Shelly Gamlielly

IML - Ex 1
Mathematical Background

Algebra Recap

1) Calculate the projection of $v=(1,2,3,4)$ on $w=(0,-1,1,2)$

ההטלה של וקטור v על וקטור w היא וקטור p באורך $\|v\|\cos\theta$ בכיוון של w:

$$p = \|v\|\cos\theta \frac{w}{\|w\|} = \frac{\langle w, v\rangle}{\|w\|^2}\bar{w}$$

$$= \frac{-2+3+8}{1+1+4}(0,-1,1,2) = \frac{\cancel{9}^{3}}{\cancel{6}_{2}}(0,-1,1,2)$$

$$= \left(0,-\frac{3}{2},\frac{3}{2},3\right)$$

2) Calculate the projection of $v=(1,2,3,4)$ on $w=(1,0,1,-1)$

$$p = \frac{\langle v,w\rangle}{\|w\|^2}\bar{w} = \frac{1+3-4}{1+1+1}(1,0,1,-1) = (0,0,0,0)$$

$= \bar{0}$ (וקטור האפס)

3) Prove the angle between two non-zero vectors $v,w\in\mathbb{R}^m$ is ± 90 iff $\langle v,w\rangle = 0$

$\Leftarrow$:

נתון כי $v,w\in\mathbb{R}^m$ אינם וקטור האפס והזווית ביניהם היא $\theta=\pm 90$

נרצה להראות כי $\langle v,w\rangle=0$

$$v=(v_1,\dots v_m) \qquad w=(w_1,\dots w_m)$$

$$\langle v,w\rangle = \sum_{i=1}^{m} v_i w_i$$

הראנו כי $\|v\|\,\|w\|\cos\theta = \langle v,w\rangle$

נבחין כי $\cos(-90)=\cos(90)=0$ ולכן $\langle v,w\rangle=0$

$\Rightarrow$:

נתון כי $\langle v,w\rangle=0$ עבור $v,w\in\mathbb{R}^m$

נרצה להראות כי הזווית בין שני הוקטורים (שאינם 0) היא ± 90.

הראנו כי $\langle v,w\rangle = \|v\|\,\|w\|\cos\theta$

נתון כי $\langle v,w\rangle=0$ ולכן $\|v\|\,\|w\|\cos\theta=0$

מכיוון ש- v ו- w אינם וקטור האפס נסיק כי $\cos\theta=0$

ולכן $\theta=\pm 90$

$$\|v-w\|^2 = \langle v-w, v-w\rangle = \|v\|^2+\|w\|^2-2\langle v,w\rangle \quad *$$

$$\|v-w\|^2 = \|v\|^2+\|w\|^2-2\|v\|\,\|w\|\cos\theta$$

$$\Downarrow$$

$$-\cancel{2}\|v\|\,\|w\|\cos\theta = -\cancel{2}\langle v,w\rangle$$

4) Prove that Orthonormal matrices are isometric transformations. That is, let $T: V\to W$ be some linear transformation and A the corresponding matrix. Then if A is orthonormal then $\forall x\in V$ $\|Ax\|_2 = \|x\|_2$

יהי $T: V \to W$ טרנספורמציה ליניארית ו-A המטריצה המתאימה

מכיוון שלכל $u,v \in V, c \in \mathbb{R}$ מתקיים
$$T(u+v) = T(u) + T(v)$$
$$T(cu) = cT(u)$$

בנוסף A היא אורתונורמלית כאשר A היא מטריצה ריבועית
שהעמודות שלה הן וקטורי יחידה שאורתוגונלים זה לזה
והשורות שלה הן וקטורי יחידה שאורתוגונלים זה לזה

יהי $x \in V$ ונזכיר כי
$$\|x\|_2 = \left(\sum_{i=1}^{m} |x_i|^2\right)^{\frac{1}{2}}$$
I_m וקטור היחידה

A מטריצה אורתוגונלית $\Leftarrow AA^T = A^TA = I$

$$\|Ax\|_2^2 = \langle Ax, Ax \rangle = \langle \sum_j a_{ij}x_j, \sum_j a_{ij}x_j \rangle$$
$$= \langle A^TAx, x \rangle$$
A אורתוגונלית ולכן $A^TA = I$
$$= \langle x, x \rangle = \|x\|^2$$

או מתוך אחר:

$$\boxed{\begin{aligned} \|Ax\|_2^2 &= \langle Ax, Ax \rangle = (Ax)^T Ax = x^TA^TAx \\ &= x^T I x = x^Tx = \langle x, x \rangle = \|x\|_2^2 \end{aligned}}$$
A אורתוגונלית

SVD

5) Assume A is invertible. Write a formula for the inverse of A using only the matrices U, D, V where UDV^T is an SVD decomposition of A.
Explain why knowing the SVD decomposition of matrix is useful

עפ"י המשפט מההרצאה: $A \in M_{m \times d}(\mathbb{R})$ מטריצה כלשהי
$$\left\{\begin{array}{l} U \in M_{m \times m}(\mathbb{R}) \\ V \in M_{d \times d}(\mathbb{R}) \end{array}\right.$$ מטריצות אורתונורמליות $\Longrightarrow U^TU = UU^T = I$, $V^TV = VV^T = I$

$D \in M_{m \times d}(\mathbb{R})$ מטריצה אלכסונית עם ערכים אי שליליים

$$A = UDV^T$$

ולכן
$$A^{-1} = VD^{-1}U^T$$

$$D^{-1} = \begin{pmatrix} \frac{1}{s_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{s_n} \end{pmatrix}$$

כאשר
שכן D מטריצה אלכסונית
ולכן הערכים על האלכסון של D^{-1}
הם s_i^{-1} לכל i

נזכיר כי מכיוון ש-U ו-V הן אורתוגונליות מתקיים כי $V^T = V^{-1}$, $U^T = U^{-1}$

נוכיח זאת פורמלית:
$$A^{-1}A = (VD^{-1}U^T)(UDV^T)$$
$$= VD^{-1}(U^TU)DV^T$$
U אורתוגונלית $= VD^{-1}IDV^T$
אלכסונית $= VD^{-1}DV^T$
$$= VIV^T$$
$$= VV^T = I$$
V אורתוגונלית

$\Longrightarrow A^{-1} = VD^{-1}U^T$ היא ההופכית של A

(5) כיוון SVD מאפשר לנו לרשום כל מטריצה A כ- USV^T ורואים מעכשיו אומר לנו שאנו יכולים לחשוב על הפעולה של A על $\tilde{x}$ כשלושה צעדים:

- סיבוב – הכפל ב-V^T שלא משנה את האורך של $\tilde{x}$
- מתיחה – מכיוון הכפל ב-D כל x_i (הרכיב ה-i) נמתח פי s_i (האיברים על האלכסון של D), הערכים הסינגולריים
- סיבוב – כפל ב-U

$A\tilde{x} = (UDV^T)\tilde{x}$ אלו בעצם 3 טרנספורמציות ליניאריות

וייראה כך

$$\tilde{x} \xrightarrow[\text{rotation}]{V^T} V^T\tilde{x} \xrightarrow[\text{Streching}]{D} DV^T\tilde{x} \xrightarrow[\text{rotation}]{U} UDV^T\tilde{x}$$

כמו כן, נוכל למצוא מיפוי הפוך ע"י כך שנעשה פעולה הפוכה של כל אחת מהצעדים שהצגנו. כלומר בהתחלה נכפיל ב-U^T (זוהי הפעולה ההפוכה של הסיבוב), ואז נבטל את המתיחה ע"י כפל ב-$\frac{1}{s_i}$ לאורך הצירים, ולסיום נבטל את הסיבוב ע"י כפל ב-V.

$$\tilde{x} \xleftarrow{V} V^T\tilde{x} \xleftarrow{D^{-1}} DV^T\tilde{x} \xleftarrow{U^T} UDV^T\tilde{x}$$

$$\tilde{x} = A^{-1}A\tilde{x} = (VD^{-1}U^T)(UDV^T)\tilde{x}$$

נבחין כי הפירוק של המטריצה A לרכיבים היה שימושי שכן אף קל לראות אם המטריצה ההפוכה לא קיימת אם אחד הערכים הסינגולריים $s_i=0$, אז D^{-1} לא קיימת מכיוון שהערכים על האלכסון יהיו $\frac{1}{s_i}=\frac{1}{0}\to\infty$.

נסכם הכל אם למטריצה A יש ערכים סינגולריים שהם 0 אז הכפלה ב-A תגרום לנו לאבד מידע. הרי שהיא לוקחת רכיב מסוים ומכפילה ב-0, ואז לא ניתן יהיה לשחזר את המידע, ואז אין דרך למצוא מיפוי הפוך ל-$A\tilde{x}$ כדי למצוא את $\tilde{x}$ המקורי. האופציה היחידה היא לשחזר את הרכיבים של $\tilde{x}$ שלא אובדו כתוצאה מההכפלה ב-0.

הסבר מדוע SVD שימושי

(6) Find an SVD of $C = \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix}$

- calculate C^TC

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

- Deduce V and D

$$\det(C^TC - \lambda I) = \begin{vmatrix} 26-\lambda & 18 \\ 18 & 74-\lambda \end{vmatrix} = (26-\lambda)(74-\lambda) - 324$$

$$= 1924 - 100\lambda + \lambda^2 - 324$$
$$= \lambda^2 - 100\lambda + 1600$$
$$= (\lambda - 20)(\lambda - 80)$$

$\lambda_1 = \underline{\underline{20}},\ \lambda_2 = \underline{\underline{80}}$ הע״ע הם $\Leftarrow$

נעבור (נמצא) את הווקטורים העצמיים:

$$C^TC - 20I = \begin{pmatrix} 6 & 18 \\ 18 & 54 \end{pmatrix} = 6 \begin{pmatrix} 1 & 3 \\ 3 & 9 \end{pmatrix}$$

$$\Rightarrow 6 \begin{pmatrix} 1 & 3 \\ 3 & 9 \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{matrix} V_1 + 3V_2 = 0 \\ 3V_1 + 9V_2 = 0 \end{matrix} \Rightarrow V_1 = -3V_2 \Rightarrow V_{\lambda_1} = \begin{pmatrix} -3 \\ 1 \end{pmatrix} \frac{1}{\sqrt{10}}$$

נדאג שיהיה וקטור יחידה (נורמל)

$$C^TC - 80I = \begin{pmatrix} -54 & 18 \\ 18 & -6 \end{pmatrix} = 6 \begin{pmatrix} -9 & 3 \\ 3 & -1 \end{pmatrix}$$

$$\Rightarrow 6 \begin{pmatrix} -9 & 3 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{matrix} -9V_1 + 3V_2 = 0 \\ 3V_1 - V_2 = 0 \end{matrix} \Rightarrow V_2 = 3V_1 \Rightarrow V_{\lambda_2} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \frac{1}{\sqrt{10}}$$

$$V = \begin{pmatrix} | & | \\ V_{\lambda_1} & V_{\lambda_2} \\ | & | \end{pmatrix} = \begin{pmatrix} \frac{-3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{pmatrix}$$ לפרק

$$D = \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} = \begin{pmatrix} 2\sqrt{5} & 0 \\ 0 & 4\sqrt{5} \end{pmatrix}$$

ואכן D מטריצה אלכסונית עם ערכים סינגולרים על האלכסון
V מטריצה אורתוגונלית שמורכבת מהווקטורים העצמיים הנורמליים

- Find U using $CV = UD$

$$CV = \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix} \frac{1}{\sqrt{10}} \begin{pmatrix} -3 & 1 \\ 1 & 3 \end{pmatrix} = \frac{1}{\sqrt{10}} \begin{pmatrix} -10 & 20 \\ 10 & 20 \end{pmatrix}$$

$$= \sqrt{10} \begin{pmatrix} -1 & 2 \\ 1 & 2 \end{pmatrix} = \underbrace{\begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{= U} \underbrace{\begin{pmatrix} 2\sqrt{5} & 0 \\ 0 & 4\sqrt{5} \end{pmatrix}}_{= D}$$

ומצאנו את U מטריצה אורתוגונלית   הבנו

7) $C_0 = A^TA$, $A \in M_{(n \times m)}(\mathbb{R})$

$(v_1, \dots, v_n)$ ו"ע שמתאימים לע"ע $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$

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

$$\lim_{k \to \infty} b_k = \pm v_1 \qquad \boxed{B}$$

נראה כי

$$C_0 = A^TA = (\underset{\substack{\uparrow \\ A \text{ של } SVD}}{U\Sigma V^T})^T (U \Sigma V^T)$$

$$= V\Sigma^T U^T U \Sigma V^T \underset{\substack{\uparrow \\ U \text{ אורתוגונלית}}}{=} V\Sigma^T\Sigma V^T \underset{\substack{\uparrow \\ \Sigma \text{ אלכסונית}}}{=} V\Sigma^2V^T$$

$\Leftarrow$ כל הע"ע של המטריצה $\Sigma' = \Sigma^2$ מפירוק EVD גדולים מ-0
שכן מפירוק SVD של A יש רק ערכים אי שליליים
באלכסון הראשי. ועל כן Σ' יהיו רק ערכים חיוביים

נראה כי לכל $k \in \mathbb{N}$ מתקיים
$$b_k = \frac{C_0^k b_0}{\|C_0^k b_0\|}$$

עבור $k=1$ עפ"י הגדרה
$$b_1 = \frac{C_0 b_0}{\|C_0 b_0\|} = \frac{C_0^1 b_0}{\|C_0^1 b_0\|}$$

נניח נכונות הטענה עבור k ונוכיח עבור $k+1$

$$b_{k+1} = \frac{C_0 b_k}{\|C_0 b_k\|} \underset{\substack{\uparrow \\ \text{מהנחת} \\ \text{האינדוקציה}}}{=} \frac{C_0 \cdot C_0^k b_0}{\|C_0 C_0^k b_0\|} = \frac{C_0^{k+1} b_0}{\|C_0^{k+1} b_0\|}$$

כנדרש.

כעת

$$C_0^k b_0 = (V\Sigma^2V^T)^k b_0 = V\Sigma^{2k}V^T b_0$$

$$\Sigma^{2k} \underset{\substack{\uparrow \\ \text{מטריצה אלכסונית}}}{=} \begin{pmatrix} \lambda_1^{2k} & & 0 \\ & \ddots & \\ 0 & & \lambda_n^{2k} \end{pmatrix} = \lambda_1^{2k} \begin{pmatrix} 1 & & & \\ & \frac{\lambda_2^{2k}}{\lambda_1^{2k}} & & \\ & & \ddots & \\ & & & \frac{\lambda_n^{2k}}{\lambda_1^{2k}} \end{pmatrix}$$

נסמן $2k = m$

$$= \lambda_1^m \begin{pmatrix} 1 & & & \\ & \left(\frac{\lambda_2}{\lambda_1}\right)^m & & \\ & & \ddots & \\ & & & \left(\frac{\lambda_n}{\lambda_1}\right)^m \end{pmatrix} \underset{\substack{\uparrow \\ \text{לכל } j \geq 2 \\ \text{מתקיים} \\ \frac{\lambda_j}{\lambda_1} < 1 \\ \left(\frac{\lambda_j}{\lambda_1}\right)^k \xrightarrow{k \to \infty} 0}}{=}$$

$$\xrightarrow{k \to \infty} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix} = Q \quad \text{נסמן}$$

$$\Rightarrow b_k = \frac{C_0^k b_0}{\|C_0^k b_0\|} = \frac{V\Sigma^{2k}V^T b_0}{\|V\Sigma^{2k}V^T b_0\|} \xrightarrow{k \to \infty} \frac{VQV^T b_0}{\|VQV^T b_0\|}$$

$$= \frac{v_1 v_1^T b_0}{\|v_1 v_1^T b_0\|} = \frac{v_1 v_1^T \sum_{i=1} a_i v_i}{\|v_1 v_1^T \sum a_i v_i\|} = \frac{a_1 v_1}{\|a_1 v_1\|} \underset{\substack{\uparrow \\ v_1 \text{ מנורמל}}}{=} \frac{a_1}{\|a_1\|} v_1 \underset{a_1 \neq 0}{=} \pm v_1$$

Multivariate Calculus

8) Let $x \in \mathbb{R}^n$ be a fixed v and $W \in \mathbb{R}^{n \times n}$ a fixed orthogonal matrix. Calculate the Jacobian of $f: \mathbb{R}^n \to \mathbb{R}^n$;

$$f(\sigma) = W \underset{\parallel}{diag(\sigma)} W^T x$$

$$diag(\sigma)_{ij} = \begin{cases} \sigma_i & i=j \\ 0 & i \neq j \end{cases}$$

$$\begin{pmatrix} W_{11} & & W_{n1} \\ W_{12} & W_{22} \dots & W_{n2} \\ W_{1n} & \dots & W_{nn} \end{pmatrix} \begin{pmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & & \ddots \ \sigma_n \end{pmatrix} \begin{pmatrix} W_{11} & \dots & W_{1n} \\ \vdots & & \\ W_{n1} & \dots & W_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$= \begin{pmatrix} W_{11} & \dots & W_{n1} \\ \vdots & & \\ W_{1n} & & W_{nn} \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots \ \sigma_n \end{pmatrix} \begin{pmatrix} W_{11}x_1 & \dots & W_{1n}x_n \\ \vdots & & \\ W_{n1}x_1 & \dots & W_{nn}x_n \end{pmatrix}$$

$$= \begin{pmatrix} W_{11} & \dots & W_{n1} \\ \vdots & & \\ W_{1n} & & W_{nn} \end{pmatrix} \begin{pmatrix} W_{11}x_1\sigma_1 & \dots & W_{1n}x_n\sigma_1 \\ \vdots & & \\ W_{n1}x_1\sigma_n & \dots & W_{nn}x_n\sigma_n \end{pmatrix}$$

$$= \begin{pmatrix} W_{11}^2 x_1 \sigma_1 + \dots + W_{n1}^2 x_1 \sigma_n, & \dots, & W_{11}W_{1n}x_n\sigma_1 + \dots & W_{n1}W_{nn}x_n\sigma_n \\ \vdots & & & \vdots \\ W_{1n}W_{11}x_1\sigma_1 + \dots + W_{nn}W_{n1}x_1\sigma_n & \dots & W_{1n}^2 x_n \sigma_1 + \dots & + W_{nn}^2 x_n \sigma_n \end{pmatrix}$$

$$f(\sigma) = \sum_{i=1}^{n} \sigma_i w_i w_i^T x \Leftarrow$$

$$\Rightarrow \frac{\partial f_j(\sigma)}{\partial \sigma_i} = [w_i w_i^T x]_j$$

$\Rightarrow$ $w_i \langle w_i, x \rangle$ נקבל היעקוביאן של ה-i העמודה

$$\Rightarrow J_f = W\, diag(W^T x) = \sum_{i=1} w_i w_i^T x$$

$diag(W^T x)_{ij} = \begin{cases} w_i^T x & i=j \\ 0 & else \end{cases}$ כאשר

9) Calculate the gradient of

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

$$= \frac{1}{2} \sum_{i=1}^{d} (f_i(\tilde{\sigma}) - y_i)^2$$

$$\frac{\partial h(\tilde{\sigma})}{\partial \tilde{\sigma}_i} = \frac{1}{2} \cdot 2(f_i(\tilde{\sigma}) - y_i) \cdot \frac{\partial f_i(\tilde{\sigma})}{\partial \tilde{\sigma}_i}$$

$$\nabla h(\tilde{\sigma}) = (f(\tilde{\sigma}) - y)^T \nabla f(\tilde{\sigma})$$

$$\begin{cases} h(\tilde{\sigma}) = \frac{1}{2} \sum_{i=1}^{d} f_i^2(\tilde{\sigma}) - 2 f_i(\tilde{\sigma}) y_i + y_i^2 \\ \frac{\partial h(\tilde{\sigma})}{\partial \tilde{\sigma}_i} = \frac{1}{2} \left((2 f_i(\tilde{\sigma}) - 2y_i) \frac{\partial f_i(\tilde{\sigma})}{\partial \tilde{\sigma}_i} \right) \\ \quad = (f_i(\tilde{\sigma}) - y_i) \frac{\partial f_i(\tilde{\sigma})}{\partial \tilde{\sigma}_i} \end{cases}$$

$$\nabla h(\tilde{\sigma}) = (f(\tilde{\sigma}) - y)^T \nabla f(\tilde{\sigma})$$ משוואה

10) Calculate the Jacobian of the softmax function:

$$g(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \qquad g(z): \mathbb{R}^K \to \mathbb{R}^K$$

$$\forall j \in 1, \ldots K$$

softmax מקבלת כקלט וקטור ומחזירה וקטור ולכן נחשב את הנגזרות החלקיות $\frac{\partial g_i}{\partial z_j}$ של הפלט ה-i המתאים לקלט ה-j

$$= \frac{\partial}{\partial z_j}\left(\frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}\right)$$

נסמן

$$h_i = \sum_{k=1}^{K} e^{a_k}, \quad g_i = e^{z_i}$$

I) עבור $i=j$

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

$$= \frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}} - \frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}} \cdot \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} = g_i (1 - g_j)$$

II) עבור $i \neq j$

$$\frac{\partial}{\partial z_j}\left(\frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}}\right) = \frac{0 \cdot \sum_{i=1}^{K} e^{z_k} - e^{z_i} \cdot e^{z_j}}{\left(\sum_{k=1}^{K} e^{z_k}\right)^2}$$

$$= -\frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}} \cdot \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} = -g_i \cdot g_j$$

$$\frac{\partial g_i}{\partial z_j} = \begin{cases} g_i(1-g_j) & \text{if } i=j \\ -g_i \cdot g_j & \text{else} \end{cases} \Leftarrow$$

$$= g_i(\delta_{ij} - g_j)$$

12) $C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ $\quad S = \begin{pmatrix} 0.1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 2 \end{pmatrix} = S^T$

$$\text{COV scale} = SCS^T = SIS^T = SS^T \underset{\substack{\uparrow \\ \text{אלכסונית}}}{=} S^2$$

$$= \begin{pmatrix} 0.1^2 & 0 & 0 \\ 0 & 0.5^2 & 0 \\ 0 & 0 & 2^2 \end{pmatrix} = \begin{pmatrix} 0.01 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

$$\hat{C} = \frac{X_{centered} \cdot X^T_{centered}}{n-1} \begin{pmatrix} 0.995 & 0.005 & -0.008 \\ 0.005 & 0.998 & 0.002 \\ -0.008 & 0.002 & 1.001 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\hat{C}_{scale} = \frac{SX \cdot (SX)^T}{n-1} = \frac{SXX^TS^T}{n-1} = S\hat{C}S^T$$

13) $$\hat{C}_{orth+scale} = \frac{HSX(HSX)^T}{n-1} = \frac{HSXX^TS^TH^T}{n-1}$$

$$= H\hat{C}_{scale}H^T$$

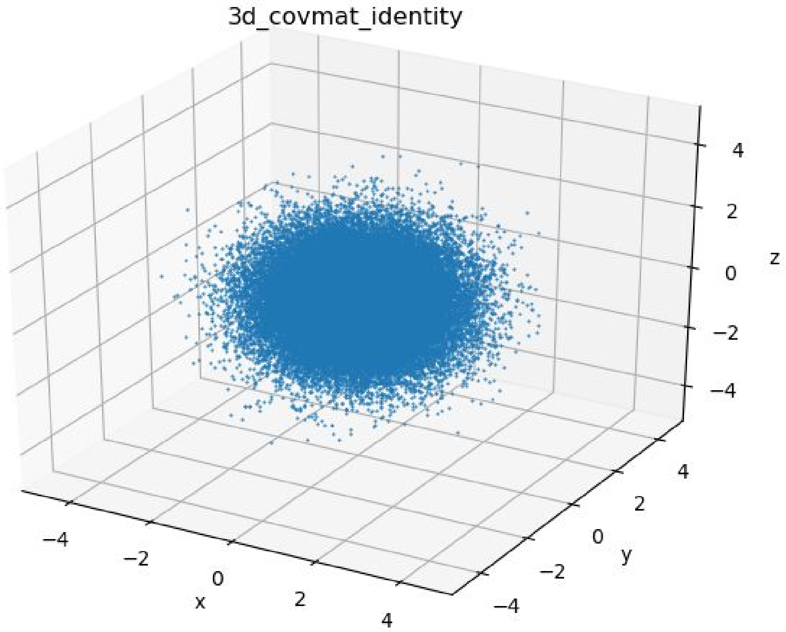
$$= HS\hat{C}S^TH^T$$

$$= HS^2H^T$$

עבור $C = I_{3\times3}$ ו-S אלכסונית

אם נציב את $\hat{C}$ נקבל תוצאות שקרובות מאוד לתוצאה
הנומריות

ex1 IML



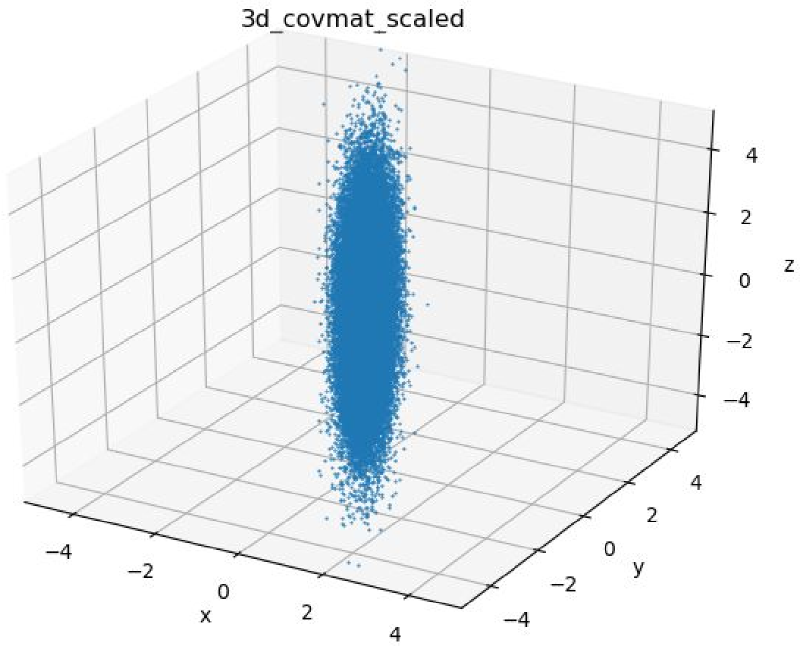
Q11:

Estimated covariance matrix of data

[[1.01499079e+00 7.00568292e-04 -1.50725425e-03]
 [7.00568292e-04 9.99304460e-01 -9.62534985e-03]
 [-1.50725425e-03 -9.62534985e-03 9.92500350e-01]]

כאשר מטריצת הקווריאנס היא מטריצת הזהות כך שקבלנו תוצאה שקרובה למציאות.

Q12:



ex1 IML

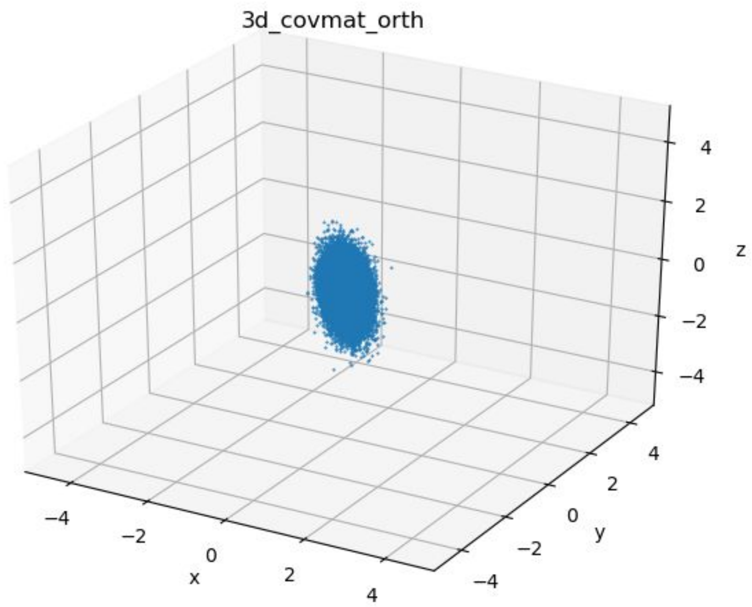
scaled covariance matrix

```
[[0.01 0.   0.  ]
 [0.   0.25 0.  ]
 [0.   0.   4.  ]]
```

scaled estimated covariance matrix

```
[[ 1.01499079e-02  3.50284146e-05 -3.01450851e-04]
 [ 3.50284146e-05  2.49826115e-01 -9.62534985e-03]
 [-3.01450851e-04 -9.62534985e-03  3.97000140e+00]]
```

Q13:



Orthogonal Matrix

```
[[-0.82500786  0.18560663  0.53377168]
 [-0.56104367 -0.38226727 -0.73423548]
 [ 0.06776447 -0.90521926  0.41950692]]
```

scaled orthogonal covariance matrix

```
[[ 1.15506765 -1.58076559  0.85312092]
 [-1.58076559  2.19608671 -1.14593873]
 [ 0.85312092 -1.14593873  0.90884564]]
```

scaled orthogonal estimated covariance matrix

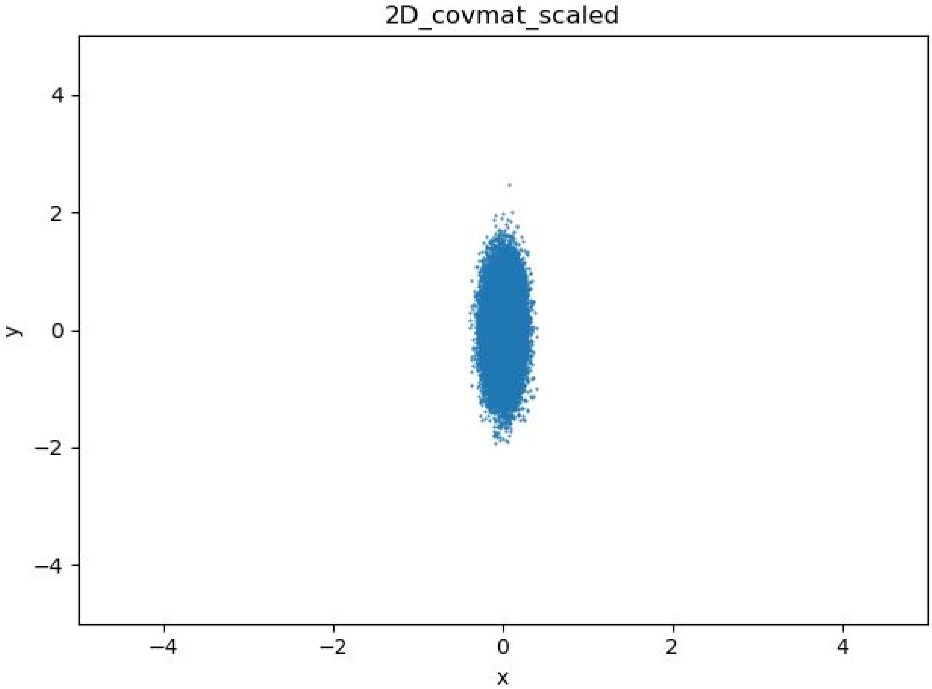
```
[[ 1.1449643  -1.5657362   0.85044579]
 [-1.5657362   2.17429968 -1.14151557]
```

ex1 IML

[0.85044579 -1.14151557 0.91071344]]

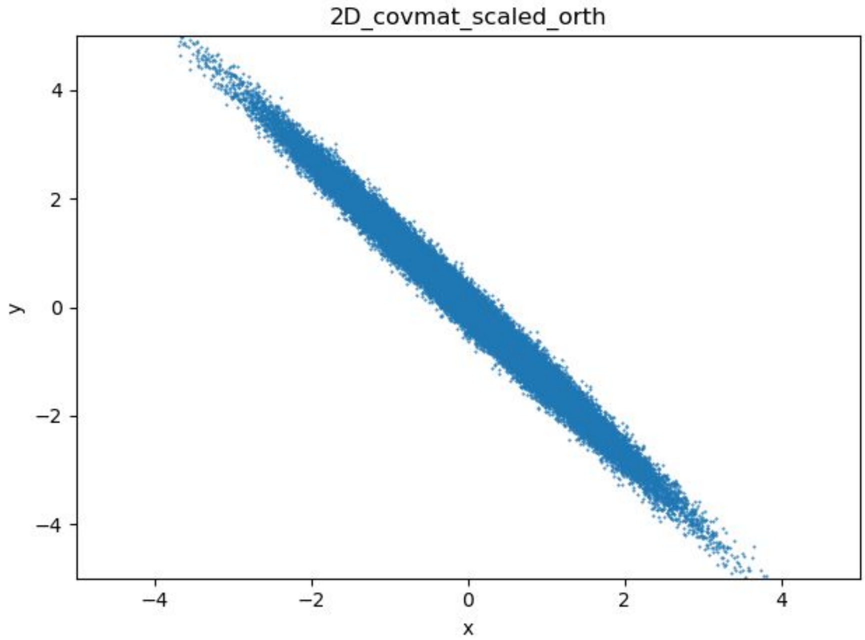
14. Plot the projection of the data to the x, y axes. What does it look like?

האליפסה התקבלה כתוצאה בטרנספורמציה של מתיחה.
ניתן לראות שמדובר באליפסה שהצירים שלה מיושרים עם הצירים x,y - כלומר הרכיבים בלתי מתואמים.



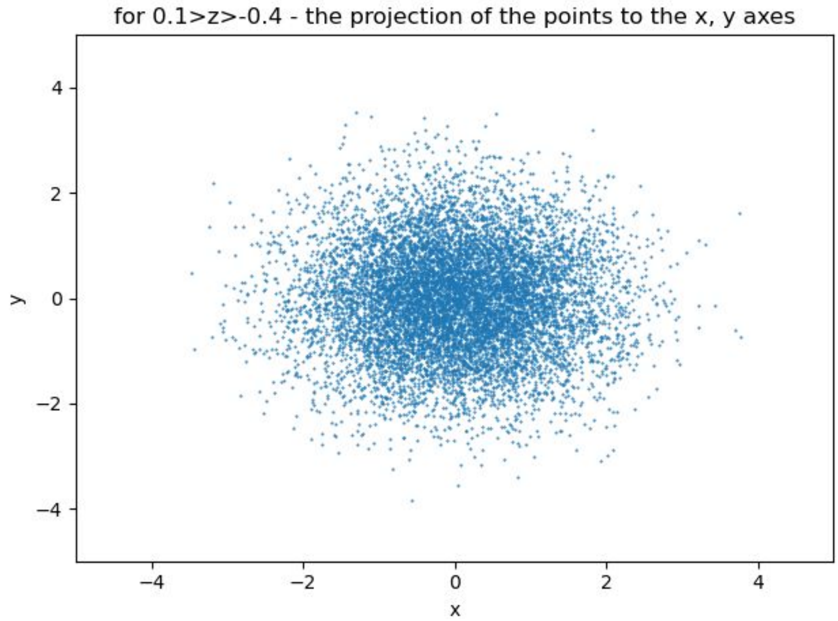
האליפסה התקבלה כתוצאה בטרנספורמציה של מתיחה וכפל במטריצה אורתוגונלית.
ניתן לראות שמדובר באליפסה שהצירים שלה אינם מיושרים עם הצירים x,y - כלומר הרכיבים מתואמים.

ex1 IML



15. Only for points where 0.1 > z >- 0.4: Plot the projection of the points to the x, y axes. What does it look like?

אנחנו רואים מעגל שציריו מיושרים לפי הצירים x,y ומרכזו בראשית.



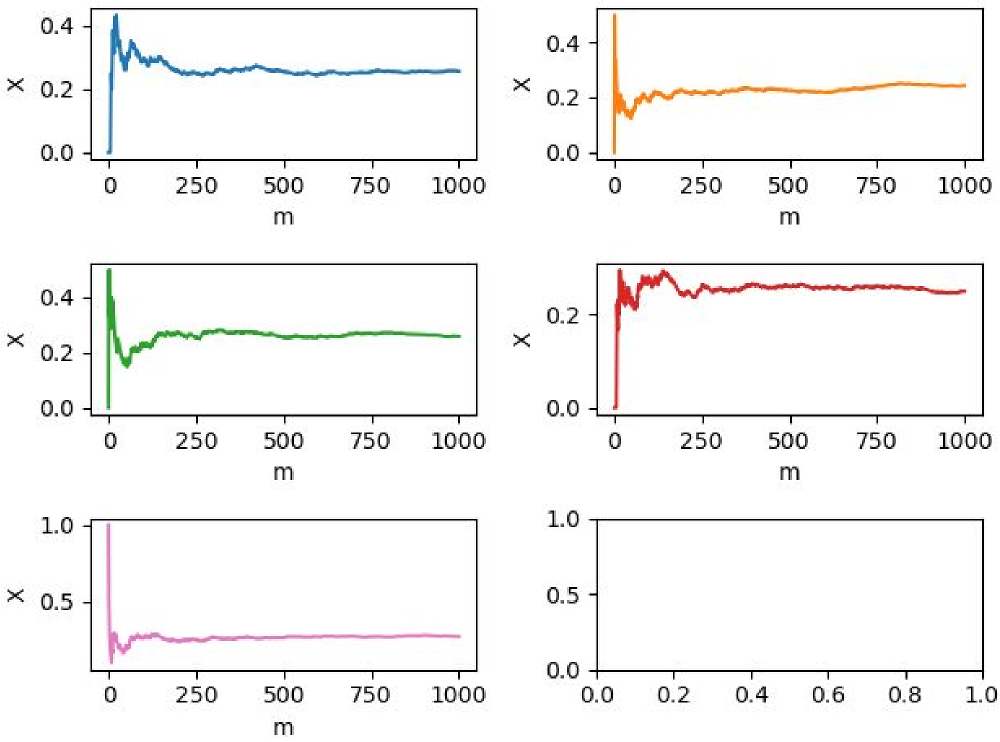
16.

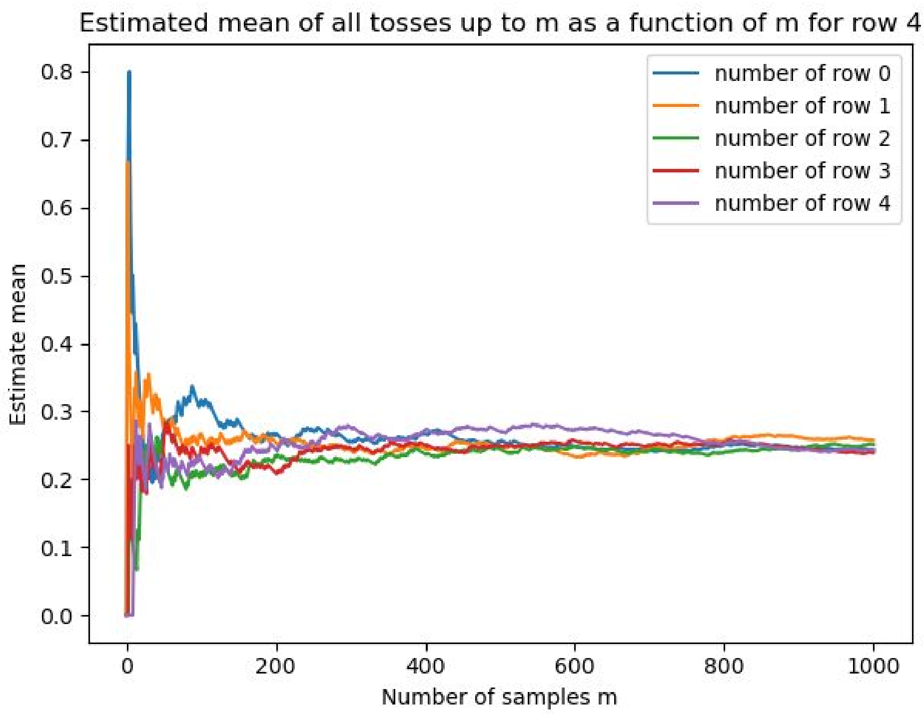
a. Plot the estimate mean of all tosses up to m with 5 plots (each row in a different color).

ככל שמספר הדגימות גדול יותר כך המשערך יהיה יותר מדוייק.

ex1 IML

עבור מ"מ שמתפלג בינומית התוחלת היא np במקרה שלנו p=0.25 n=1 ולכן כמצופה כאשר מספר ההטלות גבוה נתקרב לערך של התוחלת 0.25.



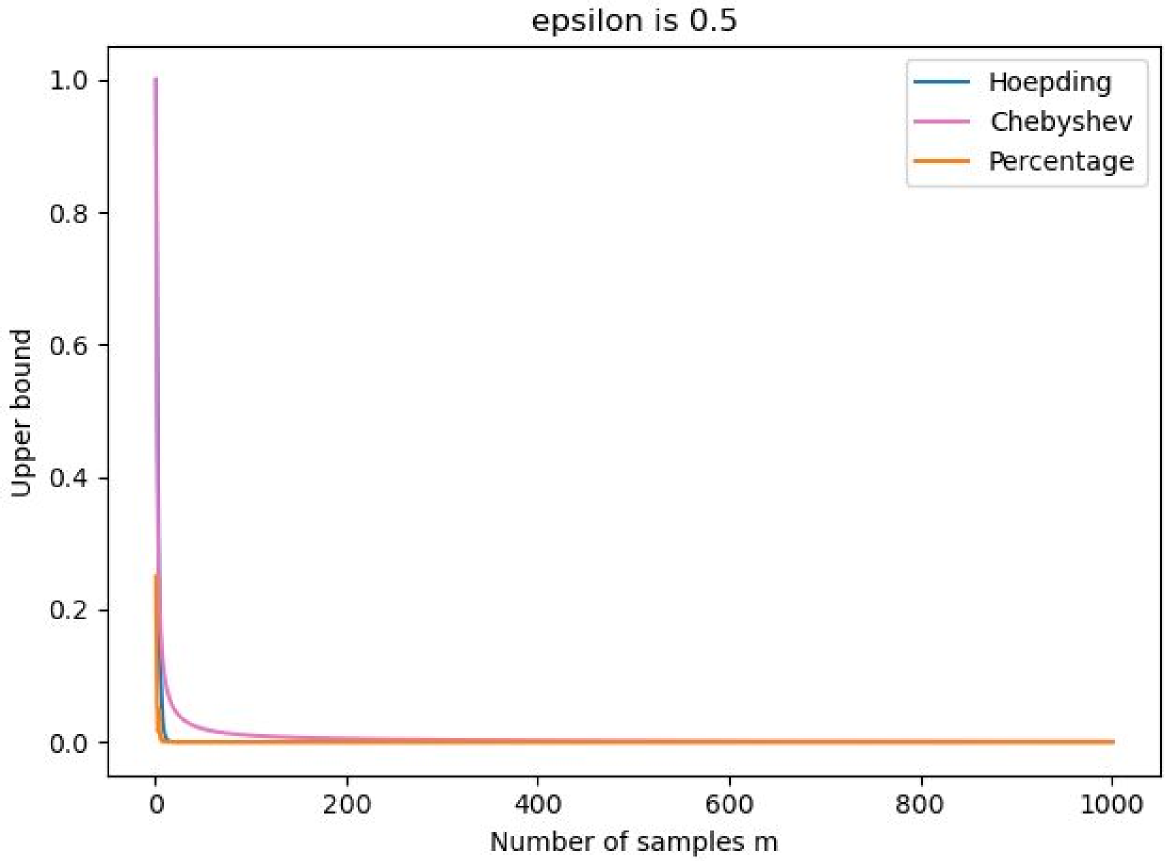


b. For each bound (Chebyshev and Hoefding seen in class) and for each epsilon ,plot the upper bound as a function of m.

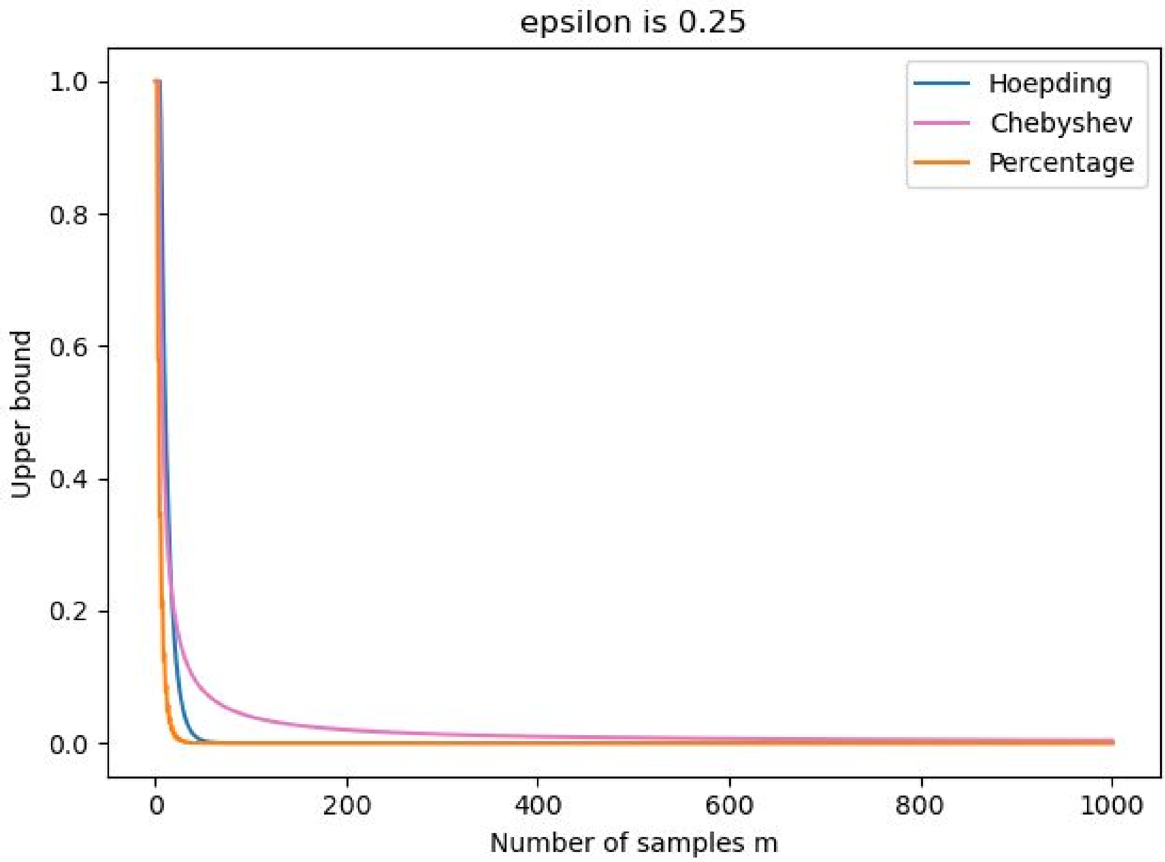
c.Plot the percentage of sequences that satisfy the equation as a function of m. What are you expecting to see in these plots?

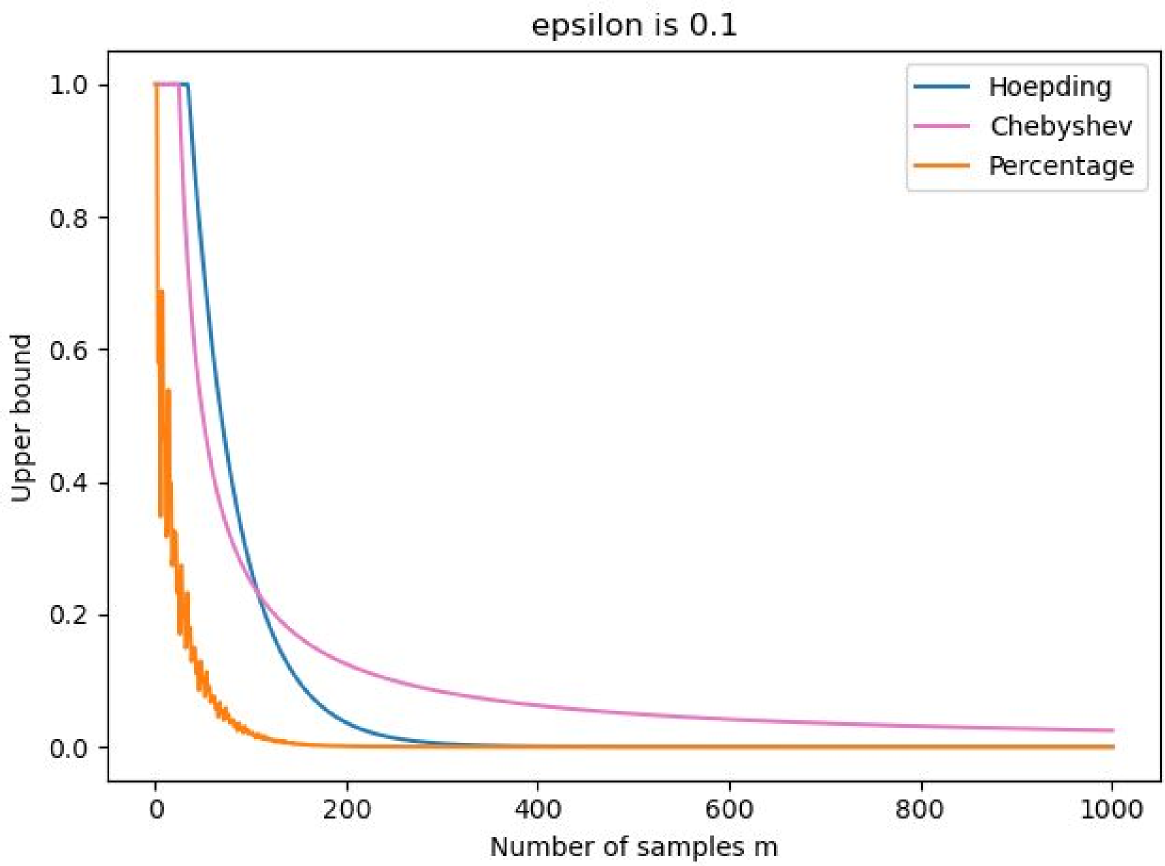
ex1 IML

באי שייויון בודקים האם המרחק בין התוחלת המשוערכת לבין התוחלת האמיתית גדול מאפסילון.
התבקשנו למצוא את כמות השורות שמספקות את האי שיויון חלקי כמות השורות בסך הכל - כלומר אחוז השורות שהמשערך טעה מעבר לטעות שהיינו מוכנים לקבל.
ניתן להבחין כי ככל שכמות הדגימות גדולה יותר כך אחוז השורות ה"רעות" קטן, וזה אכן מתיישב עם המציאות.
בנוסף לכך אפסילון משפיע על הערך של החסמים של צ'ביצ'ב והופדינג על ההסתברות שהמרחק בין הממוצע המשוערך לבין התוחלת גדול מאפסחלון. ככל שאפסילון קטן יותר (אנו דורשים דיוק רב יותר) הערך של המכנה קטן יותר וכשמחלקים במספר קטן התוצאה גדלה. לכן עבור אפסילון מאוד קטן (כלומר אנו מרשים טעות מזערית) נקבל חסם פחות הדוק. ולמעשה ככל שהאפסילון גדול יותר כך נדרשות פחות דגימות כדי להגיע לאחוז מזערי של שורות "רעות". זה מסתדר גם מבחינת החסם (מתמטית) וגם מבחינת ההיגיון.

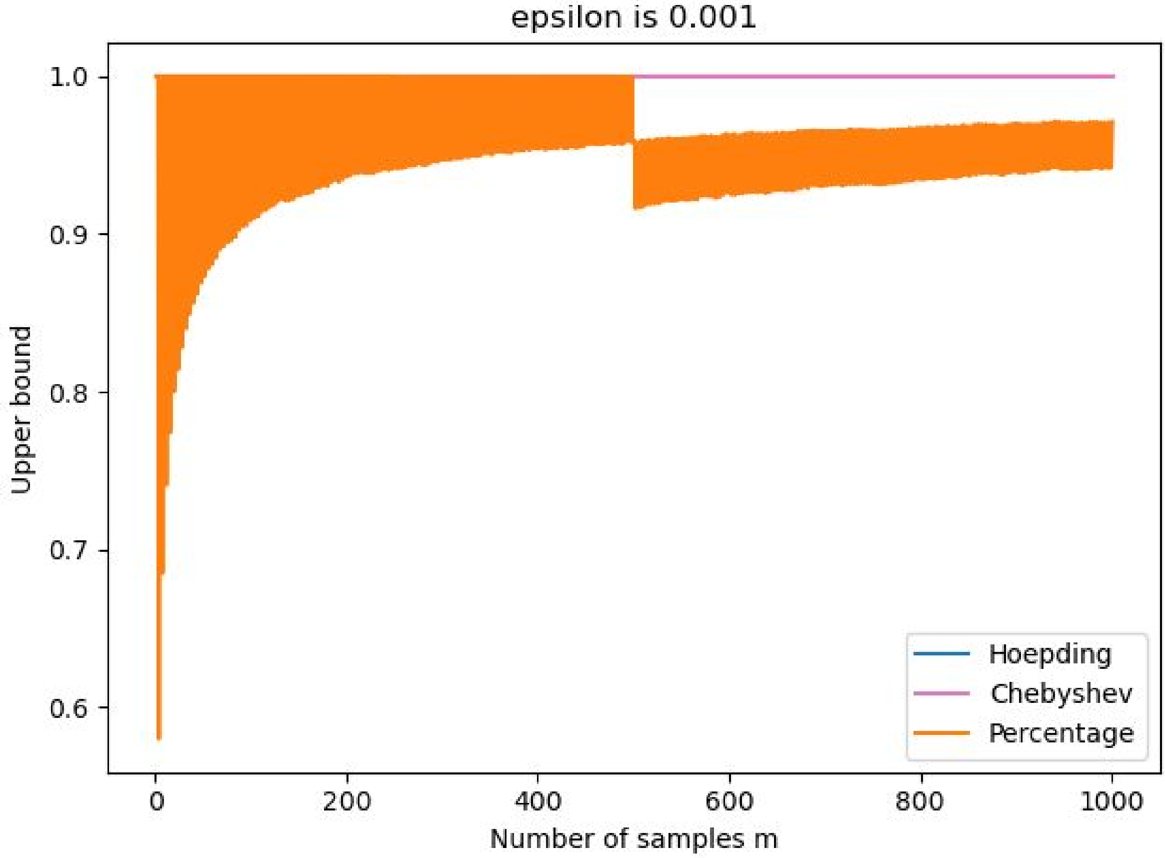


ex1 IML

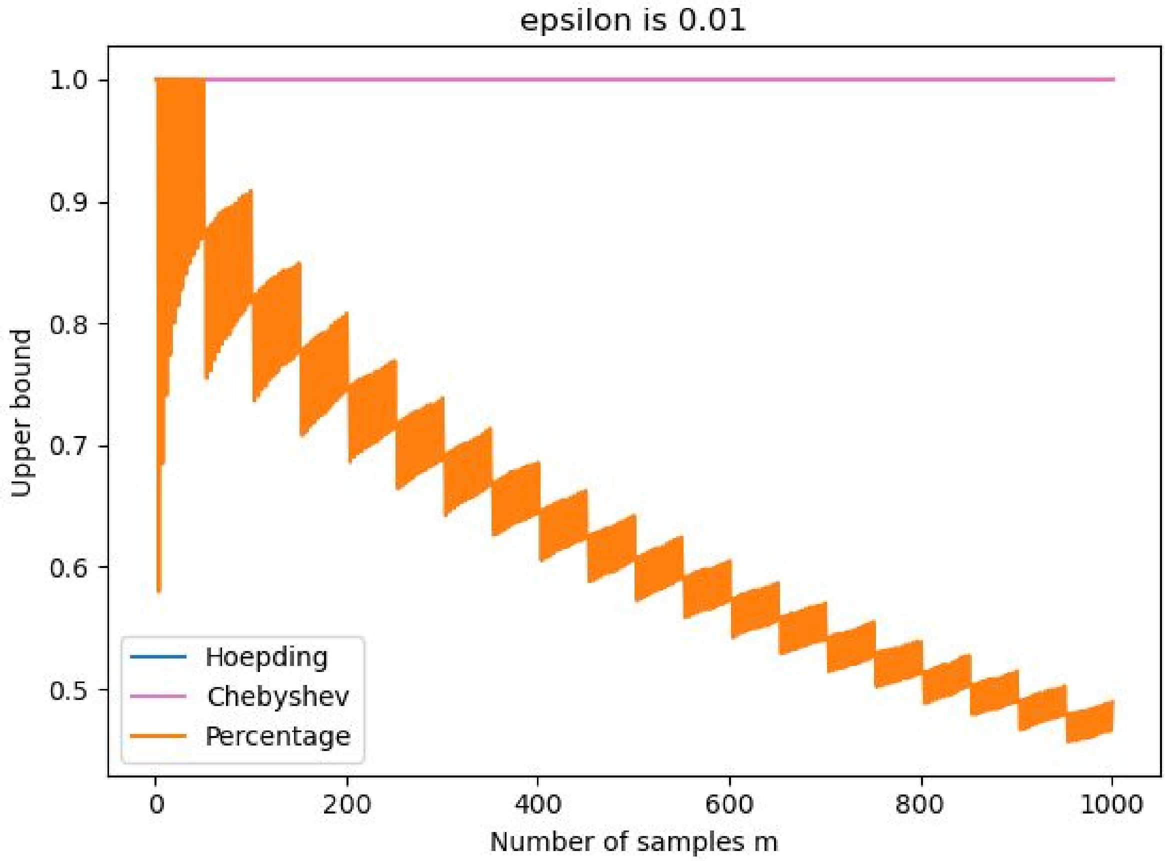

epsilon is 0.25
Hoepding
Chebyshev
Percentage
Upper bound
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
Number of samples m


epsilon is 0.1
Hoepding
Chebyshev
Percentage
Upper bound
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
Number of samples m

ex1 IML



epsilon is 0.001
1.0
0.9
0.8
0.7
0.6
Upper bound
0
200
400
600
800
1000
Number of samples m
Hoepding
Chebyshev
Percentage



epsilon is 0.01
1.0
0.9
0.8
0.7
0.6
0.5
Upper bound
0
200
400
600
800
1000
Number of samples m
Hoepding
Chebyshev
Percentage