







PODSTAWY  
NAUKI O DANYCH

Rozkłady macierzy

# DEKOMPOZYCJA DANYCH



# Rozkład macierzy na wartości własne

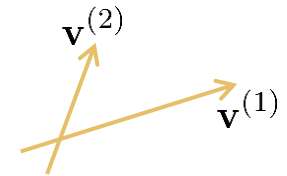
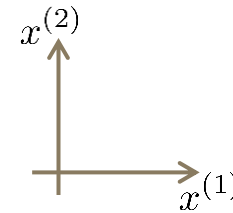
## *Eigendecomposition*

$$\begin{matrix} & m \\ m & \boxed{\mathbf{A}} \end{matrix}$$



$$\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i$$

$$\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{D}$$



$$\mathbf{A} \begin{bmatrix} | & | & \dots & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_m \\ | & | & \dots & | \end{bmatrix} = \begin{bmatrix} | & | & \dots & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_m \\ | & | & \dots & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_m \end{bmatrix}$$

$$\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$$

# Rozkład macierzy na wartości osobliwe

## *Singular Value Decomposition*

$$AV = UD$$

$$\begin{matrix} & n \\ m & \boxed{A} \end{matrix} = \boxed{U} \boxed{D} \boxed{V^T}$$

wersja pełna

$$U^T U = I_m$$

$$V^T V = I_n$$

$$\boxed{A} = \boxed{U} \boxed{D} \boxed{V^T}$$

wersja zredukowana

$$A = UDV^T$$

# Rozkład macierzy na wartości osobliwe

## *Singular Value Decomposition*

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

$$\mathbf{U} = \begin{bmatrix} | & | & \dots & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_m \\ | & | & \dots & | \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

$$\mathbf{V}^T = \begin{bmatrix} -\mathbf{v}_1- \\ -\mathbf{v}_2- \\ \vdots \\ -\mathbf{v}_n- \end{bmatrix}$$

	reksio2008	justysia	lordV	jednorozcem_jestem	ipeen
Stanisław Lem	7	1	4	0	0
Olga Tokarczuk	1	0	1	2	0
Jacek Dukaj	4	0	5	5	6
Danuta Wawiłow	0	5	0	0	0
Norman Davies	0	0	0	5	5
Isaak Asimov	7	0	6	0	0

A

eigenautorzy

```
U =
[ [-0.49  0.4  0.11  0.47 -0.51 -0.35]
  [-0.13 -0.07 -0.01  0.59  0.76 -0.21]
  [-0.62 -0.46 -0.02 -0.46  0.1  -0.43]
  [-0.01  0.03  0.99 -0.11  0.11  0.07]
  [-0.21 -0.67  0.06  0.42 -0.27  0.51]
  [-0.56  0.43 -0.11 -0.18  0.26  0.62]]
```

```
D=
[14.66  9.42  5.06  1.68  1.4 ]
```

```
VT=
[ [-0.68 -0.04 -0.58 -0.3  -0.33]
  [ 0.41  0.06  0.19 -0.61 -0.64]
  [-0.02  1.  -0.07  0.03  0.03]
  [ 0.44 -0.04 -0.56  0.56 -0.42]
  [-0.42  0.03  0.55  0.47 -0.55]]
```

eigenczytelnicy

```
dane = {
    'reksio2008'      : [7,1,4,0,0,7],
    'justysia'       : [1,0,0,5,0,0],
    'lordV'          : [4,1,5,0,0,6],
    'jednorozcem_jestem': [0,2,5,0,5,0],
    'ipeen'          : [0,0,6,0,5,0]
}

autorzy = [
    'Stanisław Lem',
    'Olga Tokarczuk',
    'Jacek Dukaj',
    'Danuta Wawiłow',
    'Norman Davies',
    'Isaak Asimov'
]

df = pd.DataFrame(dane, index=autorzy)

U, D, VT = np.linalg.svd(df, full_matrices=True)

print(f"""U = \n{np.round(U,2)}\n
D=\n{np.round(D,2)}\n
VT=\n{np.round(VT,2)}""")
```

	reksio2008	justysia	lordV	jednorozcem_jestem	ipeen
Stanisław Lem	7	1	4	0	0
Olga Tokarczuk	1	0	1	2	0
Jacek Dukaj	4	0	5	5	6
Danuta Wawiłow	0	5	0	0	0
Norman Davies	0	0	0	5	5
Isaak Asimov	7	0	6	0	0

eigenautorzy

U =

[ [-0.49 0.4 0.11 0.47 -0.51 -0.35]  
 [-0.13 -0.07 -0.01 0.59 0.76 -0.21]  
 [-0.62 -0.46 -0.02 -0.46 0.1 -0.43]  
 [-0.01 0.03 0.99 -0.11 0.11 0.07]  
 [-0.21 -0.67 0.06 0.42 -0.27 0.51]  
 [-0.56 0.43 -0.11 -0.18 0.26 0.62]]

D=

[14.66 9.42 5.06 1.68 1.4 ]

VT=

[ [-0.68 -0.04 -0.58 -0.3 -0.33]  
 [0.41 0.06 0.19 -0.61 -0.64]  
 [-0.02 1. -0.07 0.03 0.03]  
 [0.44 -0.04 -0.56 0.56 -0.42]  
 [-0.42 0.03 0.55 0.47 -0.55]]

eigenczytelnicy

A

Czy istnieje związek między osobą a lubianymi przez nią *autorami*?

- preferencje czytelnika = kombinacja liniowa  $\mathbf{v}_i$
- ocena autora = kombinacja liniowa  $\mathbf{u}_i$



THE END