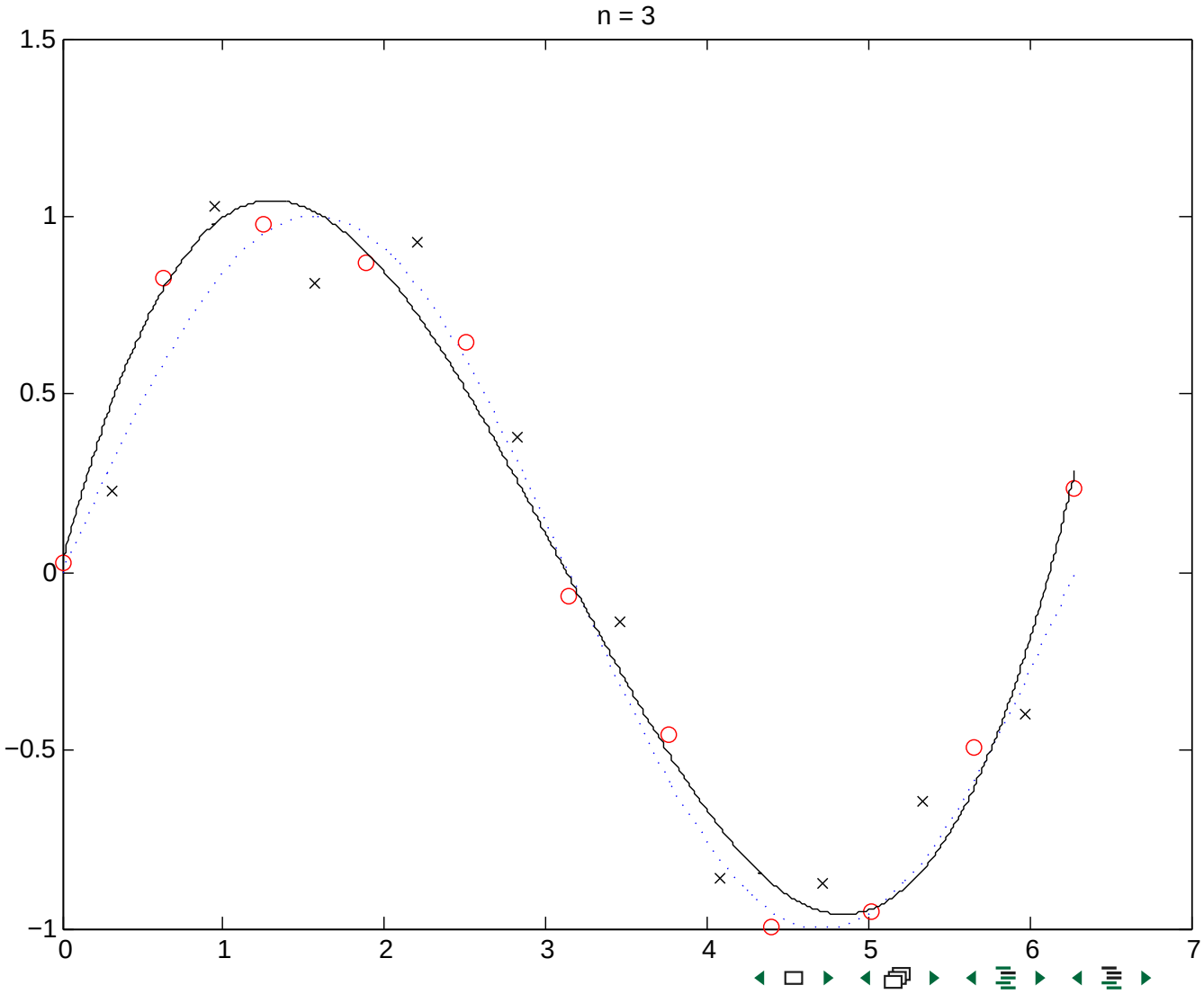
1. Aproksymacja parametryczna

c. Porównanie aproksymacji wielomianami coraz wyższych stopni



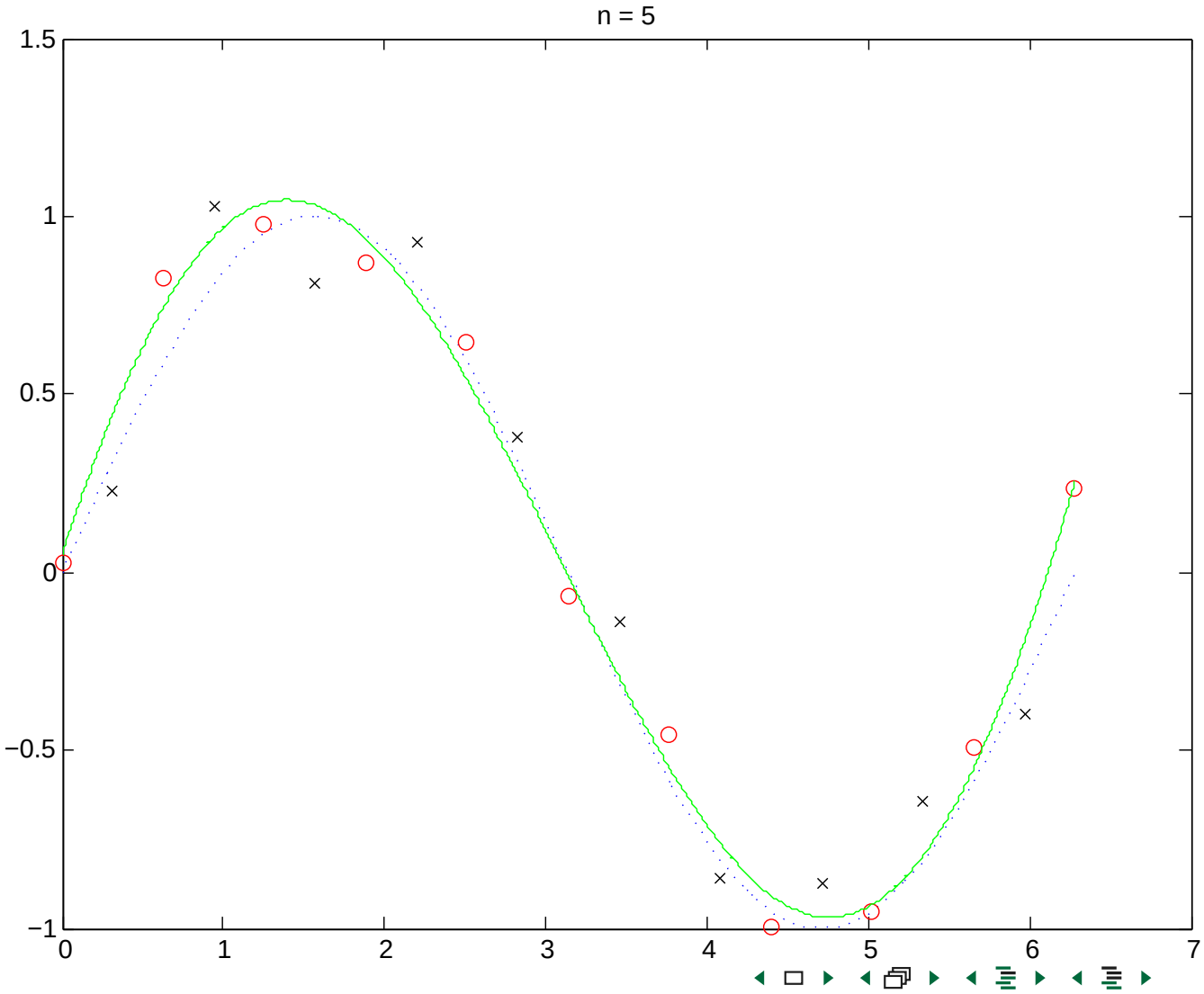
1. Aproksymacja parametryczna

c. Porównanie aproksymacji wielomianami coraz wyższych stopni



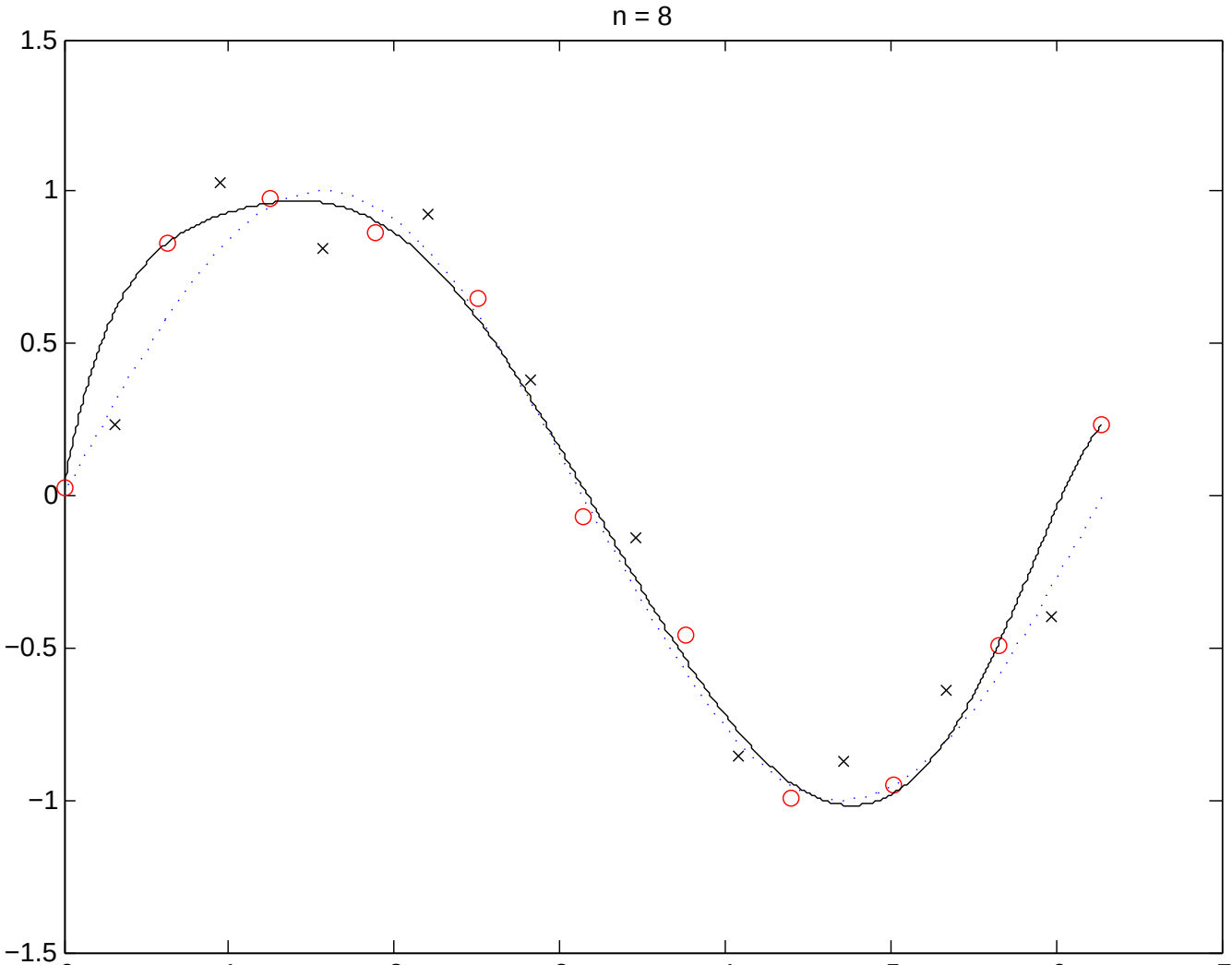
1. Aproksymacja parametryczna

c. Porównanie aproksymacji wielomianami coraz wyższych stopni



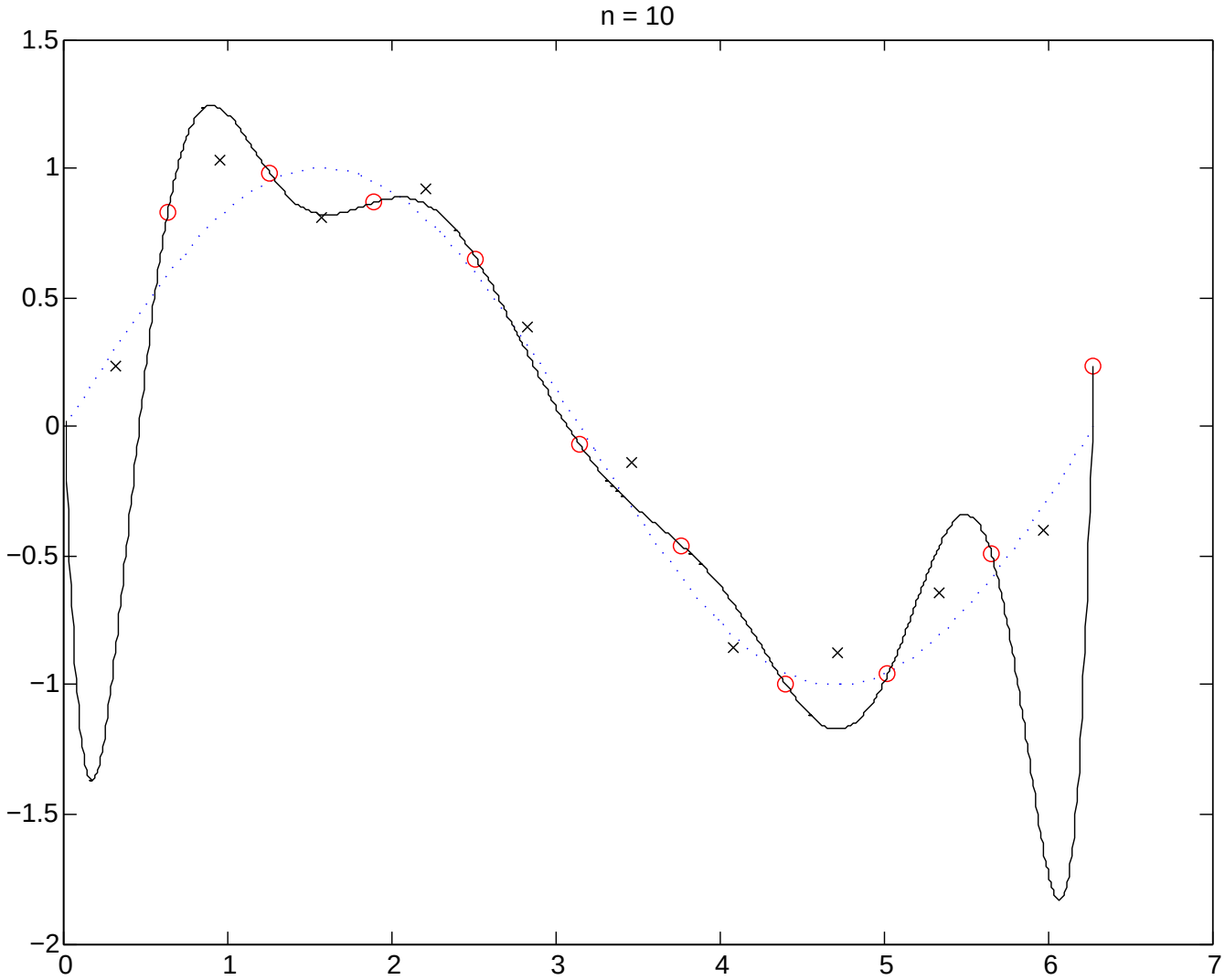
1. Aproksymacja parametryczna

c. Porównanie aproksymacji wielomianami coraz wyższych stopni



1. Aproksymacja parametryczna

c. Porównanie aproksymacji wielomianami coraz wyższych stopni



1. Aproksymacja parametryczna

c. Wygładzanie funkcji aproksymującej

Idea:

- Brzytwa Ockhama (XIV w.): „Nie warto czynić przy pomocy wielu tego, co daje się zrobić przy pomocy jednego”.
- Hamilton (XIX w.):
  „Bytów nie należy mnożyć bez konieczności”,

Realizacja:

- Zwiększanie stopnia wielomianu dopóki nie wzrasta błąd dla danych testujących.
  W praktyce często stosowana **walidacja krzyżowa**.
- Modyfikacja miary błędu

$$\tilde{E}_2(\mathrm{a}) = \frac{1}{m+1}\sum_{k=0}^{m}\left[y_k - f_n(x_k)\right]^2 + \lambda\|\mathrm{a}\|_2,$$

λ – współczynnik wagowy

1. Aproksymacja parametryczna

c. Nadmierne dopasowanie – *overfitting* – kilka uwag

Kiedy ma miejsce **„nadmierne dopasowanie”** („przeuczenie”):

- gdy liczba stopni swobody modelu przekracza zawartość informacyjną danych,
- skala błędu modelu jest mała w porównaniu z oczekiwanym poziomem szumu w danych,
- uczeń może ”wymyślić” prawidłowości, które w rzeczywistości nie mają miejsca, a są efektem przypadkowych błędów w danych uczących.

Ciekawostka: W psychiatrii odpowiednikiem nadmiernego dopasowania mogą być urojenia paranoiczne.

2. Aproksymacja nieparametryczna

Powyższe przypadki to **aproksymacja parametryczna** – polega na wyznaczaniu najlepszego zbioru **parametrów** funkcji aproksymującej.
Popatrzmy szerzej:

- Jeżeli celem aproksymacji jest
  - znalezienie zależności (funkcji) $\longrightarrow$ parametryczna,
  - znalezienie wartości w wybranym punkcie $\longrightarrow$ nieparametryczna.
- Jeżeli liczba danych aproksymowanych i liczba zmiennych jest
  - mała $\longrightarrow$ stosujemy raczej parametryczną,
  - duża $\longrightarrow$ zastosujemy nieparametryczną.

Aproksymację funkcjami sklejanymi można traktować jako „rozwiązanie pośrednie”.

2. Przykłady aproksymacji nieparametrycznej

Średnia krocząca.

Filtrowanie średnią kroczącą

- Zbiór danych ma N par (x_i, y_i), $i = 1, 2, \ldots, N$.
- Dana y_i w punkcie x_i jest zastępowana wartością filtrowaną $\bar{y}_i$

$$\bar{y}_i = \frac{1}{2n+1}\left(y_{i-n} + y_{i-n+1} + \ldots + y_i + \ldots + y_{i+n-1} + y_{i+n}\right)$$

- W pobliżu granic zbioru, dla $i \leqslant n$ przyjmuje się $n = i - 1$ oraz dla $i \geqslant N$, $n = N - i$.

2. Przykłady aproksymacji nieparametrycznej

Regresja lokalna LOWESS i LOESS.

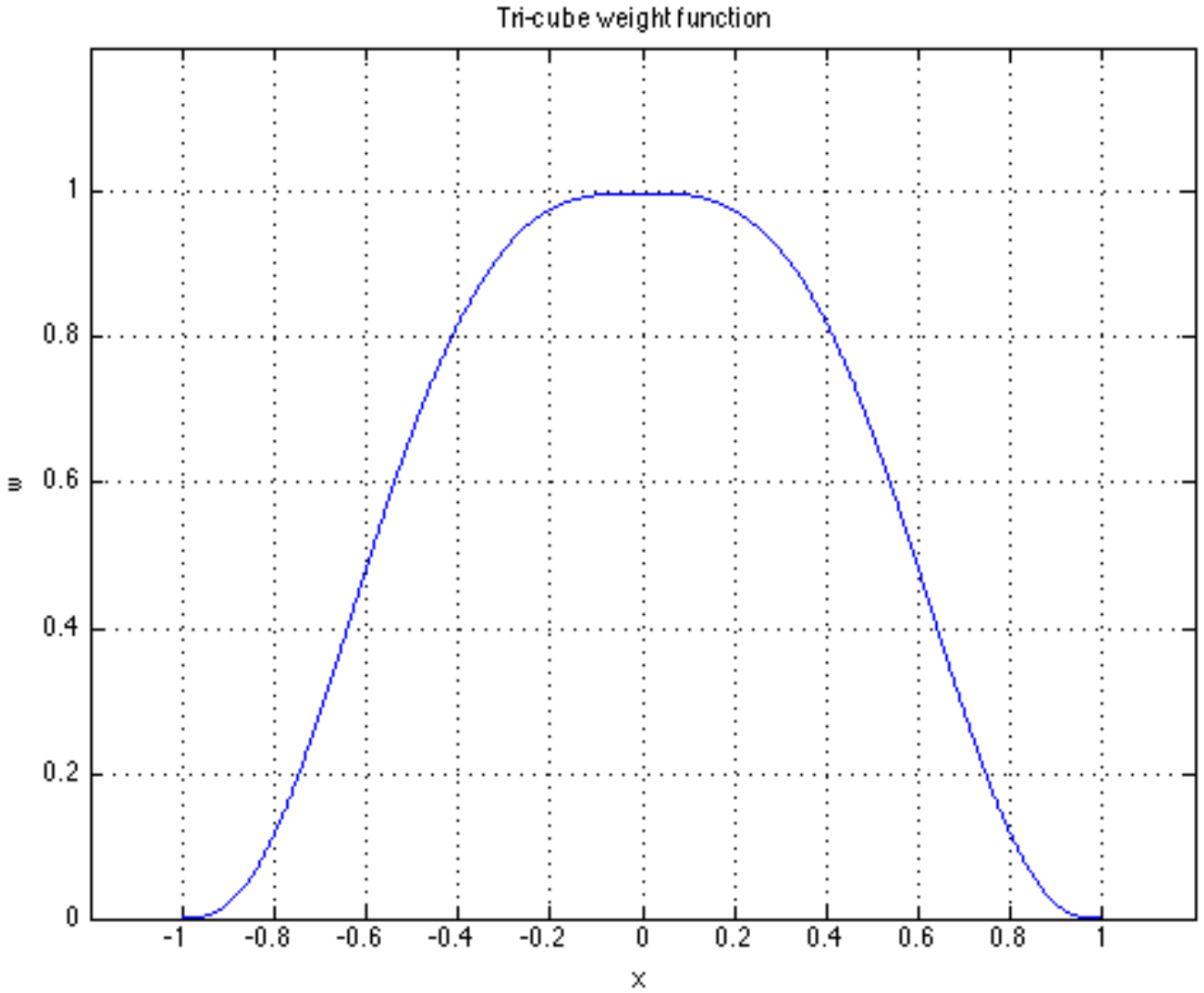
AGH

LOWESS – „locally weighted scatterplot smoothing”.

- cel - wyznaczenie wartości przybliżonej (oczekiwanej) w wybranym punkcie x_i,
- należy wyznaczyć lokalny podzbiór danych S_i (dane nie muszą tworzyć regularnej siatki),
- stopień wielomianu lokalnej regresji $p_i(x)$: 0 – 2,
- funkcja wagowa: *tri-cube weight function*
  $w(x) = \left(1 - \left|\frac{x - x_i}{d_i(x)}\right|^3\right)^3$, $d_i(x)$ - średnica podzbioru S_i,
- dane z S_i z wagą w są aproksymowane wielomianem p_i,
- wynik: wartość wielomianu $p_i(x_i)$.

2. Przykłady aproksymacji nieparametrycznej

Regresja lokalna LOWESS - funkcja wagowa



2. Przykłady aproksymacji nieparametrycznej

Robust Local Regression

W przypadku konieczności odfiltrowania danych „odstających" (*outliers*) stosuje się np. „krzepką / odporną" regresję lokalną.

Etapy::

- LOWESS,
- obliczenia residuów r_i dla każdej danej,
- obliczenie mediany M residuów,
- wyznaczenie dodatkowej wagi dla danych

$$w_i = \begin{cases} \left(1 - \left(\frac{r_i}{6M}\right)^2\right)^2, & r_i < 6M \\ 0, & r_i \geqslant 6M \end{cases}$$

- powtórna regresja lokalna.