Introduction to Machine Learning Course

Exercise 4
The Elections Challenge – Generative Modeling and Clustering

The elections are getting closer to the finish line. At this point the picture is getting clearer and party leaders (as well as journalists) begin to speculate on the day after the election. The discussion is mainly revolving around possible coalitions. Here are some of the questions that the campaign managers pose

- What will be a steady coalition (one that its voters are most similar) – show why this coalition will be more stable than other coalitions.
    - Terms for a "steady coalition"
        - Over 51% of the votes, relatively homogeneous with respect to the participating parties, and very much different from the opposition
- Identify the factor (voters' characteristic) which by manipulating you are most likely to change which party will win the election
- What would be a group of factors that by being able to manipulate, it will allow you to create a strong coalition per your choice?
    - see the definition of a steady coalition

Looking at these challenges, it became apparent to you that the best way to tackle these tasks are via generative modeling and clustering (why?).

Mandatory Assignment

You should submit a process that starts from loading and preparing the data, and up to the completion of the tasks detailed below. It is permitted to reuse anything from the processes you have submitted in the previous exercise (actually, it is even recommended)

A Python process that can handle the following prediction tasks

- Train the necessary models and make the calculation that will help you (manually) determine a steady coalition
- For each party, identify the leading features (hint: less than 8 features per party)

Such a process should implement and execute the following

1. Load the prepared training set
2. Train at least one generative model and one clustering model
    a. Each training should be done via cross-validation on the training set, to maximize performance of the model while avoiding overfitting
3. Load the prepared test set
4. Apply the trained models on the test set and check performance

Please submit

1. The Python script file that implements the modeling (training and testing) part
2. A documentation that
    a. Explains your process and any significant decision/insight you would like to share
    b. Suggest a steady coalition (one that its voters are most similar)
        i. Detail the calculations made to reach this decision
        ii. Explain why this coalition will be more stable than other coalitions
    c. For each party provide the leading features
    d. Identify the factor (voters' characteristic) which by manipulating you are most likely to change which party will win the elections
        i. Explicitly identify the manipulation needed (e.g. increase voters yearly income), and the new winning party
            1. handle this task strictly from a technical perspective, i.e. please ignore the semantic of the features
            2. Explain how did you identify these key factors
    e. Identify a group of factors that by being able to manipulate, it will allow you to either strengthen the coalition you suggested, or construct a stronger coalition
        i. Explicitly identify the manipulation needed and the resulted coalition
            1. handle this task strictly from a technical perspective, i.e. please ignore the semantic of the features
            2. Explain how did you identify these key factors

Comments

- This is probably the most creative task in the Election Challenge.
  There are no wrong answers; the answers will be evaluated based on the validity, evaluation criteria, creativity and novelty. Therefore, it is advisable that you will spend time in considering multiple options, and provide a thorough explanation.
- The aim of the "coalition" drill is to lead you to gain a "spatial" perspective of the parties' "locations", connections, and interlocks
- No bonus this time, everything is mandatory.
    - If you'll choose to implement an automation of (some of) the manual decisions, you won't get extra points, but you will gain my eternal admiration (for whatever it's worth …)

This exercise can be submitted in pairs!

- **You should submit only one copy but remember to document who are the contributors**
- **"No Couples/Triplets Swapping" during the semester**
    - **At least not without my formal approval**