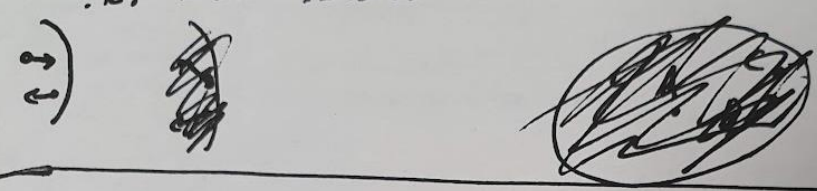


שיטות כמותיות במדעי הרפואה ב' - 274166 - מועד א 24/7/2023 תשפ"ג


תיקונים בזמן הבחינה:

פגזות דינמיות 333. הקואורנטים A, B, C
נמוגה. הציון לא.



שאלת מוסקת סגורה. אין סדר בקטגוריות

ציון נוסף



דף נוסחאות – 1 מתור 2

ערכים עצמיים במטריצה $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ הם $\lambda_{1,2} = \frac{1}{2} [\tau \pm \sqrt{\tau^2 - 4\Delta}]$

כמו כן מתקיים: כאשר

$$\Delta = ad - bc$$

$$\tau = a + d$$

$$\lambda_1 + \lambda_2 = \tau$$

$$\lambda_1 \lambda_2 = \Delta$$

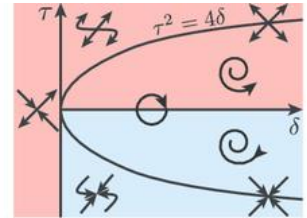
פתרון המערכת $\dot{x} = Ax$ הוא $x(t) = c_1 e^{\lambda_1 t} v_1 + c_2 e^{\lambda_2 t} v_2$ (כאשר $\lambda_1 \neq \lambda_2$)
 נוסחת אוילר

הקשר בין ערך עצמי מרוכב לתדר
 זמן מחזור:

$$e^{i\theta} = \cos \theta + i \sin \theta$$

$$\lambda = \alpha \pm i\omega$$

$$\omega = 2\pi f = \frac{2\pi}{T}$$



$$\begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix}$$

היעקוביאן של המערכת הוא $\begin{cases} \dot{x} = f(x, y) \\ \dot{y} = g(x, y) \end{cases}$

פתרון המערכת הלינארית: $\frac{dx}{dt} = -qx + u(t)$

הוא $x(t) = x(0)e^{-qt} + \int_0^t e^{-q(t-s)} u(s) ds$

ובמקרה ש u לא תלוי בזמן, מקבלים: $x(t) = x(0)e^{-qt} + \frac{u}{q}(1 - e^{-qt})$

פתרון אינטגרל של פונקציה מעריכית (אקספוננט): $\int_a^b e^{cx} dx = \frac{1}{c}(e^{cb} - e^{ca})$

טריגונומטריה:

$$\sin\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2}, \sin\left(\frac{\pi}{2}\right) = 1, \sin(\pi) = 0, \cos\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2}, \cos\left(\frac{\pi}{2}\right) = 0, \cos(\pi) = -1$$

$$\frac{d}{dx} [\sin(ax)] = a \cos(ax)$$

דף נוסחאות – 2 מתוך 2

מרחק אוקלידי בין נקודות x, y עבור n מימדים:

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

קורלציית פירסון מוגדרת כ:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

וקורלציית ספירמן מוגדרת כ:

$$\rho = \frac{S_{xy}}{S_x S_y} = \frac{\frac{1}{n} \sum_{i=1}^n (R(x_i) - \overline{R(x)}) \cdot (R(y_i) - \overline{R(y)})}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^n (R(x_i) - \overline{R(x)})^2\right) \cdot \left(\frac{1}{n} \sum_{i=1}^n (R(y_i) - \overline{R(y)})^2\right)}}$$

כאשר $R(x_i)$ הינו דירוג הערך של x_i בין כל ה- n דוגמאות

חישוב סילואט silhouette עבור נתון מסויים i

$$s(i) = \begin{cases} 1 - a(i)/b(i), & \text{if } a(i) < b(i) \\ 0, & \text{if } a(i) = b(i) \\ b(i)/a(i) - 1, & \text{if } a(i) > b(i) \end{cases}$$

כאשר $a(i)$ הוא המרחק הממוצע בין הנתון i לבין כל שאר הנתונים באותו קלאסטר.

ואילו $b(i)$ הוא המינימום על גבי כל שאר הקלאסטרים של: ממוצע המרחקים בין הנתון i לבין כל הנתונים בקלאסטר מסויים אחר.

שיטות כמותיות במדעי הרפואה ב' – 274166 – מועד א 24/7/2023 תשפ"ג

גרסה מספר: 111

שם הסטודנט _____

ת.ז. _____

סטודנט/ית יקר/ה,

המבחן כולל: 9 שאלות אמריקאיות ו- 2 שאלות פתוחות

ניקוד: 6 נקודות לכל שאלה אמריקאית, 26 נקודות לשאלה פתוחה של דינמיות, 20 נקודות לשאלה פתוחה של רב-מימד.

חומר עזר: מחשבון

יש לענות על השאלות הפתוחות על גבי טופס המבחן בלבד – בצירוף מדבקה במקום המיועד. את מחברות הטייטה לא מגישים, ואין להן כל תוקף לבדיקת המבחן או לערעור.

אנא קרא/י בתשומת לב את ההוראות:

- הדבק/י את מדבקת הברקוד במקום המיועד לכך. (במידה ואין בידך מדבקה אנא רשום את מספר ת.ז. במקום המיועד)
- יש למלא את טופס התשובות בעט כדורי בלבד על ידי סימון ברור!
- יש לסמן תשובה אחת בלבד, סימון שתי תשובות כנכונות תחשב כשגיאה גם אם אחת מהן נכונה.
- בכל מקרה של אי התאמה בין טופס הקידוד לסימון בשאלון הבחינה - טופס הקידוד הוא הקובע.

משך הבחינה: 3 שעות

ב ה צ ל ח ה !!!

אני (שם ומשפחה) _____ ת.ז. _____ נבחן/ת בבחינה זו, מצהיר/ה כי לא אעתיק ולא אשתמש באמצעים המנוגדים לטוהר הבחינות בטכניון.

חתימה _____ .

שאלה 1

חוקרת מבצעת ניסוי לבדוק את השפעת אנטיביוטיקה על זן מסוים של חיידקים בצלחת פטרי. החיידקים מתרבים, אך במקביל האנטיביוטיקה הורגת אותם. מכיוון שכמות המזון בצלחת מוגבלת, ככל שריכוז החיידקים עולה כך קצב התרבותם פוחת, ומעבר לריכוז מסוים ההתרבות עוצרת.

V – ריכוז החיידקים

R – ריכוז האנטיביוטיקה

$$\lambda, p, c > 0$$

נתמקד בריכוז החיידקים בלבד. אילו מהמשוואות הבאות מתארת את השינוי בריכוז החיידקים?

א. $\frac{dV}{dt} = V \left(\lambda - \frac{\lambda}{p} V \right) - cVR$

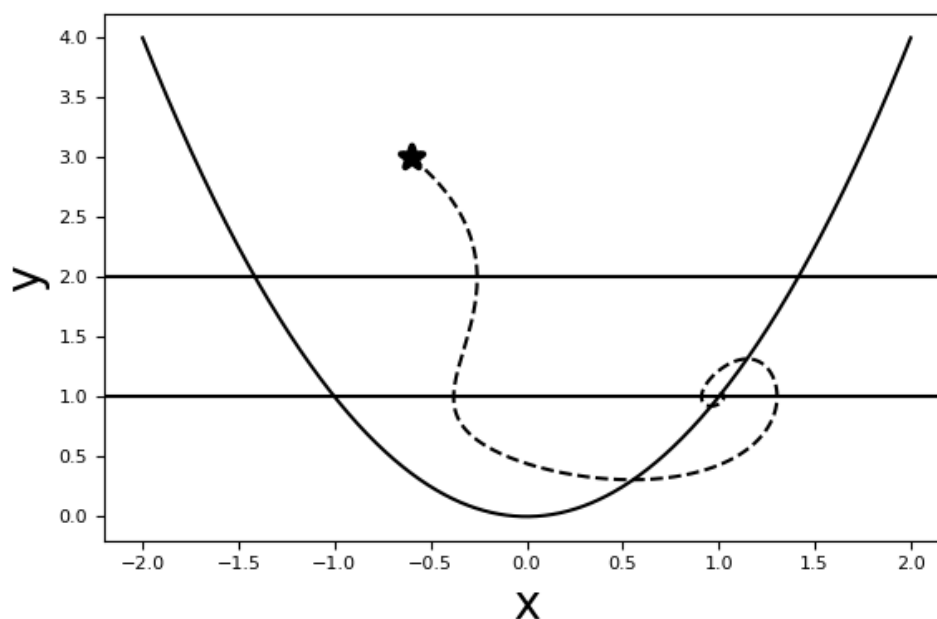
ב. $\frac{dV}{dt} = V \left(\lambda + \frac{\lambda}{p} V \right) - cVR$

ג. $\frac{dV}{dt} = V \left(\lambda - \frac{\lambda}{p} V \right) + cVR$

ד. $\frac{dV}{dt} = V \left(\lambda - \frac{\lambda}{p} VR \right) - cVR$

שאלה 2

נתונה מערכת דינמית דו-מימדית. באיור שלפניכם מופיעים עקומי האפס של המערכת (קווים מלאים), וכן מסלול אחד של מערכת זו (מתחיל בסימון הכוכב, ומסומן בקו מקוקו).



ביחרו את המשפט הנכון ביותר:

- א. הפרבולה היא עקום האפס של x
- ב. הקו האופקי העליון הוא עקום האפס של x והקו האופקי התחתון הוא עקום האפס של y
- ג. אם נקודת ההתחלה היא $(-2, 1.5)$ התנועה בהתחלה תהיה לכיוון שמאל למטה
- ד. אם נקודת ההתחלה היא $(-2, 2.5)$ התנועה בהתחלה תהיה לכיוון ימין למעלה

שאלה 3

נתונה המטריצה:

$$A = \begin{pmatrix} 5 & -1 \\ 6 & 0 \end{pmatrix}$$

ידוע כי $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ הוא וקטור עצמי של המטריצה. מהו הערך העצמי שמתאים לו?

- א. -3
- ב. -2
- ג. 2
- ד. 3

שאלה 4

בחרו את המשפט הנכון ביותר

- א. כפי שלמדנו בקורס קודם, רגרסיה לינארית היא פונקציה מהצורה: $f(x) = a_1 \cdot x_1 + a_2 \cdot x_2 + \dots$. לכן, רגרסיה לינארית היא דוגמה לאלגוריתם לא מונחה מכיוון שאינה מקבלת תיוגים כקלט בזמן החיזוי על המידע החדש.
- ב. ככל הניתן, נעדיף להשתמש באלגוריתמים לא מונחים גם אם קיימים תיוגים. זאת מכיוון שבמידע אמיתי קשה לסמוך על התיוגים שניתנו לנו. בנוסף, אלגוריתמים לא מונחים נותנים לנו אפשרויות בחירה רבות יותר (דנדוגרמה, פונקציית מרחק, linkage וכו').
- ג. בלמידה לא מונחית טיב התוצאה שלנו (לדוגמא, האישכול שלמדנו) מגיע מהנחות מוקדמות שלנו על מבנה הדאטא ומה מגדיר קבוצה, בעוד בלמידה מונחית טיב התוצאה שלנו נובע מהתיוגים.
- ד. KMEANS הוא אלגוריתם למידה מונחית מכיוון וצריך לבחור את K מראש.

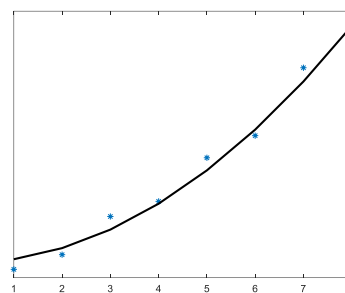
שאלה 5

תרכובת נסיונית שמטרתה להאיץ דיפרנציאציה של תאים ניתנה ל-50 תרביות, אך רק 21 מתוכן אכן הראו דיפרנציאציה מואצת. במטרה לגלות מה ההבדל בין תרביות שהגיבו לתרופה לאלו שלא, חושבו 8 מדדים ונערך מבחן סטטיסטי בין הקבוצות. ערך ה-p-value לכל מדד מופיע בטבלה:

מדד	טמפרטורת מינימום	טמפרטורת מקסימום	רמת סוכר לפני הטיפול	רמת סוכר אחרי הטיפול	כמות תאים בתרבית	רמת חשיפה לאור	גיל התרבית	צפיפות התרבית
p-value	0.02	0.07	0.009	0.006	0.075	0.002	0.09	0.015

הוגדר סף מובהקות $\alpha = 0.05$

בנוסף, שערך FDR בעזרת פרמוטציות הראה את הגרף הבא (בכונה אין מספרים בציר ה-Y):



כאשר הקו השחור מייצג קירוב של הקשר בין מספר הפרמטרים המובהקים n ל-FDR על ידי:

$$FDR = 0.01 + 0.0025 n^2$$

לפניכם מספר משפטים, כאשר רק אחד **לא נכון**. סמנו את המשפט **השגוי**:

- ההסתברות לטעות אחת לפחות מסוג 1 היא 0.663
- ללא תיקון להיפותזות מרובות ישנם 5 מדדים מובהקים
- עם תיקון **בונפרוני** להיפותזות מרובות ישנם 2 מדדים מובהקים
- עבור FDR 5% המדד "צפיפות התרבית" יהיה במדדים המובהקים, אך המדד "טמפרטורת מינימום" לא

שאלה 6

כחלק מבחירת התוכנית האומנותית לטקס "רופאים מצטיינים 2023", רופא שקדן החליט לעשות מעשה ולהשתמש בכלים אותם למד בקורס שיטות כמותיות בכדי לחלק את הרופאים במחלקה לקבוצות, כאשר חברי כל קבוצה נהנים מתחומים אומנותיים דומים (ושונים משאר הקבוצות).

לשם כך, הרופא אסף את המידע המופיע בטבלה מטה, כאשר כל נתון מייצג קטגוריה כלשהיא. לדוגמא במוסיקה, 1 גא'ז, 2 קלאסי, ...

בסדרות, ER 1, House 2, ...

לפניכם דוגמה לחלק מתוך הנתונים שנאספו

רופא א	רופא ב	רופא ג	
1	4	2	מוסיקה
3	3	8	סרטים
2	4	4	סדרות
5	5	5	סרטוני טיק-טוק

⋮

מהי פונקציית המרחק (בין רופאים) המתאימה ביותר עבור המשימה הנ"ל ?

- א. מרחק האמינג.
- ב. מרחק אוקלידי.
- ג. קורלציית פירסון.
- ד. קורלציית ספירמן.

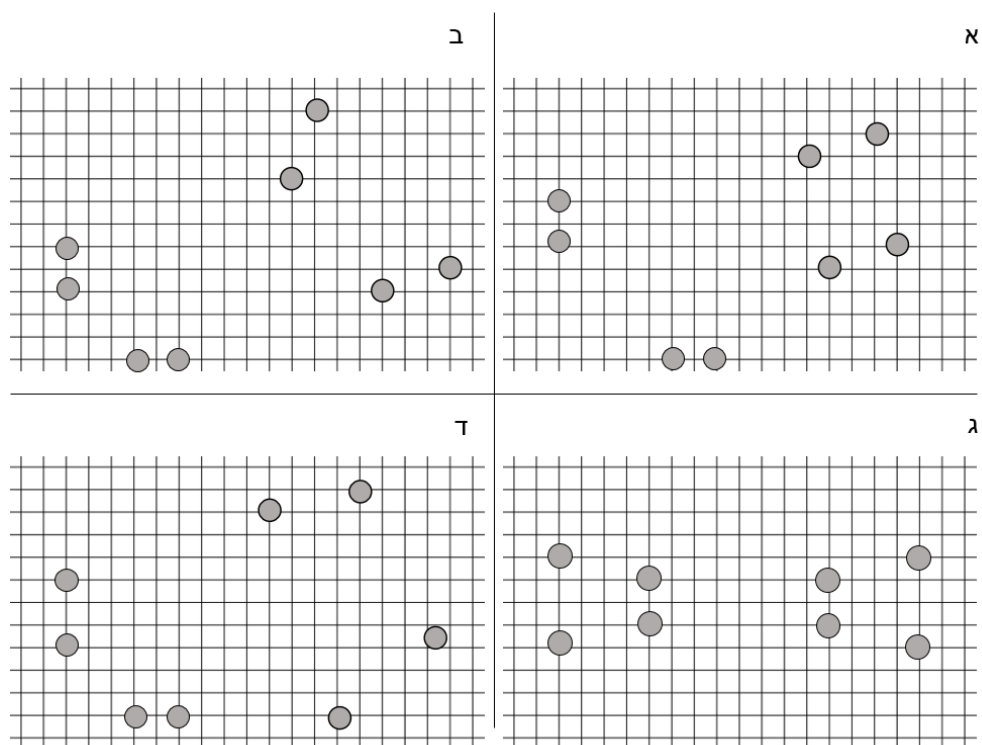
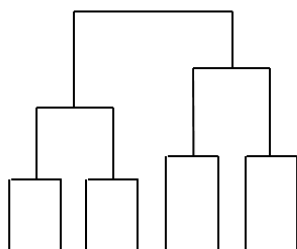
שאלה 7

בחרו את המשפט הנכון ביותר:

- א. ניתן לבנות דנדוגרמה גם עבור אישכול היררכי וגם עבור K-means
- ב. אלגוריתם Kmeans תלוי בהגדרת פונקציית מרחק בין נקודות. לעומת זאת, אישכול היררכי תלוי רק בפונקציית ה-linkage
- ג. שימוש בתנאי התחלה שונים כמה פעמים ב k-means נועדו לעזור לנו להגיע למינימום גלובלי.
- ד. באלגוריתם K MEANS, המינימום שניתן להגיע אליו עבור כל K הוא אותו מינימום.

שאלה 8

מהו מערך הדגימות המתאים לדנדוגרמה המוצגת משמאל? הניחו centroid linkage .



שאלה 9

גורם שעתוק הינו גן שביטוי גבוה שלו בתא גורם לביטוי גבוה של גנים אחרים (אחד או יותר). לצורך גילוי גורמי שעתוק חוקרים הביאו לביטוי חסר/יתר של גנים שונים, ומדדו האם כתוצאה מכך גנים אחרים הראו ביטוי חסר/יתר. כדי לנתח את התוצאות הוצע להשתמש ברשת מכוונת, שבה כל גן הוא צומת, וקשתות מחברות בין גנים שהחוקרים שינו לגנים שהשתנו כתוצאה מכך.

מה המשפט הנכון ביותר?

- א. ברשת זו סינגלטון יכול להיות גורם שעתוק
- ב. ברשת זו צומת עם דרגת כניסה גבוהה יכול להיות גורם שעתוק המשפיע על גנים רבים
- ג. רשת זו תהיה בהכרח scale free ותכיל מספר רב של גורמי שעתוק המשפיעים כל אחד על גן אחד.
- ד. ברשת זו גורם שעתוק בעל Centrality גבוה יכול להשפיע ישירות על הרבה גנים
- ה. ברשת זו קוטר הגרף מגביל את מספר גורמי השעתוק שיכולים להיות מעורבים בתהליך ביולוגי בודד

מדבקת ברקוד

שיטות כמותיות במדעי הרפואה ב' – 274166 – מועד א 24/7/2023 תשפ"ג

גרסה מספר: 111 - שאלות פתוחות

שם הסטודנט _____

ת.ז. _____

סטודנט/ית יקר/ה,

חלק זה כולל 2 שאלות פתוחות

ניקוד : 6 נקודות לכל שאלה אמריקאית, 26 נקודות לשאלה פתוחה של דינמיות, 20 נקודות לשאלה פתוחה של רב-מימד.

חומר עזר : מחשבון

יש לענות על השאלות הפתוחות על גבי טופס זה בלבד – **בצירוף מדבקה במקום המיועד**. את מחברות הטיוטה לא מגישים, ואין להן כל תוקף לבדיקת המבחן או לערעור.

פתוחה דינמיות

נתונה המערכת הדינמית הבאה:

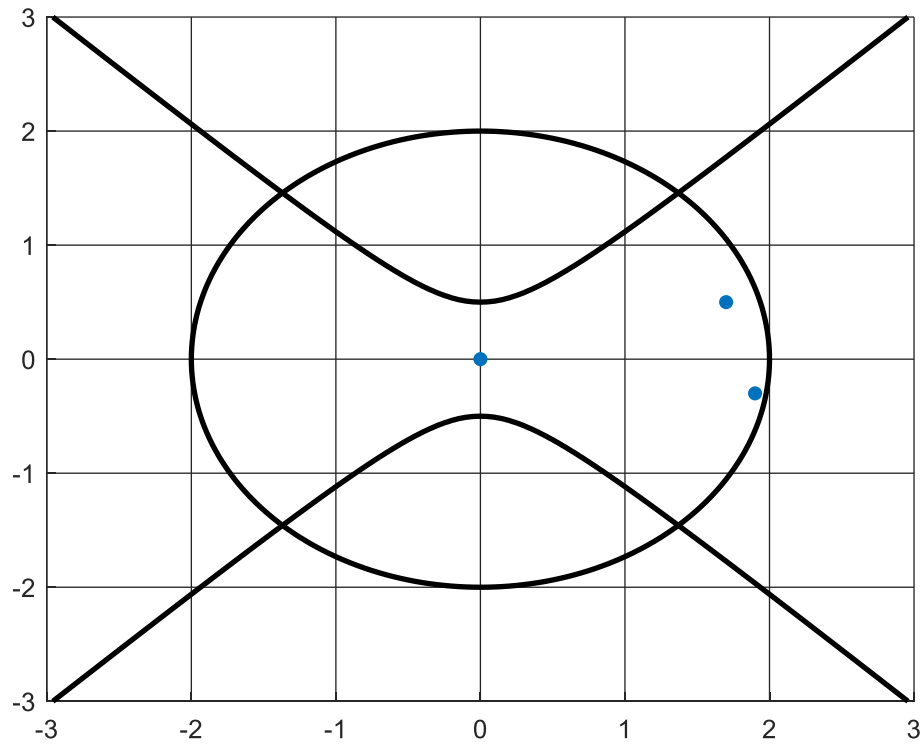
$$\begin{aligned}\dot{x} &= x^2 + y^2 - 4 \\ \dot{y} &= \epsilon (x^2 - y^2 + 0.25) \\ \epsilon &= 1\end{aligned}$$

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.

- 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 $\gamma(t)$ עבור המסלול שמתחיל בנקודה B.

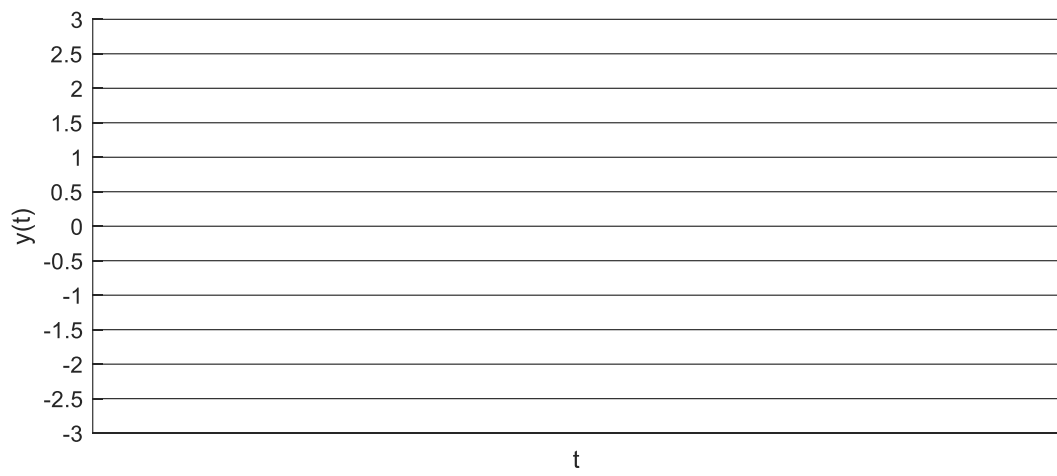
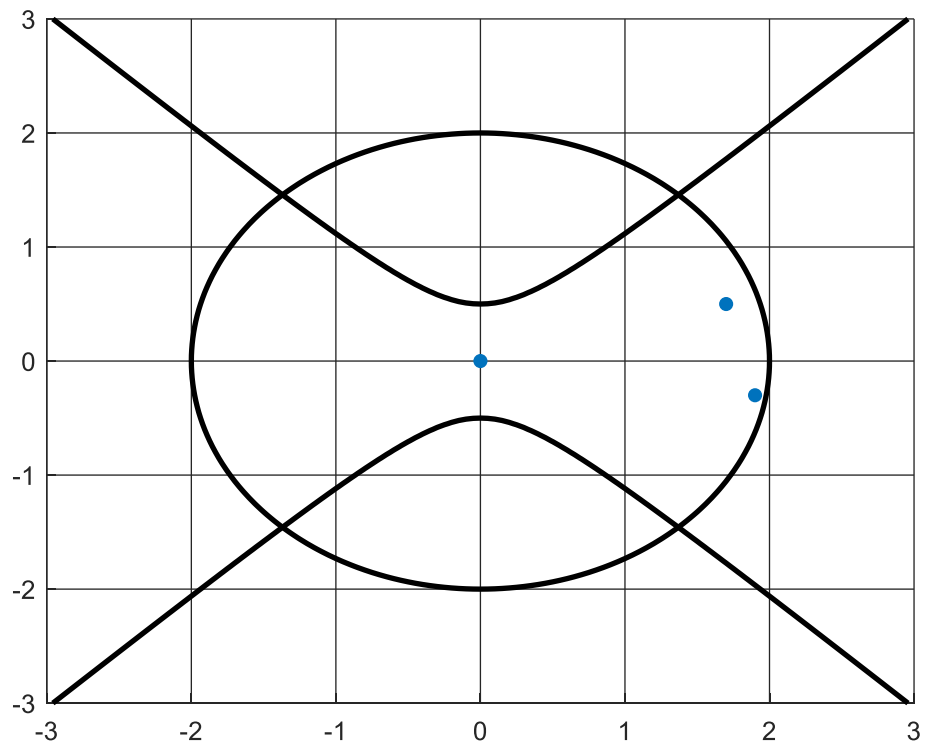
4. כעת משנים את הערך של אפסילון ל $\epsilon = 0.001$.

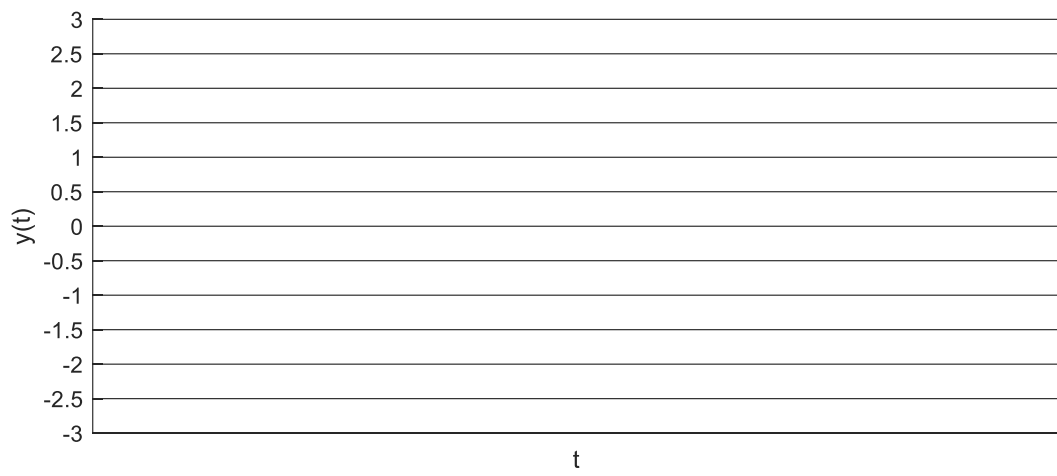
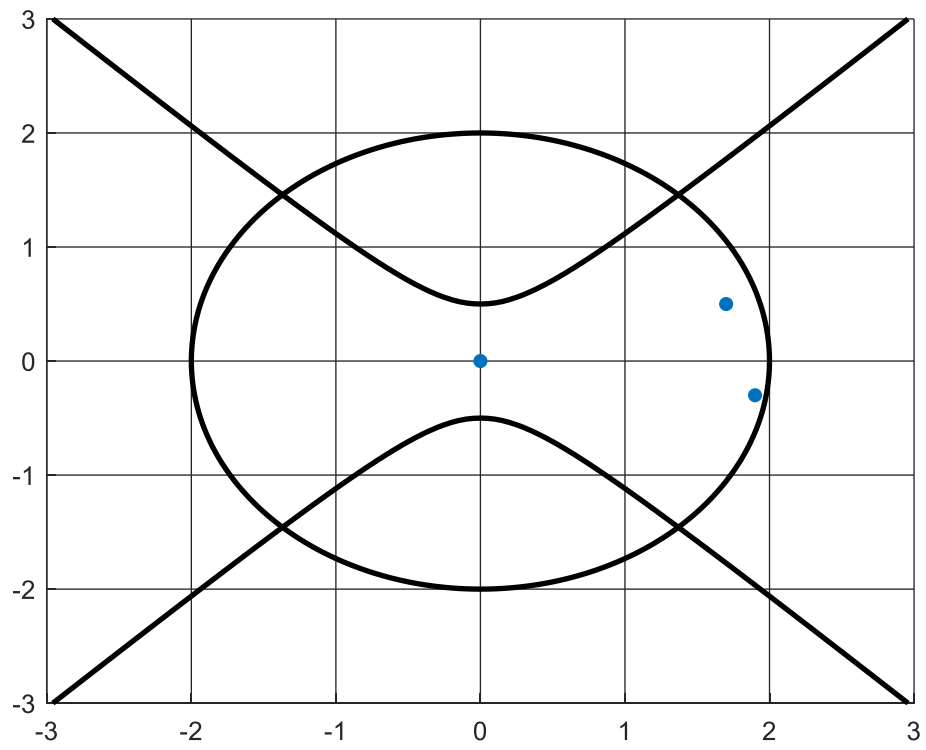
ציירו את המסלול המתחיל בנקודה C.

יש לצייר רק במרחב הפאזה. אין צורך לצייר שוב את $\gamma(t)$

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.





מקום נוסף לכתוב תשובות לשאלה פתוחה דינמיות

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

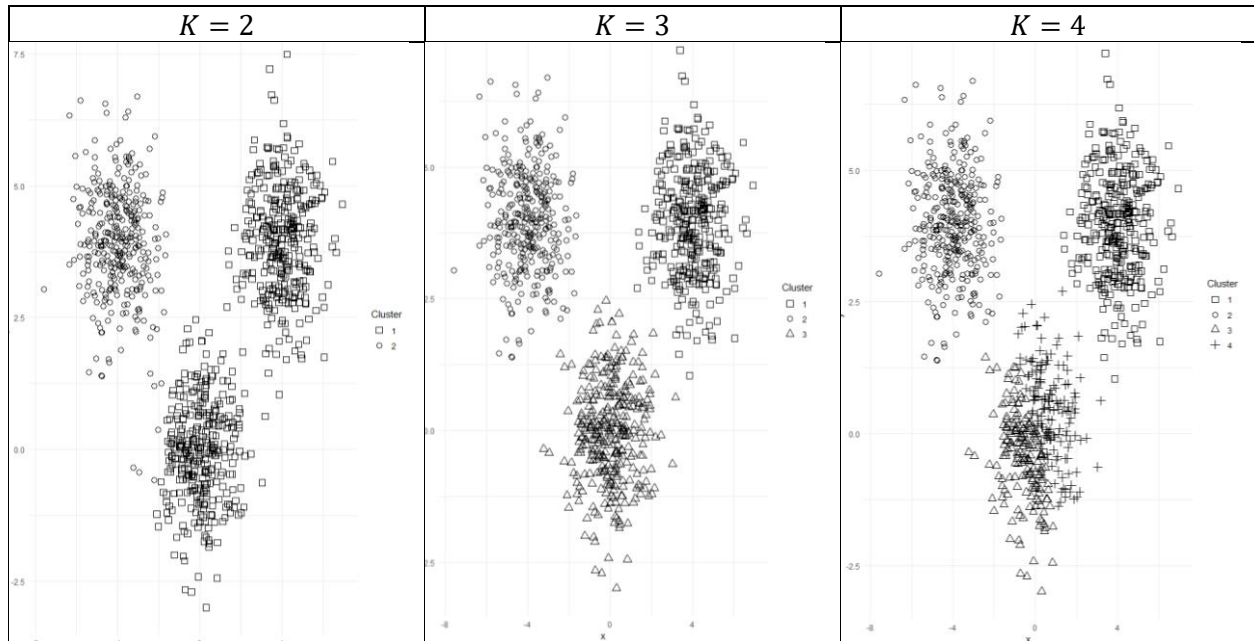
1. באיזו שיטת linkage (complete או single) נמליץ לצוות ליישם על מנת שתתקבלנה 2 קבוצות זהות בגודלן?

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.

(המשך בעמוד הבא)

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 האירים שנמצאים בראשית עמוד זה. הקפידו שברור לחלוטין איפה הסימונים, ולאיזה קלאסטר הם מתייחסים.

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

4. לפניכם 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.

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?

