

Sprawozdanie z laboratorium nr 6
Aproksymacja liniowa i nieliniowa.

Emilia Mączka, Marcin Sawczuk, Daniel Warloch, Weronika Wisz

Spis treści

1	Aproksymacja liniowa - dane dokładne (niezaburzone błędem losowym:	4
1.1	Który algorytm wyznaczania parametrów funkcji aproksymującej jest najdokładniejszy?	4
1.2	Który algorytm należy wybrać dla aproksymacji liniowej ale funkcją inną niż wielomian?	4
1.3	Wielomian którego stopnia daje najmniejszy błąd aproksymacji?	4
1.4	Czy im wyższy stopień wielomianu tym mniejszy błąd? Czy błąd aproksymacji zmaleje do zera?	5
2	Aproksymacja liniowa - dane z błędem losowym	5
2.1	Czy odległość punktów (danych) od wielomianu aproksymującego zawsze maleje ze wzrostem stopnia wielomianu?	5
2.2	Aproksymacja liniowa - dane realne. Wielomian którego stopnia najlepiej przybliży zmiany temperatury w czasie dnia? Czy decyduje wielkość błędu aproksymacji?	5
2.3	Czym się kierować przy doborze stopnia wielomianu aproksymującego gdy nie mamy wiedzy o aproksymowanej zależności?	5
3	Jak wykryć nadmierne dopasowanie	6
3.1	Jaki stopień wielomianu należy wybrać widząc rezultat kilku eksperymentów (uruchomień skryptu)? Odpowiedź nie musi być liczbą – może być przedziałem.	6
3.2	Czy zastosowany w ćwiczeniu podział danych na zbiór ćwiczebny i testowy jest zgodny z regułami?	7
3.3	Ewentualnie: Dodać więcej danych i zaproponować lepszy podział.	7
3.4	Ilościowe rozpoznawanie nadmiernego. Jak sformułować prosty algorytm właściwego doboru stopnia wielomianu - a ogólniej - liczby parametrów funkcji aproksymującej? Uwaga: użyłem słowa „prosty” dla zaznaczenia, że algorytm może porównywać wartości liczbowe, a nie np. kształty wykresów.	8
3.5	Na czym polega eksperyment walidacji krzyżowej?	8
4	Problem wyboru klasy funkcji aproksymującej	8
4.1	Czym się kierować w przypadku danych z pliku dane_przeplywu.m?	8
4.2	Które funkcje można brać pod uwagę:	8
4.3	Które z powyższych można zastosować w aproksymacji liniowej? Ew. jakie zmiany w algorytmie w stosunku do algorytmu dla wielomianów?	9
4.4	Czy możliwe jest użycie funkcji wymiernej w zadaniu aproksymacji liniowej? Ew. z jakimi ograniczeniami (tzn. czy każdą funkcję wymierną)?	9

4.5	Aproksymacja nieliniowa - zalety i wady w stosunku do liniowej (krótkie wyliczenie)	9
-----	--	---

1 Aproksymacja liniowa - dane dokładne (niezaburzone błędem losowym:

1.1 Który algorytm wyznaczania parametrów funkcji aproksymującej jest najdokładniejszy?

Po uruchomieniu skryptu "algorytmy.m" widzimy, że dla aproksymacji wielomianami niższych stopni błędy numeryczne algorytmów wyznaczania optymalnych wartości parametrów funkcji aproksymującej są na podobnym poziomie. Zwiększając jednak stopień wielomianu obserwujemy, że największy błąd jest dla algorytmu z macierzą pseudoinwersji Moore'a-Penrose'a i obliczaniem macierzy odwrotnej, następnie z wykorzystaniem matlabowego „lewego dzielenia” przez macierz współczynników $X^T X$, a najniższy błąd gwarantuje algorytm z wykorzystaniem bibliotecznej funkcji polyfit. Dla dwóch pierwszych algorytmów pojawia nam się również ostrzeżenie: "Warning: Matrix is close to singular or badly scaled. Results may be inaccurate.", czyli możemy wykonywać obliczenia na źle uwarunkowanych macierzach, co może prowadzić do dużych błędów.

Wniosek: Najdokładniejszy jest algorytm z wykorzystaniem bibliotecznej funkcji polyfit.

1.2 Który algorytm należy wybrać dla aproksymacji liniowej ale funkcją inną niż wielomian?

Należy wybrać algorytm z wykorzystaniem bibliotecznej funkcji polyfit. Rozważamy aproksymację liniową, czyli funkcja kosztu zależy liniowo od parametrów funkcji. Niezależnie od tego jaką funkcją będziemy aproksymować zadaniem algorytmu jest optymalizacja parametrów tej funkcji, czyli rozwiązujemy wielomian. Funkcja polyfit zapewnia nam najmniejszy błąd numeryczny, co jest opisane w punkcie powyżej.

1.3 Wielomian którego stopnia daje najmniejszy błąd aproksymacji?

W skrypcie "generator_danych_1.m" sprawdzamy dla jakiego stopnia wielomianu błąd będzie najmniejszy. Maksymalny stopień wielomianu jaki rozważamy to $n=15$ i dla niego właśnie obserwujemy najmniejszy błąd $=6.4302e-12$. Jeżeli mamy dane niezaburzone to zwiększając stopień wielomianu aproksymującego zmniejszamy błąd aproksymacji. Gdybyśmy jednak mieli dane zaburzone zwiększanie stopnia wielomianu nie zapewniałoby nam zmniejszenia błędu i mogłoby powodować jego zwiększanie.

1.4 Czy im wyższy stopień wielomianu tym mniejszy błąd? Czy błąd aproksymacji zmaleje do zera?

Tak jak było napisane w punkcie wyżej, w przypadku danych niezaburzonych zwiększając stopień wielomianu aproksymującego zmniejszamy błąd aproksymacji. Błąd aproksymacji nie zmaleje jednak do zera ze względu na obecność błędu numerycznego.

2 Aproksymacja liniowa - dane z błędem losowym

2.1 Czy odległość punktów (danych) od wielomianu aproksymującego zawsze maleje ze wzrostem stopnia wielomianu?

Odległość punktów od wielomianu aproksymującego zawsze maleje wraz ze wzrostem stopnia wielomianu. Im stopień wyższy, tym łatwiej przybliżyć funkcję, sprawić, by przechodziła przez dane punkty. Z drugiej strony zbyt duży stopień wielomianu może doprowadzić do nadmiernego dopasowania.

2.2 Aproksymacja liniowa - dane realne. Wielomian którego stopnia najlepiej przybliży zmiany temperatury w czasie dnia? Czy decyduje wielkość błędu aproksymacji?

Wielomian stopnia 4-5 jest najlepszy. Wielkość błędu aproksymacji nie decyduje o tym. Mały błąd nie oznacza, że przybliżenie jest dobre, wskazuje on jedynie na to, że odległość wykresu od punktów jest mała. Należy kierować się wykresem. Stopień 4 odpowiednio aproksymuje zależność, wykres jest dość zbliżony do punktów. Niepotrzebna jest idealna aproksymacja, wystarczy odpowiedni kształt wykresu. Brak oscylacji i w miarę dobre przybliżenie jest wystarczające.

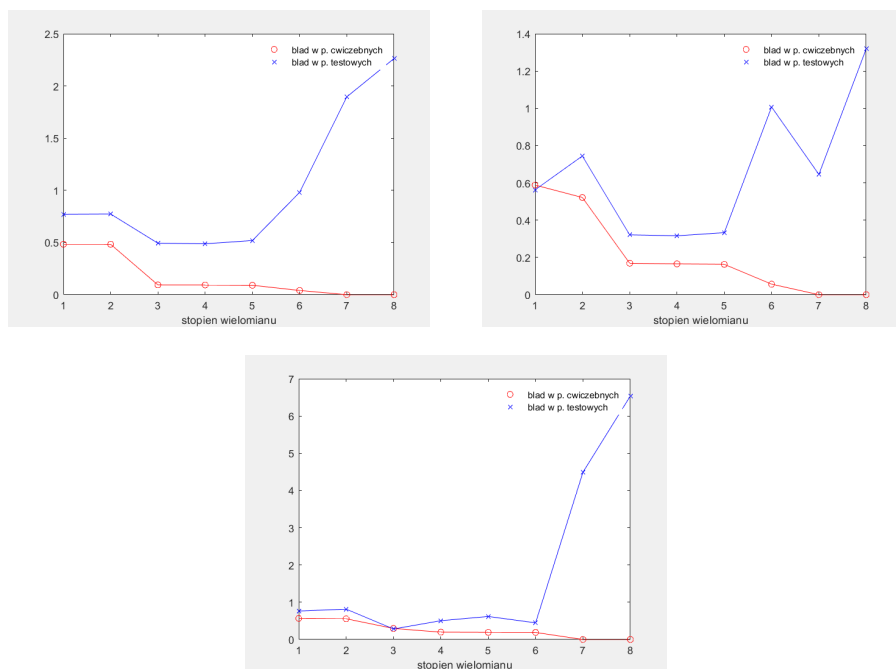
2.3 Czym się kierować przy doborze stopnia wielomianu aproksymującego gdy nie mamy wiedzy o aproksymowanej zależności?

Wielomian powinien dobrze przybliżać, uogólniać dane. Jego stopień nie powinien być za duży, gdyż zwiększając go, funkcja lepiej przechodzi przez punkty, ale niekoniecznie powoduje to zmniejszenia się błędu. Dochodzi wręcz do nadmiernego dopasowania danych, co w efekcie nie da najlepszego rozwiązania. Jednocześnie wybierając wielomian o za niskim stopniu, nie otrzymujemy odpowiednio dopasowanego wykresu, trudno będzie go dostosować do posiadanych danych. Należy dobrać taki wielomian, który jedynie zachowa odpowiedni kształt wykresu.

3 Jak wykryć nadmierne dopasowanie

3.1 Jaki stopień wielomianu należy wybrać widząc rezultat kilku eksperymentów (uruchomień skryptu)? Odpowiedź nie musi być liczbą – może być przedziałem.

W związku z losowością naszych danych, aproksymacja może się różnie zachowywać - należy wybrać przedział.



Rysunek 1: Wykresy odległości wielomianu od zbioru ćwiczebnego i testowego.

Na podstawie wykresów możemy przyjąć, że najlepszym wyborem stopnia wielomianu będzie przedział $< 3; 5 >$.

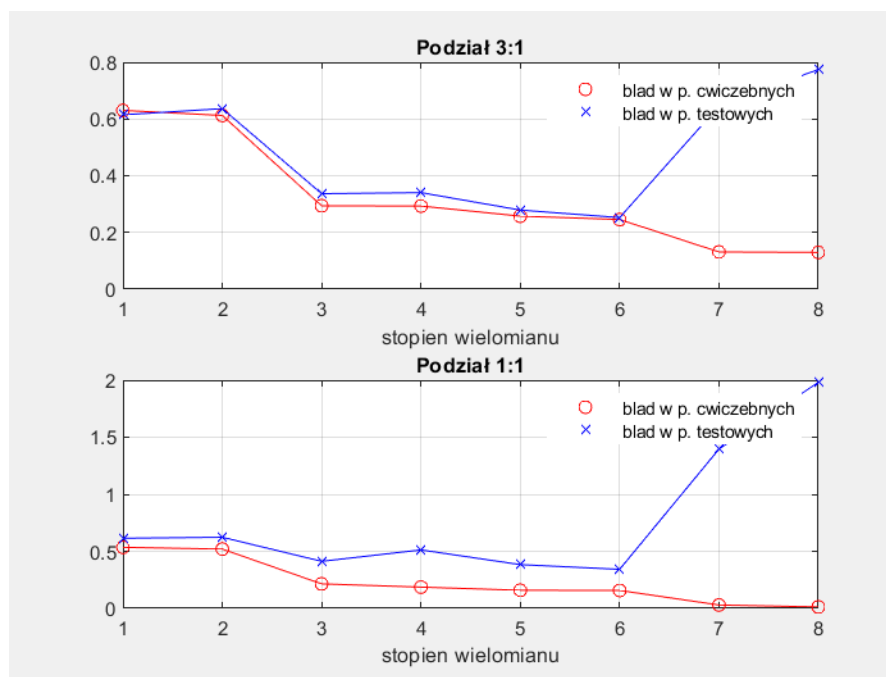
3.2 Czy zastosowany w ćwiczeniu podział danych na zbiór ćwiczebny i testowy jest zgodny z regułami?

Istnieją 2 reguły podziału danych:

1. Nie występują regularności – aby dane testowe były nieznanne tj. abyśmy nie mogli przyjąć z góry najlepszego algorytmu.
2. Danych ćwiczebnych jest 3-4 razy więcej niż testowych – jak najwięcej danych potrzebnych do uczenia i wystarczająca ich liczba do testowania

Natomiast w naszym ćwiczeniu występują regularności, a dane są w stosunku 1:1, więc nie spełnia reguł.

3.3 Ewentualnie: Dodać więcej danych i zaproponować lepszy podział.



Rysunek 2: Wykresy odległości wielomianu od zbioru ćwiczebnego i testowego przy podziale 3:1 i 1:1.

Z wykresu wynika, że lepszym podziałem danych jest podział 3:1.

3.4 Ilościowe rozpoznawanie nadmiernego. Jak sformułować prosty algorytm właściwego doboru stopnia wielomianu - a ogólniej - liczby parametrów funkcji aproksymującej? Uwaga: użyłem słowa „prosty” dla zaznaczenia, że algorytm może porównywać wartości liczbowe, a nie np. kształty wykresów.

Używany przez nas podział pozwala nam na co prawda na uzyskanie zbioru uczącego i testowego, ale nie możemy mieć przy nim pewności, że losowość nie dokonała się w nim w szczęśliwy dla klasyfikatora sposób, który wykaże jego wyższą jakość. Dla bezpieczeństwa i większej rzeczowości naszych badań musimy dokonać wielokrotnego podziału. Teoretycznie moglibyśmy wielokrotnie dokonać tego rodzaju podziału, aby zdobyć wiele par testowych i uczących, ale w takim podejściu istnieje duże prawdopodobieństwo występowania tych samych próbek w różnych zbiorach uczących. Metodą, która niweluje tego rodzaju problem jest walidacja krzyżowa.

3.5 Na czym polega eksperyment walidacji krzyżowej?

W podejściu krzyżowym dzielimy losowo zbiór danych na n możliwie równych sobie podzbiorów. Uzyskujemy dzięki temu n par, w których każdy z podzbiorów raz występuje jako zbiór uczący, a pozostała, połączona część zbioru jest wykorzystywana jako zbiór testowy.

4 Problem wyboru klasy funkcji aproksymującej

4.1 Czym się kierować w przypadku danych z pliku `dane_przeplywu.m`?

W przypadku danych z tego pliku najlepiej kierować się kształtem uzyskanego wykresu. Przypomina on wykres funkcji wykładniczej, co mówi nam, że najłatwiej będzie użyć funkcji wykładniczej oraz aproksymacji nieliniowej.

4.2 Które funkcje można brać pod uwagę:

Aby mieć do czynienia z aproksymacją nieliniową, powinniśmy skorzystać na przykład z jednej z następujących funkcji:

1. $\exp(at)$
2. $a + \exp(bt)$
3. $a \cdot \exp(bt - c)$

4. $\frac{\sum_{i=0}^n a_i t^i}{\sum_{i=0}^m b_i t^i}$ **
5. $\sum_{k=0}^n a_k \exp(b_k t)$

** - Tę funkcję można wykorzystać jedynie pod warunkiem, gdy nie jest możliwe wyciągnięcie parametru przed nawias, a następnie jego skrócenie.

4.3 Które z powyższych można zastosować w aproksymacji liniowej? Ew. jakie zmiany w algorytmie w stosunku do algorytmu dla wielomianów?

W tym przypadku mamy do czynienia z układem równań nieliniowych, z tego powodu nie możemy skorzystać z algorytmu polyfit.

4.4 Czy możliwe jest użycie funkcji wymiernej w zadaniu aproksymacji liniowej? Ew. z jakimi ograniczeniami (tzn. czy każdą funkcję wymierną)?

Tak, możliwe jest wykorzystanie funkcji wymiernej w zadaniu z aproksymacją liniową, lecz musi ona spełniać następujący warunek: Zależność od parametrów musi być liniowa.

4.5 Aproksymacja nieliniowa - zalety i wady w stosunku do liniowej (krótkie wyliczenie)

Główną zaletą metody nieliniowej jest możliwość wykorzystania jej do przedstawiania skomplikowanych układów, w postaci wielu nieskomplikowanych funkcji co znacząco upraszcza obliczenia.

Niestety opisana wyżej zaleta jest także jej wadą, ponieważ powoduje ona znaczący wzrost kosztu obliczeniowego w porównaniu do aproksymacji liniowej.