תרגיל 1
Thursday, March 10, 2022 17:20
שם: הדר חבר
ת.ז: 206996761

תרגיל 1 IML - שאלות תאורטיות

שאלה 1: תהי $T: V \to W$ טרנספורמציה לינארית ו- A מטריצה מייצגת.
אם A היא מטריצה אורתוגונלית, אז לכל $x \in V$ $\|x\| = \|Ax\|$.

<u>הוכחה:</u>

מכיוון ש- A מטריצה אורתוגונלית, אז ע"פ ההגדרה מתקיים לכל $u,v \in V$: $\langle u,v \rangle = \langle Au, Av \rangle$
כאשר $\langle \cdot,\cdot \rangle$ היא המכפלה הפנימית המתאימה לנורמה הנתונה.

נרצה להוכיח $\|v\| = \|Av\|$ לכל $v \in V$.

לפי ההגדרה של נורמה שמושרית ע"י מכפלה פנימית, מתקיים: $\|v\| = \sqrt{\langle v,v \rangle}$, $\|Av\| = \sqrt{\langle Av, Av \rangle}$

ולכן לפי הגדרת מה"פ אורתוגונלית נקבל שמתקיים:

$$\|v\| = \sqrt{v \cdot v} = \sqrt{Av \cdot Av} = \|Av\|.$$

שאלה 2:

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

רוצים למצוא מטריצות אורתוגונליות U, V ומטריצה אלכסונית Σ כך
ש $A = U \Sigma V^t$. ראינו בתרגול שמתקיים

$$A^T A = (U\Sigma V^t)^t (U\Sigma V^t) = V \Sigma^t \underbrace{U^t U}_{I} \Sigma V^t = V \Sigma^2 V^t$$

לכן אם נחשב את $A^T A$ ונחשב את פירוק ה-EVD שלה (פירוק לפי ע"ע) אז
נדע את Σ^2 ואת V.

$$A^T A = \overset{3\times 2}{\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{bmatrix}} \overset{2\times 3}{\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix}} = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix}$$

נמצא ערכים עצמיים ע"י פתרון המשוואה
$$\det(A^t A - \lambda I) = 0$$

כי אם λ ע"ע של B אז קיים $v \neq 0$
כך ש: $Bv = \lambda v$
$$\Downarrow$$
$$(B - \lambda I)v = 0$$
וזה אומר ש $\det(B - \lambda I) = 0$
כי היא לא הפיכה!
קיים פתרון לא
טריוויאלי למשוואה.

$$\det(A^t A - \lambda I) = \begin{vmatrix} 2-\lambda & 0 & 2 \\ 0 & 2-\lambda & -2 \\ 2 & -2 & 4-\lambda \end{vmatrix} =$$

$$= (2-\lambda)\begin{vmatrix} 2-\lambda & -2 \\ -2 & 4-\lambda \end{vmatrix} + 2\begin{vmatrix} 0 & 2-\lambda \\ 2 & -2 \end{vmatrix} = (2-\lambda)\Big((2-\lambda)(4-\lambda)-4\Big) + 2\Big(-2(2-\lambda)\Big)$$

$$= (2-\lambda)\Big((2-\lambda)(4-\lambda)-4-4\Big) = (2-\lambda)(\lambda^2-6\lambda) = (2-\lambda)\lambda(\lambda-6)$$

לכן הע"ע הם $2,0,6$

נמצא וקטור עצמי לכל ע"ע: עבור 0 נחפש $A^tAx=0$

$$\begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow X_0 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

עבור $\lambda=2$: נחפש וקטור שמאפס את A^tA-2I

$$\begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & -2 \\ 2 & -2 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & -1 & 1 \end{bmatrix} \Rightarrow X_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\lambda=6: \begin{bmatrix} -4 & 0 & 2 \\ 0 & -4 & -2 \\ 2 & -2 & -2 \end{bmatrix} \to \begin{bmatrix} -2 & 0 & 1 \\ 0 & -2 & -1 \\ 1 & -1 & -1 \end{bmatrix} \to \begin{bmatrix} 0 & -2 & -1 \\ 0 & -2 & -1 \\ 1 & -1 & -1 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 0 & 0 \\ 0 & -2 & -1 \\ 1 & -1 & -1 \end{bmatrix} \Rightarrow X_6 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$$

מצאנו בסיס של ו"ע, כעת נרצה להפוך אותו לבסיס אורתונורמלי בשביל לבנות את המטריצה V שהיא אמורה להיות מטריצה אורתוגונלית (לכן העמודות שלה יהוו בסיס א"נ) נשים לב שהוקטורים בבסיס שמצאנו כבר מאונכים זה לזה, כי ו"ע עם ע"ע שונים הם תמיד מאונכים, לכן צריך רק לנרמל כל אחד מהם.

$$V_0 = \frac{X_0}{\|X_0\|} = \begin{bmatrix} -\sqrt{3}/3 \\ \sqrt{3}/3 \\ \sqrt{3}/3 \end{bmatrix}, \quad V_2 = \frac{X_2}{\|X_2\|} = \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \\ 0 \end{bmatrix}, \quad V_6 = \frac{X_6}{\|X_6\|} = \begin{bmatrix} \sqrt{6}/6 \\ -\sqrt{6}/6 \\ \sqrt{6}/3 \end{bmatrix}$$

לכן המטריצה Σ^2 היא מטריצת הע"ע:

$$\Sigma^2 = \begin{bmatrix} 6 & & \\ & 2 & \\ & & 0 \end{bmatrix} \Rightarrow \Sigma = \begin{bmatrix} \sqrt{6} & & \\ & \sqrt{2} & \\ & & 0 \end{bmatrix}$$

והמטריצה V היא המטריצה $V = [v_6\ v_2\ v_0]$

$$V = \begin{bmatrix} \sqrt{6}/6 & \sqrt{2}/2 & -\sqrt{3}/3 \\ -\sqrt{6}/6 & \sqrt{2}/2 & \sqrt{3}/3 \\ \sqrt{6}/3 & 0 & \sqrt{3}/3 \end{bmatrix}$$

כעת מצאנו את Σ, V ונשאר למצוא את המטריצה U:

$$A\cdot V = U\Sigma = \begin{bmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{bmatrix}\begin{bmatrix} \sqrt{6} & & \\ & \sqrt{2} & \\ & & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{6}\cdot u_1 & \sqrt{2}\cdot u_2 & 0 \end{bmatrix}$$

$$AV = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} \sqrt{6}/6 & \sqrt{2}/2 & -\sqrt{3}/3 \\ -\sqrt{6}/6 & \sqrt{2}/2 & \sqrt{3}/3 \\ \sqrt{6}/3 & 0 & \sqrt{3}/3 \end{bmatrix} = \begin{bmatrix} 0 & \sqrt{2} & 0 \\ \sqrt{6} & 0 & 0 \end{bmatrix}$$

$$\Rightarrow \quad u_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad u_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad u_3 = ?$$

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

נרצה למצוא את x, y: נחשב

$$\begin{bmatrix} 0 & 1 & x \\ 1 & 0 & y \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ x & y \end{bmatrix} = \begin{bmatrix} 1+x^2 & xy \\ xy & 1+y^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

לכן $x=y=0$.

ואכן מתקיים כנדרש:

$$A = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sqrt{6} & & \\ & \sqrt{2} & \\ & & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \sqrt{6}/6 & \sqrt{2}/2 & -\sqrt{3}/3 \\ -\sqrt{6}/6 & \sqrt{2}/2 & \sqrt{3}/3 \\ \sqrt{6}/3 & 0 & \sqrt{3}/3 \end{bmatrix}^T}_{V}$$

שאלה 3: תהי A מטריצה $A \in \mathbb{R}^{m\times n}$. נגדיר $C_0 = A^TA$. יהיו $\lambda_1, \dots, \lambda_n$ הערכים העצמיים של C_0, עם הו"ע $v_1, \dots, v_n$ בהתאמה. (וקטורים מנורמלים)
נניח ש $\lambda_1 > \lambda_2$, ונגדיר $b_k \in \mathbb{R}^n$ באופן רקורסיבי הבא:

$$b_0 = \sum_{i=1}^{n} a_i v_i \qquad b_{k+1} = \frac{C_0 b_k}{\|C_0 b_k\|}$$

b_0 הוא צ"ל כלשהו של הבסיס האורתונורמלי, עם $a_1 \neq 0$.

צריך להוכיח שמתקיים: $\lim_{k\to\infty} b_k = \pm v_1$.

<u>הוכחה:</u> נפתח את ההגדרה הרקורסיבית באופן איטרטיבי:

$$b_{k+1} = \frac{C_0 b_k}{\|C_0 b_k\|} = \frac{C_0 \cdot \frac{C_0 b_{k-1}}{\|C_0 b_{k-1}\|}}{\|C_0 b_k\|} = \frac{C_0^2 b_{k-1}}{\|C_0 b_k\| \, \|C_0 b_{k-1}\|} = \frac{C_0^3 b_{k-2}}{\|C_0 b_k\| \cdot \|C_0 b_{k-1}\| \cdot \|C_0 b_{k-2}\|}$$

באופן אינדוקטיבי ניתן להראות שמתקיים:

$$b_{k+1} = \frac{C_0^{k+1} b_0}{k}$$

באופן אינדוקטיבי ניתן להראות שמתקיים:

$$b_{k+1} = \frac{C_0^{k+1} b_0}{\prod_{i=0}^{k} \|C_0 b_i\|}$$

ובצורה שקולה: לכל $k \in \mathbb{N}$

$$b_k = \frac{C_0^{k} b_0}{\prod_{i=0}^{k-1} \|C_0 b_i\|}$$

נשים לב שמכיוון ש $b_k = \frac{C_0 b_{k-1}}{\|C_0 b_{k-1}\|}$ אז b_k הוא וקטור מנורמל לכל $k \in \mathbb{N}$ ולכן

מתקיים:

$$1 = \|b_k\| = \frac{\|C_0^k b_0\|}{\prod_{i=0}^{k-1} \|C_0 b_i\|} \Rightarrow \prod_{i=0}^{k-1} \|C_0 b_i\| = \|C_0^k b_0\| \Rightarrow b_k = \frac{C_0^k b_0}{\|C_0^k b_0\|}$$

מכיוון ש- C_0 היא מטריצה סימטרית, קיים לה פירוק EVD. אם נגדיר $U = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}$

ו: $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$ אז מתקיים $C_0 = UDU^T$. כעת

$$C_0^k = (UDU^T)^k = \underbrace{(UDU^T)(UDU^T)\cdots(UDU^T)}_{k \text{ פעמים}}$$

נשים לב $U^T U = I$ כי U אורתוגונלית, ולכן

$$C_0^k = (UD)(U^TU)D(U^TU)D(U^TU)D\cdots(U^TU)(DU^T) = UD^kU^T$$

D אלכסונית ולכן:

$$D^k = \begin{bmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{bmatrix}$$

$$C_0^k = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{bmatrix} \begin{bmatrix} - v_1 - \\ \vdots \\ - v_n - \end{bmatrix} = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} \cdot \begin{bmatrix} -\lambda_1^k v_1- \\ -\lambda_2^k v_2- \\ \vdots \\ -\lambda_n^k v_n- \end{bmatrix} = \sum_{i=1}^{n} \lambda_i^k \cdot v_i \otimes v_i$$

$$C_0^k b_0 = \left(\sum_{i=1}^{n} \lambda_i^k v_i \otimes v_i\right) \cdot b_0 = \sum_{i=1}^{n} \lambda_i^k (v_i \otimes v_i) b_0 = \sum_{i=1}^{n} \lambda_i^k (v_i \otimes v_i) \cdot \sum_{j=1}^{n} a_j v_j = \sum_{i=1}^{n} \lambda_i^k \sum_{j=1}^{n} a_j (v_i \otimes v_i) v_j$$

נסתכל על המכפלה: $(v_i \otimes v_i) v_j$:

אם נסמן $v_i = [x_1 \cdots x_n]$. אז עבור $i=j$:

$$(v_i \otimes v_i) \cdot v_i = \begin{bmatrix} x_1^2 & x_1 x_2 & \cdots & x_1 x_n \\ x_2 x_1 & x_2^2 & & \vdots \\ \vdots & & \ddots & \\ x_n x_1 & \cdots & & x_n^2 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = (x_1^2 + \dots + x_n^2) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = v_i$$

ועבור $i \neq j$ נקבל שמכיוון שכל שורה במטריצה $v_i \otimes v_i$ היא כפל סקלר של הוקטור v_i, ומכיוון ש $v_i \perp v_j$, אז $(v_i \otimes v_i) \cdot v_j = 0$. לכן:

$$C_0^k b_0 = \sum_{i=1}^{n} \lambda_i^k a_i v_i$$

$$C_0^k b_0 = \lambda_1^k \cdot \sum_{i=1}^{n} \left(\frac{\lambda_i}{\lambda_1}\right)^k \cdot a_i v_i$$

כאשר $k\to\infty$, אז מכיוון שלפי הנתון λ_1 גדול ממש מכל שאר הע"ע, אז מתקיים $\lim_{k\to\infty}\left(\frac{\lambda_i}{\lambda_1}\right)^k = 0$. עבור $i\neq 1$. $\Leftarrow$

$$\lim_{k\to\infty}\frac{C_0^k b_0}{\lambda_1^k} = a_1 v_1$$

בחלוקה לביטוי מקודם:

$$b_k = \frac{C_0^k b_0}{\|C_0^k b_0\|} = \frac{\frac{C_0^k b_0}{|\lambda_1^k|}}{\left\|\frac{C_0^k b_0}{\lambda_1^k}\right\|}$$

לכן קיבלנו שבגבול b_k מתכנס לוקטור שבהצגה שלו לפי הבסיס האו"נ $B=(v_1,\dots,v_n)$, כל הרכיבים הם אפס מלבד הרכיב של v_1.

נשים לב שלכל $k\in\mathbb{N}$, b_k הוא וקטור מנורמל, כיוון שלפי ההגדרה הרקורסיבית הוא מתקבל מחלוקה של וקטור בנורמה שלו. ומכיוון שאנחנו יודעים שמתקיים $\|v_1\|=1$ כי מדובר בבסיס או"נ אז המקדם של הוקטור v_1 בהצגה של הגבול חייב להיות ± 1. כלומר בסה"כ

$$\lim_{k\to\infty} b_k = \pm v_1$$ כנדרש. ∎

שאלה 4: יהי $x\in\mathbb{R}^n$ וקטור קבוע, ותהי $U\in\mathbb{R}^{n\times n}$ מטריצה אורתוגונלית. רוצים לחשב את היעקוביאן של הפונקציה f המוגדרת בתור:

$f:\mathbb{R}^n\to\mathbb{R}^n$

$$f(y) = U\cdot \mathrm{diag}(y)\,U^T x$$

כאשר $\mathrm{diag}(y)$ היא המטריצה האלכסונית

$$\mathrm{diag}(y) = \begin{bmatrix} y_1 & & & \\ & y_2 & & \\ & & \ddots & \\ & & & y_n \end{bmatrix}$$

לכל $i\in[n]$ נרצה לכתוב מפורשות את $f_i(y)$, כלומר את הקואורדינטה ה-i של $f(y)$, כפונקציה של $y_1,\dots,y_n$. נסמן $x=(x_1,\dots,x_n)^T$. ונסמן $B = U\,\mathrm{diag}(y)\,U^T$.

$$f_i(y) = f(y)_i = (Bx)_i = R_i(B)\cdot x$$

כאשר $R_i(B)$ היא השורה ה-i של המטריצה B.
לכן נרצה למצוא ביטוי ל- $R_i(B)$:

$$\mathrm{diag}(y)\cdot U^T = \begin{bmatrix} y_1 & & \\ & \ddots & \\ & & y_n\end{bmatrix}\begin{bmatrix} -u_1- \\ \vdots \\ -u_n-\end{bmatrix} = \begin{bmatrix} -y_1u_1- \\ -y_2u_2- \\ \vdots \\ -y_nu_n-\end{bmatrix}$$

$$U\,\mathrm{diag}(y)\,U^T = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{bmatrix}\begin{bmatrix} -y_1u_1- \\ \vdots \\ -y_nu_n-\end{bmatrix} = \sum_{i=1}^n y_i\, u_i\otimes u_i = B$$

$$R_i(B) = \sum_{j=1}^n y_j R_i(u_j\otimes u_j) = \sum_{j=1}^n y_j u_{ij} u_j$$

$$R_i(u_j\otimes u_j) = u_{ij}\cdot u_j$$

$$f_i(y_1,\dots,y_n) = R_i(B)\cdot x = \sum_{j=1}^n y_j u_{ij} u_j^T x$$

אם נרצה לגזור את הפונקציה f_i לפי המשתנה y_j, נקבל שכל המקדמים מתאפסים מלבד המקדם של y_j, כלומר

$$J_{ij}(f) = \frac{\partial f_i}{\partial y_j} = u_{ij}u_j^T x$$

אם נרצה לגזור את הפונקציה f_i לפי המשתנה y_j, נקבל שכל המקדמים מתאפסים מלבד המקדם של y_j, כלומר

$$J_y(f)_{ij} = \frac{\partial f_i}{\partial y_j} = u_{ij} u_j^T x$$

כאשר u_{ij} הכניסה ה-ij של U, ו-u_j העמודה ה-j של U.

שאלה 5: עבור וקטור y קבוע מגדירים פונקציה $h(\sigma) = \frac{1}{2}\| f(\sigma) - y \|^2$

נגדיר פונקציה $g: \mathbb{R}^n \to \mathbb{R}$ $g(\sigma) = \frac{1}{2}\| \sigma - y \|^2$ ונרצה להביע את היעקוביאן של g.

מתקיים $h = g \circ f$ ולכן לפי כלל השרשרת המורחב מתקיים

$$J_\sigma(h) = J_\sigma(g \circ f) = J_{f(\sigma)}(g) \cdot J_\sigma(f).$$

בשביל לחשב את היעקוביאן של g, נרצה לגזור את g לפי כ"א מהמשתנים $\sigma_1, \dots, \sigma_n$ (היעקוביאן יהיה וקטור כיוון ש-g היא פונ' ל-$\mathbb{R}$)

$$g(\sigma) = \frac{1}{2}\sum (\sigma_i - y_i)^2 \Rightarrow \frac{\partial g}{\partial \sigma_i} = \frac{1}{2} \cdot 2(\sigma_i - y_i) \cdot 1 = \sigma_i - y_i$$

כלומר $J_\sigma(g) = [\sigma_1 - y_1, \dots, \sigma_n - y_n]$

$$J_{f(\sigma)}(g) = [f(\sigma)_1 - y_1, \dots, f(\sigma)_n - y_n]$$

פיתחנו קודם את הביטוי

$$f_i(y_1, \dots, y_n) = R_i(B) \cdot X = \sum_{j=1}^n y_j u_{ij} u_j^T x$$

לכן:

$$\left(J_{f(\sigma)}(g)\right)_i = \sum_{j=1}^n \sigma_j u_{ij} u_j^T x - y_i$$

בנוסף ראינו את מהו היעקוביאן של f:

$$\left(J_\sigma(f)\right)_{ij} = u_{ij} u_j^T x$$

$$\Rightarrow J_\sigma(g \circ f)_i = J_{f(\sigma)}(g) \circ C_i\left(J_\sigma(f)\right) = \sum_{k=1}^n \left(J_{f(\sigma)}(g)\right)_k \cdot \left(J_\sigma(f)\right)_{ki}$$

$$= \sum_{k=1}^n \left(\sum_{j=1}^n \sigma_j u_{kj} u_j^T x - y_k \right) \circ u_{ki} u_i^T x$$

נשים לב שהיעקוביאן של h הוא וקטור שורה באורך n, והכניסה ה-i שלו מתקבלת מהביטוי הנ"ל.

6. Calculate the Jacobian of the softmax function $S : \mathbb{R}^d \to [0,1]^k$ **שאלה 6:**

e^{x_j}

שאלה 6:

6. Calculate the Jacobian of the softmax function $S:\mathbb{R}^d \to [0,1]^k$

$$S(\mathbf{x})_j = \frac{e^{x_j}}{\sum_{l=1}^{k} e^{x_l}}$$

$$\frac{\partial S(x)_i}{\partial x_j} = J_x(S)_{ij}$$

נרצה לגזור את הקואורדינטה ה-i של $S(X)$ לפי המשתנה x_j

$$S(X)_i = \frac{e^{x_i}}{\sum_{l=1}^{k} e^{x_l}}$$

אם $i \neq j$ אז המונה לא תלוי בכלל ב-x_j. אם נסמן את המכנה כ-$f(x_j)$ אז:

$$\left(\frac{1}{f(x)}\right)' \to -\frac{f'(x)}{(f(x))^2} \qquad \frac{\partial S(x)_i}{\partial x_j} = \frac{-e^{x_i}\cdot e^{x_j}}{\left(\sum_{l=1}^{k} e^{x_l}\right)^2}$$

כאשר $i=j$ אז גם במונה וגם במכנה מופיע המשתנה x_j שרוצים לגזור לפיו. נגזור לפי נגזרת של מנה

$$\frac{\partial S(x)_i}{\partial x_i} = \frac{e^{x_i}\cdot\sum_{l=1}^{k} e^{x_l} - e^{x_i}\cdot e^{x_i}}{\left(\sum_{l=1}^{k} e^{x_l}\right)^2} = \frac{e^{x_i}\cdot\sum_{\substack{l=1\\ l\neq i}}^{k} e^{x_l}}{\left(\sum_{l=1}^{k} e^{x_l}\right)^2}$$

בסה"כ:

$$J_x(S)_{ij} = \begin{cases} \dfrac{-e^{x_i}\cdot e^{x_j}}{\left(\sum_{l=1}^{k} e^{x_l}\right)^2} & : \quad i\neq j \\[2ex] \dfrac{e^{x_i}\cdot\sum_{\substack{l=1\\ l\neq i}}^{k} e^{x_l}}{\left(\sum_{l=1}^{k} e^{x_l}\right)^2} & : \quad i=j \end{cases}$$

שאלה 7:

7. Let $f:\mathbb{R}^d \to \mathbb{R}$ be defined as $f(x,y) = x^3 - 5xy - y^5$. Calculate the Hessian of f.

עבור $f: R^2 \to R$ ההסיאן מוגדר בתור מטריצה

$$\begin{bmatrix} \frac{\partial^2 f}{\partial^2 x} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial^2 y} \end{bmatrix}$$

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

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 (x^3 - 5xy - y^5)}{\partial x \partial y} = \frac{\partial(-5x - 5y^4)}{\partial x} = -5$$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 (x^3 - 5xy - y^5)}{\partial x \partial y} = \frac{\partial(-5x - 5y^4)}{\partial x} = -5$$

$$\frac{\partial^2 f}{\partial^2 y} = \frac{\partial^2 (x^3 - 5xy - y^5)}{\partial^2 y} = \frac{\partial(-5x - 5y^4)}{\partial y} = -20y^3$$

⇐ קיבלנו את מטר' ההסיאן הבאה:

$$\begin{bmatrix} 6x & -5 \\ -5 & -20y^3 \end{bmatrix}$$

שאלה 8:

8. Let $x_1, x_2, \ldots \overset{iid}{\sim} \mathcal{P}$ be a sample of infinity size drawn from some probability distribution function $\mathcal{P}$ with finite expectation and variance. Show that the sample mean estimator $\hat{\mu}_n = \frac{1}{n}\sum x_i$ calculated over the first n samples is a consistent estimator. Hint: for any given fixed value of $n \in \mathbb{N}$ bound from above the probability of deviating more than ε.

לפי הגדרה, בשביל להוכיח שהאומד $\hat{\mu}_n$ הוא קונסיסטנטי, צריך להראות שלכל $\varepsilon>0$ מתקיים

$$\lim_{n\to\infty} \mathbb{P}(|\hat{\mu}_n(x_1,\dots,x_n) - \mu| > \varepsilon) = 0$$

כאשר μ היא התוחלת האמיתית של ההתפלגות $\mathcal{P}$.

נוכיח זאת ב-2 שלבים: קודם כל נוכיח שמתקיים $\mathbb{E}[\hat{\mu}_n(x_1,\dots,x_n)] = \mu$, כלומר שהאומד $\hat{\mu}_n$ הוא בלתי-מוטה, ולאחר מכן נחסום את ההסתברות הנ"ל בעזרת א"ש צ'בישב.

מהנתון, התוחלת של המ"מ X_i המתאים לדגימה ה-i מקיימת $\mathbb{E}(X_i)=\mu$. (כי כל הדגימות הן i.i.d כלומר המ"מ המתאימים הם ב"ת ושווי-התפלגות). לכן:

$$\mathbb{E}(\hat{\mu}_n(x_1,\dots,x_n)) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\cdot\sum_{i=1}^{n}\mu = \mu$$

כעת, לפי א"ש צ'בישב מתקיים לכל $\varepsilon>0$

$$\mathbb{P}(|\hat{\mu}_n - \mathbb{E}(\hat{\mu}_n)| > \varepsilon) \le \frac{Var(\hat{\mu}_n)}{\varepsilon^2}$$

אם נסמן σ^2 את השונות של ההתפלגות $\mathcal{P}$, אז מתקיים $Var(X_i) = \sigma^2$ לכל $i \in \mathbb{N}$. כעת נחשב את השונות של האומד $\hat{\mu}_n$:

$$Var(\hat{\mu}_n(x_1,\dots,x_n)) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

מכיוון ש $\{X_i\}$ הם ב"ת בזוגות, אז ניתן לפתוח את ה-var בצורה לינארית.

לפי צ'בישב מתקבל לכל $\varepsilon>0$:

$$\mathbb{P}(|\hat{\mu}_n(x_1,\dots,x_n) - \mu| > \varepsilon) \le \frac{\frac{\sigma^2}{n}}{\varepsilon^2} = \frac{\sigma^2}{n\cdot\varepsilon^2} \xrightarrow[n\to\infty]{} 0$$

לכן $\lim_{n\to\infty} \mathbb{P}(|\hat{\mu}_n(x_1,\dots,x_n) - \mu| > \varepsilon) = 0$, ולכן $\hat{\mu}_n$ הוא אכן קונסיסטנטי כנדרש. ■

שאלה 9:

9. Let $\mathbf{x}_1, \ldots, \mathbf{x}_m \overset{iid}{\sim} \mathcal{N}(\mu, \Sigma)$ be m observations sampled i.i.d from a multivariate Gaussian with expectation of $\mu \in \mathbb{R}^d$ and a covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$. Derive the log-likelihood function of $\mathcal{N}(\mu, \Sigma)$. Hint: follow the approach used to derive the likelihood function for the univariate case.

ראשית נבין את הנוסחה להתפלגות גאוסית רב-ממדית: אם $X \sim N(\mu, \Sigma)$ אז

$$f(x) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \cdot \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

נרצה לפתח נוסחה לפונקציית הנראות של ההתפלגות הנ"ל, כלומר פונ' שבהינתן דגימות $x_1, \ldots, x_n$ של X, מתאימה לכל θ את ההסתברות לקבל את הדגימות הנתונות תחת הפרמטרים θ.
אם נסתכל על דגימה בודדת x_i, אז פונקציית הנראות היא: $\left(\theta = (\hat{\mu}, \hat{\Sigma})\right)$

$$L(\theta | x_i) = f_\theta(x_i) = \frac{1}{\sqrt{(2\pi)^d |\hat{\Sigma}|}} \cdot \exp\left(-\frac{1}{2}(x_i - \hat{\mu})^T \hat{\Sigma}^{-1} (x_i - \hat{\mu})\right)$$

מכיוון שהדגימות הן i.i.d, אז הנראות של θ מסוים תחת סדרה של דגימות $x_1, \ldots, x_n$, תהיה שווה למכפלת ההסתברויות לקבל כל דגימה בנפרד בהינתן שההתפלגות היא עם פרמטר θ. כלומר

$$L(\theta | x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

פונקציית ה log-likelihood המתקבלת היא:

$$\log L(\theta | x_1, \ldots, x_n) = \log\left(\prod_{i=1}^{n} f_\theta(x_i)\right) = \sum_{i=1}^{n} \log f_\theta(x_i) =$$

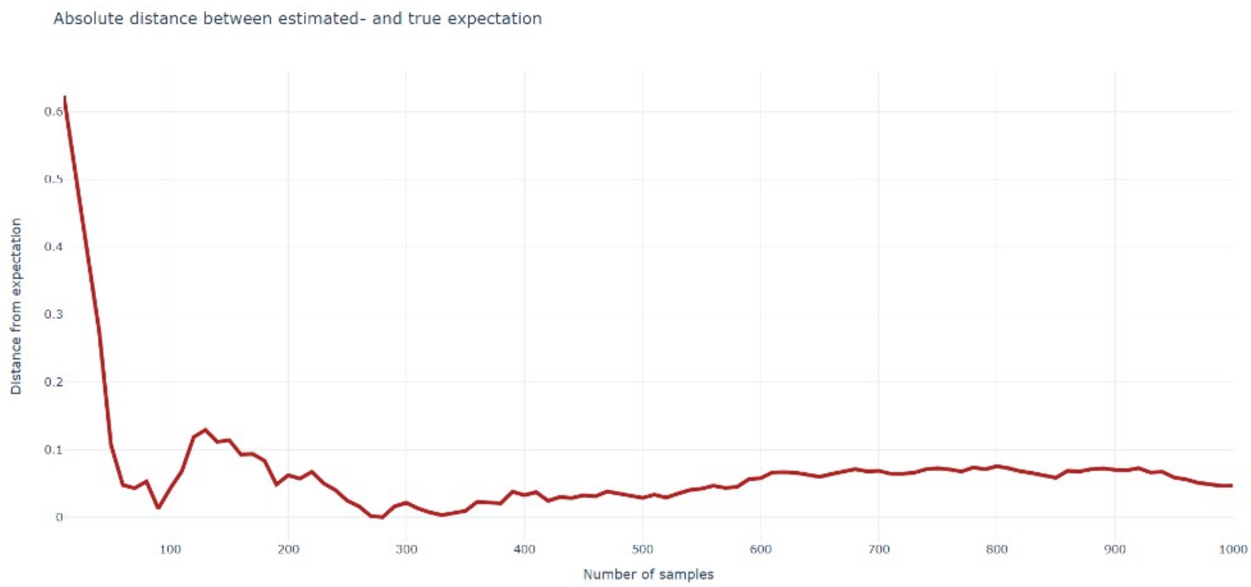
$$= n \log\left(\frac{1}{\sqrt{(2\pi)^d |\hat{\Sigma}|}}\right) + \sum_{i=1}^{n} \left(-\frac{1}{2}(x_i - \hat{\mu})^T \hat{\Sigma}^{-1} (x_i - \hat{\mu})\right)$$

גרפים לחלק המעשי:

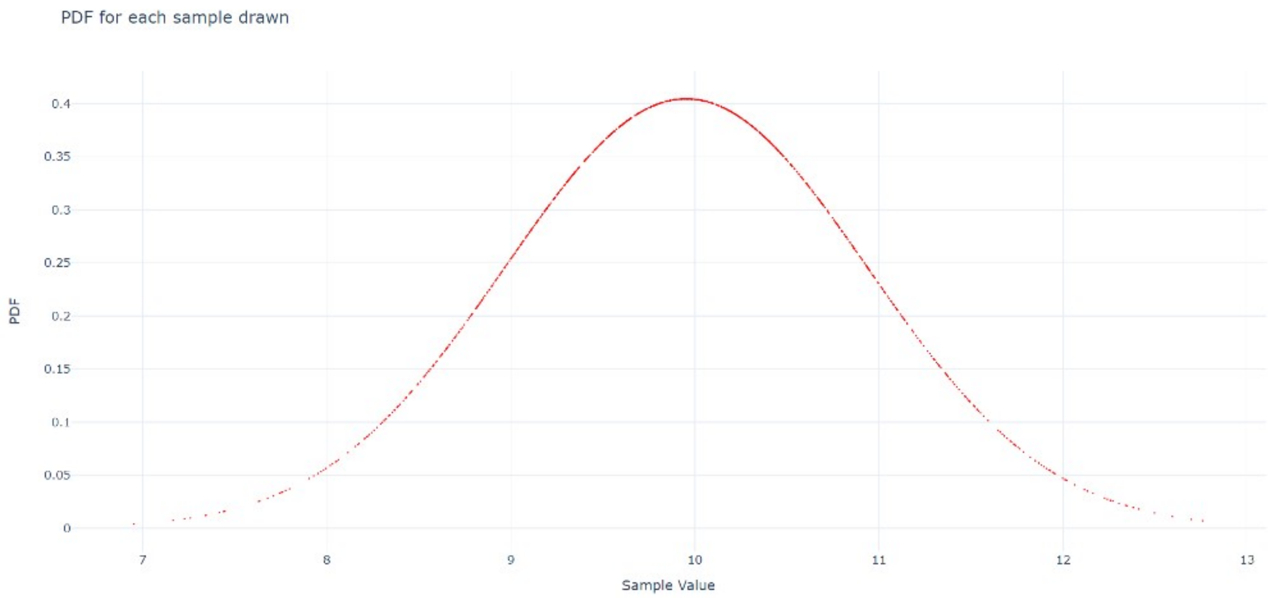
שאלה 1:

(0.9752096659781323 , 9.954743292509804)

שאלה 2:



שאלה 3:



מהגרף ניתן לראות שפונקציית הצפיפות המתקבלת היא בצורה של ההתפלגות נורמלית עם ממוצע שסוטה רק במעט מהערך האמתי של 10. זה הגיוני כיוון שהדגימות אכן הגיעו מהתפלגות נורמלית והתוחלת של פונקציית ההתפלגות ממנה נדגמו הייתה 10.

שאלה 4:

[-0.02282878 0.04313959 3.9932571 -0.02038981]

[[0.9166760774999685, 0.16634443759233586, -0.030275629165979034, 0.4628827064594544], [0.16634443759233586, 1.974182797387516, -0.005877889086032535, 0.04557631130140597],
[-0.030275629165979034, -0.005877889086032535, 0.9796027125646046, -0.020366857161867046], [0.4628827064594544, 0.04557631130140597, -0.020366857161867046, 0.9725372963099224]]

שאלה 5:

מהגרף למדנו שאכן הערך המקסימלי של הlog-likelihood מתקבל באזור של f1=0, f3=4. אלו אכן הפרמטרים האמיתיים של ההתפלגות ממנה הדגימות הגיעו, ולכן הגיוני שלערכים אלו תהיה נראות מקסימלית.

שאלה 6:

The f1, f3 values which yielded maximal likelihood: -0.050 3.969