

מעבדה לסטטיסטיקה 52568 - 2019-20, מטלה 5. להגשה והצגה ב-1.12

תיאור המשימה:

המעבדה עוסקת בהצגה גלובלית של מפלגות וקלפיות באמצעות הורדת מימד בשיטת Principal Component Analysis (PCA). בנוסף לקבצי תוצאות בחירות ספטמבר 2019 יש להוריד ולהשתמש במשימה זו בקובץ בחירות אפריל 2019 לפי ישובים

1. [15 נק] הציגו תמונה של **המפלגות** במישור הנפרש ע"י שני המרכיבים הראשיים המובילים (PCA) עבור **קובץ הקלפיות** בבחירות ספטמבר 2019 ע"י scatter plot. כלומר, כל מפלגה ניתן לייצג ע"י וקטור באורך 10359 המכיל את שכיחויות ההצבעה למפלגה בכל קלפי - לאחר טרנספורמציה PCA המפלגה תיוצג ע"י וקטור באורך 2, כלומר נקודה במישור. הוסיפו טקסט המתאר את שם המפלגה ליד כל נקודה. האם הקרבה בין מפלגות מצופה/מפתיעה? האם תוכלו לתת משמעות לשני הצירים?
2. [25 נק] הציגו תמונה של **הקלפיות** במישור הנפרש ע"י שני המרכיבים הראשיים המובילים (PCA) עבור **קובץ הקלפיות** בבחירות ספטמבר 2019. כלומר, כל קלפי ניתן לייצג ע"י וקטור באורך 10 המכיל את שכיחויות ההצבעה לכל המפלגות הגדולות בקלפי - לאחר טרנספורמציה PCA הקלפי תיוצג ע"י וקטור באורך 2, כלומר נקודה במישור. בנוסף, בחרו שתי ערים עם לפחות 10 קלפיות והציגו את הקלפיות שלה בנקודות, בצבע אחר עבור כל עיר. האם ישנן קלפיות הנראות חריגות ביחס לעיר? בדקו אילו קלפיות נמצאות בחלקים שונים של התמונה ונסו לתת משמעות לצירים.
3. [60 נק] השוו את תוצאות הבחירות הראשונות ב-2019 (אפריל) לתוצאות הבחירות השניות ב-2019 (ספטמבר) באופן הבא:
 - א. הציגו תמונה של הישובים במישור הנפרש ע"י שני המרכיבים הראשיים המובילים (PCA) עבור **קובץ הישובים** בבחירות ספטמבר 2019, כאשר שטחי עיגולי הישובים פרופורציוניים לגודלם.
 - ב. חשבו עבור כל ישוב את שכיחות ההצבעה לכל מפלגה עבור 14 המפלגות הגדולות בבחירות אפריל 2019.
 - ג. כדי להתמודד עם השינויים בשמות והרכבי המפלגות בין שתי מערכות הבחירות, חשבו מתוך השכיחויות של 14 המפלגות הגדולות מה היו השכיחויות המשוערות של 10 המפלגות הגדולות כפי שהן קיימות כיום, באופן הבא: (עבור שאר המפלגות יש פשוט להעתיק את השכיחויות)

מפלגות באפריל 2019	מפלגה בספטמבר 2019
(איחוד מפלגות הימין)*0.5	עוצמה יהודית
ימין חדש + (זהות + איחוד מפלגות הימין)*0.5	ימינה
הליכוד + כולנו + (זהות)*0.5	הליכוד
עבודה + גשר	עבודה גשר
רעם בלד + חדש תעל	הרשימה המשותפת

ד. חשבו עבור וקטורי השכיחויות שקיבלתם בסעיף הקודם את הטלתם למישור הנפרש ע"י 2 המרכיבים הראשיים המובילים מסעיף א, והוסיפו אותם בצבע אחר לתמונה בסעיף א.

ה. עבור כל אחד מהישובים עם מעל 10,000 בוחרים בספטמבר 2019, הוסיפו לתמונה חץ המתחיל בעיגול של הישוב בבחירות אפריל 2019 ומסתיים בעיגול הישוב בבחירות ספטמבר 2019. תוכלו להשתמש בפקודה `plt.arrow` מתוך `matplotlib`. תארו את המתקבל - אילו מגמות מראים החצים?

ו. מצאו את שלושת הישובים עבורם השינוי בין שתי מערכות בחירות היה מקסימלי - כאשר השינוי מוגדר להיות המרחק הריבועי בין ייצוגי הישוב במישור הנפרש ע"י שני המרכיבים הראשיים המובילים ב-2 מערכות הבחירות. הוסיפו חצים ל-3 ישובים אלו כמו בסעיף ה' אך בצבע אחר וציינו את שמות הישובים ליד החצים. בנוסף, עבור כל אחד מ-3 הישובים ציירו גרף ובו `bar-plot` כפול המתאר את שכיחויות ההצבעה ל-10 המפלגות בישוב ב-2 מערכות הבחירות (לכל מפלגה 2 עמודות בצבעים שונים). תארו את השינויים העיקריים בהצבעה בישובים אלו.

הערות:

- חשבו על עיצוב הגרפים. תנו כותרת לצירים, שימו לב לאורך הצירים.
- השתמשו בצבעים, עובי נקודה, וכו' כדי להדגיש נקודות חשובות.
- מותר לכם להיות יצירתיים; נסו לחשוב על שיטות אחרות להצגת השינוי בין מערכות הבחירות. למשל, אפשר למצוא ישובים בהם המגמה היתה שונה מאשר המגמה בכלל הישובים.