מבוא למערכות לומדות (67577) | תרגיל 1

שם: אלון בינר | ת"ז: 318418118
April 2, 2023

שאלה 2

אלגברה לינארית

(1) <u>פיתרון:</u>

תהא A מטריצה אורתוגונלית.
יהא $x \in V$.
אזי:

$$||Ax|| = \sqrt{\langle Ax, Ax \rangle} \overset{(1)}{=} \sqrt{(Ax) \cdot (Ax)} = \sqrt{(Ax)^t (Ax)} = \sqrt{(x^t A^t Ax)} \overset{(2)}{=} \sqrt{(x^t Ix)} =$$

$$= \sqrt{(x^t x)} = \sqrt{\langle x, x \rangle} = ||x||$$

שוויון (1) נובע מכך שקבענו בקורס שמכפלה פנימית בה נשתמש היא המכפלה הסקלרית.
שוויון (2) נובע מאורתוגונליות A.

(2) <u>פיתרון:</u>

בהינתן מטריצה $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix}$ נחשב את ה-SVD של A באופן הבא:
נמצא מטריצות U, Σ ו-V^T, כאשר U ו-V הן מטריצות אורתוגונליות ו-Σ מטריצה אלכסונית.
ראשית, נחשב:

$$A^T A = \left(U \Sigma V^T\right)^T \left(U \Sigma V^T\right) = \left(\left(U \Sigma\right) V^T\right)^T \left(U \Sigma V^T\right) = \left(V \left(U \Sigma\right)^T\right) \left(U \Sigma V^T\right) =$$

$$= \left(V \Sigma^T U^T\right) \left(U \Sigma V^T\right) = V \Sigma^T U^T U \Sigma V^T = V \Sigma^T \Sigma V^T$$

כשהשוויון האחרון נובע מאורתוגונליות U.
בנוסף:

$$A^T A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix}$$

כעת נמצא את הערכים העצמיים של A. לשם כך, נחשב את הפולינום האופייני של A, נשווה אותו לאפס ונקבל שהשורשים של הפולינום הם הערכים העצמיים של A:

$$det\left(A^T A - \lambda I\right) = det\left(\begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}\right) = det\left(\begin{bmatrix} 2-\lambda & 0 & 2 \\ 0 & 2-\lambda & -2 \\ 2 & -2 & 4-\lambda \end{bmatrix}\right) =$$

$$= -\lambda^3 + 8\lambda^2 - 12\lambda = -\lambda\left(\lambda^2 - 8\lambda + 12\right) = -\lambda\left(\lambda^2 - 2\lambda - 6\lambda + 12\right) =$$

$$= -\lambda\left(\lambda\left(\lambda-2\right)-6\left(\lambda-2\right)\right) = -\lambda\left(\lambda-6\right)\left(\lambda-2\right)$$

$$\Downarrow$$

$$det\left(A-\lambda I\right)=0 \iff -\lambda\left(\lambda-6\right)\left(\lambda-2\right)=0$$

$$\Downarrow$$

$$\lambda \in \{0,2,6\}$$

כעת נחשב את הוקטורים העצמיים של A, לשם כך נמצא $v \neq 0$ המקיימים את המשוואה: $\left(AA^T-\lambda I\right)v=0$ לכל $\lambda \in \{0,2,6\}$:

$$AA^T-\lambda I = \begin{bmatrix} 2-\lambda & 0 & 2 \\ 0 & 2-\lambda & -2 \\ 2 & -2 & 4-\lambda \end{bmatrix}$$

לכן, עבור $\lambda=0$ מתקיים:

$$AA^T-\lambda I = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix}$$

ניקח ו"ע $v_1 = \begin{bmatrix} \frac{2}{\sqrt{2^2+0^2+2^2}} \\ \frac{0}{\sqrt{2^2+0^2+2^2}} \\ \frac{2}{\sqrt{2^2+0^2+2^2}} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{4+0+4}} \\ 0 \\ \frac{2}{\sqrt{4+0+4}} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{8}} \\ 0 \\ \frac{2}{\sqrt{8}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

עבור $\lambda=2$ מתקיים:

$$AA^T-\lambda I = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & -2 \\ 2 & -2 & 2 \end{bmatrix}$$

ניקח ו"ע $v_2 = \begin{bmatrix} 0 \\ 0 \\ \frac{2}{\sqrt{0^2+0^2+2^2}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \frac{2}{\sqrt{0+0+4}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

עבור $\lambda=6$ מתקיים:

$$AA^T-\lambda I = \begin{bmatrix} -4 & 0 & 2 \\ 0 & -4 & -2 \\ 2 & -2 & -2 \end{bmatrix}$$

ניקח ו"ע $v_3 = \begin{bmatrix} \frac{-4}{\sqrt{(-4)^2+0^2+2^2}} \\ 0 \\ \frac{2}{\sqrt{(-4)^2+0^2+2^2}} \end{bmatrix} = \begin{bmatrix} \frac{-4}{\sqrt{20}} \\ 0 \\ \frac{2}{\sqrt{20}} \end{bmatrix} = \begin{bmatrix} \frac{-2}{\sqrt{5}} \\ 0 \\ \frac{1}{\sqrt{5}} \end{bmatrix}$

מכך נסיק כי:

$$V = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{-2}{\sqrt{5}} \\ 0 & 0 & 0 \\ \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{5}} \end{bmatrix}, \Sigma = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix}$$

כעת, נחשב את U.

נוכל לחשב את עמודות U לפי מה שראינו בתרגול:

$$t_1 = \frac{1}{\sqrt{6}} \cdot Av_3 = \begin{bmatrix} -\sqrt{\frac{2}{15}} \\ 0 \end{bmatrix} \Rightarrow u_1 = \frac{t_1}{\parallel t_1 \parallel} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$t_2 = \frac{1}{\sqrt{2}} \cdot Av_2 = \begin{bmatrix} 0 \\ \sqrt{\frac{2}{3}} \end{bmatrix} \Rightarrow u_2 = \frac{t_2}{\parallel t_2 \parallel} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

לכן :

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

(3) <u>פיתרון:</u>

יהיו $v \in \mathbb{R}^n, u \in \mathbb{R}^m$. נגדיר $A = v \otimes u$.
על-פי הגדרה $A = vu^T$ ולכן $A \in \mathbb{R}^{n\times m}$.
אם נסמן $v = \begin{bmatrix} v_1 \\ ... \\ v_n \end{bmatrix}, u = \begin{bmatrix} u_1 \\ ... \\ u_m \end{bmatrix}$
אז מתקיים :

$$vu^T = \begin{bmatrix} v_1 \\ ... \\ v_n \end{bmatrix} \begin{bmatrix} u_1 & ... & u_m \end{bmatrix} = \begin{bmatrix} v_1u_1 & v_1u_2 & ... & v_1u_m \\ v_2u_1 & v_2u_2 & ... & v_2u_m \\ ... & ... & ... & ... \\ v_nu_1 & v_nu_2 & ... & v_nu_m \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ ... \\ v_n \end{bmatrix} u_1 & \begin{bmatrix} v_1 \\ v_2 \\ ... \\ v_n \end{bmatrix} u_2 & ... & \begin{bmatrix} v_1 \\ v_2 \\ ... \\ v_n \end{bmatrix} u_m \end{bmatrix}$$

לכל $t, k \in [m]$ שונים זה מזה מתקיים :
אם $u_t = 0$ אז העמודה ה-t של A שווה ל-0 ולכן תלויה לינארית בשאר העמודות של המטריצה.
אחרת, קיים סקלר $\frac{u_k}{u_t} \neq 0$ כך ש-

$$\begin{bmatrix} v_1 \\ ... \\ v_n \end{bmatrix} u_t \cdot \frac{u_k}{u_t} = \begin{bmatrix} v_1 \\ ... \\ v_n \end{bmatrix} u_k$$

לכן, כל וקטור בעמודה ה-t תלוי לינארית בוקטור העמודה ה-k.
לכן $rank\left(A\right) = rank\left(vu^T\right) = 1$.

(4) <u>פיתרון:</u>

יהא $(u_i)_{i=1}^n$ בסיס אורתונורמלי כלשהו ויהא $x \in \mathbb{R}^n$ שמתקבל כצירוף לינארי של איברי הבסיס.
יהא $k \in [1, n]$.
צריך להראות שהמקדם של u_k שווה ל-$\langle x, u_k\rangle$.
נסמן את המקדם של u_k ב-a_k.
אזי :

$$\langle x, u_k\rangle = \left\langle \sum_{j=1}^n a_j \cdot u_j, u_k \right\rangle = \sum_{j=1}^n \langle a_j \cdot u_j, u_k\rangle = \sum_{j=1}^n a_j \langle u_j, u_k\rangle \overset{(1)}{=} a_k$$

כאשר שוויון (1) נובע מאורתונורמליות הבסיס $(u_i)_{i=1}^n$.

אינפי בכמה משתנים

(5) <u>פיתרון:</u>

יהא $x \in \mathbb{R}^n$ קבוע ותהא $U \in \mathbb{R}^{n\times n}$ מטריצה אורתוגונלית קבועה.
צריך לחשב את היעקוביאן של $f : \mathbb{R}^n \to \mathbb{R}^n$ המוגדרת על-ידי

$$f\left(\sigma\right) = U \cdot diag\left(\sigma\right) U^T x$$

לכל $\sigma \in \mathbb{R}^n$.

$$J\left(f\left(\sigma\right)\right)=\begin{bmatrix} \frac{\partial f_1(\sigma)}{\partial \sigma_1} & ... & \frac{\partial f_1(\sigma)}{\partial \sigma_n} \\ ... & ... & ... \\ \frac{\partial f_n(\sigma)}{\partial \sigma_1} & ... & \frac{\partial f_n(\sigma)}{\partial \sigma_n} \end{bmatrix}$$

יהיו $i,j \in [n]$, אזי:

$$J\left(f\left(\sigma\right)\right)_{i,j}=\frac{\partial f_i\left(\sigma\right)}{\partial \sigma_j}=$$

$$f_i\left(\sigma\right)=\left(U\cdot diag\left(\sigma\right)U^Tx\right)_i=$$

$$Udiag\left(\sigma\right)U^Tx=\begin{bmatrix} u_1^1 & ... & u_n^1 \\ ... & ... & ... \\ u_1^n & ... & u_n^n \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & ... & \\ & & \sigma_n \end{bmatrix}\begin{bmatrix} u_1^T \\ ... \\ u_n^T \end{bmatrix}x=\begin{bmatrix} u_1^1\sigma_1 & u_2^1\sigma_2 & ... & u_n^1\sigma_n \\ u_1^2\sigma_1 & ... & ... & u_n^2\sigma_n \\ ... & ... & ... & ... \\ u_1^n\sigma_1 & ... & ... & u_n^n\sigma_n \end{bmatrix}\begin{bmatrix} u_1^T \\ ... \\ u_n^T \end{bmatrix}x=$$

$$=\begin{bmatrix} u_1\sigma_1 & ... & u_n\sigma_n \end{bmatrix}\begin{bmatrix} u_1^T \\ ... \\ u_n^T \end{bmatrix}x=\left(\sum_{i=1}^n u_i\sigma_i u_i^T\right)x=\left(\sum_{i=1}^n \sigma_i u_i u_i^T\right)x=\left(\sum_{i=1}^n \sigma_i\left(u_iu_i^T\right)\right)x=$$

$$=\left(\sigma_1\left(u_1u_1^T\right)+\sigma_2\left(u_2u_2^T\right)+...+\sigma_n\left(u_nu_n^T\right)\right)x=$$

$$=\left(\sigma_1\left(u_1u_1^T\right)\right)x+\left(\sigma_2\left(u_2u_2^T\right)\right)x+...+\left(\sigma_n\left(u_nu_n^T\right)\right)x=$$

$$=\sum_{i=1}^n\left(\sigma_i\left(u_iu_i^T\right)x\right)$$

$$\Downarrow$$

$$\forall k\in[n]\quad f_k\left(\sigma\right)=\left(\sum_{i=1}^n\left(\sigma_i\left(u_iu_i^T\right)x\right)\right)_k=\left(\sum_{i=1}^n\left(\sigma_i\left(u_iu_i^T\right)x\right)_k\right)=$$

$$=\left(\sum_{i=1}^n\sigma_i\left(u_iu_i^Tx\right)_k\right)$$

לכן לכל $k,l\in[n]$ מתקיים:

$$J\left(f\left(\sigma\right)\right)_{k,l}=\frac{\partial f_k\left(\sigma\right)}{\partial \sigma_l}=\left(u_lu_l^Tx\right)_k$$

כלומר,

$$J\left(f\left(\sigma\right)\right)=\begin{bmatrix} \left(u_1u_1^Tx\right)_1 & ... & \left(u_nu_n^Tx\right)_1 \\ ... & ... & ... \\ \left(u_1u_1^Tx\right)_n & ... & \left(u_nu_n^Tx\right)_n \end{bmatrix}=\begin{bmatrix} u_1u_1^Tx & ... & u_nu_n^Tx \end{bmatrix}$$

(6) <u>פיתרון:</u>

$$\nabla\left(h\left(\sigma\right)\right)=\nabla\left(\frac{1}{2}||f\left(\sigma\right)-y||^{2}\right)$$

$$h\left(\sigma\right)=\frac{1}{2}||f\left(\sigma\right)-y||^{2}=\frac{1}{2}||\left[\begin{array}{c}\left(f\left(\sigma\right)-y\right)_{1}\\ ...\\ \left(f\left(\sigma\right)-y\right)_{n}\end{array}\right]||^{2}=\frac{1}{2}\sum_{i=1}^{n}\left(f\left(\sigma\right)-y\right)_{i}^{2}=$$

מתקיים כי :

$$||f\left(\sigma\right)-y||^{2}=\left\langle f\left(\sigma\right)-y,f\left(\sigma\right)-y\right\rangle =\left\langle f\left(\sigma\right)-y,f\left(\sigma\right)\right\rangle -\left\langle f\left(\sigma\right)-y,y\right\rangle =$$

$$=\left\langle f\left(\sigma\right),f\left(\sigma\right)\right\rangle -\left\langle y,f\left(\sigma\right)\right\rangle -\left(\left\langle f\left(\sigma\right),y\right\rangle -\left\langle y,y\right\rangle \right)=$$

$$=\left\langle f\left(\sigma\right),f\left(\sigma\right)\right\rangle -\left\langle y,f\left(\sigma\right)\right\rangle -\left\langle f\left(\sigma\right),y\right\rangle +\left\langle y,y\right\rangle =$$

$$=||f\left(\sigma\right)||^{2}-\left\langle y,f\left(\sigma\right)\right\rangle -\left\langle y,f\left(\sigma\right)\right\rangle +||y||^{2}=$$

$$=||f\left(\sigma\right)||^{2}-2\left\langle y,f\left(\sigma\right)\right\rangle +||y||^{2}$$

נגדיר פונקציה :

$$g\left(\sigma\right)=\frac{1}{2}\left(||\sigma||^{2}-2\left\langle y,\sigma\right\rangle +||y||^{2}\right)$$

לכן :

$$||f\left(\sigma\right)||^{2}-2\left\langle y,f\left(\sigma\right)\right\rangle +||y||^{2}=\left(g\left(f\left(\sigma\right)\right)\right)=\left(g\circ f\right)\left(\sigma\right)=h\left(\sigma\right)$$

אזי לפי משפט שראינו בתרגול מתקיים כי :

$$J_{x}\left(h\right)=J_{x}\left(g\circ f\right)=J_{f(x)}\left(g\right)J_{x}\left(f\right)$$

מתקיים ש :

$$J_{x}\left(g\right)=\frac{\partial}{\partial x}\left(\frac{1}{2}\left(||x||^{2}-2\left\langle y,x\right\rangle +||y||\right)\right)=\frac{\partial}{\partial x}\left(\frac{1}{2}\left(\left\langle x,x\right\rangle -2\left\langle y,x\right\rangle +\left\langle y,y\right\rangle \right)\right)=$$

$$=\frac{\partial}{\partial x}\left(\frac{1}{2}\left(x^{T}x-2y^{T}x+y^{T}y\right)\right)=\frac{1}{2}\left(2x^{T}-2y^{T}\right)=\left(x^{T}-y^{T}\right)=\left(x-y\right)^{T}$$

ולכן :

$$J_{f(\sigma)}\left(g\right)=\left(f\left(\sigma\right)-y\right)^{T}$$

לכן

$$J_{\sigma}\left(h\right)=\left(f\left(\sigma\right)-y\right)^{T}J_{\sigma}\left(f\right)$$

(7) <u>פיתרון:</u>

עבור פונקציית ה-Softmax, אותה נסמן ב-S, המוגדרת על-ידי:

$$\forall j \in [k] \quad S(x)_j = \frac{e^{x_j}}{\sum_{k=1}^{d} e^{x_k}}$$

לכל $i \in [k], j \in [d]$ מתקיים כי:

$$(J_x(S))_{i,j} = \frac{\partial S_i(x)}{\partial x_j}$$

אם $i = j$ אז לפי מה שראינו בתרגול:

$$\frac{\partial S_i(x)}{\partial x_i} = S(x)_i \left(1 - S(x)_i\right)$$

אחרת, $i \neq j$.
נסמן ב-f_i וב-g את הפונקציות $f_i(x) = e^{x_i}$ ו-$g(x) = \sum_{k=1}^{d} e^{x_k}$ בהתאמה.
אזי $\frac{\partial f_i(x)}{\partial x_j} = 0, \frac{\partial g(x)}{\partial x_j} = e^{x_j}$ ולכן

$$\frac{\partial S_i(x)}{\partial x_j} = \frac{\partial \frac{f_i}{g}(x)}{\partial x_j} = \frac{\partial \frac{f_i(x)}{g(x)}}{\partial x_j} = \frac{\frac{\partial f_i(x)}{\partial x_j} \cdot g(x) - f_i(x) \cdot \frac{\partial g(x)}{\partial x_j}}{(g(x))^2} =$$
$$= \frac{0 \cdot g(x) - e^{x_i} \cdot e^{x_j}}{\left(\sum_{k=1}^{d} e^{x_k}\right)^2} = \frac{-e^{x_i}}{\sum_{k=1}^{d} e^{x_k}} \cdot \frac{e^{x_j}}{\sum_{k=1}^{d} e^{x_k}} = -(S(x))_i \cdot (S(x))_j$$

לכן:

$$J_x(S) = \begin{bmatrix} S(x)_1 \cdot (1 - S(x)_1) & -S(x)_1 \cdot S(x)_2 & \ldots & -S(x)_1 \cdot S(x)_n \\ \ldots & \ldots & \ldots & \ldots \\ -S(x)_n \cdot S(x)_1 & \ldots & -S(x)_n \cdot S(x)_{n-1} & S(x)_n \cdot (1 - S(x)_n) \end{bmatrix}$$

(8) <u>פיתרון:</u>

תהא $f : \mathbb{R}^2 \to \mathbb{R}$ המוגדרת ע"י $f(x,y) = x^3 - 5xy - y^5$.
אזי:

$$\frac{\partial f}{\partial x \partial x} = \frac{\partial}{\partial x}\left(3x^2 - 5y\right) = 6x$$

$$\frac{\partial f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(-5x - 5y^4\right) = -5$$

$$\frac{\partial f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(3x^2 - 5y\right) = -5$$

$$\frac{\partial f}{\partial y \partial y} = \frac{\partial}{\partial y}\left(-5x - 5y^4\right) = -20y^3$$

נסיק אפוא ש:

$$H_{(x,y)} f = \begin{bmatrix} 6x & -5 \\ -5 & -20y^3 \end{bmatrix}$$

קמירות

(9) פיתרון:

צריך להראות ש-$C := \bigcap_{i\in I} C_i$ היא קבוצה קמורה כאשר $\{C_i : i \in I\}$ הוא אוסף קבוצות קמורות.
נוכיח את הטענה באינדוקציה על $|I|$.
בסיס:
עבור $|I| = 0$, מתקיים ש-$I = \emptyset$ לכן $\bigcap_{i\in I} C_i = \emptyset$. אזי $C = \emptyset$ וההגדרה לקבוצה קמורה מתקיימת עבורה באופן ריק.
עבור $|I| = 1$, מתקיים ש-$\bigcap_{i\in I} C_i = C_1$, לכן $C = C_1$. היות ש-$C_1 \in \{C_i : i \in I\}$ אז היא קבוצה קמורה לכן C היא קבוצה קמורה.

צעד: נניח שהטענה מתקיימת עבור אוסף קבוצות קמורות בגודל $|I| - 1$.
כלומר עבור $C' := \bigcap_{i\in I'} C_i$ כאשר $\{C_i : i \in I'\}$ הוא אוסף קבוצות קמורות בגודל $|I| - 1$.
נוכיח ש-$C = \bigcap_{i\in I} C_i$ היא גם קבוצה קמורה, כאשר $I = I' \cup \{i_{|i|}\}$. הוכחה זו מספיקה עבור קבוצה I מסוג זה משום שמהנחת האינדוקציה נובע שלכל קבוצה I קיימת קבוצה קמורה $C \backslash C_{i_{|I|}}$ עבורה $I' = I \backslash \{i_{|I|}\}$ היא קבוצה קמורה.
נסמן $I' = \{i_1, i_2, ..., i_{|I|-1}\}$.
נשים לב כי נוכל לכתוב את C בתור $C' \cap C_{i_I}$. מהנחת האינדוקציה נובע ש-C' היא קבוצה קמורה. בנוסף, נתון לנו ש-C_{i_I} היא קבוצה קמורה. נפעיל שוב את הנחת האינדוקציה ונסיק מכך ש-$C' \cap C_{i_I}$ היא גם קבוצה קמורה.

(10) פיתרון:

צריך להראות שסכום הוקטורי $C_1 + C_2 := \{c_1 + c_2 : c_1 \in C_1, c_2 \in C_2\}$, כאשר C_1 ו-C_2 הן קבוצות קמורות, הוא גם קבוצה קמורה.
יהיו $u, v \in C_1 + C_2$. אזי קיימים u_1, u_2 עבורם $u = u_1 + u_2$ וגם $u_1 \in C_1, u_2 \in C_2$.
בנוסף, ישנם v_1, v_2 עבורם $v = v_1 + v_2$ וגם $v_1 \in C_1, v_2 \in C_2$.

יהא $\alpha \in [0, 1]$.
צריך להראות ש-$(1-\alpha) u + \alpha v \in C_1 + C_2$.
מתקיים כי:

$$(1-\alpha) u + \alpha v = (1-\alpha)(u_1 + u_2) + \alpha (v_1 + v_2) =$$

$$= (1-\alpha) u_1 + (1-\alpha) u_2 + \alpha v_1 + \alpha v_2 =$$

$$= (1-\alpha) u_1 + \alpha v_1 + (1-\alpha) u_2 + \alpha v_2$$

היות ש-C_1 קבוצה קמורה ו-$u_1, v_1 \in C_1$ אזי $(1-\alpha) u_1 + \alpha v_1 \in C_1$.
היות ש-C_2 קבוצה קמורה ו-$u_2, v_2 \in C_2$ אזי $(1-\alpha) u_2 + \alpha v_2 \in C_2$.

יהיו C_1, C_2 קבוצות קמורות ויהיו $u_1, v_1 \in C_1$ ו-$u_2, v_2 \in C_2$.
אזי לכל $\alpha \in [0, 1]$ מתקיים ש-$(1-\alpha) u_1 + \alpha v_1 \in C_1$ וגם $(1-\alpha) u_2 + \alpha v_2 \in C_2$.
לכן $(1-\alpha) u + \alpha v \in C_1 + C_2$ לכן $C_1 + C_2$ היא קבוצה קמורה.

(11) פיתרון:

צריך להראות שלכל סקלר λ ולכל קבוצה קמורה C, מתקיים ש-$\lambda C := \{\lambda c : c \in C\}$.
יהא λ סקלר ותהא C קבוצה קמורה.
יהיו $u, v \in \lambda C$ ויהא $\alpha \in [0, 1]$.
אזי $(1-\alpha) u + \alpha v \in C$.
נחשב:

$$((1-\alpha) u + \alpha v) = \left((1-\alpha) \lambda \frac{u}{\lambda} + \alpha \lambda \frac{v}{\lambda}\right) = \lambda \left((1-\alpha) \frac{u}{\lambda} + \alpha \frac{v}{\lambda}\right)$$

היות ש-$u, v \in \lambda C$ אזי $\frac{u}{\lambda}, \frac{v}{\lambda} \in C$ לכן $(1-\alpha) \frac{u}{\lambda} + \alpha \frac{v}{\lambda} \in C$.
לכן $\lambda \left((1-\alpha) \frac{u}{\lambda} + \alpha \frac{v}{\lambda}\right) \in \lambda C$.

תאוריית ההערכה

(12) <u>פיתרון:</u>

יהיו $\{x_i\}_{i=1}^{\mathbb{N}}$ דגימות של משתנים מקריים שווי התפלגות בלתי-תלויים שמתפלגים מעל פונקציית התפלגות $\mathcal{P}$. $\mathcal{P}$ היא פונקציית התפלגות בעלת תוחלת סופית ושונות סופית.
יהא $n \in \mathbb{N}$ קבוע.
צריך להראות שמשערך ממוצע הדגימות, $\hat{\mu}_n$, המוגדר על-ידי $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, הוא קונסיסטנטי.
יהא $\varepsilon > 0$.
צריך להראות ש-$lim_{n\to\infty}\left(\mathcal{P}\left(|\hat{\mu}_n - \mu| > \varepsilon\right)\right) = 0$.
מתקיים כי:

$$\mathcal{P}\left(|\hat{\mu}_n - \mu| > \varepsilon\right) \overset{(1)}{\leq} \frac{Var\left(\hat{\mu}_n\right)}{\varepsilon^2} \overset{(2)}{=} \frac{\frac{\sigma^2}{n}}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 n}$$

כאי-שוויון (1) הוא אי-שוויון צ'בישב ושוויון (2) נובע מהחישוב של משערך ממוצע הדגימות עבור n דגימות, $\hat{\mu}_n$, שאליו הגענו בהרצאה.

לכן:

$$lim_{n\to\infty}\left(\mathcal{P}\left(|\hat{\mu}_n - \mu| > \varepsilon\right)\right) = lim_{n\to\infty}\left(\frac{\sigma^2}{\varepsilon^2 n}\right) = 0$$

כשהשוויון האחרון נובע מאריתמטיקה של גבולות.

(13) <u>פיתרון:</u>

יהיו $x_1, ..., x_m \overset{iid}{\sim} \mathcal{N}\left(\mu, \Sigma\right)$ m תצפיות שנדגמות באופן בלתי תלוי עם התפלגות גאוסיאנית רבת משתנים עם תוחלת של $\mu \in \mathbb{R}^d$ ומטריצת שונויות משותפות $\Sigma \in \mathbb{R}^{d\times d}$. נספק ביטוי עבור פונקציית ה-$log - likelihood$ של $\mathcal{N}\left(\mu, \Sigma\right)$.
מתקיים כי:

$$\mathcal{L}\left(\mu, \Sigma | x_1, ..., x_m\right) = f_{\mu,\Sigma}\left(x_1, ..., x_m\right) = \prod_{i=1}^{m} f_{\mu,\Sigma}\left(x_i\right) =$$

$$= \prod_{i=1}^{m}\left(\frac{1}{\sqrt{\left(2\pi\right)^d |\Sigma|}} exp\left(-\frac{1}{2}\cdot\left(x_i - \mu\right)^T \Sigma^{-1}\left(x_i - \mu\right)\right)\right) =$$

$$= \left(\frac{1}{\sqrt{\left(2\pi\right)^d |\Sigma|}}\right)^m exp\left(\sum_{i=1}^{m}\left(-\frac{1}{2}\cdot\left(x_i - \mu\right)^T \Sigma^{-1}\cdot\left(x_i - \mu\right)\right)\right) =$$

$$= \left(\left(2\pi\right)^d |\Sigma|\right)^{-\frac{m}{2}} exp\left(-\frac{1}{2}\sum_{i=1}^{m}\left(\left(x_i - \mu\right)^T \Sigma^{-1}\left(x_i - \mu\right)\right)\right)$$

אזי:

$$log\left(\mathcal{L}\left(\mu, \Sigma | x_1, ..., x_m\right)\right) = log\left(\left(\left(2\pi\right)^d |\Sigma|\right)^{-\frac{m}{2}} exp\left(-\frac{1}{2}\sum_{i=1}^{m}\left(\left(x_i - \mu\right)^T \Sigma^{-1}\left(x_i - \mu\right)\right)\right)\right) =$$

$$= log\left(\left(\left(2\pi\right)^d |\Sigma|\right)^{-\frac{m}{2}}\right) + log\left(exp\left(-\frac{1}{2}\sum_{i=1}^{m}\left(\left(x_i - \mu\right)^T \Sigma^{-1}\left(x_i - \mu\right)\right)\right)\right) =$$

$$= -\frac{m}{2}\left(log\left(\left(2\pi\right)^d |\Sigma|\right)\right) - \frac{1}{2}\sum_{i=1}^{m}\left(\left(x_i - \mu\right)^T \Sigma^{-1}\left(x_i - \mu\right)\right) =$$

$$= -\frac{m}{2}\left(log\left((2\pi)^d\right) + log\left(|\Sigma|\right)\right) - \frac{1}{2}\sum_{i=1}^{m}\left((x_i-\mu)^T\,\Sigma^{-1}\,(x_i-\mu)\right) =$$

$$= -\frac{m}{2}log\left((2\pi)^d\right) - \frac{m}{2}log\left(|\Sigma|\right) - \frac{1}{2}\sum_{i=1}^{m}\left((x_i-\mu)^T\,\Sigma^{-1}\,(x_i-\mu)\right) =$$

$$= -\frac{md}{2}log\left(2\pi\right) - \frac{m}{2}log\left(|\Sigma|\right) - \frac{1}{2}\sum_{i=1}^{m}\left((x_i-\mu)^T\,\Sigma^{-1}\,(x_i-\mu)\right)$$

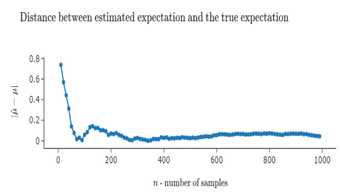
<u>חלק פרקטי</u>:

<u>שאלה 1</u>:

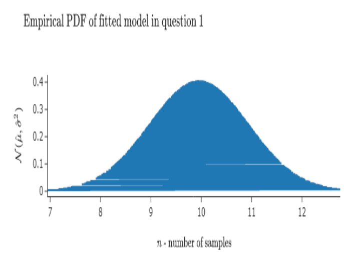
פיתרון:

```
(9.954743292509804, 0.9752096659781323)
```

<u>שאלה 2</u>: גרף:



<u>שאלה 3</u>: גרף:



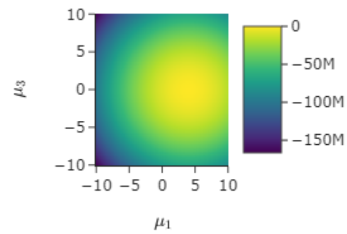
היינו מצפים לראות גרף פעמון שבו פונקציית ה pdf מקבלת ערכים גבוהים יותר ככל שמספר הדגימות קרוב יותר לתוחלת (שהיא 10)

<u>שאלה 4</u>: פיתרון:

```
[-0.02282878 -0.04313959  3.9932571  -0.02038981]
[[ 0.91578223  0.16622123 -0.03423861  0.46244021]
 [ 0.16620092  1.97225175 -0.00986527  0.04555112]
 [-0.03022252 -0.00582887  0.97462985 -0.0203261 ]
 [ 0.46244265  0.04557387 -0.02433975  0.97158515]]
```

<u>שאלה 5</u>: גרף:

Log-likelihood of expectation dependi



<u>שאלה 6</u>: פיתרון:

```
Maximum Likelihood:  [-0.05  3.97]
```