

Sprawozdanie

Zajęcia: Matematyka konkretna

Prowadzący: prof. dr hab. Vasyl Martsenyuk

Ćw 4

23.10.2025

Temat: Metoda PCA na podstawie SVD

Wariant: 2

Patrycja Kubica
Informatyka II stopień,
stacjonarne,
2 semestr,

1. Zadanie dotyczy obliczenia środka, osi głównych oraz kątu obrotu danych dwuwymiarowych z pliku .csv zgodnie z wariantem zadania

Wariant 2.csv

2. Kod

```
import numpy as np
import matplotlib.pyplot as plt

# Konfiguracja wykresu
plt.rcParams['figure.figsize'] = [12, 6]

# Wczytanie danych z pliku CSV
data = np.loadtxt('2.csv', delimiter=',')

# Obliczenie środka danych (mean)
Xavg = np.mean(data, axis=1)

# Dane odjęte o średnią
B = data - Xavg[:, np.newaxis]

# Wyznaczenie głównych osi (PCA) przez SVD
U, S, VT = np.linalg.svd(B / np.sqrt(data.shape[1]), full_matrices=False)

# Kąt obrotu głównej osi względem osi x
theta = np.arctan2(U[1,0], U[0,0])
theta_deg = np.degrees(theta)
print(f"Środek danych: {Xavg}")
print(f"Główne osie (U):\n{U}")
print(f"Odchylenia standardowe (S): {S}")
print(f"Kąt obrotu głównej osi względem osi x: {theta_deg:.2f}°")

# Rysowanie danych i PCA
fig, ax = plt.subplots()
ax.plot(data[0,:], data[1:], '.', color='k')
ax.grid()
plt.xlim(np.min(data[0,:])-1, np.max(data[0,:])+1)
plt.ylim(np.min(data[1,:])-1, np.max(data[1,:])+1)
ax.set_aspect('equal')

# Elipsy
theta_circle = 2 * np.pi * np.arange(0,1,0.01)
for k in [1,2,3]:
    Xstd = U @ np.diag(S) @ np.array([np.cos(theta_circle), np.sin(theta_circle)])
    ax.plot(Xavg[0] + k*Xstd[0,:], Xavg[1] + k*Xstd[1:], '-', color='r', linewidth=2)
```

```

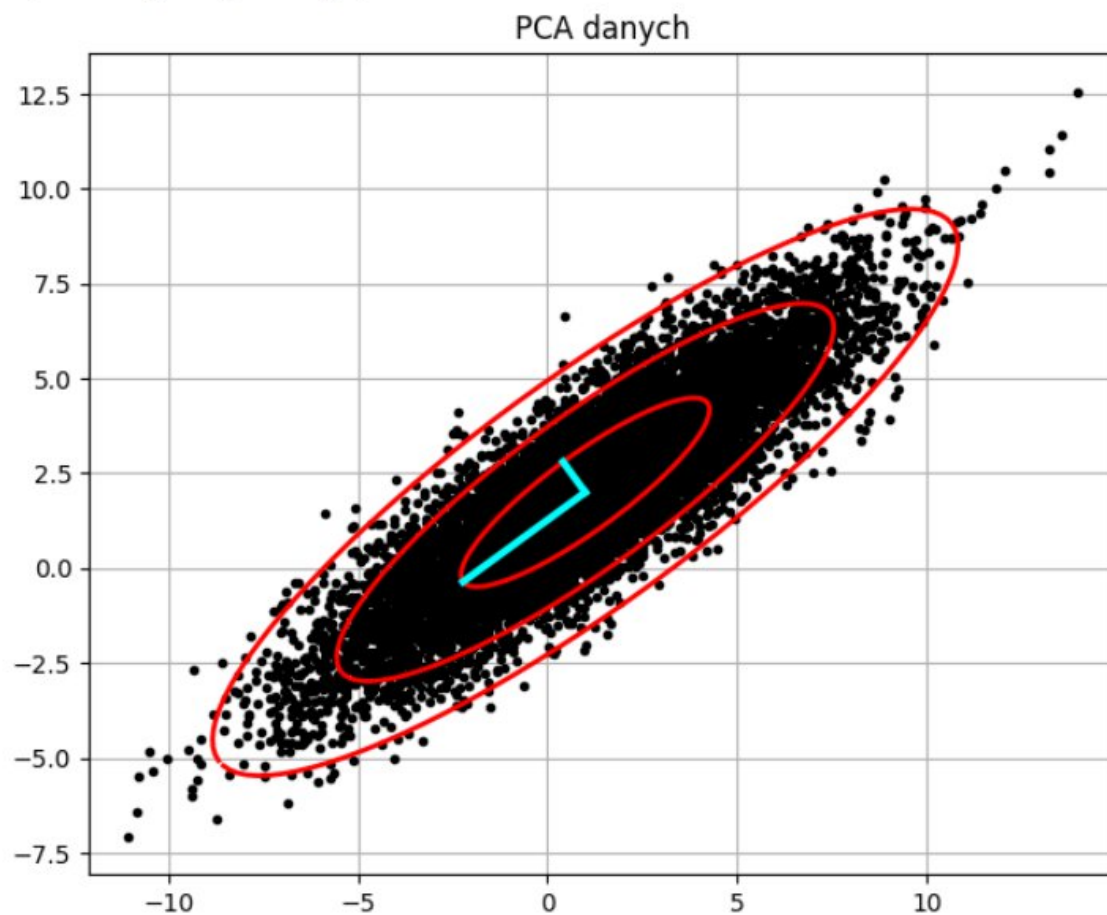
# Główne osie
for i in range(2):
    ax.plot([Xavg[0], Xavg[0] + U[0,i]*S[i]],
            [Xavg[1], Xavg[1] + U[1,i]*S[i]],
            '-', color='cyan', linewidth=3)

plt.title("PCA danych")
plt.show()

```

3. Wynik

Środek danych: [1.00130733 2.00033545]
 Główne osie (U):
 [[-0.80751036 -0.58985339]
 [-0.58985339 0.80751036]]
 Odchylenia standardowe (S): [3.99579262 0.99279451]
 Kąt obrotu głównej osi względem osi x: -143.85°



4. Wnioski

- Środek danych wynosi około $[1.0, 2.0]$, co wskazuje na punkt centralny zbioru. Dane są wokół niego symetrycznie rozłożone.
- Główna oś PCA wskazuje kierunek największej zmienności, z odchyleniem standardowym około 4, natomiast druga oś ma mniejszą zmienność (~ 1). Dane mają więc wyraźny, wydłużony kształt elipsoidalny.
- Kąt głównej osi względem osi X wynosi około -144° , co pokazuje orientację rozkładu danych w przestrzeni. Wizualizacja elips i osi PCA ułatwia obserwację kierunku i zakresu zmienności.