WINF, lab. Nr 14. Sem. 2025Z

MATLAB (cz. 3) & PYTHON

ESTYMACJA WSPÓŁCZYNNIKA TRANSMISJI KORONAWIRUSA

Opracowanie: Maciej Linczuk, maciej.linczuk@pw.edu.pl, Grzegorz Nieradka, grzegorz.nieradka@pw.edu.pl

1. Wstęp teoretyczny

W ramach ćwiczenia dokonamy estymacji współczynnika transmisji koronawirusa oraz będziemy przewidywać rozwój pandemii na podstawie danych z Wikipedii.

Celem ćwiczenia jest implementacja wzorów podanych w tym wstępie, a nie analiza zaproponowanej metody estymacji.

Niech $x[n]$ oznacza zdiagnozowaną ilość zakażeń danego dnia numer *n*. Przyjmujemy założenie, że każdego kolejnego dnia wykrywamy k krotnie więcej zakażeń niż dnia poprzedniego. Liczbę k nazywamy współczynnikiem transmisji wirusa. Jeżeli k >1 to pandemia rozwija się i przybywa zakażonych, jeżeli k <1 to pandemia zanika, ubywa zakażonych. Czyli:

$$\begin{cases} x[2] = kx[1] \\ x[3] = kx[2] \\ \quad .... \\ x[N] = kx[N-1] \end{cases} \tag{1}$$

Chcemy znaleźć rozwiązanie powyższego układu równań. Oczywiście nie istnieje k spełniające ten układ równań, ponieważ ten układ jest sprzeczny (więcej równań niż niewiadomych). Poszukujemy liczby $\tilde{k}$ która najlepiej spełnia ten układ równań.

Poszukujemy $\tilde{k}$ która minimalizuje funkcję błędu $ERk(k)$:

$$\begin{aligned} \tilde{k} &= \arg\min_k \left(ERk(k)\right) \\ ERk(k) &= \left(kx[1]-x[2]\right)^2 + \left(kx[2]-x[3]\right)^2 + \ldots + \left(kx[N-1]-x[N]\right)^2 \\ ERk(k) &= \sum_{n=1}^{N-1}\left(kx[n]-x[n+1]\right)^2 \end{aligned} \tag{2}$$

Aby wyznaczyć minimum tej funkcji policzę pochodną po k tej funkcji, a następnie przyrównam do zera.

$$\begin{aligned} \frac{\partial ERk(k)}{\partial k} &= \sum_{n=1}^{N-1} 2\left(kx[n]-x[n+1]\right)x[n] = 0 \\ 2k\sum_{n=1}^{N-1} x^2[n] &= 2\sum_{n=1}^{N-1} x[n]x[n+1] \end{aligned} \tag{3}$$

Z równania (3) wynika, że liczba $\tilde{k}$ jest równa:

$$\tilde{k} = \frac{\sum_{n=1}^{N-1} x[n]x[n+1]}{\sum_{n=1}^{N-1} x^2[n]} \tag{4}$$

Wzór (4) możemy wykorzystać jako estymator współczynnika transmisji wirusa. Obliczona wartość $\tilde{k}$ jest estymatą współczynnika transmisji wirusa dla naszych danych.

Za pomocą wzoru (4) możemy obliczyć estymatę $\tilde{k}$ współczynnika transmisji wirusa. Powstaje pytanie: jak wyznaczyć estymatę ilości zakażeń danego dnia. Liczby $x[n]$ to wartości – ilości wykrytych przypadków danego dnia. Okazuje się, że wykorzystując liczby $x[n], n = 1,..,N$ oraz model rozprzestrzeniania się wirusa zapisany w (1) możemy dokładniej wyznaczyć ilość zakażeń danego dnia. W tym celu modyfikuję układ równań (1) do następującej postaci:

$$\begin{cases} x[2] = k^1 \tilde{x}[1] \\ x[3] = k^2 \tilde{x}[1] \\ \quad .... \\ x[N] = k^{N-1} \tilde{x}[1] \end{cases} \tag{5}$$

Symbol $\tilde{x}[n]$ oznacza estymatę prawdziwej ilości zakażeń n-tego dnia pandemii. Rozwiązanie układu równań (5) nie istnieje, możemy , jak poprzednim razem, znaleźć rozwiązanie przybliżone, minimalizujące funkcję błędu:

$$\begin{gathered} \tilde{x}[1] = \arg\min_{x[1]}(ERx) \\ ERx = \left(k^1\tilde{x}[1] - x[2]\right)^2 + \left(k^2\tilde{x}[1] - x[3]\right)^2 + ... + \left(k^{N-1}\tilde{x}[1] - x[N]\right)^2 \\ ERx = \sum_{n=1}^{N-1}\left(k^n\tilde{x}[1] - x[n+1]\right)^2 \end{gathered} \tag{6}$$

Minimalizuję ten błąd licząc pochodną po $\tilde{x}[1]$ którą następnie przyrównuję do zera.

$$\begin{gathered} \frac{\partial ERx}{\partial \tilde{x}[1]} = \sum_{n=1}^{N-1} 2\left(k^n\tilde{x}[1] - x[n+1]\right)\cdot k^n = 0 \\ 2\sum_{n=1}^{N-1} k^n \cdot k^n \cdot \tilde{x}[1] = 2\sum_{n=1}^{N-1} k^n \cdot x[n+1] \end{gathered} \tag{7}$$

Z równania (7) wynika następujące metoda estymacji ilości zakażeń pierwszego dnia pandemii:

$$\tilde{x}[1] = \frac{\sum_{n=1}^{N-1} k^n \cdot x[n+1]}{\sum_{n=1}^{N-1} k^n \cdot k^n} \tag{8}$$

Wykorzystując $\tilde{k}$ oraz estymatę ilości zakażeń pierwszego dnia $\tilde{x}[1]$ wyznaczam ilość zakażeń w kolejnych dniach zgodnie ze wzorem:

$$\tilde{x}[2] = k^{1}\tilde{x}[1]$$
$$\tilde{x}[3] = k^{2}\tilde{x}[1]$$
$$\ldots$$
$$\tilde{x}[N] = k^{N-1}\tilde{x}[1] \qquad (9)$$

Wykorzystując wzory (4), (8) i (9) można wyznaczyć współczynnik transmisji wirusa oraz dokładniejsze wartości liczby zakażeń danego dnia. Warto dodać, że powyższe wzory są słuszne tylko przy założeniu modelu rozprzestrzeniania się pandemii zgodnie ze wzorem (1). W przypadku nie spełnienia modelu (1) powyższe obliczenia nie mają zastosowania.

2. Zadania do wykonania – MATLAB (2 ptk)

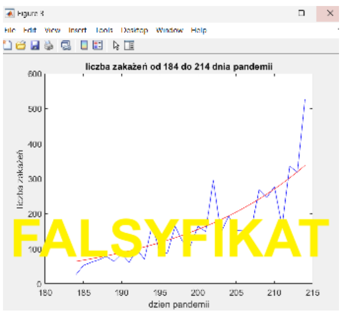
W pliku covid1.txt znajduje się liczba wykrytych przypadków koronawirusa w Województwie Mazowieckim w kolejnych dniach. Pierwsza liczba to liczba przypadków z dnia 08.03.2020. Do analizy wykorzystamy dane od 184 do 214 dnia pandemii.

Utwórz nowy katalog i skopiuj do niego pliki: covid1.txt oraz covid_script_WINF_2025.m . Otwórz plik covid_script_WINF_2025.m w Matlabie. Uruchom go. Na ekranie otrzymasz wczytane dane do dalszej analizy. Następnie:

- Wyznacz współczynnik k zgodnie ze wzorem (4). Współczynnik k należy policzyć za pomocą jednego polecenia, wykorzystując obliczenia wektorowe.
- Wyznacz estymator ilości zakażeń pierwszego analizowanego dnia (czyli 184 dnia pandemii) na podstawie wzoru (8). W tym celu należy w pętli obliczyć wektor k^n, a następnie za pomocą jednego polecenia, wykorzystując obliczenia wektorowe wyznaczyć estymatę x[1] .
- Wyznacz, za pomocą za pomocą jednego polecenia, wykorzystując obliczenia wektorowe estymaty liczby zakażeń w kolejnych dniach pandemii zgodnie ze wzorem (9). Przedstaw uzyskane wyniki na wykresie

Uwagi do ćwiczenia część MATLAB:

- W sprawozdaniu zamieścić uzyskane wykresy. Do sprawozdania dołączyć skrypt, za pomocą którego uzyskano wyniki. Należy wykorzystać/modyfikować skrypt covid_script_WINF_2025.m
- Całe ćwiczenie sprowadza się do implementacji trzech wzorów (nr (4), (8() i (9)) i skorzystania z przygotowanego skryptu do prezentacji wyników. **Należy korzystać z obliczeń wektorowych, a nie pętli !!!**
- Wykorzystane źródło danych: https://pl.wikipedia.org/wiki/Statystyki_pandemii_COVID-19_w_Polsce , Zakładka: **Dane statystyczne w województwach w 2020.** Wykorzystano dane dla Województwa Mazowieckiego.
- Dla porównania zamieszczam uzyskane wyniki:
  Współczynnik transmisji wirusa: 1.0572
  Liczba zakażeń pierwszego dnia: 63.6152
  Wykres:



3. Zadania do wykonania – PYTHON (2 ptk)

Celem tej części ćwiczenia jest wykonanie w PYTHON tych samych obliczeń, co w części MATLAB. W tym celu:

- wczytaj plik covid1.txt, w którym znajduje się liczba wykrytych przypadków koronawirusa w Województwie Mazowieckim w kolejnych dniach. Ponieważ do analizy wykorzystamy dane od 184 do 214 dnia pandemii, przepisz do nowej listy interesujące nas dane. Pamiętaj, że MATLAB indeksuje tablice od 1, a PYTHON od 0.
- Wyznacz estymator ilości zakażeń pierwszego analizowanego dnia (czyli 184 dnia pandemii) na podstawie wzoru (8).
- Wyznacz estymaty liczby zakażeń w kolejnych dniach pandemii zgodnie ze wzorem (9).
- Przedstaw uzyskane wyniki na wykresie

Na zakończenie:

Po wykonaniu tego ćwiczenia prosimy o zamieszczenie w sprawozdaniu kilku informacji:

- Która część była łatwiejsza – MATLAB czy PYTHON ? Na wykonanie której części trzeba poświęcić więcej czasu i o ile ?
- Które środowisko umożliwia łatwiejszą analizę danych z tego ćwiczenia ?
- Którego środowiska powinno być więcej na studiach: MATLABa czy PYTHONa ?

Nie zapomnij wypełnić ankiety przedmiotowej na serwerze USOS !

Ponieważ jest to już ostanie laboratorium z WINF, to życzymy

POWODZENIA NA STUDIACH

i do zobaczenia na kolejnych przedmiotach z informatyki.

Maciej Linczuk

Grzegorz Nieradka