SPRAWOZDANIE

Zajęcia: Matematyka konkretna

Prowadzący: prof. dr hab. Vasyl Martsenyuk

|  |  |
| --- | --- |
| Laboratorium Nr 7  Data 11.01.2025  Temat: "Praktyczne zastosowanie analizy skupień (clustering) do zbiorów danych”  Wariant 7 | Daria Krecichwost  Informatyka  II stopień, stacjonarne,  1 semestr, gr. 1b |

# Sprawozdanie – Wieloliniowa regresja z użyciem macierzy pseudoodwrotnej

## 1. Cel ćwiczenia

Celem ćwiczenia było obliczenie współczynników regresji w równaniu postaci:  
y = a \* x₁ + b \* x₂  
z wykorzystaniem macierzy pseudoodwrotnej (Moore’a–Penrose’a). Zadanie dotyczyło regresji liniowej wielowymiarowej, czyli przewidywania jednej zmiennej zależnej (y) na podstawie dwóch zmiennych niezależnych (x₁, x₂).

## 2. Metoda

Macierz pseudoodwrotna jest często używana do rozwiązywania układów równań liniowych w przypadku:  
- układów nadokreślonych (więcej równań niż niewiadomych),  
- układów, dla których nie istnieje dokładne rozwiązanie.  
  
Rozwiązanie metodą najmniejszych kwadratów:  
x = A⁺y  
  
gdzie:  
- A – macierz danych wejściowych (kolumny: x₁, x₂),  
- y – wektor wartości oczekiwanych,  
- A⁺ – macierz pseudoodwrotna A.

## 3. Kod źródłowy (Python / Jupyter Notebook)

import numpy as np  
import matplotlib.pyplot as plt  
  
# Dane przykładowe  
x1 = np.array([1, 2, 3, 4, 5])  
x2 = np.array([2, 1, 0, 1, 2])  
y = np.array([2, 3, 6, 7, 8])  
  
# Tworzenie macierzy A  
A = np.column\_stack((x1, x2))  
  
# Obliczanie pseudoodwrotności  
A\_pinv = np.linalg.pinv(A)  
  
# Obliczanie współczynników regresji  
x = A\_pinv @ y  
a, b = x  
print(f"Współczynniki: a = {a:.4f}, b = {b:.4f}")  
  
# Obliczanie y przewidzianych  
y\_pred = A @ x  
  
# Wykres  
plt.plot(y, 'ro-', label='Rzeczywiste y')  
plt.plot(y\_pred, 'bo--', label='Przewidziane y')  
plt.xlabel('Indeks')  
plt.ylabel('y')  
plt.legend()  
plt.title('Wieloliniowa regresja: y = a\*x1 + b\*x2')  
plt.grid(True)  
plt.show()

## 4. Wnioski

Z użyciem macierzy pseudoodwrotnej uzyskano przybliżone rozwiązanie układu regresji. Metoda pozwala skutecznie estymować współczynniki regresji nawet przy danych z szumem. Możliwość alternatywnej formuły z użyciem np.linalg.pinv zapewnia elastyczność i niezawodność.

## 5. Link do repozytorium GitHub

https://github.com/DariaKrecichwostQA/StudiaMagisterskie