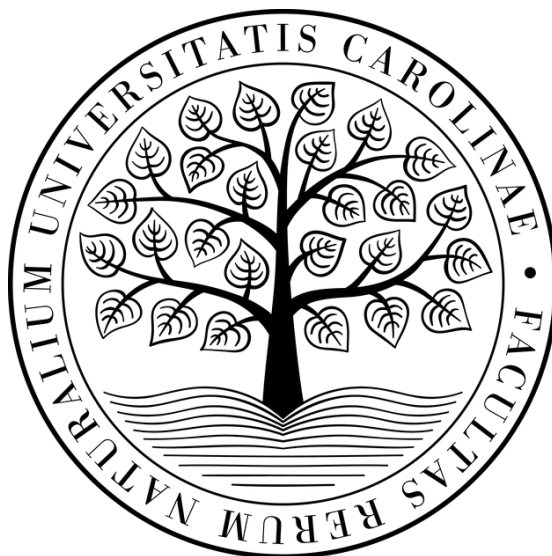


**UNIVERZITA KARLOVA**  
**PŘÍRODOVĚDECKÁ FAKULTA**



**Analýza hlavních komponent**  
**Geoinformatika**

Eliška Králová, Teodor Sorokáč

1. NGKDZP

Praha, 2024

## Zadání

S využitím oblíbeného programovacího prostředí či výpočetního sw (Python, Matlab, Excel, ...) vytvořte dva příklady dvourozměrných datových sad (např. po 20 pozorováních), na nichž ukážete význam transformace hlavních komponent. V prvním případě bude po transformaci první hlavní komponenta obsahovat alespoň 70% informace datového souboru. Ve druhém případě bude vliv transformace minimální (obsah informace v původních a transformovaných osách se nebude lišit více než o 10 %). V obou případech spočítejte vlastní čísla a vlastní vektory kovarianční matice.

# Vypracování

Celá úloha byla řešena v SW Visual Studio Code. a napsána v programovacím jazyce Python. Do dvou datových sad bylo vybráno 20 bodů tak, aby vypočítaná první hlavní komponenta nesla požadovaná procenta informace. U toho byl využíván předpoklad, že hlavní komponenta bude nést 100 % informace, jestliže body leží na jedné přímce.

Nejprve bylo nutné pro tyto body vypočítat kovarianční matici, pomocí níž se poté spočítaly vlastní čísla a vlastní vektory. Vlastní vektor udává směr a vlastní číslo udává obsah informace nesené hlavními komponentami. Přesné množství informace jednotlivých komponent se zjistí tak, že se vezme vlastní číslo a vydělí se sumou všech vlastních čísel. Nakonec byly body vizualizovány pomocí bodové grafu (scatter plot).

První hlavní komponenta první datové sady měla obsahovat alespoň 70 % informace. Body této datové sady, jejich kovarianční matice, vlastní čísla a vektory, spolu s přesným množstvím nesené informace a rozmístěním bodů jsou následující.

$X = [1, 2, 3, 4, -4, 6, 7, 24, 9, 20, 11, -13, 13, 14, 5, 17, 17, 18, 13, 20]$

$Y = [-1, 6, -3, 5, -5, 6, 2, 8, 9, 10, -1, 12, -2, -4, 15, 6, -1, 5, 9, 4]$

kovarianční matice:

```
[[83.71315789  0.78947368]
 [0.78947368  32.10526316]]
```

vlastní čísla:

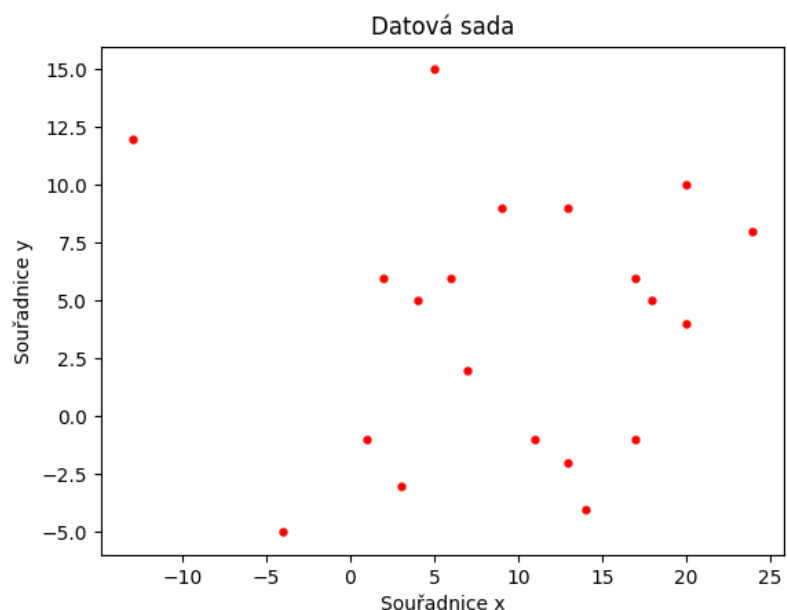
```
[83.72523207  32.09318898]
```

vlastní vektory:

```
[[0.99988307 -0.01529217]
 [0.01529217  0.99988307]]
```

První hlavní komponenta obsahuje:

72.29 %



Obsah informace v druhé datové sadě má první komponenta obsahovat více než 90 %. Body této datové sady, jejich kovarianční matice, vlastní čísla a vektory, spolu s přesným množstvím nesené informace a rozmístěním bodů jsou následující.

$X = [1, -2, 3, 4, 5, 6, -7, 8, -9, 10, 11, 12, -13, 14, -15, -16, 17, 18, 19, 20]$

$Y = [1, 2, -3, -4, 3, 6, -5, 8, 7, 10, 1, 12, -11, 14, 0, -14, 7, 8, 19, 20]$

kovarianční matice:

```
[[131.58947368  81.72105263]
 [81.72105263  79.82947368]]
```

vlastní čísla:

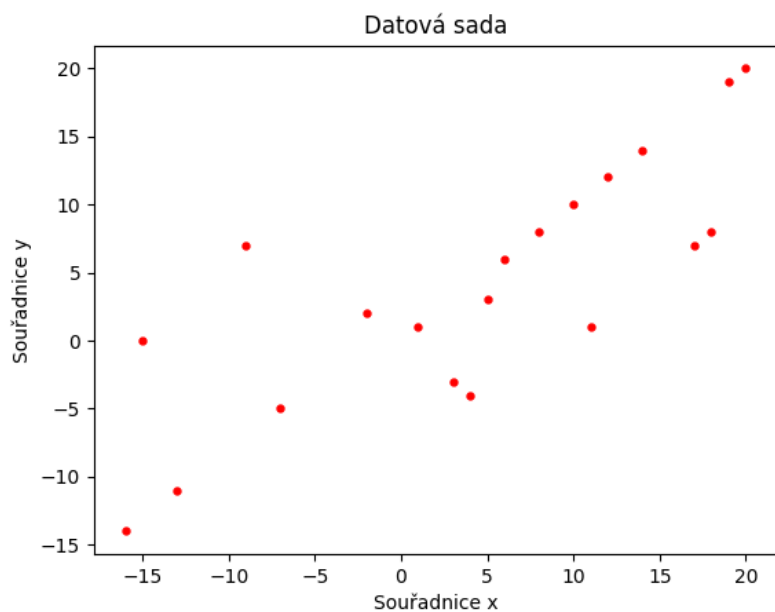
```
[191.43405178  19.99489559]
```

vlastní vektory:

```
[[0.8068012  -0.590823]
 [0.590823   0.8068012]]
```

První hlavní komponenta obsahuje:

90.54 %



Jelikož zadání požadované informace pro druhou datovou sadu nebylo zcela jasné, byla vytvořena třetí datová sada. Tato sada má v první hlavní komponentě obsaženo méně než 10 %. Body této datové sady, jejich kovarianční matice, vlastní čísla a vektory, spolu s přesným množstvím nesené informace a rozmístěním bodů jsou následující.

$X = [1, 2, 3, 4, -4, 6, 7, 24, 9, 20, 11, -13, 13, 14, 5, 17, 17, 18, 13, 20]$

$Y = [-1, 16, -3, 35, -5, 76, 2, 68, 9, 10, -11, 12, -12, -40, 15, 16, -21, 5, 19, 44]$

kovarianční matice:

$\begin{bmatrix} 83.71315789 & 31.11052632 \\ 31.11052632 & 769.27368421 \end{bmatrix}$

vlastní čísla:

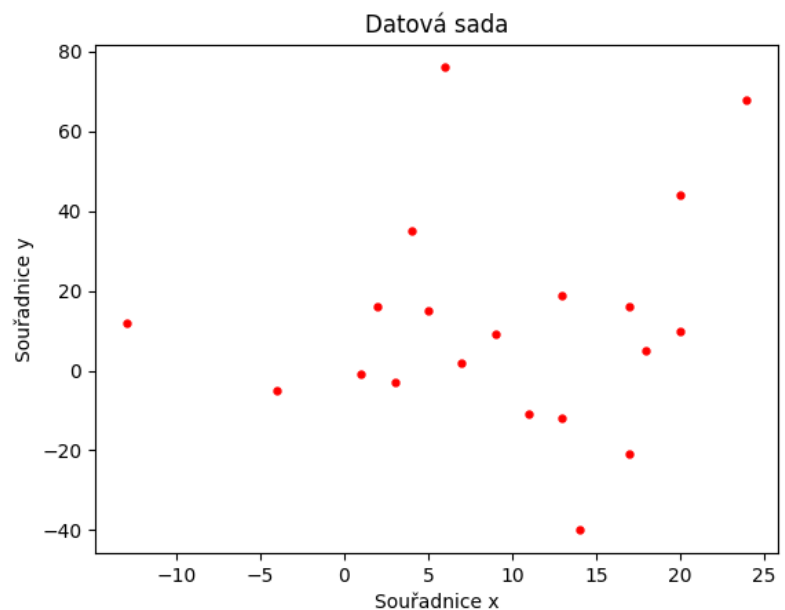
$[82.30426716 \quad 770.68257495]$

vlastní vektory:

$\begin{bmatrix} -0.99897614 & -0.04524026 \\ 0.04524026 & -0.99897614 \end{bmatrix}$

První hlavní komponenta obsahuje:

9.65 %



# Zdroje

NumPy (2024): NumPy reference – numpy.cov.

<https://numpy.org/doc/stable/reference/generated/numpy.cov.html> [cit. 24.11.2024]

NumPy (2024): NumPy reference – numpy.linalg.eig.

<https://numpy.org/doc/stable/reference/generated/numpy.linalg.eig.html> [cit. 24.11.2024]

W3schools (2024): Python – Matplotlib tutorial.

[https://www.w3schools.com/python/matplotlib\\_intro.asp](https://www.w3schools.com/python/matplotlib_intro.asp) [cit. 27.11.2024].