**עיבוד אותות תרגיל 8**

שאלה 1)

1. השונות של כל מאפיין היא 2.5. השונות מחושבות ע"י חישוב השונות של כל עמודה:

Var([-2,-1,0,1,2]) = 2.5. שני הוקטורים זהים. (חישבנו את השונות במטלאב, וזוהי שונות מתוקננת. השונות הלא מתוקננת היא 2).

1. הכיוון הוא באלכסון בזווית של 45 מעלות מציר ה-x כי בכיוון זה השונות המקסימלית. מכיוון שהנקודות מונחות בדיוק על הישר y=x (כל הנקודות הן בעלות ערכי x ו-y שווים). הקרולציה היא מושלמת.
2. אחד. מכיוון שוקטור אחד תלוי ליניארית בשני ולכן הוא לא תורם שום מידע חדש, והשונות המוסברת שלו היא 0.
3. במידה ולא מנרמלים, כל וקטור מקורי הוא בעל שונות של 2.5. מכיוון שוקטור אחד הוא מיותר, ומסביר 0 שונות, השונות שלו PC1 תהיה 5, כי הוא מסביר 100 אחוז מהשונות המקורית. במידה ומנרמלים בעזרת zscore, הערך הוא 2.

שאלה 2)

1. המדינות שנראות קרובות: פורטוגל ויוון, אנגליה ושוודיה, גרמניה ופינלנד וספרד ואיטליה. גם ישראל וגם דנמרק נראות בודדות.
2. מהגרף המצורף ניתן ללמוד שה-PC הראשון הוא המשמעותי ביותר (מסביר יותר מ-60 אחוז מהשונות). 3 ה-PC הראשונים מסבירים את רוב השונות (בערך 90%).
3. לפי PC1, הפרמרטים (features) החשובים ביותר שמסבירים את הדתא הם – שוקולד, תבלינים, שום (בצורה קולרטיבית חיובית), בשר ודגים (בצורה קורלטיבית שלילית) וקפה הכי פחות משפיע.



1. לפי הטבלה, המאפיינים (fearutres) שנמצאים בקורלציה חיובית חזקה הם: תבלינים-שום, ודגים-בשר. המאפיינים שנמצאים בקרולציה שלילית חזקה הם: בשר-שום, בשר-תבלינים ודגים-תבלינים.

שאלה 3)

1. השונות של המאפיין הראשון: 0, השונות של המאפיינים השני והשלישי: 1 ושני שלישים. סכום השונויות הוא 3 ושליש.
2. בין המאפיינים השני והשלישי יש קורלציה מושלמת. (המאפיין השלישי הוא המאפיין השני ועוד 1, כלומר סכום של המאפיינים השני והראשון).
3. יש PC אחד שהשונות שלו שונה מאפס, מכיוון שישנו מאפיין אחד ללא שונות, שלא תורם לנו שום מידע. וישנם עוד שני מאפיינים בעלי קורלציה מושלמת ולכן נוכל להסתפק באחד מהם.
4. הערך עצמי יהיה 3 ושליש כי הוא מהווה מאה אחוז מהשונות המקורית.
5. נצפה שהמשקל ש PC1 יתן למאפיין הראשון הוא 0, כי הוא לא תורם שום שונות.
6. המשקלים של המאפיין השני והשלישי יהיו שווים, מכיוון שאחרי הנרמול יש להם שונות שווה והם תורמים מידע באופן שווה.
7. לא, מכיוון שממילא אנחנו מקבלים את הוקטורים העצמיים והערכים העצמים ממוינים בסדר יורד של הערכים העצמיים. לכן הסדר לא משנה.
8. אם לא ננרמל מאפיינים בעלי שונות שונה, המאפיינים שיש להם שונות גבוהה יותר ישפיעו יותר על הPC ויקבלו משקלים יותר גבוהים.