

תיאור המשימה:

המעבדה עוסקת באמידת ותיאור מעבר קולות בין מפלגות בין שתי מערכות בחירות. יש להשתמש בקבצי תוצאות הבחירות על פי קלפיות בבחירות ספטמבר 2019 ומרץ 2020.

1. יהיו $n_{ij}^{(a)}$, $n_{ij}^{(b)}$ מספרי הקולות למפלגה j בקלפי i במערכות בחירות 2019, 2020 עבור הקלפיות המשותפות, נתונים במטריצות $N^{(a)}$, $N^{(b)}$ עבור 10 המפלגות הגדולות ביותר ב-2019 ו-9 המפלגות הגדולות ביותר ב-2020.
א. מצאו את המטריצה M^* בגודל 10×9 בשיטת הריבועים הפחותים ע"י מזעור ה-loss הריבועי הבא:

$$M^* = \operatorname{argmin}_M \|N^{(a)}M - N^{(b)}\|_F^2$$
כאשר $\| \cdot \|_F^2$ היא נורמת פרובניוס של מטריצה, כלומר סכום הריבועים של כל ערכי המטריצה. הציגו heatmap של המטריצה, כאשר שמות המפלגות מופיעים בשורות ובעמודות. סדרו את המפלגות בשורות ובעמודות כך שערכי m_{jk}^* הגבוהים ביותר יהיו קרוב ככל האפשר לאלכסון.
ב. אפסו כל ערך m_{jk}^* הקטן מ-0.5% וכן נרמלו כל שורה ב- M^* כך שסכומה יהיה 1 כנדרש ממטריצה מעבר. הציגו שוב את המטריצה כמו בסעיף א. נתחו את השינויים המשמעותיים בהצבעה על פי מטריצה זו - התמקדו בשינויים שאינם צפויים (כלומר פרט למעבר ממפלגה לעצמה או כתוצאה מאיחוד מפלגות).
2. חשבו שוב את המטריצה M^* בשיטת הריבועים הפחותים, אבל הפעם הוסיפו למטריצות $N^{(a)}$, $N^{(b)}$ עמודה המתארת את מספר בעלי זכות הבחירה שלא הצביעו לאף מפלגה בכל אחת ממערכות הבחירות, בהתאמה. (כאן M^* תהיה בגודל 11×10). הציגו את המטריצה כמו בשאלה 1 לאחר האיפוס ונרמול של סעיף ב. נתחו את הערכים שנוספו עבור השורה ועמודה שהוספו. כמו כן בדקו ודווחו האם וכמה הוספת השורה והעמודה השפיעה על אמידת ערכי המעברים בין המפלגות הגדולות (כלומר בשאר השורות והעמודות) לעומת סעיף 1.ב.
3. חשבו שוב את המטריצה M^* אבל הפעם בשיטת Non-Negative-Least-Squares, כלומר מזעור ה-loss הריבועי מסעיף 1.א. אבל תחת האילוץ שכל האיברים m_{jk}^* הם אי שליליים. תוכלו להשתמש בפונקציית `nnls` מתוך המודול `scipy`. הציגו את המטריצה כמו בשאלה 1 לאחר האיפוס ונרמול של סעיף ב. בדקו ודווחו האם הוספת אילוצי האי-שליליות לאופטימיזציה השפיעה על אמידת ערכי המעברים בין מפלגות לעומת סעיף 1.ב.
4. חשבו את מטריצת השאריות $Res = N^{(a)}M^* - N^{(b)}$ עבור המודל שנאמד בשאלה 2. עבור כל מפלגה (כולל "מפלגת" אלו שלא הצביעו), חשבו את השארית הריבועית הממוצעת על פני כל הקלפיות וציירו bar-plot של השאריות הריבועיות הממוצעות עם שמות המפלגות. עבור אילו מפלגות המודל לא מצליח לחזות בצורה טובה את ההצבעה עבורן בבחירות 2020 (b) מתוך בחירות 2019 (a)?
5. [בנוסף] חזרו על הצגת המטריצה הנאמדת M^* עבור השאלות 1-3, אבל הפעם באמצעות Sankey plot: [/https://plot.ly/python/sankey-diagram](https://plot.ly/python/sankey-diagram)
כאשר המפלגות בבחירות 2019 (a) הן בצד שמאל ובבחירות 2020 (b) בצד ימין.

הערות:

- חשבו על עיצוב הגרפים. תנו כותרת לצירים, שימו לב לאורך הצירים.
- השתמשו בצבעים, עובי נקודה, וכו' כדי להדגיש נקודות חשובות.