

**Deadline: Fr. Feb 17, 10:00** Submit in the post boxes at Samelsonplatz or upload a .pdf file via LearnWeb.

## 1. Political Compass Analysis (PCA) (20 points)

In this tutorial, we will have some fun and use Principal Component Analysis (PCA) to recreate a variant of the infamous [political compass](#). We will use data provided by the [Wahl-O-Mat](#), which is an online tool created by the [Bundeszentrale für politische Bildung \(bpb\)](#) (German Federal Agency for Civic Education). At each major election all parties get asked a catalog of questions and their answers are recorded. Citizens can go through the same questionnaire and see how their answers align with the different political parties.

**A. [2p]** Obtain the questionnaire data from the German Federal Election 2021. <https://www.bpb.de/themen/wahl-o-mat/bundestagswahl-2021/337541/download/>. Perform the following preprocessing:

- Encode the answers as a numerical variable (“stimme nicht zu” = -1, “neutral”=0, “stimme zu”= +1).
- Extract two subsets of the data: (1): the answers from the major parties represented in the German Parliament<sup>1</sup>, and (2) the answers from all the parties. (You should get a DataFrame of size  $6 \times 38$  and  $38 \times 38$  respectively)

**B. [2p]** Explain why it is important to **always** center the data first before performing PCA.

**C. [2p]** Explain under which circumstances one should also scale<sup>2</sup> the data before performing PCA. Why would or wouldn't you scale the data from exercise 1A?

**D. [4p]** Using the major parties, perform PCA with  $k = 2$  components. Visualize it as a labeled 2d scatter plot.

**E. [2p]** Come up with an interpretation for the PCA-axes. Re-normalize the PCA-components by multiplying them with  $\pm 1$ , such that AfD is in the first quadrant. Explain why it is ok to do so.

**F. [2p]** Repeat the experiments, now using all parties that answered the questionnaire. Comment on the change of relative position of the parties with respect to each other compared to 1E.

**G. [6p]** Exercise 1F revealed an obvious flaw in the methodology we used. Repeat experiment 1E and 1F, but this time using weighted PCA (wPCA). For 1E weigh each datapoint by the relative number of seats the party has in the German parliament. For 1F, use the relative number of votes<sup>3</sup> each party obtained. The official results can be found at <https://www.bundeswahlleiter.de/bundestagswahlen/2021/ergebnisse/bund-99.html>. Comment on the result compared to 1E and 1F respectively.

**Note:** Weighted PCA should minimize the loss  $\sum_n w_n \|x_n - \text{decoder}(\text{encoder}(x_n))\|^2$ , where  $\sum_n w_n = 1$  and  $w_n \geq 0$ . See lecture slides 1, section 4 for a discussion of case weights for linear regression.

<sup>1</sup>“CDU/CSU”, “SPD”, “GRÜNE”, “AfD”, “FDP” and “DIE LINKE”

<sup>2</sup>i.e. divide by the standard deviation after centering

<sup>3</sup>Use the “Zweitstimme”.