Challenge 3: Self-Organizing Maps

Unsupervised Machine Learning

Answer the following questions to the best of your ability. Be sure to show *all* code in-line, in addition to full written responses to each of the questions. Write in complete sentences where appropriate. **A complete submission is a single rendered PDF submitted to Canvas**.

You will be using the 2016 American National Election Pilot Study data (`anes_2016.csv`). You'll be exploring whether we can pick up on differences in a "social question space" along a dimension of *identity*. For this exercise, the identity is either "American" or "racial," on the basis of responses to a respective question (more below). A naive expectation is that people's identities should be present in their responses to social questions.

Note: if you get stuck, remember to always look to the package help files (e.g., `?som`), as well as class notes.

The Social Issue Questions

Below are the 14 social issue questions along with scales and variable code names to be used in the analysis. Question wording and response categories were copied and pasted from the ANES 2016 Pilot Study Questionnaire.

- `vaccine` - "Do you favor, oppose, or neither favor nor oppose requiring children to be vaccinated in order to attend public schools?" (7 point from favor a great deal (1) to oppose a great deal (7))
- `autism` - "How likely or unlikely is it that vaccines cause autism?" (6 point from Extremely likely (1) to Extremely unlikely (6))
- `birthright_b` - "Do you favor, oppose, or neither favor nor oppose children of unauthorized immigrants automatically getting citizenship if they are born in this country?" (7 point from Favor a great deal (1) to Oppose a great deal (7))
- `forceblack` - "How often do you think police officers use more force than is necessary under the circumstances when dealing with BLACK people?" (5 point from Never (1) to Very often (5))
- `forcewhite` - "How often do you think police officers use more force than is necessary under the circumstances when dealing with WHITE people?" (5 point from Never (1) to Very often (5))
- `stopblack` - "How often do to think police officers stop BLACK people on the street without a good reason?" (5 point from Never (1) to Very often (5))
- `stopwhite` - "How often do to think police officers stop WHITE people on the street without a good reason?" (5 point from Never (1) to Very often (5))
- `freetrade` - "Do you favor, oppose, or neither favor nor oppose the U.S. making free trade agreements with other countries?" (7 point from Favor a great deal (1) to Oppose a great deal (7))
- `aa3` - "Do you favor, oppose, or neither favor nor oppose allowing universities to increase the number of underrepresented minority students studying at their schools by considering race along with other factors when choosing students?" (7 point from Favor a great deal (1) to Oppose a great deal (7))

- `warmdo` - "Do you think the federal government should be doing more about rising temperatures, should be doing less, or is it currently doing the right amount? (7 point from Should be doing a great deal more (1) to Should be doing a great deal less (7))
- `finwell` - "Do you think people's ability to improve their financial well-being is now better, worse, or the same as it was 20 years ago?" (7 point from A great deal better (1) to A great deal worse (7))
- `childcare` - "Do you favor an increase, decrease, or no change in government spending to help working parents pay for CHILD CARE when they can't pay for it all themselves?" (7 point from Increase a great deal (1) to Decrease a great deal (7))
- `healthspend` - "Do you favor an increase, decrease, or no change in government spending to help people pay for HEALTH INSURANCE when they can't pay for it all themselves?" (7 point from Increase a great deal (1) to Decrease a great deal (7))
- `minwage` - "Should the minimum wage be raised, kept the same, lowered but not eliminated, or eliminated altogether?" (4 point from Raised [1], Kept the same [2], Lowered [3], Eliminated [4])

The Task

Explore whether responses to these social questions are grouped along "identity"-specific lines. To do so, you will be exploring weak/strong American identity (as these are American respondents), and then weak/strong racial identity.

More specifically, you will build a self-organizing map of the responses to the above "social" questions. Then, you will explore the diagnostics of the map. Then, you will plot the output layer with color according to weak/strong identities for American first, and race second. Follow the prompts below to complete this task.

The Questions

1. (10 points) Create a dichotomous feature for each identity question: `amer_ident` and `race_ident` (consider `ifelse()`). The question wording is, *How important to your identity is [being American/your race]?* For each of these, the response categories are: 1 Extremely important, 2 Very important, 3 Moderately important, 4 A little important, 5 Not important at all. So your first task is to create a new feature (e.g., `strong_amer_ident`), where 1 = strong identity (`amer_ident` = 1 or 2) and 0 = weak identity (`amer_ident` = 3, 4, or 5). Of note, this is an imperfect measure to be sure, but it gets to the substantive bottom line of this simple case.
2. (20 points) Build a self-organizing map based on *only* the above-listed social questions. Note: think carefully about the *scale* of the responses. The grid can be specified however you'd like, and hyperparameters can be tuned however you'd like. Just make decisions and justify them as you go. *Hint*: You might consider the `kohonen` package in R (though there are many others), or the `minisom` package in Python.
3. (20 points) Diagnose the output of your self-organizing map **visually.** *Hint*: consider looking at the help documentation for plotting options (e.g., counts, learning rate, etc.). *Discuss the model in a few sentences.*
4. (20 points) Color the output layer ("mapping") by your dichotomous feature for weak/strong **American** identity. Discuss the output in a few sentences. For example, do you see grouping along respondents with similar senses of American identity? Why or why not do you think? Any surprising patterns? And so on.
5. (20 points) Color the output layer ("mapping") by your dichotomous feature for weak/strong **race** identity. Discuss the output in a few sentences. For example, do you see grouping along respondents with similar senses of racial identity? Why or why not do you think? Any surprising patterns? And so on.

6. (10 points) Offer a few concluding thoughts comparing the patterns in responses to social questions across these two conceptualizations of identity. Talk about similarities and/or differences, what we can learn, how these patterns comport with your loose expectations, and so on. *Just a few sentences will suffice.*