

# Volby

Lukáš Hromadník

## 1 Úkol 1

Jednotlivé politické strany  $\{a_1, \dots, a_{25}\}$ , kde  $a_i \in \mathbb{R}^{100}$  je možné chápat jako body v prostoru odpovědí na referenční programové otázky. Chtěli bychom tento prostor vhodně vizualizovat. To uděláme tak, že body  $a_i$  proložíme **afinním** podprostorem dimenze 2, tak aby součet kvadrátů vzdáleností původních  $a_i$  a promítnutých bodů  $a'_i$  byl minimální. Následně promítnuté body zobrazíme v souřadnicích báze nalezeného afinního podprostoru.

1. Formulujte optimalizační problém.

Hledáme bázi afinního podprostoru dimenze 2, jehož body  $a'_i$  mají od bodů  $a_i$  nejmenší vzdálenost. Pro body  $a'_i$  a  $a_i$  tedy platí, že

$$\min \|a'_i - a_i\|, i \in [1, 25].$$

2. Vyřešte optimalizační problém.

Prostor, ve kterém se nacházejí body  $a_i, i \in [1, 25]$  posuneme pomocí těžiště těchto bodů do počátku a spektrálním rozkladem získáme matici vlastních vektorů  $V$  a matici vlastních čísel  $D$ . S využitím věty 7.3 získáme bázi hledaného podprostoru tak, že vezmeme poslední 2 sloupce matice  $V$  (tj. dva vlastní vektory příslušející dvěma největším vlastním číslům). Tyto vektory tvoří ortonormální bázi hledaného podprostoru.

Kritériem je nazvána hodnota stopy matice  $A$ , což je součet  $k$  nejmenších vlastních čísel matice  $A$ . S využitím věty 7.2 získáme vzorec

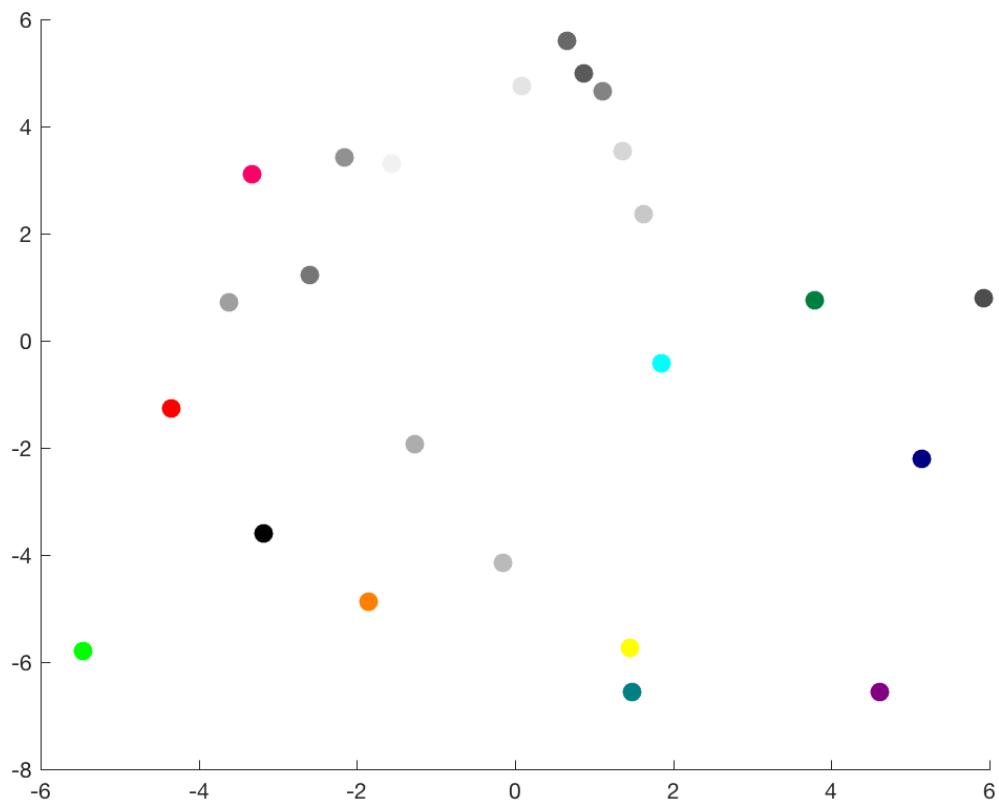
$$\min \{ \text{tr}(X^T A X) \mid X \in \mathbb{R}^{n \times k}, X^T X = \mathbf{I} \} = \lambda_1 + \dots + \lambda_k$$

S využitím Matlabu

```
1 sum(sum(D(1:end - 2, :)))
```

získáváme  $\text{tr}(A) = 1098,1$ .

3. Najděte ortonormální bázi nalezeného afinního podprostoru  $\text{span}\{a'_1, \dots, a'_{25}\} + a_0$  a jednotlivé politické strany zobrazte v souřadnicích této báze (jejíž vektory ztotožníte s osami 2D grafu). Zakreslené body obarvěte barvou strany (podle `T.color`).



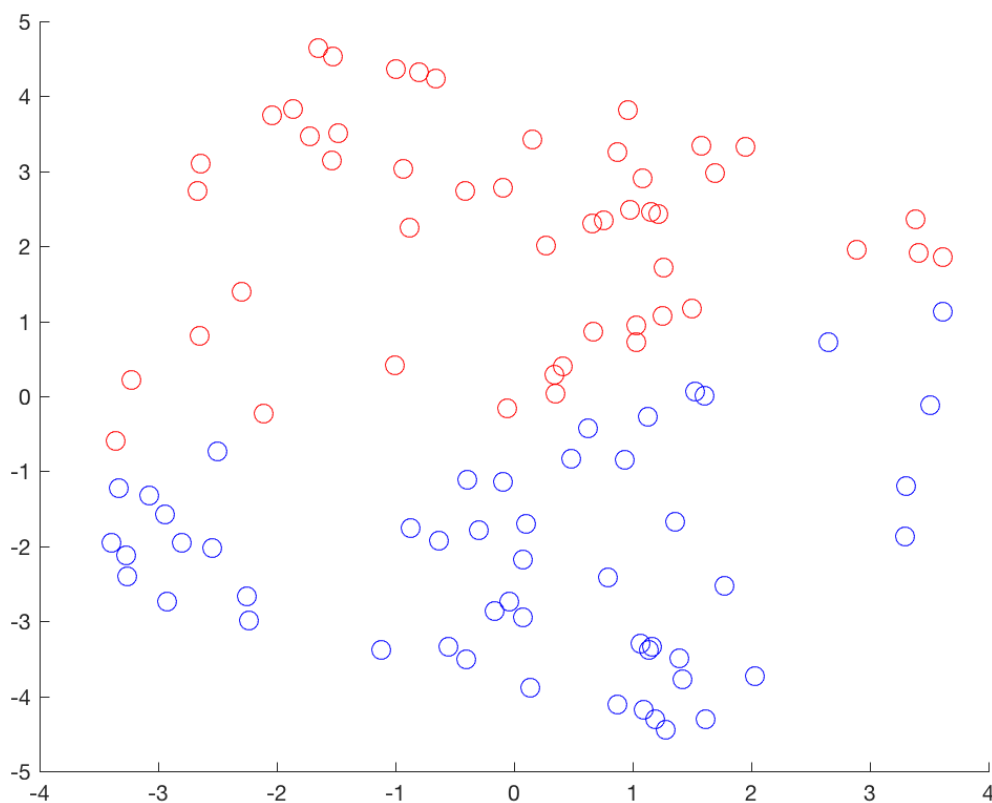
4. Interpretujte výsledek.

Dle výsledného grafu se strany podle svislé osy pravděpodobně rozdělily na pravicové a levicové. Rozdělení podle vodorovné osy není zřejmé.

## 2 Úkol 2

Vizualizovat tato data je možné i z opačného pohledu. Jednotlivé otázky  $\{b_1, \dots, b_{100}\}$ , kde  $b_i \in \mathbb{R}^{25}$  jsou body v prostoru politických stran. Otázky zobrazíme opět v prostoru dimenze 2.

1. Postupujte obdobně jako v minulém příkladu a zakresejte jednotlivé otázky v souřadnicích ortonormální báze dimenze 2 prostoru proložený ve smyslu nejmenších čtverců. Výsledné body obarvete podle toho, zda většina politických stran odpovídá ano.



2. Interpretujte výsledek.

Vizualizací otázek do podprostoru dimenze 2 došlo ke grafickému oddělení jednotlivých otázek podle toho, jestli na ně strany častěji odpovídaly ano či ne. Červeně obarvené jsou ty otázky, na které většina politických stran odpovídala ano a modře jsou otázky s většinovým podílem záporných odpovědí.

3. Je možné nalézt ortonormální bázi hledaného podprostoru a souřadnice promítnutých bodů v této bázi využitím výpočtu z předchozího úkolu? Pokud ano, vysvětlete.

V druhém úkolu se pracuje se stejnou vstupní maticí, avšak tato matice je transponovaná. Pro řešení tohoto úkolu lze využít stejný postup jako v případě Úkol č. 1 s menšími úpravami, ale finální výsledky se rozhodně liší.

## 3 Příloha

### 3.1 Zdrojový kód pro řešení úkolu č. 1

```
1 load volby_2017
2
3 teziste = sum(T.data, 2) / size(T.data, 2);
4 body_v_pocatku = T.data - teziste;
5
6 A = body_v_pocatku * body_v_pocatku';
7 [V, D] = eig(A);
8 dim2 = V(:,end-1:end);
9 strany = (body_v_pocatku' * dim2);
10 barvicky = vec2mat(cell2mat(T.color)', 3);
11 scatter(strany(:,1), strany(:,2), 75, barvicky, 'filled');
12 saveas(gcf, 'strany.png');
```

### 3.2 Zdrojový kód pro řešení úkolu č. 2

```
1 load volby_2017
2
3 teziste = sum(T.data) / size(T.data, 1);
4
5 body_v_pocatku = T.data - teziste;
6
7 A = body_v_pocatku' * body_v_pocatku;
8 [V, D] = eig(A);
9 dim2 = V(:,end-1:end);
10 otazky = (body_v_pocatku * dim2);
11
12 odpovedi = sum(T.data, 2);
13 barvicky = zeros(size(odpovedi, 1), 3);
14 for i = 1:size(odpovedi)
15     if odpovedi(i) > 0
16         barvicky(i, 1) = 1;
17     else
18         barvicky(i, 3) = 1;
19     end
20 end
21 scatter(otazky(:,1), otazky(:,2), 75, barvicky);
22 saveas(gcf, 'otazky.png');
```