

שיטות כמותיות במדעי הרפואה ב' - בחינה לדוגמא תשפ"ב

גרסה מספר:

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. את מחברות הטיטה 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 _____ .

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

$$\Delta = ad - bc$$

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

$$\tau = a + d$$

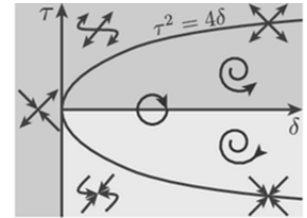
$$\lambda_1 \lambda_2 = \Delta$$

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

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

.1

נתונה המערכת:

$$\frac{dx}{dt} = 2x \cdot \left(1 - \frac{x}{2}\right) - xy$$

$$\frac{dy}{dt} = 3y \cdot \left(1 - \frac{y}{3}\right) - 2xy$$

מי מהבאים הוא עקום אפס של γ ?

א. ציר ה-X

ב. ציר ה-Y

ג. $y = 2 - x$

ד. $y = 3 + 2x$

פתרון: נבדוק מתי הנגזרות הרלוונטיות שווה לאפס:

$$\frac{dy}{dt} = 3y \cdot \left(1 - \frac{y}{3}\right) - 2xy = 0$$
$$y[3 - y - 2x] = 0$$

יש שני פתרונות:

$$y = 0$$

$$y = 3 - 2x$$

אחד מהם זה ציר ה-X.

אפשרות אחרת היא להציב את ארבעת התשובות במשוואה של dy/dt ולבדוק אם זה מתאפס.

.2

נתונה מערכת דו-מימדית לינארית:

$$\frac{dx}{dt} = \begin{pmatrix} 10 & 0 \\ -5 & -5 \end{pmatrix} x$$
$$x(0) = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$$

מה יהיה הערך של x בזמן $t=1$ באמצעות פתרון נומרי עם $dt = 0.5$?

- א. $\begin{pmatrix} 6 \\ -10 \end{pmatrix}$
- ב. $\begin{pmatrix} 11 \\ -25 \end{pmatrix}$
- ג. $\begin{pmatrix} 25 \\ 25 \end{pmatrix}$
- ד. $\begin{pmatrix} 36 \\ 0 \end{pmatrix}$
- ה. $\begin{pmatrix} 121 \\ 45 \end{pmatrix}$

פתרון: נשתמש בהגדרה של הנגזרת:

$$\frac{dx}{dt} = \lim_{dt \rightarrow 0} \frac{x(t + dt) - x(t)}{dt}$$

נתעלם מהגבול, ונרשום את זה עבור dt :

$$x(t + dt) = x(t) + \frac{dx}{dt} dt$$

במקרה שלנו, אנחנו יודעים גם את dt וגם את $\frac{dx}{dt}$:

$$x(t + 0.5) = x(t) + \begin{pmatrix} 10 & 0 \\ -5 & -5 \end{pmatrix} x(t) * 0.5$$

כעת נתקדם מ $t=0$ עד ל $t=1$, כלומר שני צעדים.

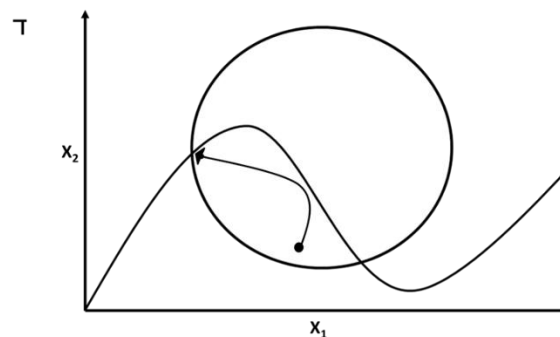
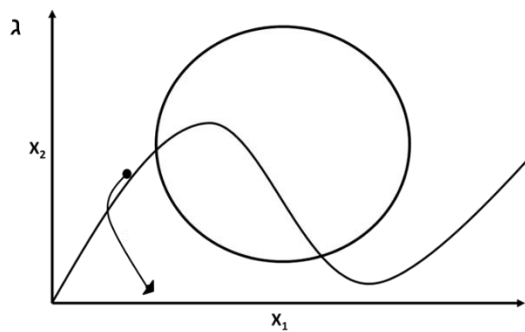
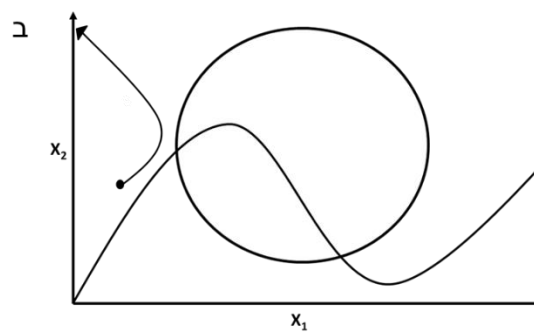
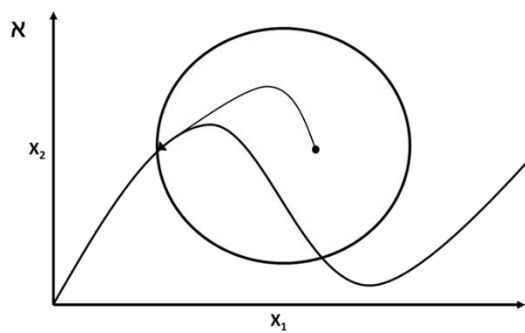
$$x(0.5) = \begin{pmatrix} 1 \\ 5 \end{pmatrix} + \begin{pmatrix} 10 & 0 \\ -5 & -5 \end{pmatrix} \begin{pmatrix} 1 \\ 5 \end{pmatrix} * 0.5 = \begin{pmatrix} 1 \\ 5 \end{pmatrix} + \begin{pmatrix} 5 \\ -15 \end{pmatrix} = \begin{pmatrix} 6 \\ -10 \end{pmatrix}$$

$$x(1) = \begin{pmatrix} 6 \\ -10 \end{pmatrix} + \begin{pmatrix} 10 & 0 \\ -5 & -5 \end{pmatrix} \begin{pmatrix} 6 \\ -10 \end{pmatrix} * 0.5 = \begin{pmatrix} 6 \\ -10 \end{pmatrix} + \begin{pmatrix} 30 \\ 10 \end{pmatrix} = \begin{pmatrix} 36 \\ 0 \end{pmatrix}$$

נתונה מערכת דינמית דו מימדית.

בכל אחד מהאיורים מופיעים שני עקומי האפס של המערכת (עיגול ועקום בצורת האות "מ" בכתב עגול), וכן מופיע מסלול אחד (מסומן בעיגול וחץ).

איזה מסלול מבין המסלולים אפשרי?



פתרון: לא ניתן לשנות כיוון (שמאל-ימין או למעלה-למטה) ללא חציית עקום אפס.

לכן רק מקרה ג אפשרי.

איזה מהמשפטים הבאים בנוגע לרשתות **בכּוּן**?

- הסרת ה-hub בעל ה-degree הגבוהה ביותר יפרק את הרשת לכמות הכי גדולה של תתי רשתות.
- תתכן רשת שבה יש צומת אחד שהוא גם בעל ה-degree הגבוהה ביותר, גם בעל ה-centrality הגבוה ביותר.
- ניתן לדעת כמה קליקות יש ברשת מתוך היסטוגרמה של ה-degrees שיש לכל צומת.
- אם ברשת לא מכוונת יש מעגל – הוספת כיווניות לקשתות בהכרח לא תסיר את המעגל מהרשת.

פתרון: ב. למשל – רשת בצורת כוכב. צומת אחת שמחוברת לכל שאר הצמתים. כל שאר הצמתים מחוברות רק לצומת המרכזי.

לפניכם טבלה עם ציונים של 5 סטודנטים, ב-4 קורסים. מצאו את הסטודנט/ית **הרחוקה ביותר** מהסטודנט חיים, כשדמיון מוגדר לפי **קורלציית ספירמן**.

פיזיולוגיה של התא	פיזיולוגיה מערכתית	אלגברה ליניארית	אותות א	
87	95	76	88	חיים
90	68	88	89	אופירה
76	87	88	95	אחמד
85	72	87	80	חן
61	80	56	62	יעקב

- אופירה
- אחמד
- חן
- יעקב
- אי אפשר לדעת

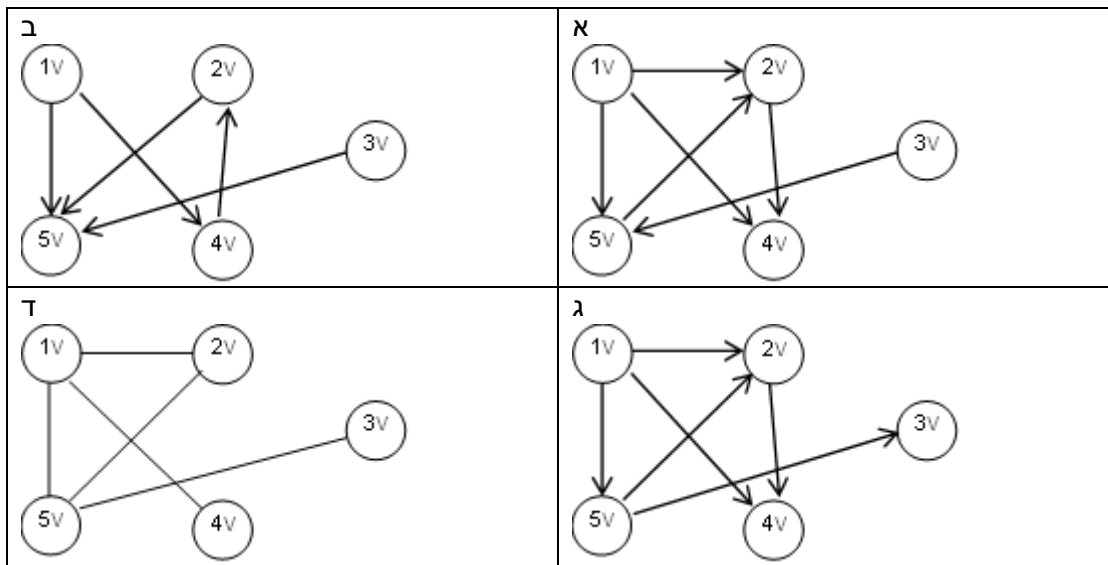
פתרון: קורלציית ספירמן הופכת את המספרים לדירוגים, ולכן הטבלה תהפוך ל

פיזיולוגיה של התא	פיזיולוגיה מערכתית	אלגברה ליניארית	אותות א	
3	1	4	2	חיים
1	4	3	2	אופירה
4	3	2	1	אחמד
2	4	1	3	חן
3	1	4	2	יעקב

כעת ניתן לראות כי הדירוגים של חן הפוכים לאלו של חיים, ולכן התשובה היא ג

6 איזה מהגרפים מתאימים למטריצת שכנות?

0 1 0 1 1
 0 0 0 1 0
 0 0 0 0 1
 0 0 0 0 0
 0 1 0 0 0



פתרון: ראשית, המטריצה לא סימטרית, ולכן הגרף מכון. ניתן לראות שהצומת הראשונה לא מקבלת חצים, אבל מוציאה שלושה – כלומר או א או ג.
 ההבדל בין א לבין ג הוא מה קורה לצומת 3. ניתן לראות שצומת 3, כמו צומת 1, לא מקבלת חצים. ולכן התשובה היא א

איזה מהמשפטים הבאים נכון לגבי תיקון Bonferroni?

- א. משנים את הקריטריון למובהקות לפי מספר ההשוואות
- ב. בתיקון בונפרוני, משתמשים ברנדומיזציה של התוצאות
- ג. התיקון מגדיל את מספר הטעויות מסוג 1.
- ד. התיקון משפיע על סטיית התקן של ההשוואות.

פתרון: א. זו ההגדרה.

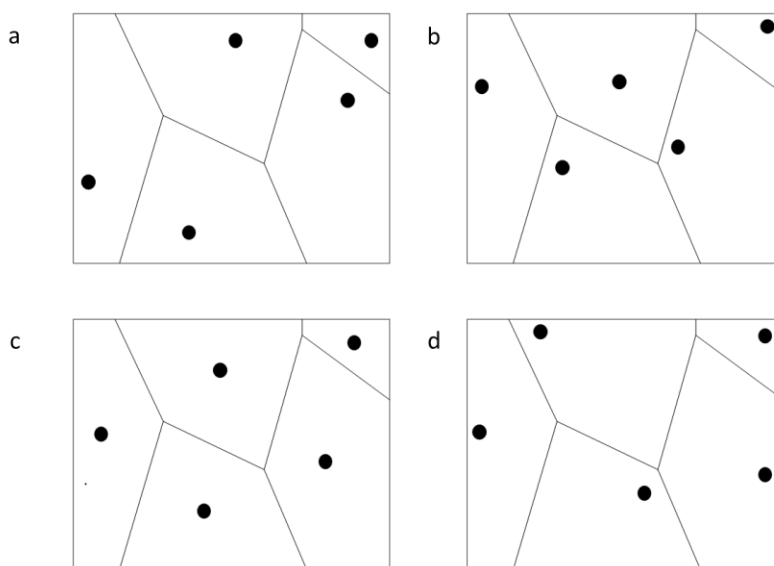
חוקר ביצע ניתוח גורמים ראשיים (PCA) על נתונים רב מימדים ומצא שני גורמים ראשיים המסבירים את רוב השונות של הנתונים. לאחר ביצוע האנליזה התברר לו שחלה טעות בקידוד הנתונים, וכדי לתקנה יש להכפיל את כלל הנתונים במספר 3.25. איך צפוי התיקון להשפיע על תוצאות האנליזה?

- א. התיקון לא ישפיע על תוצאות האנליזה, כי המשקל היחסי של כל מימד לא משתנה.
- ב. התיקון לא ישפיע על תוצאות האנליזה, כי בעקבותיו משתנה המשקל היחסי של כל מימד.
- ג. התיקון ישפיע על תוצאות האנליזה, כי המשקל היחסי של כל מימד לא משתנה.
- ד. התיקון ישפיע על תוצאות האנליזה, כי בעקבותיו משתנה המשקל היחסי של כל מימד.

פתרון: כל גורם מתאר יחס בין המימדים. אם מכפילים את כלל הנתונים באותו מספר, המשקל היחסי של כל מימד לא משתנה. ולכן גם תוצאות האנליזה לא משתנות.

לפניכם 4 תרשימים שמכילים 5 אשכולות . הקווים מגדירים את הגבולות בין האשכולות והנקודות מתארות את הצנטרואידים שלהם.

איזה מבין התרשימים יכול לתאר מיקום נכון של הצנטרואידים המתקבלים משיטת kmeans?



a. א.

b. ב.

c. ג.

d. ד.

ה. כל התרשימים אפשריים.

פתרון: הגבול ב KMEANS הוא באמצע בין שני צנטרואידים, כי השיוך נקבע לפי הצנטרואיד הקרוב ביותר.

זה מתקיים רק באיור A.

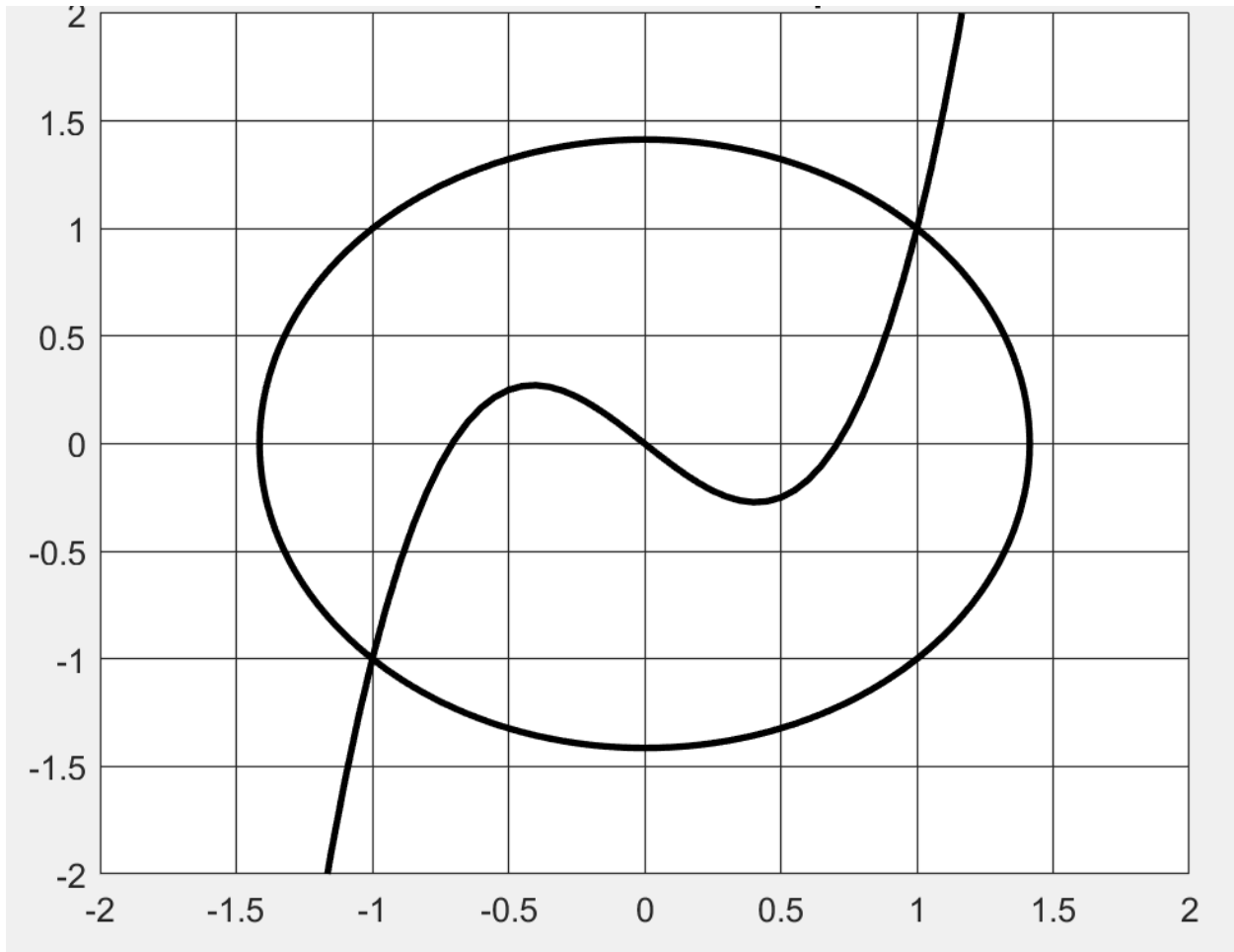
שאלה פתוחה - דינמיות

נתונה המערכת הדינמית הבאה:

$$\dot{x} = f(x, y)$$

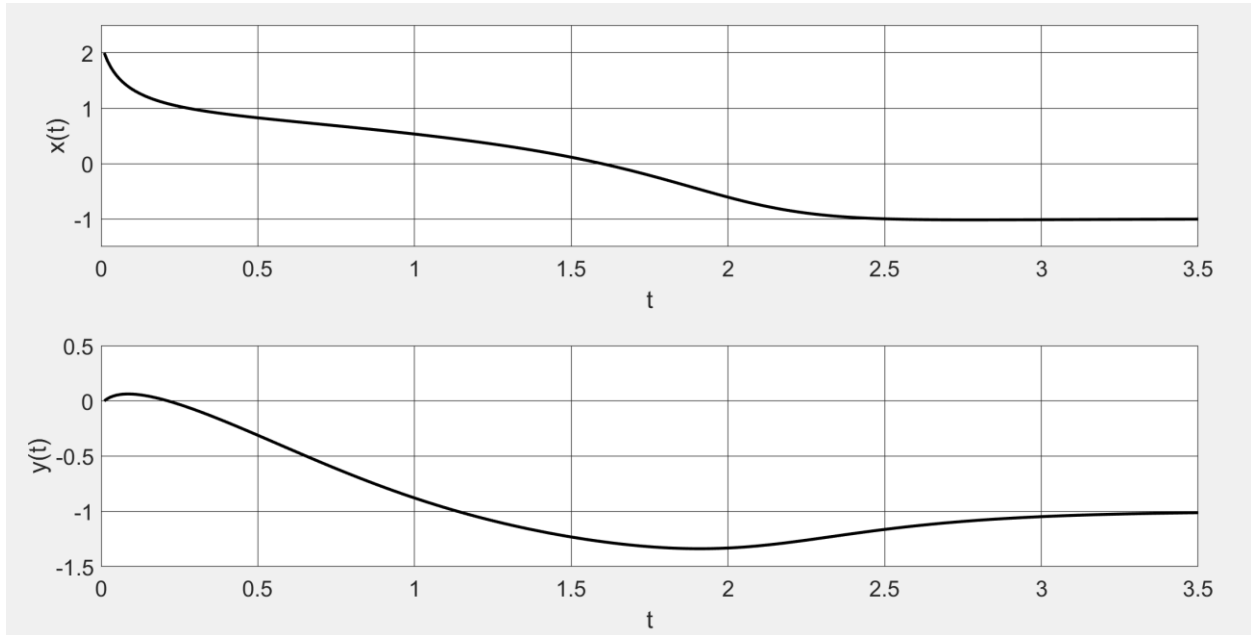
$$\dot{y} = g(x, y)$$

כמו כן, בציור שלפניכם מופיעים עקומי האפס של המערכת (בשחור)

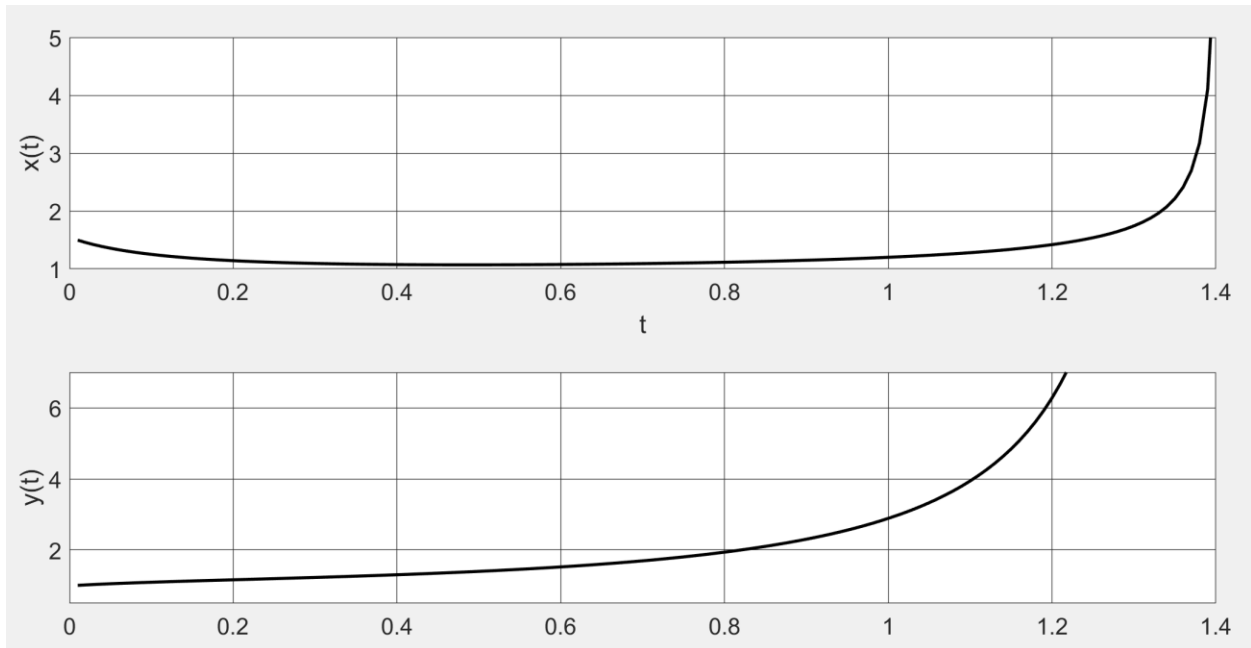


בנוסף, לפניכם גרפים $x(t)$, $y(t)$ עבור שני מסלולים שונים.

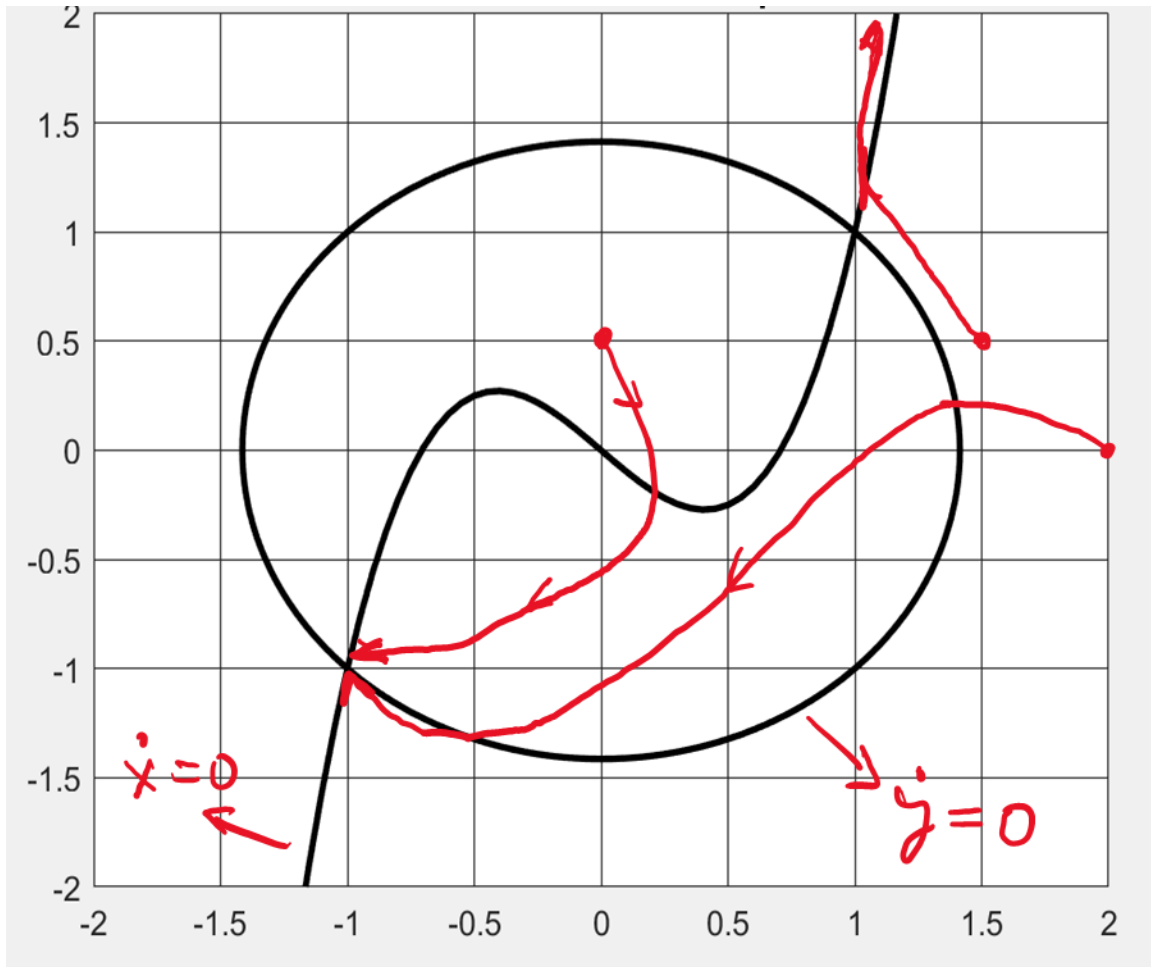
מסלול ראשון:



מסלול שני:



1. ציירו את שני המסלולים על גבי מרחב הפאזה.
2. ציינו איזה עקום אפס מתאים לאיזה משתנה
3. ציירו את המסלול המתחיל בנקודה $(0, 0.5)$



ציור המסלולים, לפי הנתונים. שמים לב מתי X או Y מגיעים למינימום/מקסימום, ואיפה המשתנה השני נמצא באותו זמן.

זיהוי עקומי אפס, בהתאם לחציית העקומים של המסלולים.

המסלול החדש, לפי כיווני הזרימה. היה אפשר גם להגיע רחוק יותר ימינה, אבל עדיין לחצות כלפי מטה את עקום האפס של X .

אתם מעוניינים לבצע קליסטור של דוגמאות באמצעות אלגוריתם k-means.

הנתונים שעומדים לרשותכם הם פרופיל ביטוי גנים במקבץ דוגמאות.

בכל סעיף, מתוארת תכונה של הנתונים (אין קשר בין הסעיפים). התכונה עלולה להיות בעייתית עבור האלגוריתם. הסבירו כיצד תשפיע התכונה על האלגוריתם, והציעו פתרון אפשרי שיאפשר לאלגוריתם להיות עמיד לבעייתיות זו.

א. אם מבצעים קליסטור של גנים לפי פרופיל הביטוי שלהם בדוגמאות (שימו לב, זו לא המטרה הסופית שלכם) ניתן בברור לראות כי הם מחולקים לתתי-קבוצות של גנים בעלות גודל מאוד שונה. הגנים המאפיינים כל קבוצה מראים קורלציה גבוהה ביותר בינם לבין עצמם.

השפעה: מספר הגנים הגדול מתת-קבוצה אחת יגרום להטיה חזקה של דמיון בין דוגמאות על סך גנים אלו תוך כדי "פיספוס" עושר הדוגמא.

פתרון: כיוון שגנים בכל תת-קבוצה בעלי קורלציה גבוהה, בחירה של "גנים ייצוגיים" מכל תת-קבוצה וקליסטור דוגמאות לפיהם יוכל לשמר על עושר הדוגמא תוך כדי אובדן מידע מינימלי.

ב. הגנים שונים זה מזה באופן משמעותי בטווח הביטוי

השפעה: הגנים בעלי טווח הביטוי הרחב ישיפעו באופן חזק מאוד על הקליסטור של הדוגמאות, למרות שטווח ביטוי לאו דווקא מצביע על חשיבות ביולוגית.

פתרון: נרמול של הגנים לטווח דומה (לדוגמא להתפלגות של ממוצע אפס, וסטיית תקן אחת) יתן משקל דומה לכל גן, אם כי, יש להזהר בהתייחסות לגנים ששונתם בין דוגמאות היא מינימלית ונובעת מרעש טכני.