

SPRAWOZDANIE

Zajęcia: Matematyka Konkretna

Prowadzący: prof. dr hab. inż. Vasyl Martsenyuk

Laboratorium Nr 4 Data 30.10.2023 Temat: SVD – PCA Wariant 6	Rafał Klinowski Informatyka II stopień, stacjonarne, 2 semestr, gr. a
---	--

1. Polecenie:

Ćwiczenie polegało na stworzeniu notatnika Jupyter w języku Python do przeprowadzenia analizy głównych składowych (PCA) przy pomocy metody SVD.

Wariant zadania: 6

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

2. Napisany program, uzyskane wyniki

Podczas implementacji ćwiczenia bazowano na projekcie udostępnionym w ramach materiałów do laboratorium.

Na początku utworzono niezbędne macierze, potrzebny kąt oraz wczytano dane z pliku tekstowego. Dane zostały zamienione na macierz (10000x2).

```
xC = np.array([2, 1])          # Center of data (mean)
sig = np.array([2, 0.5])       # Principal axes

theta = np.pi/3               # Rotate cloud by pi/3

R = np.array([[np.cos(theta), -np.sin(theta)],      # Rotation matrix
              [np.sin(theta), np.cos(theta)]]])

data = pd.read_csv('6.csv', sep=";", header=None)
data = data.to_numpy()
```

Następnie utworzono pierwszy wykres, przedstawiający ułożenie wczytanych punktów w układzie współrzędnych.

```
# Rysowanie wczytanych punktów
nPoints = data.shape[1]
print("nPoints = ", nPoints)

X = R @ np.diag(sig) @ data + np.diag(xC) @ np.ones((2,nPoints))

fig = plt.figure()
ax1 = fig.add_subplot(121)
ax1.plot(X[0,:],X[1,:], '.', color='k')
ax1.grid()
plt.xlim((-6, 8))
plt.ylim((-6,8))

plt.show()
```

Kolejnym krokiem było przeprowadzenie SVD, analogicznie jak w poprzednich ćwiczeniach laboratoryjnych.

```
# Przygotowanie do SVD
Xavg = np.mean(X,axis=1)
B = X - np.tile(Xavg,(nPoints,1)).T

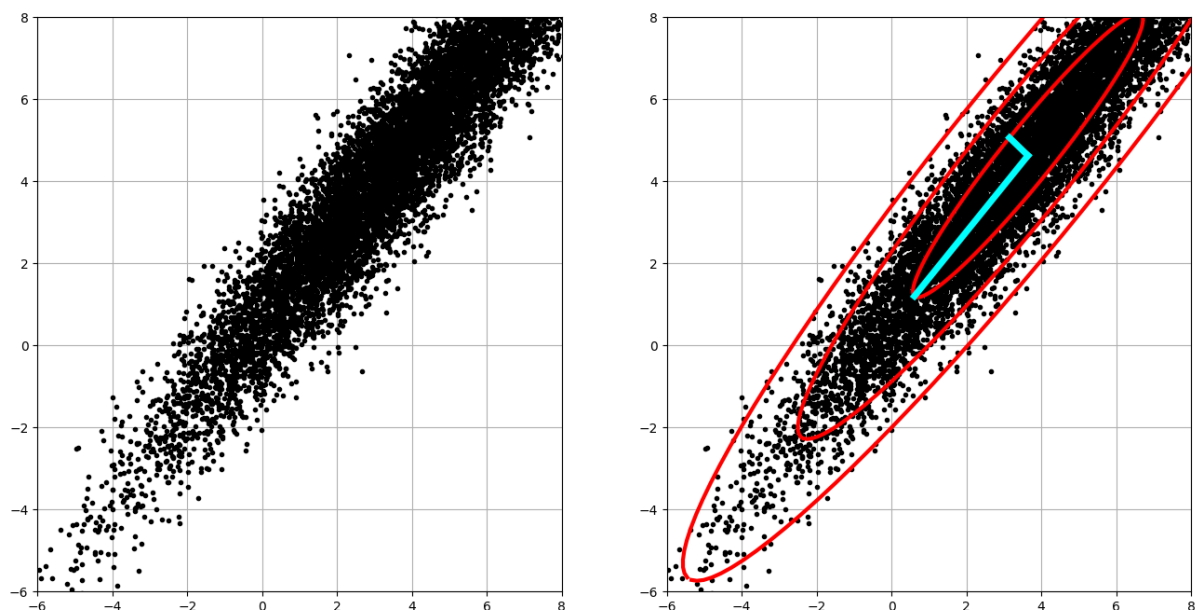
# Przeprowadzenie SVD
U, S, VT = np.linalg.svd(B/np.sqrt(nPoints),full_matrices=0)
```

Następnie na podstawie SVD przygotowano oraz obliczono PCA, zgodnie z instrukcją laboratoryjną.

```
# Przygotowanie PCA
theta = 2 * np.pi * np.arange(0,1,0.01)

# Pierwszy standardowy rozkład
Xstd = U @ np.diag(S) @ np.array([np.cos(theta),np.sin(theta)])
```

Poniżej można znaleźć wykres przedstawiający uzyskane wyniki – punkty w układzie z nałożonymi obliczonymi wartościami.



Na podstawie ćwiczenia można wyciągnąć parę wniosków:

- Domyślną implementację PCA bardzo łatwo dostosować do nowych danych – większość instrukcji jest analogiczna, różni się jedynie samymi danymi oraz ich rozmiarem
- SVD doskonale nadaje się do analizy głównych składowych – PCA

Repozytorium zawierające uzyskane wyniki wraz z niezbędnymi plikami:

<https://github.com/Stukeley/MatematykaKonkretna/tree/master/Lab4>