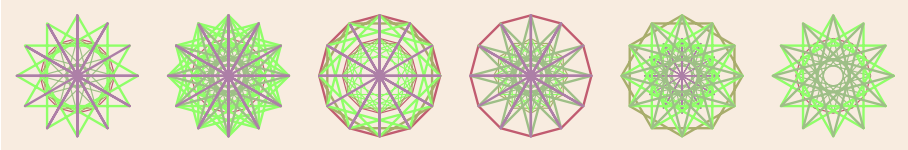
Exercise

CSC316 Machine Learning Leon Tabak

16 February 2022



A political scientist has given you a dataset that includes a column that contains integers. The integers all have values in the interval [0, 7]. The codes have the following meanings:

0 A Democrat who voted in 2020 Presidential election.

1 A Democrat who did not vote in the 2020 Presidential election.

2 A Republican who voted in 2020 Presidential election.

3 A Republican who did not vote in the 2020 Presidential election.

4 A member of a third party who voted in 2020 Presidential election.

5 A member of a third party who did not vote in the 2020 Presidential election.

6 A member of no political party who voted in 2020 Presidential election.

7 A member of no political party who did not vote in the 2020 Presidential election.

We want to prepare this data for use in a model that will predict the outcomes of future elections.

Before we work on the large dataset, we might want to do a little experiment on the side. This will help us develop and test our methods.

- Write code that creates a small Pandas DataFrame with a single column that contains random integers that all lie in the interval [0, 7].
- Add code to the program that replaces the single column with two columns.
  - One column will contain integers in the interval [0, 3]. These integers will indicate a voter's party affiliation.
  - A second column will contains integers that have the value zero or one. These integers will indicate whether or not a voter participated in the 2020 presidential election.
- Add code to the program that uses one hot encoding to replace the column that holds the information about party affiliation with four columns that each contain only zeros and ones.