# פתרון מטלה 4 — חישוביות וקוגניציה, 6119

1 בדצמבר 2025

## שאלת הכנה

### סעיף א׳

יהי וקטור $u \neq 0$ ונרצה לתאר את $u$ על־ידי בסיס אורתונורמלי $(v^l)_{l=1}^N$. נסמן $u = \sum_{l=1}^N a_l v^l$ ונבדוק אילו טענות נכונות במקרה זה.
**פתרון** שינוי הכיוון של $u$ יגרור שינוי במקדמים $a_l$ (תשובה ג'), באופן דומה גם כפל בסקלר של $u$ יגרור גם שינוי למקדמים (תשובה ד'). אם $\|u\| = 1$ אז נובע שגם $\sum_{l=1}^N a_l = 1$ (תשובה ח'). כדי לחשבם מספיק מפריסה אורתוגונלית לחשב $a_l = \langle u, v^l \rangle$ (תשובה ט').

### סעיף ב׳

תהי מטריצת קורלציה $C$ בעלת שלושה ערכים עצמיים $\lambda_1 = 2\lambda_2 = 3\lambda_3$. נחשב את השונות שתתקבל מהטלה של דוגמה על הווקטור העצמי הגדול ביותר.

**פתרון** ראינו כי הטלה כזו מניבה את ערך השונות הגדול ביותר, ולכן נוכל להסיק ללא חישוב שהתשובה היא ד'.

## שאלה 1

יהי $y : \mathbb{R}^N \to \mathbb{R}$ נוירון לינארי המקיים $y(x) = w^t x$ כאשר $w$ וקטור משקולות. נניח ש־$\mathbb{E}(x) = 0$ וכן נניח ש־$C = \mathbb{E}(xx^t)$ מטריצת הקורלציה של $y$. מטרת הרשת היא למקסם על שונות הפלט תחת אילוץ על $w$.

### סעיף א׳

נחשב את $\mathrm{Var}(y)$ כאשר $\|w\| = 1$ וקטור עצמי של $C$. נבין מה הקשר בין התוצאה לבין פתרון אופטימלי ל־PCA.

**פתרון** לפי הגדרת השונות $\mathrm{Var}(y) = \mathbb{E}((y - \mathbb{E}(y))^2)$, ולכן,

$$\mathrm{Var}(y) = \mathbb{E}(y^2) - 2(\mathbb{E}(y))^2 + (\mathbb{E}(y))^2 = \mathbb{E}(y^2) - (\mathbb{E}(y))^2$$

נעבור לחישוב ערכים אלה,

$$\mathbb{E}(y) = \mathbb{E}(w^t x) = w^t \mathbb{E}(x) = 0$$

ישירות מהנתון $\mathbb{E}(x) = 0$, וכן,

$$\mathbb{E}(y^2) = \mathbb{E}((w^t x)^t w^t x) \stackrel{(1)}{=} \mathbb{E}(w^t x x^t w) = w^t C w = w^t \lambda w = \lambda$$

כאשר $\lambda$ הערך העצמי של הווקטור $w$ ב־$C$ וכאשר המעבר $(1)$ הוכח בכיתה. אז מצאנו ש־$\mathrm{Var}(y) = \lambda$ בדיוק. ככל שהשונות גדולה יותר בקבוצה סגורה נוכל להסיק שאיבדנו פחות אינפורמציה, לכן נצפה שבחירת $w$ עם ערך עצמי הגדול ביותר תניב את איבוד המידע הקטן ביותר, וקיבלנו את בדיוק אופן הפעולה של מזעור השגיאה ב־PCA.

### סעיף ב׳

נניח ש־$N = 2$ וכן $x = (x_1, x_2)$ עבור $x_1 \sim N(0,1), x_2 \sim N(0,4)$ בלתי־תלויים.

#### i

נחשב את מטריצת הקורלציה.

**פתרון** נבחין כי $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ ולכן עבור $X \sim N(a,b)$ נקבל $\mathbb{E}(X^2) = a + b$. נתון כי $\mathrm{Cov}(x_1, x_2) = 0$, ולכן,

$$C = \mathbb{E}(xx^t) = \mathbb{E}\begin{pmatrix} x_1 x_1 & x_1 x_2 \\ x_2 x_1 & x_2 x_2 \end{pmatrix} = \begin{pmatrix} \mathbb{E}(x_1^2) & \mathbb{E}(x_1)\mathbb{E}(x_2) \\ \mathbb{E}(x_1)\mathbb{E}(x_2) & \mathbb{E}(x_2^2) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$$

#### ii

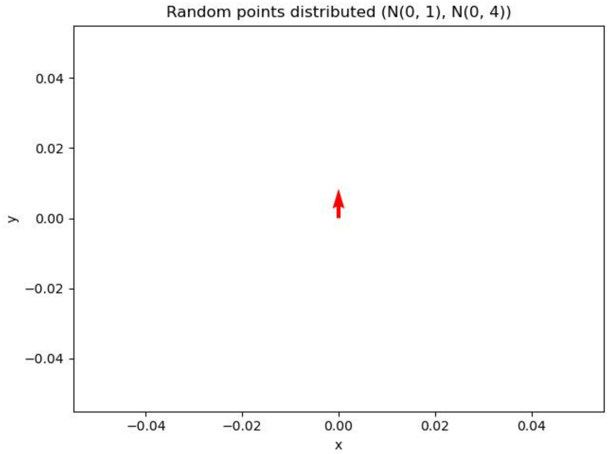
נמצא $w$ אופטימלי תחת האילוץ $\|w\| = 1$.

**פתרון** קיבלנו ש־$C$ אלכסונית וכן ש־$C(0,1)^t = 4(0,1)^t$ מחישוב ישיר, ולכן $w = (0,1)^t$ הוא האופטימלי.

#### iii

נשרטט במישור דוגמה להתפלגות אופיינית וכן נצייר את הכיוון של $w$.

**פתרון** נשרטט,




אנחנו יכולים לראות שהמרחק האופקי בין הנקודות הוא גדול מהמרחק האנכי של הנקודות, בכך מתקבל ההיגיון שמאחורי וקטור משקולות אשר תולה את כל ערכו בערך ה־y.

## iv

נחשב את אחוז השונות המוסברת.

**פתרון** הגדרנו את השונות המוסברת על־ידי $\text{Var}(y) = w^tCw = 4$ וכן הגדרנו את השונות הכוללת על־ידי $\text{Var}(x^tx) = \text{trace}(C) = 5$ ולכן אחוז השונות המוסברת הוא 80.

### v

נקבע ללא חישוב את וקטור המשקולות האופטימלי במקרים $x_1 \sim N(0,4), x_2 \sim N(0,1)$ ו־$x_1 \sim N(0,1), x_2 \sim N(0,2)$.

**פתרון** במקרה הראשון נקבל $w = (1,0)^t$ בשל סימטריה למקרה שחקרנו בסעיפים הקודמים. במקרה השני נקבל וקטור זהה לווקטור $w$ שחישבנו בסעיפים הקודמים, מאותה סיבה שקיבלנו את האחד שקיבלנו.

## vi

נניח עתה ש־$x_1, x_2 \sim N(0,1)$ בלתי־תלויים, נחשב את מטריצת הקורלציה ואת ערכיה העצמיים, נבין מה חריג במקרה זה ונחשב את וקטור המשקולות האופטימלי תחת האילוץ $w \in S(0,1)$.

**פתרון** עתה נקבל,

$$C = \begin{pmatrix} \mathbb{E}(x_1^2) & 0 \\ 0 & \mathbb{E}(x_2^2) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

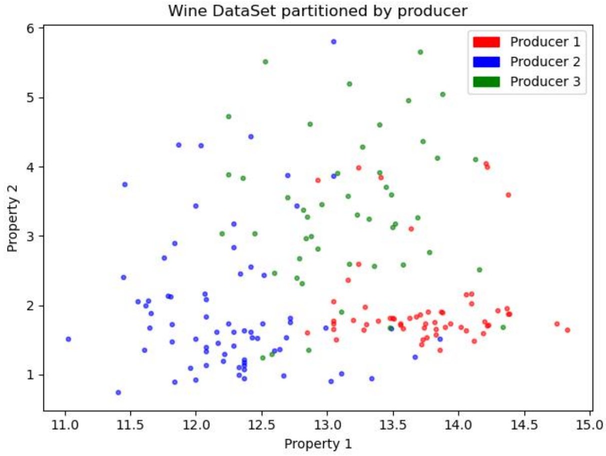
תוך שימוש בתת־סעיף i. הפעם קיבלנו מטריצה בה כל הערכים העצמיים הם 1 ואלכסונית, ולכן כל וקטור $w \in S(0,1)$ יקבל $Cw = 1 \cdot w$, כלומר נוכל לבחור כל וקטור.

# שאלה 2

לאורך התרגיל נעבוד עם dataset שמציג מידע על סוגי יינות המיוצרים על־ידי שלושה יצרני ענבים באזור מסוים באיטליה.

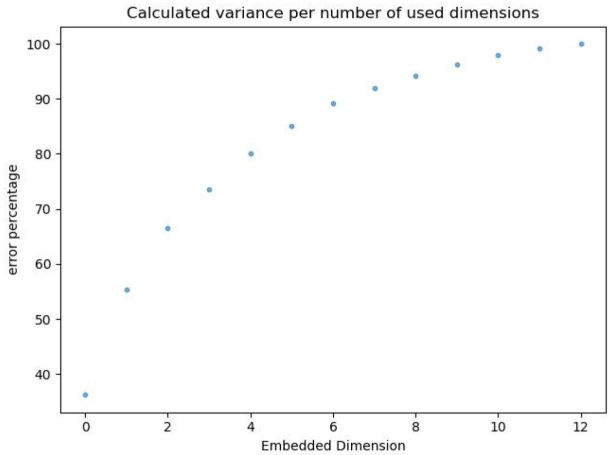
## סעיף ב׳

נציג את הווקטורים על מישור על־ידי ההעתקה $T(\bar{x}) = (x_1, x_2)$,



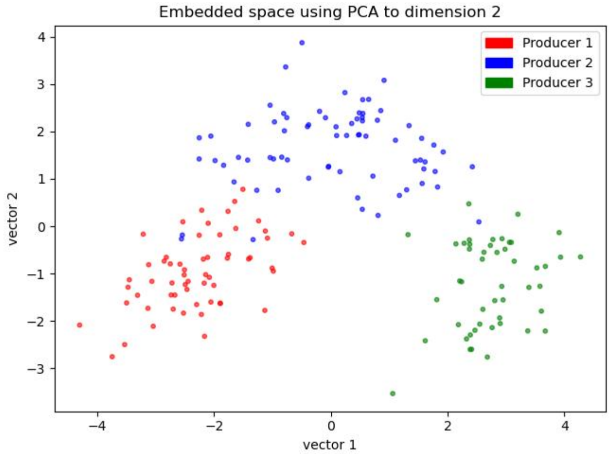

## סעיף ה׳

נציג גרף של אחוז השונות המצטברת, בציר $x$ יוצג מספר הרכיבים שאנו מטילים עליהם (בהתאם להגדרת PCA) ובציר $y$ נציג את אחוז השונות.




## סעיף ו׳

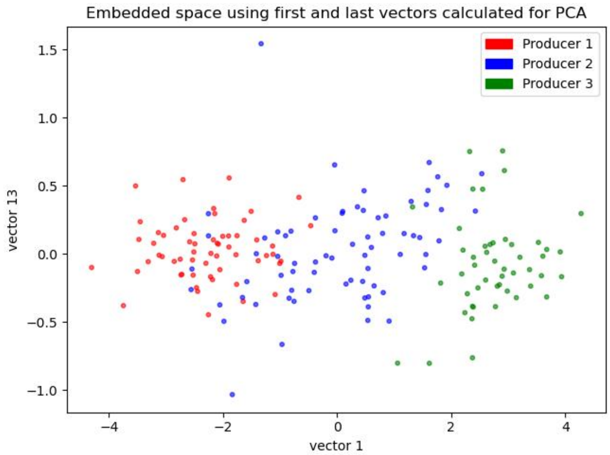
נשרטט את הנקודות כאשר הן מוטלות על־ידי שני הרכיבים הראשונים שחישבנו עם PCA עבור הווקטורים.




נבחין כי גרף זה נראה מהותית שונה מהגרף בסעיף ב', זאת שכן נקודות בצבעים שונים מופיעות באופן יותר צפוף ומרוחק משאר הנקודות. כלומר יש איזשהו מתאם שהצגה זו יוצרת בין יצרן ותכונות היין שלו, תכונות ששימוש בשתי תכונות היין הראשונות (לפי סדר) לא מצליח לתפוס.

## סעיף ז'

הפעם נשרטט גרף שמשתמש ברכיב הראשון והאחרון שחושב עם PCA ונשווה את התוצאה לתוצאות הקודמות,




בזמן שהרכיב הראשון מצליח ליצור בידול ויזואלי בין הנקודות המייצגות יין של יצרנים שונים, הציר השני משפיע באופן זניח על בידול התוצאות ולמעשה בלתי ניתן לראיה.