

פתרונות ומפתח ניקוד

תשובה	בגרסה 3	בגרסה 2	בגרסה 1	תאור	פריט
1	3	2	1	חיידקים	1
4	1	3	2	מסלול	2
3	2	1	3	עצמי	3
3	5	6	4	רגרסיה	4
1	9	8	5	FDR	5
1	6	4	6	מוזיקה	6
3	4	9	7	KMEANS	7
2	7	5	8	דנדרוגרמה	8
4 גם 5 וגם 4	8	7	9	רשת	9

פתוחה דינמית גרסה 111 (הגרסאות האחרות זה אותו דבר מסובב)

נתונה המערכת הדינמית הבאה:

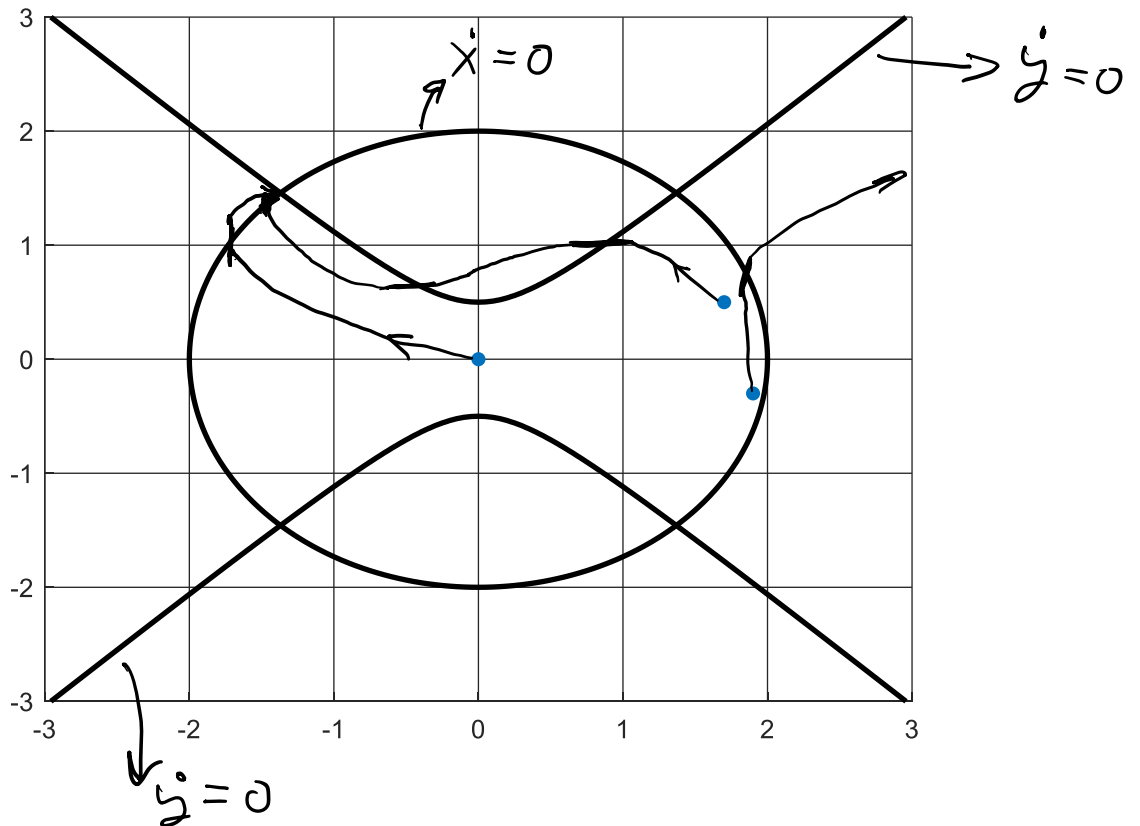
$$\begin{aligned}\dot{x} &= x^2 + y^2 - 4 \\ \dot{y} &= \epsilon (x^2 - y^2 + 0.25) \\ \epsilon &= 1\end{aligned}$$

כמו כן, בצירור שלפניכם מופיעים עקומי האפס של המערכת, וגם מסומנות עליו שלוש נקודות התחלה של מסלולים:

נקודה A (0,0)

נקודה B (1.7,0.5)

נקודה C (1.9, -0.3)



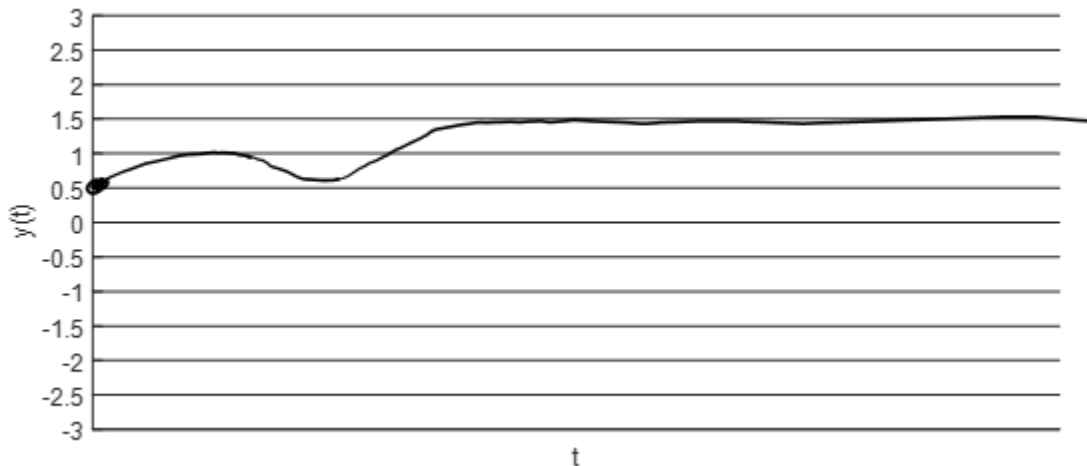
1. ציינו איזה עקום אפס מתאים לאיזה משתנה. נמקו בקצרה.

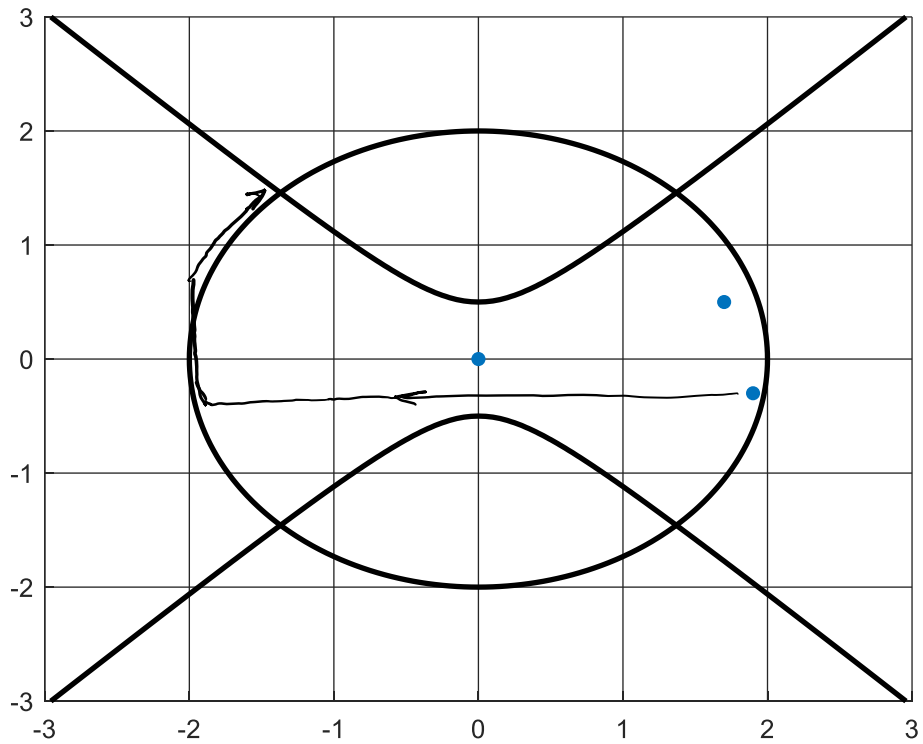
$$\dot{x} = 0 \quad -k11eN \quad N \text{ מעגל}$$

- נתון כי המסלולים המתחילים בנקודות B,A מסתיימים באותו מקום.
- כמו כן נתון כי המסלול המתחיל בנקודה C מסתיים במקום שונה מהם.
- 2. ציירו את שלושת המסלולים.
- 3. ציירו את $y(t)$ עבור המסלול שמתחיל בנקודה B.
- 4. כעת משנים את הערך של אפסילון ל $\epsilon = 0.001$.
- ציירו את המסלול המתחיל בנקודה C.
- יש לצייר רק במרחב הפאזה. אין צורך לצייר שוב את $y(t)$

דגשים והערות:

- יש לצייר מסלולים עד הגעה לגבולות מערכת הצירים, או עד הגעה לנקודת שבת יציבה.
- יש מקום בהמשך לכל החישובים.
- יש מערכות צירים כדי לרשום בהן את התשובה, ועוד נוספות למקרה שצריך לתקן.
- יש מערכות צירים נוספות בתור טיוטה – שימו לב שברור לחלוטין איפה התשובה הסופית שלכם לכל סעיף
- בכל סעיף יש להראות את כל החישובים שנחוצים כדי להגיע לתשובה הסופית. אם הגעתם למסקנה כלשהי בלי חישובים – זה גם בסדר, אבל צריך להסביר איך עשיתם זאת.
- אין להראות חישובים שלא נחוצים (אפשר לרשום אותם על טיוטה כמובן).
- יש לוודא כי התשובה הסופית כתובה באופן ברור לגמרי.





ניקוד:

3	שיוך נכון של עקומי אפס
2	נימוק נכון לשיוך
5	שלושת המסלולים נכונים באופן כללי: מגיעים למקומות הנכונים, עוברים בדרך במקומות הנכונים
2	המסלולים חוצים עקומי אפס בכיוון הנכון – בניצב לצירים
3	אין שינוי כיוון במסלולים ללא חציית עקומי אפס
5	ציור $\gamma(t)$: מתחיל במקום הנכון. עולה למקסימום של בין 0.5 ל-1.5. יורד למינימום שגדול מ-0.5. עולה לנקודת שבת ב-1.5 לא חייבים לדייק במספרים. לא חייבים להראות תנודות סביב נקודת השבת.
4	הפרדת קבועי זמן – המסלולים נכונים באופן כללי
2	הפרדת קבועי זמן – חצייה וזחילה נכונה של עקומי אפס

333	222
$\dot{x} = x^2 + y^2 - 4$ $\dot{y} = \epsilon (y^2 - x^2 - 0.25)$ $\epsilon = 1$	$\dot{x} = \epsilon (y^2 - x^2 + 0.25)$ $\dot{y} = x^2 + y^2 - 4$ $\epsilon = 1$

פתוחה רב-מימד

ארבעה (4) מטופלים הגיעו לחדר מיון עם תסמיני מחלה נדירה.

כדי להתאים את הטיפול הצוות בחר לאשכל את המטופלים (לבצע clustering).

כחלק מהערכה הקלינית של המטופלים נאספו מדדים רבים וחושבו המרחקים בין החולים:

מס' חולה	A	B	C	D
A	0			
B	1	0		
C	4	2	0	
D	10	8	3	0

1. באיזו שיטת linkage (complete או single) נמליץ לצוות ליישם על מנת שתתקבלנה 2 קבוצות זהות בגודלן?

COMPLETE

	A-B	C-D
A-B	0	
C-D	10	0

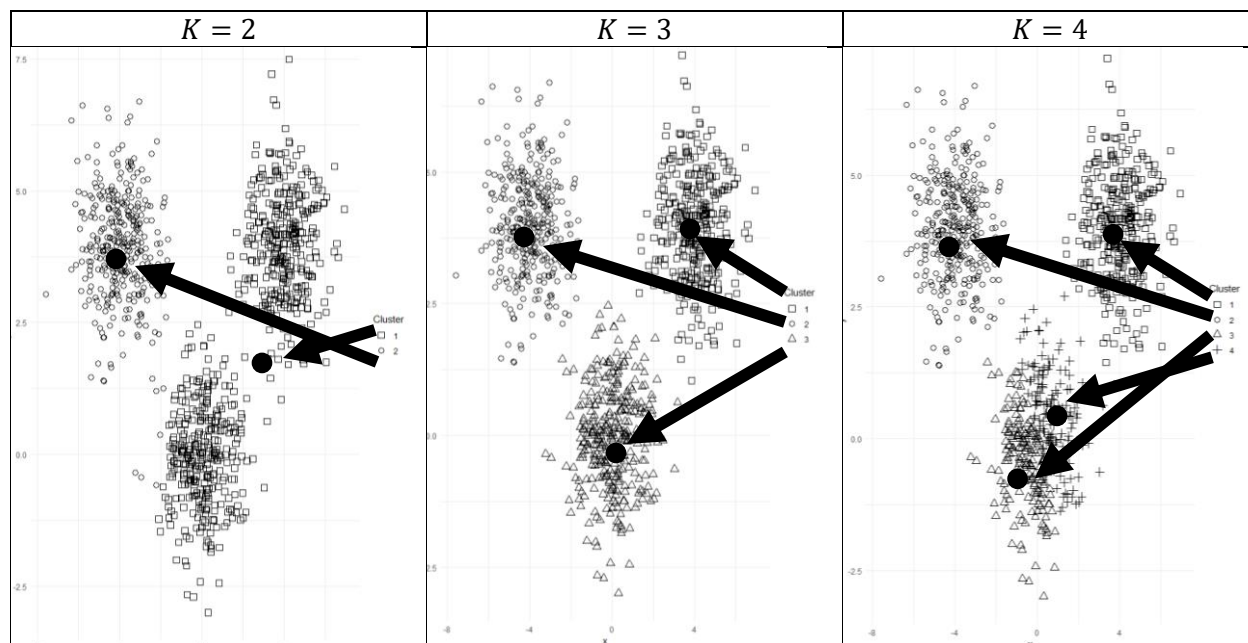
2. על מנת לקבוע את היעילות של התערבות רפואית, נשאף שהשונות בתוך הקבוצות תהיה כמה שיותר קטנה לעומת השונות בין הקבוצות. האם שיטת ה-linkage שנבחרה מתאימה להשגת המטרה הזאת? נמקו.

באופן כללי, השונות בתוך קבוצה קטנה יותר בשיטת complete linkage בהשוואה לשיטת single linkage. הסיבה לכך היא ששיטת complete linkage לוקחת בחשבון את המרחק המקסימלי בין כל זוג נקודות מאשכולות שונים בעת מיזוג אשכולות, מה שמאפשר אשכולות קומפקטיים יותר ומקובצים יותר.

single linkage, לעומת זאת, האשכולות נוצרים על סמך המרחק המינימלי בין כל שתי נקודות מאשכולות שונים. זה יכול להוביל למיזוג של אשכולות שיש להם נקודות המרוחקות זו מזו, וכתוצאה מכך שונות גדולה יותר בתוך הקבוצה.

עם התפשטות המחלה הנדירה, התקבלו למיון 1000 מטופלים נוספים והתברר כי יש לחלק אותם ליותר קלאסטרים. הוחלט להשתמש בשיטת K-means.

לפניכם התוצאות (שימו לב לכותרת שמציינת באיזה K השתמשו בכל מקרה)



3. העריכו איפה נמצא הסנטרואיד של כל קלאסטר.

עליכם לסמן את הנקודה עבור כל קלאסטר, ועבור כל ערך של K.

יש לסמן על גבי האיורים שנמצאים בראשית עמוד זה. הקפידו שברור לחלוטין איפה הסימונים, ולאיזה קלאסטר הם מתייחסים.

שימו לב, בהמשך יש את אותם איורים שוב פעם בתור טיוטה.

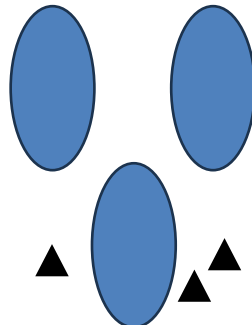
4. לפיכך ערכי ה-silhouette של כל קלסטר בכל K:

K = 2			K = 3			K = 4		
Cluster	Size	Average silhouette	Cluster	Size	Average silhouette	Cluster	Size	Average silhouette
1	659	0.44	1	333	0.68	1	332	0.63
2	340	0.73	2	333	0.69	2	333	0.68
			3	333	0.65	3	179	0.35
						4	155	0.32

. הסבירו בקצרה כיצד תקבעו מהו מספר הקלסטרים האופטימלי.

נמצא את ערכו של ה-Silhouette בכל אחד מהמצבים.
 נראה כי הערך המינימלי מתקבל תחת חלוקה ל-3 קלסטרים (0.673)

5. בהתבסס על מספר הקלסטרים האופטימלי שקבעתם בסעיף ד', מהי נקודת האתחול ה"אופטימלית" של האלגוריתם, ממנה יתכנס האלגוריתם הכי מהר? ומה עלולה להיות נקודת האתחול הגרועה ביותר?
6. בהינתן שמספר הקלסטרים האופטימלי שווה ל-3, נקודות האתחול האופטימליות הינן במרכז הכובד של קלסטר. לעומת זאת, נק' האתחול הגרועות ביותר תמוקמנה כך שהם ימצאו קלאסטר "מקומי" ויתר הקלסטרים יתערבבו זה בזה, למשל כך (אליפסות – הנתונים, משולשים – נקודות אתחול):



ניקוד: (שימו לב שהיה עדכון)

4		1
	בחירה בCOMPLETE	
3		2
	הסבר מראה הבנה של COMPLETE	
	הסבר מראה הבנה של SINGLE	
	הסבר לוגי מדוע עדיף COMPLETE	
4		3
	2=K	
	3=K	
	4=K	
6		4
	בחירה של 3=K	
	נימוק עם ממוצע silhouette	
3		5
	אופטימלי	
	גרועה	