מבוא למערכות לומדות־ תרגיל 1

מגישה זוהר בוחניק 311142293

31 במרץ 2019

שאלה 3:

חשבו את ההטלה של $v = \begin{pmatrix}1\\2\\3\\4\end{pmatrix}$ על הוקטור $w = \begin{pmatrix}0\\-1\\1\\2\end{pmatrix}$.

נחשב: $\langle v,w\rangle = 0-2+3+8 = 9$ $|w|^2 = 0+(-1)^2+1^2+2^2 = 6$

לכן ההטלה המתבקשת היא הוקטור־ $\frac{9}{6}\begin{pmatrix}0\\-1\\1\\2\end{pmatrix} = \begin{pmatrix}0\\-\frac{3}{2}\\\frac{3}{2}\\3\end{pmatrix}$

שאלה 4:

חשבו את ההטלה של $v = \begin{pmatrix}1\\2\\3\\4\end{pmatrix}$ על הוקטור $w = \begin{pmatrix}1\\0\\1\\-1\end{pmatrix}$.

נחשב: $\langle v,w\rangle = 1+0+3-4 = 0$, כלומר קיבלנו כי הוקטורים w, v מאונכים אזי ההטלה המתבקשת היא $\begin{pmatrix}0\\0\\0\\0\end{pmatrix}$

שאלה 7:

בעבור $A = UDV^T$ נייצג את המטריצה A^{-1} באופן הבא:

$$A^{-1} = \left(UDV^T\right)^{-1} = \left(V^T\right)^{-1} D^{-1}U^{-1} = VD^{-1}U^T$$

כאשר $D^{-1} = diag(\frac{1}{\sigma_1}, ..., \frac{1}{\sigma_n})$
אם נדע את צורת הSVD של המטריצה A נוכל למצוא את A^{-1} בקלות על ידי חישוב D^{-1} בצורה הנ"ל ואז מכפלת המטריצות $VD^{-1}U^T$ תביא לתוצאה הסופית המבוקשת.

שאלה 8:

מצאו את ה SVD של $C = \begin{pmatrix}5 & 5\\-1 & 7\end{pmatrix}$:

ראשית נחשב את המטריצה C^TC

$$\begin{pmatrix}5 & -1\\5 & 7\end{pmatrix}\begin{pmatrix}5 & 5\\-1 & 7\end{pmatrix} = \begin{pmatrix}26 & 18\\18 & 74\end{pmatrix}$$

כעת נמצא את הערכים העצמיים של המטריצה C^TC:- $detC^T\left(CC-\lambda I\right) = det\left(\begin{pmatrix}26-\lambda & 18\\18 & 74-\lambda\end{pmatrix}\right) = \left(26-\lambda\right)\left(74-\lambda\right) -$
$18^2 = \lambda^2 - 100\lambda + 1600 = \left(\lambda-20\right)\left(\lambda-80\right) = 0$

כלומר קיבלנו כי $\lambda_1 = 20, \lambda_2 = 80$ הערכים העצמיים של C^TC
עכשיו נחשב את הוקטורים העצמיים, זאת בעזרת מציאת בסיסים למרחבים העצמיים של C^TC:

$$C^TC - 20I = \begin{pmatrix} 6 & 18 \\ 18 & 54 \end{pmatrix}$$

נשים לב כי הטור השני הוא פי שלוש מהטור הראשון ולכן $v_1 = \begin{pmatrix} -3 \\ 1 \end{pmatrix}$ הוא בסיס מתאים עבור $\lambda = 20$

$$C^TC - 80I = \begin{pmatrix} -54 & 18 \\ 18 & -6 \end{pmatrix}$$

באופן דומה הטור הראשון הוא פי שלוש בערך מוחלט מהטור השני ולכן $v_2 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ הוא בסיס מתאים עבור $\lambda = 80$

ננרמל את הוקטורים

$$||v_1|| = \sqrt{\langle v_1, v_1 \rangle} = \sqrt{9+1} = \sqrt{10}$$

$$||v_2|| = \sqrt{\langle v_2, v_2 \rangle} = \sqrt{9+1} = \sqrt{10}$$

$$V = \begin{pmatrix} \frac{-3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{pmatrix} = \boldsymbol{V^T}$$

הערכים הסינגוליים של C^TC הם $\sigma_1 = \sqrt{20}, \sigma_2 = \sqrt{80}$ ולכן

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\sqrt{20}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\sqrt{80}} \end{pmatrix}$$

לבסוף נחשב את U בעזרת $CV = U\Sigma$

$$CV = \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix} \begin{pmatrix} \frac{-3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{pmatrix} = \begin{pmatrix} \frac{-10}{\sqrt{10}} & \frac{20}{\sqrt{10}} \\ \frac{10}{\sqrt{10}} & \frac{20}{\sqrt{10}} \end{pmatrix} = \begin{pmatrix} -\sqrt{10} & 2 \cdot \sqrt{10} \\ \sqrt{10} & 2 \cdot \sqrt{10} \end{pmatrix} = U\Sigma = U \cdot \begin{pmatrix} \sqrt{20} & 0 \\ 0 & \sqrt{80} \end{pmatrix}$$

כלומר:

$$\begin{pmatrix} -\sqrt{10} & 2 \cdot \sqrt{10} \\ \sqrt{10} & 2 \cdot \sqrt{10} \end{pmatrix} = U \cdot \begin{pmatrix} \sqrt{20} & 0 \\ 0 & \sqrt{80} \end{pmatrix}$$

$$x_{11} \cdot \sqrt{20} = -\sqrt{10} \Rightarrow x_{11} = -\frac{\sqrt{10}}{\sqrt{20}} = -\frac{1}{\sqrt{2}}$$

$$x_{12} \cdot \sqrt{80} = 2\sqrt{10} \Rightarrow x_{12} = \frac{2\sqrt{10}}{\sqrt{80}} = \frac{1}{\sqrt{2}}$$

$$x_{21} \cdot \sqrt{20} = \sqrt{10} \Rightarrow x_{21} = \frac{1}{\sqrt{2}}$$

$$x_{22} \cdot \sqrt{80} = 2\sqrt{10} \Rightarrow x_{22} = \frac{2\sqrt{10}}{\sqrt{80}} = \frac{1}{\sqrt{2}}$$

$$\Longrightarrow \boldsymbol{U} = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

שאלה 9:

תהי $A \in M_{m\times n}(\mathbb{R})$ מדרגה r, ותהי $C_0 = A^T A$. נגדיר $b_{k+1} = \frac{C_0 b_k}{||C_0 b_k||}$ כאשר b_0 מאותחל רנדומלית. צריך להראות כי $lim_{k\to\infty} b_k = v_1$ כאשר v_1 הוא וקטור סינגולרי הימני ביותר התואם לערך הסינגולרי הגדול ביותר של C_0. ניתן להניח כי $\lambda_1 > \lambda_2$ כאשר $\lambda_1,...,\lambda_n$ הערכים העצמיים של C_0 $v_1,...,v_n$ ו"ע של C_0, וכי $b_0 = \sum_{i=1}^n a_i v_i$ $a_1 \neq 0$. נייצג את C_0 על ידי EVD:
$C_0 = VDV^T$
נמצא את b_1 בעזרת b_0:

$$b_1 = \frac{C_0 b_0}{||C_0 b_0||} = \frac{VDV^T \cdot \sum_{i=1}^n a_i v_i}{||VDV^T \cdot \sum_{i=1}^n a_i v_i||}$$

$$VDV^T \cdot \sum_{i=1}^n a_i v_i = V \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} V^T \cdot V \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = V \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \cdot I_n \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = V \cdot \begin{bmatrix} \lambda_1 a_1 \\ \vdots \\ \lambda_n a_n \end{bmatrix} = \sum_{i=1}^n \lambda_i \cdot v_i \cdot a_i$$

כלומר קיבלנו:

$$b_1 = \frac{\sum_{i=1}^n \lambda_i \cdot v_i \cdot a_i}{||\sum_{i=1}^n \lambda_i \cdot v_i \cdot a_i||}$$

טענה: לכל k מתקיים

$$b_k = \frac{\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i}{||\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i||}$$

נראה באינדוקציה:
<u>בסיס</u>־ ראינו כי

$$b_1 = \frac{\sum_{i=1}^n \lambda_i \cdot v_i \cdot a_i}{||\sum_{i=1}^n \lambda_i \cdot v_i \cdot a_i||}$$

<u>צעד</u>־ נניח עבור k ונוכיח עבור $k+1$:
מהנחת האינדוקציה מתקיים:

$$b_k = \frac{\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i}{||\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i||}$$

לכן נציב:

$$b_{k+1} = \frac{C_0 b_k}{||C_0 b_k||} = \frac{VDV^T \cdot b_k}{||VDV^T \cdot b_k||}$$

מתקיים־

$$C_0 b_k = VDV^T \cdot b_k = VDV^T \cdot \frac{\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i}{||\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i||} = \frac{VDV^T \cdot V \cdot \begin{bmatrix} \lambda_1^k a_1 \\ \vdots \\ \lambda_n^k a_n \end{bmatrix}}{||\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i||} = \frac{V \cdot \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \cdot \begin{bmatrix} \lambda_1^k a_1 \\ \vdots \\ \lambda_n^k a_n \end{bmatrix}}{||\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i||}$$

$$= \frac{V \cdot \begin{bmatrix} \lambda_1^{k+1} a_1 \\ \vdots \\ \lambda_n^{k+1} a_n \end{bmatrix}}{\left|\left|\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i\right|\right|} = \frac{\sum_{i=1}^n v_i \lambda_i^{k+1} a_i}{\left|\left|\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i\right|\right|}$$

נשים לב כי

$$b_{k+1} = \frac{\frac{\sum_{i=1}^n v_i \lambda_i^{k+1} a_i}{\left|\left|\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i\right|\right|}}{\left|\left|\frac{\sum_{i=1}^n v_i \lambda_i^{k+1} a_i}{\left|\left|\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i\right|\right|}\right|\right|} = \frac{\frac{\sum_{i=1}^n v_i \lambda_i^{k+1} a_i}{\left|\left|\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i\right|\right|}}{\frac{1}{\left|\left|\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i\right|\right|} \cdot \left|\left|\sum_{i=1}^n v_i \lambda_i^{k+1} a_i\right|\right|} = \frac{\sum_{i=1}^n v_i \lambda_i^{k+1} a_i}{\left|\left|\sum_{i=1}^n v_i \lambda_i^{k+1} a_i\right|\right|}$$

כלומר הוכחנו את צעד האינדוקציה כנדרש.
כעת נשתמש בביטוי הזה של b_k על מנת לחשב את הגבול המבוקש:

$$lim_{k\to\infty} b_k = lim_{k\to\infty} \frac{\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i}{\left|\left|\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i\right|\right|} = lim_{k\to\infty} \frac{V \cdot \begin{bmatrix} \lambda_1^k a_1 \\ \vdots \\ \lambda_n^k a_n \end{bmatrix}}{\left|\left|\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i\right|\right|}$$

$$= lim_{k\to\infty} \frac{V \cdot \lambda_1^k \cdot \begin{bmatrix} 1 & & & \\ & \frac{\lambda_2^k}{\lambda_1^k} & & \\ & & \ddots & \\ & & & \frac{\lambda_n^k}{\lambda_1^k} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}}{\left|\left|\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i\right|\right|}$$

מתקיים

$$lim_{k\to\infty} \begin{bmatrix} 1 & & & \\ & \frac{\lambda_2^k}{\lambda_1^k} & & \\ & & \ddots & \\ & & & \frac{\lambda_n^k}{\lambda_1^k} \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix}$$

כי λ_1 הוא הערך העצמי הכי גדול ולכן לכל $i \neq 1$ מתקיים $lim_{k\to\infty} \left(\frac{\lambda_i}{\lambda_1}\right)^k = 0$
לכן בתצוגה $\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i$ ניתן לראות כי כאשר $k \to \infty$ אנו מקבלים כי: $\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i \to v_1 a_1 \lambda_1^k$
מכאן נובע לבסוף־

$$lim_{k\to\infty} b_k = lim_{k\to\infty} \frac{\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i}{\left|\left|\sum_{i=1}^n \lambda_i^k \cdot v_i \cdot a_i\right|\right|} = lim_{k\to\infty} \frac{v_1 a_1 \lambda_1^k}{\left|\left|v_1 a_1 \lambda_1^k\right|\right|} = v_1$$

זאת כיוון ש $||v_1|| = 1$ וכיוון ש $lim_{k\to\infty} \frac{a_1 \lambda_1^k}{|a_1|\left|\lambda_1^k\right|} = 1$ (בהנחה שהערכים חיוביים).

שאלה 14:

יהי V תת מרחב של $\mathbb{R}$ ממימד k, ויהי $v_1,...,v_k \in \mathbb{R}^d$ בסיס אורתונורמלי של V. תהי $P=\sum_{i=1}^{k} v_i v_i^T$ מטריצת הטלה לתת המרחב V.

צ.ל־ P סימטרית.

הוכחה־ ראשית נשים לב כי $v_i v_i^T$ היא מטריצה סימטרית לכל i. זאת כיוון שמתקיים $\left(v_i v_i^T\right)^T=\left(v_i^T\right)^T v_i^T=v_i v_i^T$. שנית, נשים לב כי P היא סכום של מטריצות סימטריות וסכום של שתי מטריצות סימטריות שומר על תכונת הסימטריה: יהיו A,B סימטריות אזי: $(A+B)^T=A^T+B^T=A+B$, כלומר $A+B$ סימטרית. מכאן P סימטרית.

שאלה 15:

צ.ל $v_1,...v_k$ הם וקטורים עצמיים של P עם ערך עצמי $\lambda=1$. כלומר צריך להראות כי לכל $i\in[k]$ מתקיים $P\cdot v_i=v_i$.

$$P\cdot v_i=\left(\sum_{i=1}^{k} v_i v_i^T\right)\cdot v_i=\sum_{i=1}^{k} v_i v_i^T v_i=v_1 v_1^T v_i+v_2 v_2^T v_i+....+v_i v_i^T v_i+...+v_k v_k^T v_i$$

כיוון ש $v_i\perp v_j$ לכל $i\neq j$ אזי המכפלה $v_j^T v_i=\bar{0}$, לכן נקבל סה"כ כי

$$P\cdot v_i=v_i v_i^T v_i$$

אמנם, כיוון ש v_i הוא אורתונורמלי הנורמה שלו שווה ל1, והרי $v_i^T v_i=||v_i||^2=1$, כלומר קיבלנו $P\cdot v_i=v_i$ כנדרש.

שאלה 16:

כתבו EVD אפשרי ל־P.

ראשית ניקח את הבסיס $v_1,..,v_k$ של V ונשלים אותו לבסיס אורתונומלי של $\mathbb{R}^d$ $v_1=u_1,...,v_k=u_k,u_{k+1},..,u_d$

תהי $V=\begin{bmatrix} | & | & & | \\ | & | & & | \\ u_1 & u_2 & \cdots & u_d \\ | & | & & | \\ | & | & & | \end{bmatrix}$ מטריצה אורתוגונלית, אזי $P=VDV^T$ כאשר $D=\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}\in M_{d\times d}(\mathbb{R})$ (מספר האחדות הוא לפחות k, והשאר אפסים).

שאלה 17:

הראו כי $P^2=PP^T=P^TP=P$.

ראשית P סימטרית, לכן $P=P^T$, מכאן $P^2=PP=P^TP=PP^T$. נותר להראות כי $P=P^2$. נשים לב כי P היא מטריצת הטלה, מהגדרה לאחר שמבצעים הטלה על וקטור כלשהו, אם נבצע שוב הטלה לא יקרה כלום, כלומר להטיל פעמיים זה בדיוק כמו להטיל פעם אחת ולכן $P=P^2$. בנוסף, בעזרת טענה מסעיף 18 ניתן להסיק את השיוויון:

$$(I-P)\,P=\bar{0}\Rightarrow IP-P^2=0\Rightarrow IP=P^2\Rightarrow P=P^2$$

שאלה 19:

הראו כי לכל $x\in V$ מתקיים $Px=x$. בשאלה 15 ראינו כי כל וקטור בבסיס $v_1,..,v_k$ של V מקיים $Pv_i=v_i$. עבור $x\in V$ מתקיים $x=a_1v_1+...+a_kv_k$ בעבור $a_1,...,a_k\in\mathbb{R}$ סקלרים כלשהם. לכן

$$Px=P\cdot(a_1v_1+...+a_kv_k)=P\cdot a_1v_1+...+P\cdot a_kv_k=a_1\cdot(P\cdot v_1)+...+a_k\cdot(P\cdot v_k)=$$

$$=a_1\cdot(v_1)+...+a_k\cdot(v_k)=x$$

כנדרש בשאלה.

שאלה 20:

נחשב את היעקוביאן של

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

נסמן

$$\sigma=\begin{bmatrix}\lambda_{1}\\ \\ \vdots\\ \\ \lambda_{n}\end{bmatrix}$$

$$U\cdot diag\left(\sigma\right)U^{T}x=\begin{bmatrix}| & & |\\ | & & |\\ u_{1} & \cdots & u_{n}\\ | & & |\\ | & & |\end{bmatrix}\begin{bmatrix}\lambda_{1} & & \\ & \ddots & \\ & & \lambda_{n}\end{bmatrix}\begin{bmatrix}- & - & u_{1} & - & -\\ & & \vdots & & \\ - & - & u_{n} & - & -\end{bmatrix}\begin{bmatrix}x_{1}\\ \vdots\\ x_{n}\end{bmatrix}=$$

$$=\begin{bmatrix}| & & |\\ | & & |\\ u_{1} & \cdots & u_{n}\\ | & & |\\ | & & |\end{bmatrix}\begin{bmatrix}\lambda_{1} & & \\ & \ddots & \\ & & \lambda_{n}\end{bmatrix}\begin{bmatrix}\left\langle u_{1},x\right\rangle \\ \\ \vdots\\ \\ \left\langle u_{n},x\right\rangle \end{bmatrix}=\begin{bmatrix}| & & |\\ | & & |\\ u_{1} & \cdots & u_{n}\\ | & & |\\ | & & |\end{bmatrix}\begin{bmatrix}\lambda_{1}\cdot\left\langle u_{1},x\right\rangle \\ \\ \vdots\\ \\ \lambda_{n}\cdot\left\langle u_{n},x\right\rangle \end{bmatrix}=$$

$$=\begin{bmatrix}| & & |\\ | & & |\\ u_{1} & \cdots & u_{n}\\ | & & |\\ | & & |\end{bmatrix}\left(\begin{bmatrix}\lambda_{1}\cdot\left\langle u_{1},x\right\rangle \\ 0\\ \vdots\\ \\ 0\end{bmatrix}+....+\begin{bmatrix}0\\ \\ \vdots\\ 0\\ \lambda_{n}\cdot\left\langle u_{n},x\right\rangle \end{bmatrix}\right)=$$

$$=\lambda_{1}\cdot\left\langle u_{1},x\right\rangle \cdot\begin{bmatrix}|\\ |\\ u_{1}\\ |\\ |\end{bmatrix}+.....+\lambda_{n}\cdot\left\langle u_{n},x\right\rangle \begin{bmatrix}|\\ |\\ u_{n}\\ |\\ |\end{bmatrix}=\begin{bmatrix}\sum_{i=1}^{n}\lambda_{i}\cdot\left\langle u_{i},x\right\rangle \left[u_{i}\right]_{\boldsymbol{1}}\\ \\ \vdots\\ \\ \sum_{i=1}^{n}\lambda_{i}\cdot\left\langle u_{i},x\right\rangle \left[u_{i}\right]_{\boldsymbol{n}}\end{bmatrix}$$

קיבלנו בעצם n פונקציות $f_{j}:\mathbb{R}^{n}\rightarrow\mathbb{R}$ כאשר $j\in\left[n\right]$:

$$f_{j}\left(\sigma\right)=\sum_{i=1}^{n}\lambda_{i}\cdot\left\langle u_{i},x\right\rangle \left[u_{i}\right]_{\boldsymbol{j}}$$

נחשב את הגרדיאנט של f_{j}:

$$\frac{\partial f_{j}\left(\sigma\right)}{\partial\lambda_{k}}=\left\langle u_{k},x\right\rangle \left[u_{k}\right]_{\boldsymbol{j}}$$

$$\nabla f_{j}\left(\sigma\right)=\begin{bmatrix}\left\langle u_{1},x\right\rangle \left[u_{1}\right]_{\boldsymbol{j}}\\ \\ \vdots\\ \\ \left\langle u_{n},x\right\rangle \left[u_{n}\right]_{\boldsymbol{j}}\end{bmatrix}$$

לכן לבסוף נקבל את היעקוביאן הבא:

$$\begin{bmatrix} \langle u_1, x\rangle [u_1]_1 & \cdots & \langle u_n, x\rangle [u_n]_1 \\ & \vdots & \\ \langle u_1, x\rangle [u_1]_n & \cdots & \langle u_n, x\rangle [u_n]_n \end{bmatrix} = \begin{bmatrix} (\nabla f_1(\sigma))^T \\ \vdots \\ (\nabla f_n(\sigma))^T \end{bmatrix}$$

שאלה 21:

מתקיים

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

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

צריך לחשב את הגרדיאנט של h. נגדיר $g(\bar{u}) = \frac{1}{2} ||\bar{u} - y||^2$ $g : \mathbb{R}^n \to \mathbb{R}$, אזי $h = g \circ f(\sigma)$ ומכלל השרשרת $J_\sigma(g \circ f) = J_{f(\sigma)}(g) \cdot J_\sigma(f)$

נשים לב כי את הגרדיאנט של g אנו מכירים מהכיתה כלומר $\Delta g(u) = 2 \cdot \frac{1}{2} (\bar{u} - y)^T = (\bar{u} - y)^T$, לכן

$$J_\sigma(g \circ f) = J_{f(\sigma)}(g) \cdot J_\sigma(f) = \left(\underbrace{U \cdot diag(\sigma) \cdot U^T x}_{f(\sigma)} - y \right)^T \cdot \begin{bmatrix} \langle u_1, x\rangle [u_1]_1 & \cdots & \langle u_n, x\rangle [u_n]_1 \\ & \vdots & \\ \langle u_1, x\rangle [u_1]_n & \cdots & \langle u_n, x\rangle [u_n]_n \end{bmatrix}$$

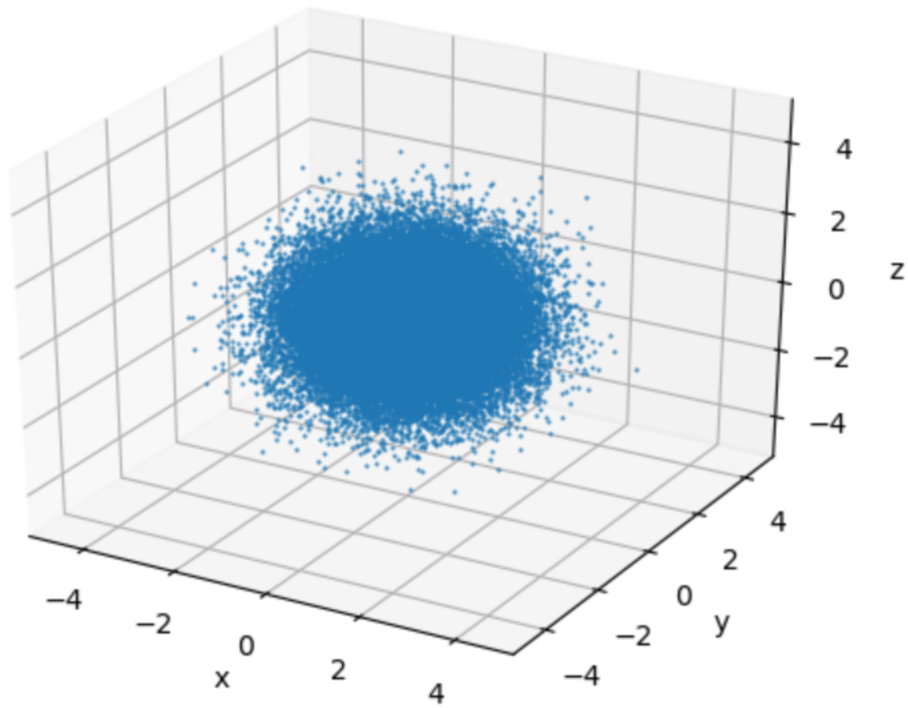
שאלה 22:

נמשיך את החישוב עבור $i \neq j$:

$$g(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$$

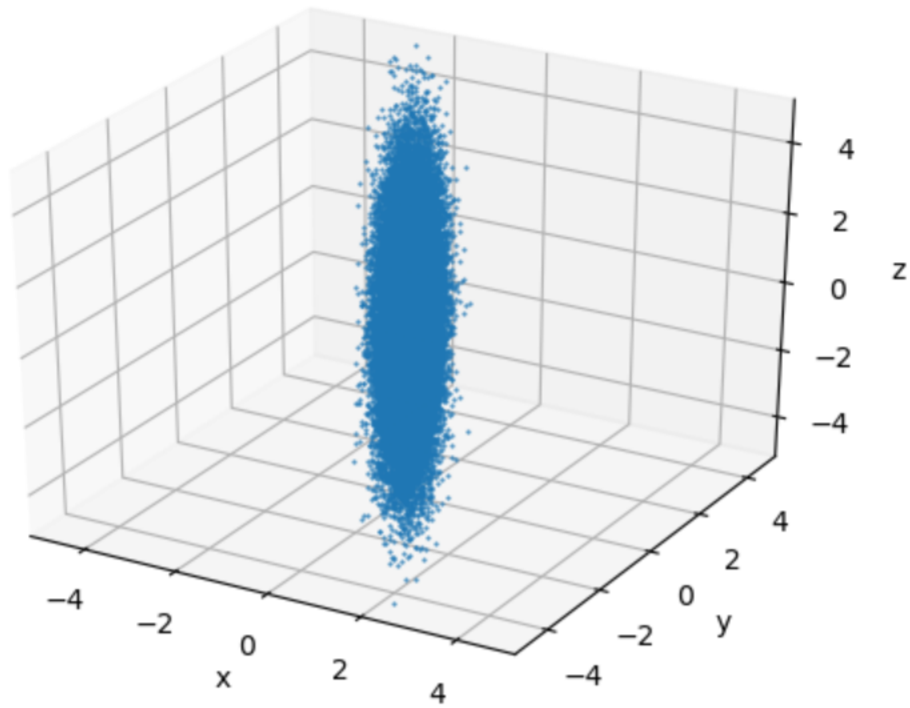
$$\frac{\partial S_i}{\partial a_j} = \frac{-e^{z_i} \cdot e^{z_j}}{\left(\sum_{k=1}^K e^{z_k}\right)^2} = -\frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}} \cdot \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} = -S_i \cdot S_j$$

שאלה 23:



בבחירה של נקודות רנדומליות קיבלנו את הגרף הנ"ל

שאלה 24:



נציג את מטריצת הcov בשתי הצורות:
אנליטית־

$$SCS^T = \begin{bmatrix} 0.01 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

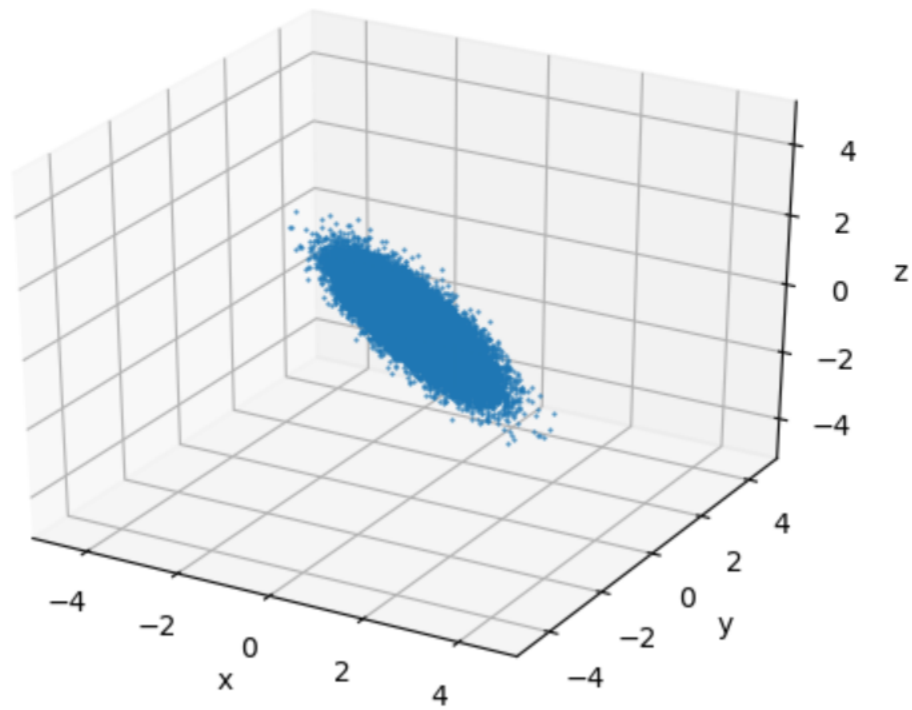
נומרית־

$$\hat{C} = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right) \left(X_i - \bar{X}\right)^T = \frac{X \cdot X^T}{50000 - 1} =$$

```
cov matrix numeric calculation:
[[ 9.92371909e-03 -1.75578080e-04 -8.59447728e-04]
 [-1.75578080e-04  2.48545643e-01 -1.11313226e-02]
 [-8.59447728e-04 -1.11313226e-02  4.03725894e+00]]
```

נראה כי עדיין אין תלות בין הx הy וה־ z. אין השפעה של עליית ה־ x על ערכי ה$-y$ או z בצורה בולטת. ניתן לראות כי מטריצת הcov אלכסונית, כלומר המשתנים המקריים בלתי מתואמים. בהצגה נומרית קיבלנו ערכים מאוד דומים למטריצת הcov בחישוב האנליטי.

שאלה 25:



נציג את מטריצת הcov בשתי הצורות:
אנליטית-

$$RCR^T =$$

```
cov matrix analytic calculation:
[[ 2.04288197 -1.88123062  0.62605709]
 [-1.88123062  1.75381475 -0.60694516]
 [ 0.62605709 -0.60694516  0.46330327]]
```

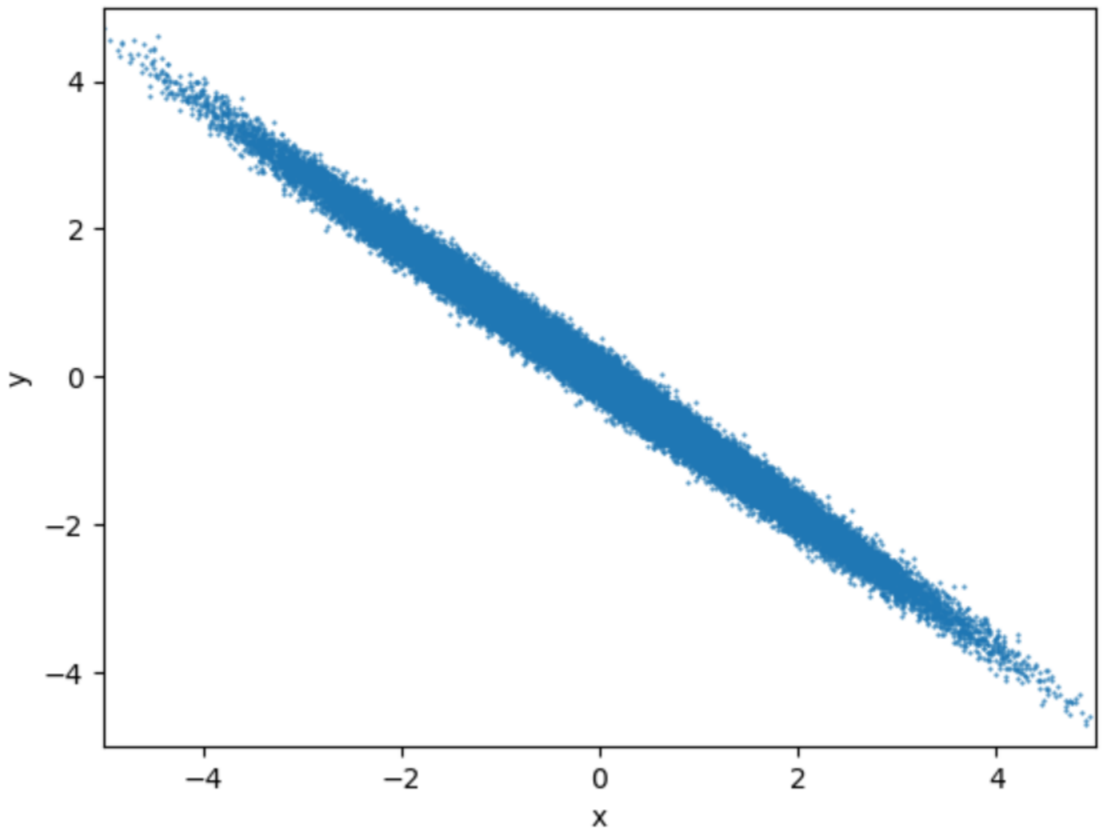
נומרית-

$$\hat{C} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)^T = \frac{X \cdot X^T}{50000-1} =$$

```
cov matrix numeric calculation:
[[ 2.03892359 -1.87860185  0.62783799]
 [-1.87860185  1.75193748 -0.60813309]
 [ 0.62783799 -0.60813309  0.46504032]]
```

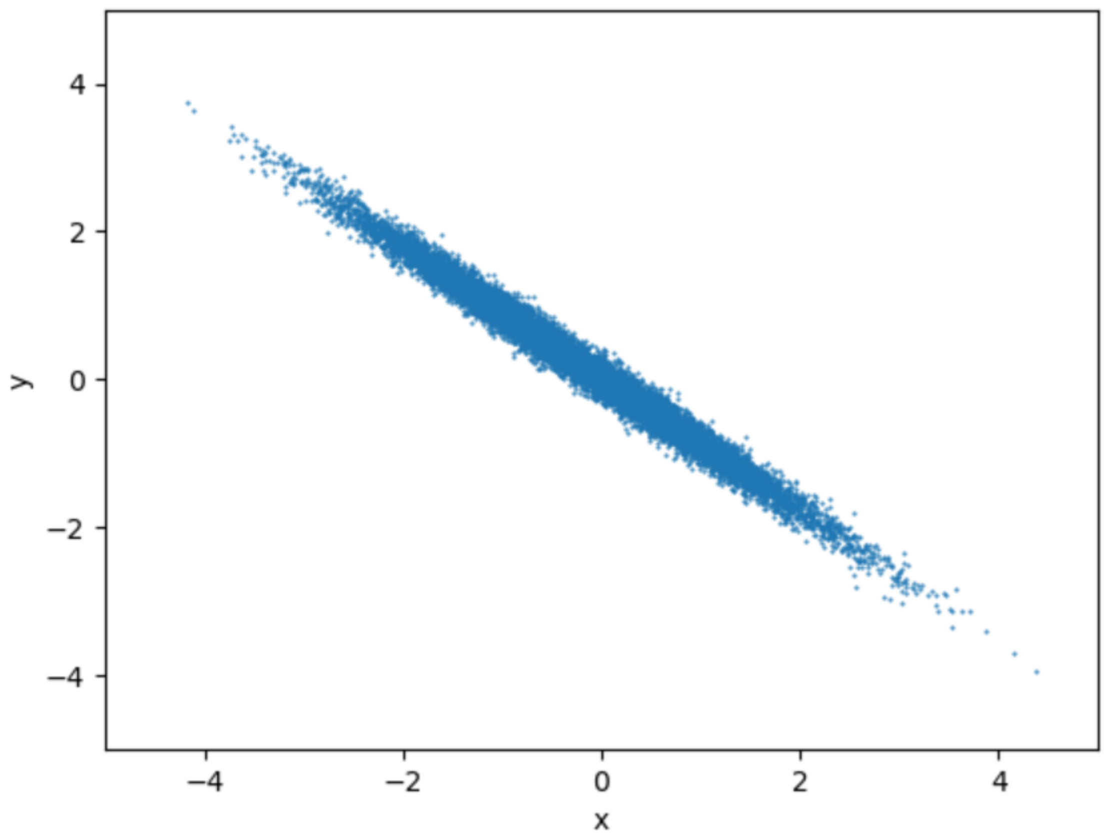
נראה כי נוצרה תלות לינארית בין הפרמטרים לאחר הסיבוב. נשים לב כי במטרצית ה-cov השתנתה כך שניתן לראות כי המשתנים לא בלתי מתואמים.

שאלה 26:



בגרף אחרי השיטוח וביטול ציר ה z ניתן לראות כי הנקודות שמרו על הצורה הלינארית שלהן, גם בדו מימד. הצפיפות המרכזית היא במרכז הגרף וככל שמתקרבים לקצוות הצפיפות יורדת

שאלה 27:



לאחר שהוספנו את התנאי קיבלנו צורה מוקטנת של הגרף הקודם. בעצם ניקינו את כל הנקודות שחורגות מהגבול הנתון ואז כששיטחנו את הנתונים לדו מימד קיבלנו כיווץ של הגרף הקודם שהיה ללא תנאים נוספים. אמנם הצפיפות נשמרה.

שאלה 28:

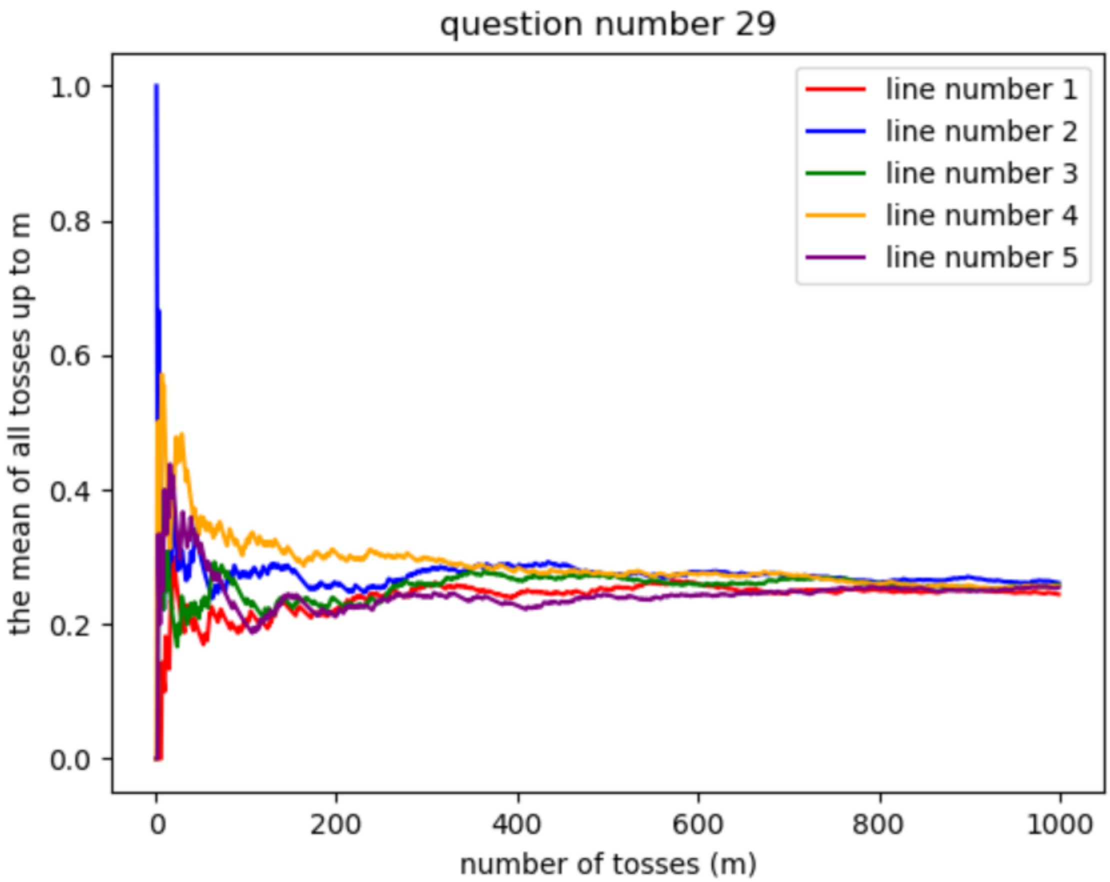
הm שמצאנו היה

$$m = \lceil \frac{1}{2\varepsilon^2} \cdot log\left(\frac{4}{\delta}\right) \rceil$$

על ידי שימוש בחסמים של הופדינג וצ'בישב מצאנו את m כך שבשימוש באלגוריתם A שראינו בכיתה נוכל לקבוע את p בדיוק הקירוב εוביטחון δ.
חסמנו בעצם את מורכבות המדגם $m\left(\varepsilon,\delta\right)$ על ידי שני חסמים, בתחילה $m\left(\varepsilon,\delta\right) \leq \lceil \frac{1}{4\cdot\varepsilon^2} \cdot \frac{1}{\delta} \rceil$, ואז שיפרנו אותו לידי $m\left(\varepsilon,\delta\right) \leq \lceil \frac{1}{2\cdot\varepsilon^2} \cdot log\left(\frac{2}{\delta}\right) \rceil$, ואז השתמשנו בחסמים להוכחה של האלגוריתם A.

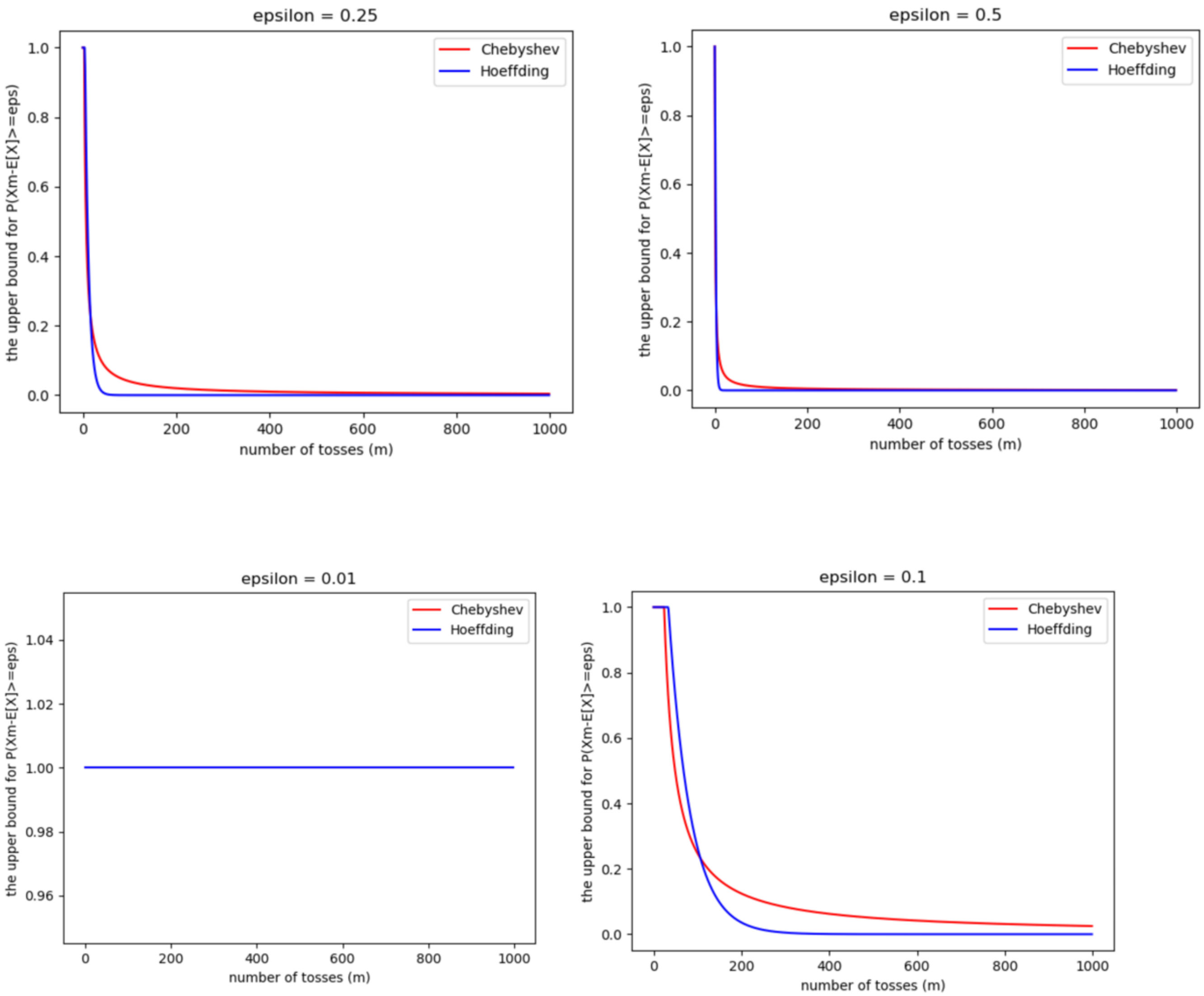
שאלה 29:

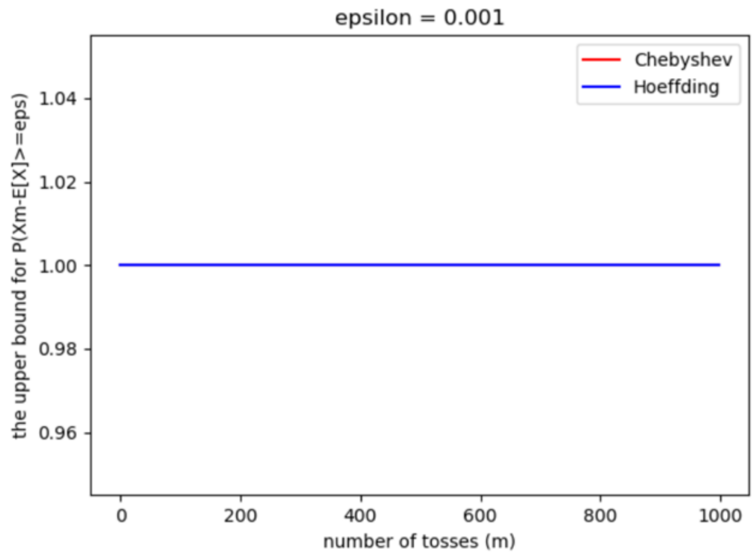
א.



נצפה שככל שm גדל כך התוצאות בין השורות השונות יהיו יותר דומות, כלומר הממוצע שיתקבל בכל שורה יהיה דומה, וכך גם קיבלנו בגרף ככל שm גדל כך הממוצעים התקרבו אחד לשני, כלומר הערך המתקבל יותר מהימן.

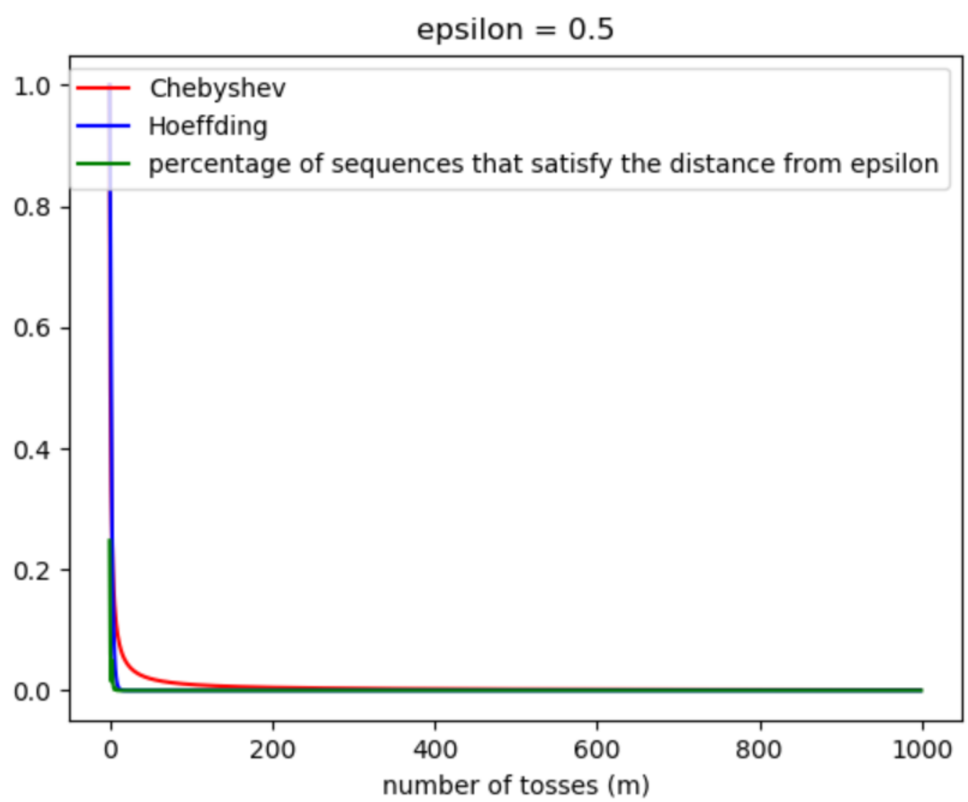
ב.

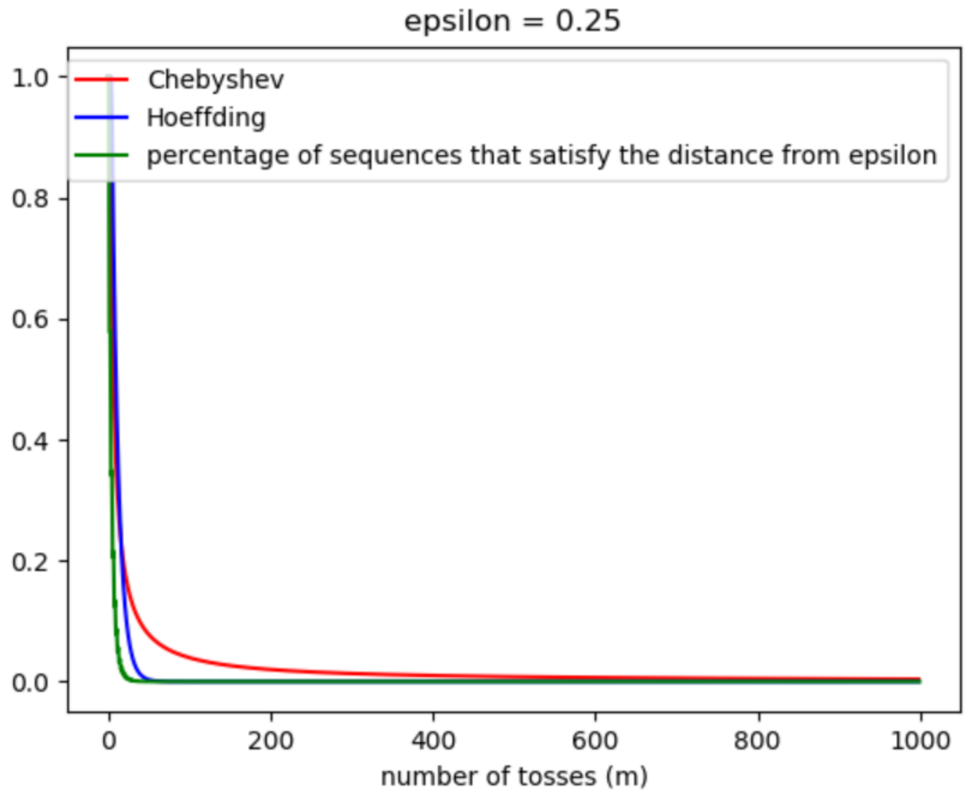

epsilon = 0.25
the upper bound for P(Xm-E[X]>=eps)
number of tosses (m)
Chebyshev
Hoeffding
epsilon = 0.5
the upper bound for P(Xm-E[X]>=eps)
number of tosses (m)
Chebyshev
Hoeffding
epsilon = 0.01
the upper bound for P(Xm-E[X]>=eps)
number of tosses (m)
Chebyshev
Hoeffding
epsilon = 0.1
the upper bound for P(Xm-E[X]>=eps)
number of tosses (m)
Chebyshev
Hoeffding

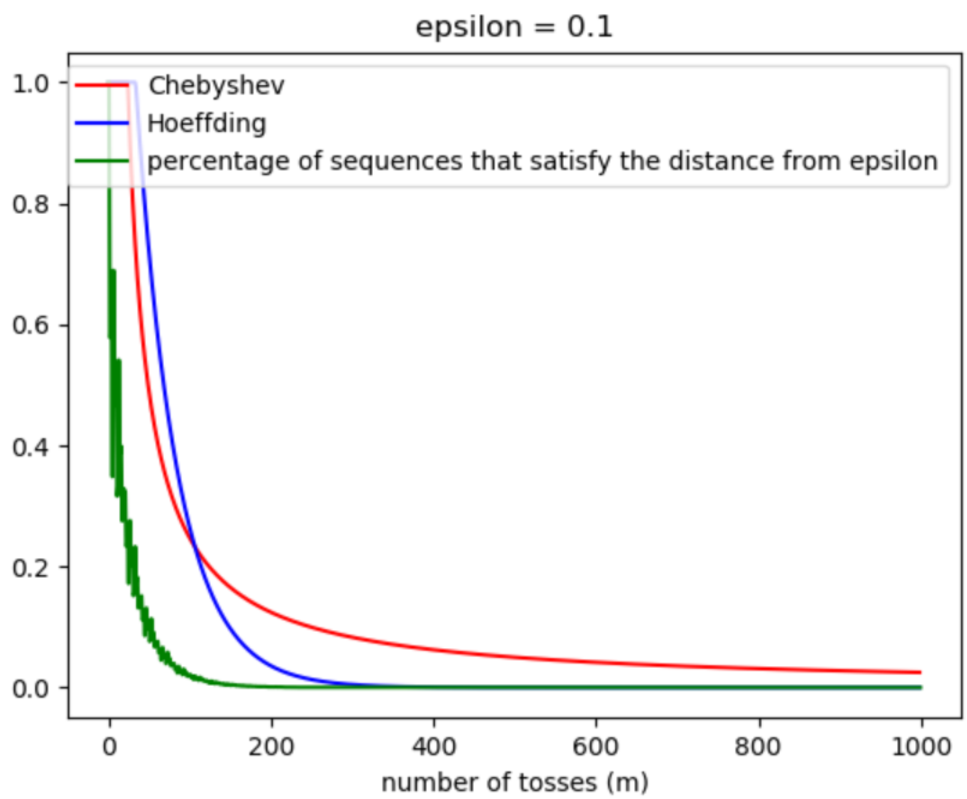


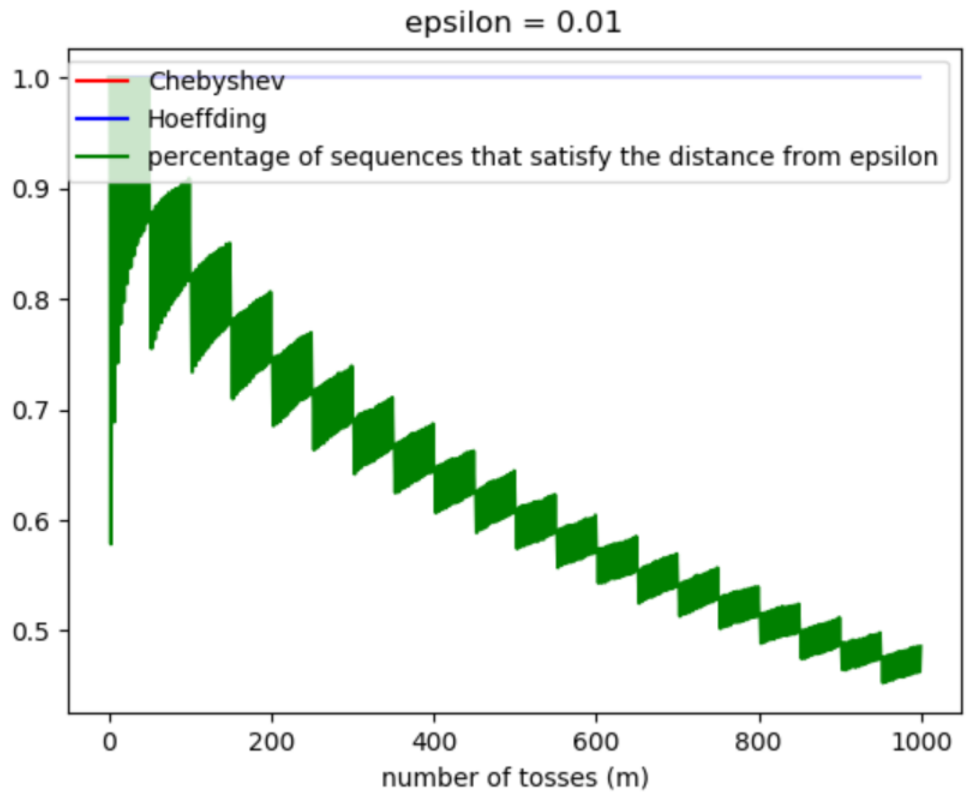
בגרפים האלו אנו רואים כי החסם של הופדינג ושל צ'בישב משתפר וקטן ככל ש-m גדל. זאת כיוון שככל שמספר הדגימות עולה, כך נצפה שממוצע הדגימות יתקרב לתוחלת וההסתברות שהמרחק הזה יהיה גדול מאפסילון קטנה יותר ויותר והחסמים יכולים להבטיח גבול עליון על הטעות, אשר הוא קטן מ־1 ויש לו משמעות. בעבור ε קטנים במיוחד ניתן לראות כי גם עבור $m = 1000$ עדיין החסמים של הופדינג וצ'בישב לא יכולים לספק חסם טוב יותר מ־ 1. זה מאוד הגיוני כי ככל שאפסילון קטן כך קשה להבטיח שההסתברות שהמרחק בין הממוצע לתוחלת יהיה קטן מאותו ה־ε.

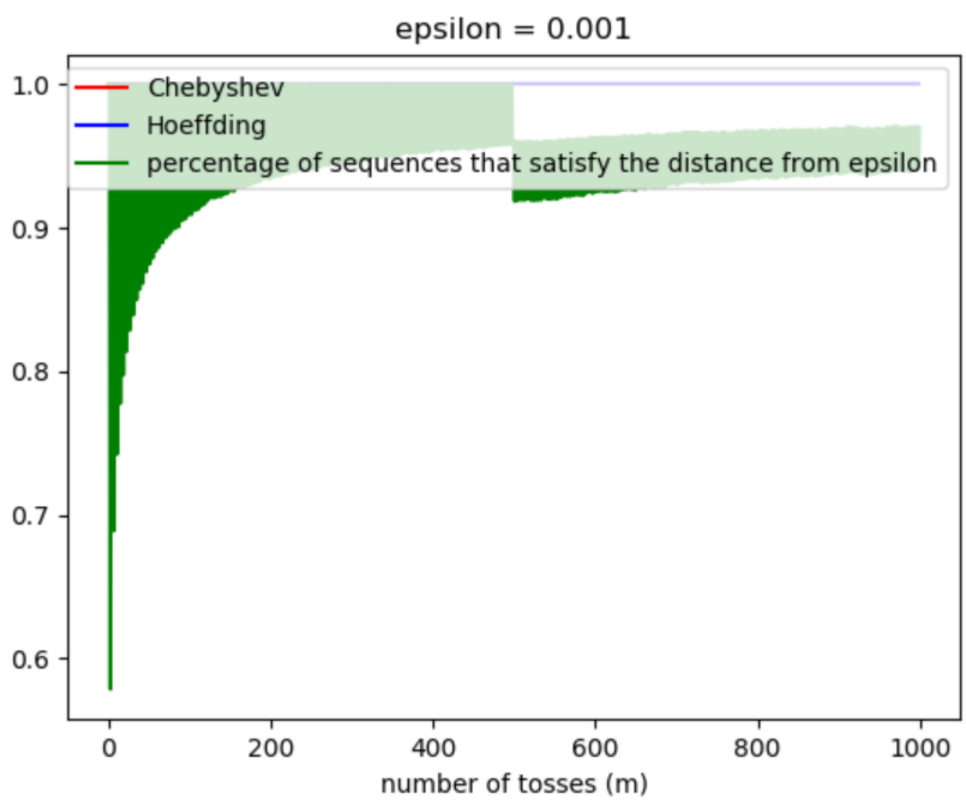
ג.




epsilon = 0.25
Chebyshev
Hoeffding
percentage of sequences that satisfy the distance from epsilon
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
number of tosses (m)


epsilon = 0.1
Chebyshev
Hoeffding
percentage of sequences that satisfy the distance from epsilon
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
number of tosses (m)





הגרפים הבאים מחשבים בנוסף בעבור כל m כמה מתוך הסבבים של הדגימות היו בממוצע רחוקים מהתוחלת של המשתנה המקרי של הטלת המטבע יותר מε. מה שנצפה לראות זה שככל שm גדל, כלומר אנו מסתכלים על יותר דגימות, לפי החוק של המספרים הגדולים נצפה שהממוצע יתקרב לתוחלת כך שיהיו כמה שפחות סבבים שלא מקיימים את התנאי (רחוקים מידי מε). נראה כי ככל

שε יותר קטן כך קשה לקיים את התנאי ואחוז הדגימות שבממוצע רחוקות מהתוחלת גדל. זה מאוד הגיוני כי ככל שאפסילון קטן כך הסבירות לטעות גדולה מאפסילון עולה.