תרגיל 1 - IML - מאי ביבי

ראינו בתרגול שההטלה של וקטור v על w היא

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

שאלה 1:

$$p=\frac{\langle v,w\rangle}{\|v\|^2}w=\frac{\begin{bmatrix}1\\2\\3\\4\end{bmatrix}\cdot\begin{bmatrix}0\\-1\\1\\2\end{bmatrix}}{\begin{bmatrix}0\\-1\\1\\2\end{bmatrix}^2}w=\frac{0-2+3+8}{0+1+1+4}w=\frac{9}{6}w$$

שאלה 2:

$$p=\frac{\langle v,w\rangle}{\|v\|^2}w=\frac{\begin{bmatrix}1\\2\\3\\4\end{bmatrix}\cdot\begin{bmatrix}1\\0\\1\\-1\end{bmatrix}}{\begin{bmatrix}1\\0\\1\\-1\end{bmatrix}^2}w=\frac{1+0+3-4}{1+0+1+1}w=\frac{0}{3}w=\begin{bmatrix}0\\0\\0\\0\end{bmatrix}$$

שאלה 3:

נרצה להראות שעבור שני וקטורים $v,w\in\mathbb{R}^m$ שונים מאפס מתקיים שהזווית ביניהם שווה ± 90 אם ורק אם $v^Tw=0$. הוכחה:
$\Longleftarrow$: נתון שהזווית θ בין v ל-w היא ± 90. בתרגול הוכחנו:

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

במקרה שלנו, $\theta=\pm 90$, כלומר $\cos\theta=0$ וכן המכפלה הפנימית היא המכפלה הפנימית הסטנדרטית, כלומר $\langle v,w\rangle=v^Tw$ ולכן מתקיים:

$$v^Tw=\langle v,w\rangle=\|v\|\,\|w\|\cos\theta=\|v\|\,\|w\|\cdot 0=0$$

$$\Longrightarrow v^Tw=0$$

$\Longleftarrow$: נתון $v^T w = 0$. אז כמו קודם:

$$\langle v, w\rangle = v^T w = 0 = \|v\| \, \|w\| \cos\theta$$

$$\Longrightarrow 0 = \cos\theta \Longrightarrow \theta = \pm 90$$

שאלה 4:

תהי $T : V \to W$ העתקה לינארית ו־A המטריצה המתאימה להעתקה. נוכיח כי אם A אורתוגונלית אז מתקיים $\forall x \in V \;\; \|Ax\| = \|x\|$.
הוכחה:
נתון A אורתוגונלית ולכן $A^T A = A^T A = I$. יהי $x \in V$.

$$\|x\| = \sqrt{x \cdot x} = \sqrt{x^T x}$$

$$\|Ax\| = \sqrt{Ax \cdot Ax} = \sqrt{(Ax)^T (Ax)} = \sqrt{x^T A^T (Ax)} = \sqrt{x^T \left(A^T A\right) x} = \sqrt{x^T I x} = \sqrt{x^T x} = \|x\|$$

כדרוש.

שאלה 5:

יהיה $A = U\Sigma V^T$ פירוק SVD של A מטריצה הפיכה. כיוון ש־U, V הן אורתוגונליות, $U^T = U^{-1}$ וכן $V^T = V^{-1}$. נשים לב שעבור $B = V\Sigma^{-1}U^T$ מתקיים

$$AB = \left(U\Sigma V^T\right)\left(V\Sigma^{-1}U^T\right) = \left(U\Sigma\right)\left(V^T V\right)\left(\Sigma^{-1}U^T\right) = \left(U\Sigma\right) I \left(\Sigma^{-1}U^T\right)$$

$$= U\left(\Sigma\Sigma^{-1}\right)U^T = UIU^T = UU^T = I$$

ולכן $A^{-1} = B = V\Sigma^{-1}U^T$. כיוון ש־ Σ אלכסונית קל לחשב את ההופכית שלה:

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{11}^{-1} & & 0 \\ & \ddots & \\ 0 & & \Sigma_{dd}^{-1} \end{bmatrix}$$

כיוון שאנו יודעים את SVD של A ניתן לחשב את A^{-1} ביעילות.

שאלה 6:

נמצא SVD של המטריצה

$$A = U\Sigma V^T = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix}$$

נשתמש בשתי המשוואות:

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

$$AV = U\Sigma$$

$V\Sigma^T\Sigma V^T$ זה לכסון של $A^T A$ ולכן אם נמצא את הוקטורים העצמיים של $A^T A$, נוכל למצוא את V, כי הם העמודות שלה.

$$A^T = \begin{bmatrix} 5 & -1 \\ 5 & 7 \end{bmatrix}$$

$$B=A^TA=\begin{bmatrix}5&-1\\5&7\end{bmatrix}\begin{bmatrix}5&5\\-1&7\end{bmatrix}=\begin{bmatrix}26&18\\18&74\end{bmatrix}$$

נמצא את הערכים העצמיים של A^TA:

$$0=\det(XI-B)=\det\left(\begin{bmatrix}X&0\\0&X\end{bmatrix}-\begin{bmatrix}26&18\\18&74\end{bmatrix}\right)=\det\left(\begin{bmatrix}X-26&-18\\-18&X-74\end{bmatrix}\right)$$

$$=(X-26)(X-74)-18^2=X^2-100X+1600=(X-80)(X-20)$$

$$\Longrightarrow X=80,20$$

נמצא ו"ע עבור ע"ע 20:

$$Bx=20x\Longrightarrow(B-20I)x=0$$

$$\begin{bmatrix}6&18\\18&54\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=0$$

נדרג:

$$\begin{bmatrix}6&18\\18&54\end{bmatrix}\to\begin{bmatrix}1&3\\18&54\end{bmatrix}\to\begin{bmatrix}1&3\\0&0\end{bmatrix}$$

$$x_1+3x_2=0\overset{x_1=t}{\Longrightarrow}x_2=-\frac{1}{3}t\Longrightarrow t\begin{bmatrix}1\\\frac{-1}{3}\end{bmatrix}$$

לכן עבור למשל $t=-3$, $\begin{bmatrix}-3\\1\end{bmatrix}$ הוא וקטור עצמי לע"ע 20. נרצה וקטור יחידה לכן נחלק בנורמה ונקבל את הוקטור העצמי $v_1=\begin{bmatrix}\frac{-3}{\sqrt{10}}\\\frac{1}{\sqrt{10}}\end{bmatrix}$.

נמצא וקטור עצמי שמתאים לערך עצמי 80:

$$Bx=80x\Longrightarrow(B-80I)x=0$$

$$\begin{bmatrix}-54&18\\18&-6\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=0$$

נדרג:

$$\begin{bmatrix}-54&18\\18&-6\end{bmatrix}\to\begin{bmatrix}-54&18\\0&0\end{bmatrix}\to\begin{bmatrix}1&\frac{-1}{3}\\0&0\end{bmatrix}$$

$$x_1-\frac{1}{3}x_2=0\overset{x_1=t}{\Longrightarrow}x_2=3t\Longrightarrow t\begin{bmatrix}1\\3\end{bmatrix}$$

לכן עבור למשל $t=3$, $\begin{bmatrix}1\\3\end{bmatrix}$ הוא וקטור עצמי לע"ע 80. נרצה וקטור יחידה לכן נחלק בנורמה ונקבל את הוקטור העצמי $v_2=\begin{bmatrix}\frac{1}{\sqrt{10}}\\\frac{3}{\sqrt{10}}\end{bmatrix}$. נקבל

$$V=\begin{bmatrix}v_1&v_2\end{bmatrix}=\begin{bmatrix}\frac{-3}{\sqrt{10}}&\frac{1}{\sqrt{10}}\\\frac{1}{\sqrt{10}}&\frac{3}{\sqrt{10}}\end{bmatrix}=\frac{1}{\sqrt{10}}\begin{bmatrix}-3&1\\1&3\end{bmatrix}$$

$$\Sigma^2=\begin{bmatrix}20&0\\0&80\end{bmatrix}\Longrightarrow\Sigma=\begin{bmatrix}\sqrt{20}&0\\0&\sqrt{80}\end{bmatrix}=\begin{bmatrix}\sqrt{2}\sqrt{10}&0\\0&2\sqrt{2}\sqrt{10}\end{bmatrix}=$$

כעת נמצא את U:

$$AV=U\Sigma\Longrightarrow AV\Sigma^{-1}=U\Sigma\Sigma^{-1}=U$$

$$\Sigma^{-1}=\begin{bmatrix}\frac{1}{\sqrt{2}\sqrt{10}} & 0\\ 0 & \frac{1}{2\sqrt{2}\sqrt{10}}\end{bmatrix}$$

$$AV\Sigma^{-1}=\frac{1}{\sqrt{10}}\begin{bmatrix}5 & 5\\ -1 & 7\end{bmatrix}\begin{bmatrix}-3 & 1\\ 1 & 3\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}\sqrt{10}} & 0\\ 0 & \frac{1}{2\sqrt{2}\sqrt{10}}\end{bmatrix}=\frac{1}{\sqrt{10}}\begin{bmatrix}-\frac{10}{\sqrt{2}\sqrt{10}} & \frac{20}{2\sqrt{2}\sqrt{10}}\\ \frac{10}{\sqrt{2}\sqrt{10}} & \frac{20}{2\sqrt{2}\sqrt{10}}\end{bmatrix}=$$

$$=\begin{bmatrix}-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}=U$$

שאלה 7:

תהי $A\in\mathbb{R}^{m\times n}$, $C_0=A^TA$. יהיו $\lambda_n\geq\ldots\geq\lambda_1$ ע"ע של C_0 ו־ $v_1,\ldots,v_n$ ו"ע מנורמלים מתאימים. $\lambda_1>\lambda_2$. נגדיר $b_{k+1}=\frac{C_0b_k}{\|C_0b_k\|}$ ו־ $b_0=\sum_{i=1}^n\alpha_iv_i$ עבור $\alpha_i\in\mathbb{R}$, $\alpha_1\neq 0$. נרצה להראות $\lim\limits_{k\to\infty}b_k=\pm v_1$. נפתח את נוסחאת הרקורסיה

$$b_{k+1}=\frac{C_0b_k}{\|C_0b_k\|}=C_0^2\frac{\frac{b_{k-1}}{\|C_0b_{k-1}\|}}{\left\|\frac{C_0^2b_{k-1}}{\|C_0b_{k-1}\|}\right\|}\overset{*}{=}C_0^2\frac{\frac{b_{k-1}}{\|C_0b_{k-1}\|}}{\frac{\left\|C_0^2b_{k-1}\right\|}{\|C_0b_{k-1}\|}}=\frac{C_0^2b_{k-1}}{\left\|C_0^2b_{k-1}\right\|}=\ldots=\frac{C_0^{k+1}b_0}{\left\|C_0^{k+1}b_0\right\|}$$

$$\Longrightarrow b_k=\frac{C_0^kb_0}{\left\|C_0^kb_0\right\|}$$

המעבר $*$ מהומוגניות של נורמה. מתקיים

$$C_0^kb_0=C_0^k\sum_{i=1}^n\alpha_iv_i=\sum_{i=1}^n\alpha_iC_0^kv_i=\sum_{i=1}^n\alpha_i\left(C_0\right)^{k-1}C_0v_i=$$
$$\overset{*}{=}\sum_{i=1}^n\alpha_i\left(C_0\right)^{k-1}\lambda_iv_i=\sum_{i=1}^n\alpha_i\lambda_i\left(C_0\right)^{k-2}C_0v_i=$$
$$=\sum_{i=1}^n\alpha_i\lambda_i^2\left(C_0\right)^{k-2}v_i=\ldots=\sum_{i=1}^n\alpha_i\lambda_i^kv_i=$$
$$=\alpha_1\lambda_1^kv_1+\sum_{i=2}^n\alpha_i\lambda_i^kv_i\overset{\alpha_1\neq 0}{=}\alpha_1\lambda_1^k\left(v_1+\sum_{i=2}^n\frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^kv_i\right)$$

המעבר $*$ מכך ש־ v_i הוא ו"ע המתאים לע"ע λ_i.
כיוון ש־ $\lambda_1 > \lambda_2 \geq ... \geq \lambda_n$ מתקיים $\frac{\lambda_i}{\lambda_1} < 1$ עבור $2 \leq i \leq n$ ולכן $\lim\limits_{k\to\infty} \left(\frac{\lambda_i}{\lambda_1}\right)^k = 0$.

$$\begin{aligned}
\lim_{k\to\infty} b_k = \lim_{k\to\infty} \frac{C_0^k b_0}{\left\|C_0^k b_0\right\|} &= \lim_{k\to\infty} \frac{\alpha_1 \lambda_1^k \left(v_1 + \sum_{i=2}^n \frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k v_i\right)}{\left\|\alpha_1 \lambda_1^k \left(v_1 + \sum_{i=2}^n \frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k v_i\right)\right\|} = \\
&= \lim_{k\to\infty} \frac{\alpha_1 \lambda_1^k \left(v_1 + \sum_{i=2}^n \frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k v_i\right)}{\left|\alpha_1 \lambda_1^k\right| \left\|\left(v_1 + \sum_{i=2}^n \frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k v_i\right)\right\|} = \\
&= \lim_{k\to\infty} \pm \frac{\left(v_1 + \sum_{i=2}^n \frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k v_i\right)}{\left\|\left(v_1 + \sum_{i=2}^n \frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k v_i\right)\right\|} = \\
&= \pm \frac{\left(v_1 + \sum_{i=2}^n \frac{\alpha_i}{\alpha_1} \cdot 0 \cdot v_i\right)}{\left\|\left(v_1 + \sum_{i=2}^n \frac{\alpha_i}{\alpha_1} \cdot 0 \cdot v_i\right)\right\|} = \pm \frac{v_1}{\|v_1\|} \\
&\overset{*}{=} \pm \frac{v_1}{1} = \pm v_1
\end{aligned}$$

כאשר מעבר $*$ מכך ש־v_1 וקטור יחידה.

שאלה 8:

נתונה $f(\sigma) = U \mathrm{diag}(\sigma) U^T x$. נכתוב

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

מתקיים

$$f(\sigma) = \begin{pmatrix} f_1(\sigma) & \dots & f_n(\sigma) \end{pmatrix}^T = \begin{pmatrix} \left[\sum_{i=1}^n \sigma_i u_i u_i^T x\right]_1 & \dots & \left[\sum_{i=1}^n \sigma_i u_i u_i^T x\right]_n \end{pmatrix}^T$$

לכן הנגזרת החלקית של $\left[\sum_{i=1}^n \sigma_i u_i u_i^T x\right]_j$ לפי σ_k היא $\left[u_k u_k^T x\right]_j$ (כי עבור $i \neq k$ הנגזרת היא אפס וכאשר $i = k$ הנגזרת היא $u_k u_k^T x$).
אז עבור $i, j \in [n]$

$$\frac{\partial f_j(\sigma)}{\partial \sigma_i} = \left[u_i u_i^T x\right]_j$$

ולכן

$$J_\sigma(f) = \begin{bmatrix} \left[u_1 u_1^T x\right]_1 & \cdots & \left[u_n u_n^T x\right]_1 \\ \vdots & \ddots & \vdots \\ \left[u_1 u_1^T x\right]_n & \cdots & \left[u_n u_n^T x\right]_{1n} \end{bmatrix} = \begin{bmatrix} | & & | \\ u_1 u_1^T x & \cdots & u_n u_n^T x \\ | & & | \end{bmatrix} = U \mathrm{diag}\left(U^T x\right)$$

המעבר האחרון נובע מכך שכפל במטריצה אלכסונית מכפיל את העמודה ה־i בערך ה־i באלכסון. במקרה שלנו, העמודה u_i מוכפלת בסקאלר $u_i^T x$.

שאלה 9:

נחשב את הגרדיאנט של הפונקציה $h(\sigma) = \frac{1}{2}\|f(\sigma) - y\|^2$. נסמן $g(\sigma) = \frac{1}{2}\|\sigma\|^2$, $\ell(\sigma) = f(\sigma) - y$. מתקיים $h(\sigma) = (g \circ \ell)(\sigma)$.
ראינו בתרגול $\nabla g(\sigma) = \frac{1}{2} \cdot 2\sigma = \sigma$. נשתמש בכלל השרשת

$$(\nabla h(\sigma))^T = J_\sigma(h) = J_\sigma(g \circ \ell) = J_{\ell(\sigma)}(g)\, J_\sigma(\ell) = (\nabla g(\ell(\sigma)))^T J_\sigma(\ell) =$$

$$= (\ell(\sigma))^T \cdot J_\sigma(\ell) = (f(\sigma) - y)^T \cdot J_\sigma(f)$$

$$\Longrightarrow \nabla h(\sigma) = J_\sigma(f)^T (f(\sigma) - y)$$

שאלה 10:

נחשב את היעקוביאן של פונקציית softmax, $S : \mathbb{R}^d \to (0,1)^d$

$$S(a)_i = \frac{e^{a_i}}{\sum_{k=1}^d e^{a_k}}$$

כמו בתרגול נסמן $g_i(a) = e^{a_i}$, $h(a) = \sum_{k=1}^d e^{a_k}$. יהיו $i, j \in [d]$.

$$\frac{\partial S_i(a)}{\partial a_j} = \frac{\partial}{\partial a_j} \frac{e^{a_i}}{\sum_{k=1}^d e^{a_k}} = \frac{\partial}{\partial a_j} \frac{g_i(a)}{h(a)}$$

הנגזרת של h לפי a_j היא e^{a_j}. הנגזרת של g_i לפי a_j היא e^{a_j} כאשר $i = j$ ו־0 כאשר $i \neq j$. עבור $i = j$ מתקיים

$$\frac{\partial}{\partial a_j} \frac{e^{a_i}}{\sum_{k=1}^d e^{a_k}} = \frac{\partial}{\partial a_i} \frac{e^{a_i}\left(\sum_{k=1}^d e^{a_k}\right) - e^{a_i}e^{a_i}}{\left(\sum_{k=1}^d e^{a_k}\right)^2} = \frac{e^{a_i}\left(\left(\sum_{k=1}^d e^{a_k}\right) - e^{a_i}\right)}{\left(\sum_{k=1}^d e^{a_k}\right)^2}$$

$$= \frac{e^{a_i}}{\sum_{k=1}^d e^{a_k}} \frac{\left(\sum_{k=1}^d e^{a_k}\right) - e^{a_i}}{\sum_{k=1}^d e^{a_k}} = S_i(a)(1 - S_i(a))$$

עבור $i \neq j$

$$\frac{\partial}{\partial a_j} \frac{e^{a_i}}{\sum_{k=1}^d e^{a_k}} = \frac{\partial}{\partial a_j} \frac{0\left(\sum_{k=1}^d e^{a_k}\right) - e^{a_i}e^{a_j}}{\left(\sum_{k=1}^d e^{a_k}\right)^2} = \frac{-e^{a_i}e^{a_j}}{\left(\sum_{k=1}^d e^{a_k}\right)^2}$$

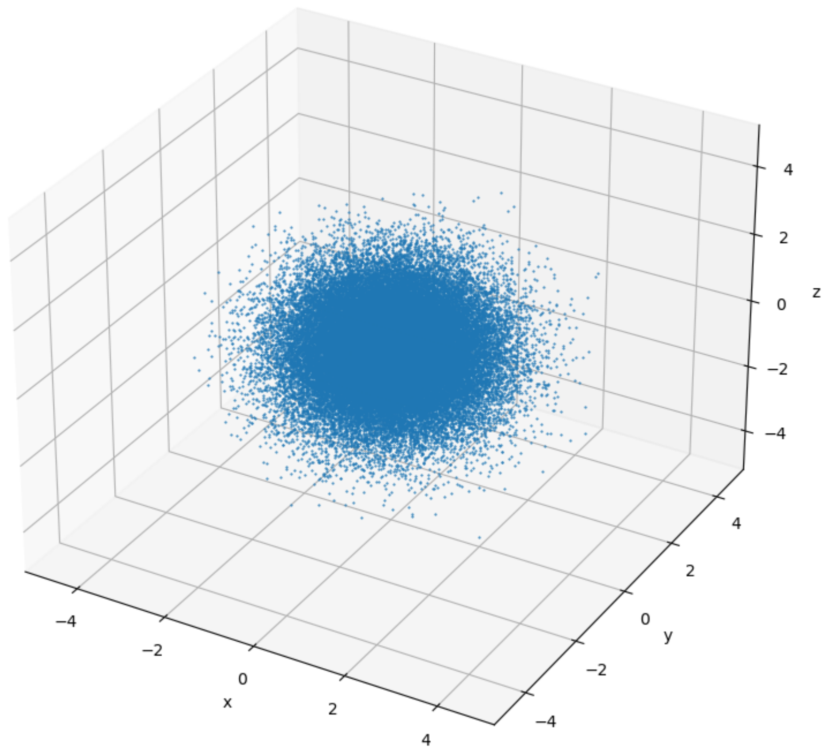
$$= \frac{-e^{a_i}}{\sum_{k=1}^d e^{a_k}} \frac{e^{a_j}}{\sum_{k=1}^d e^{a_k}} = -S_i(a)S_j(a)$$

סה"כ קיבלנו

$$J_a(S)_{ij} = \begin{cases} -S_i(a)S_j(a) & i \neq j \\ S_i(a)(1 - S_i(a)) & i = j \end{cases}$$

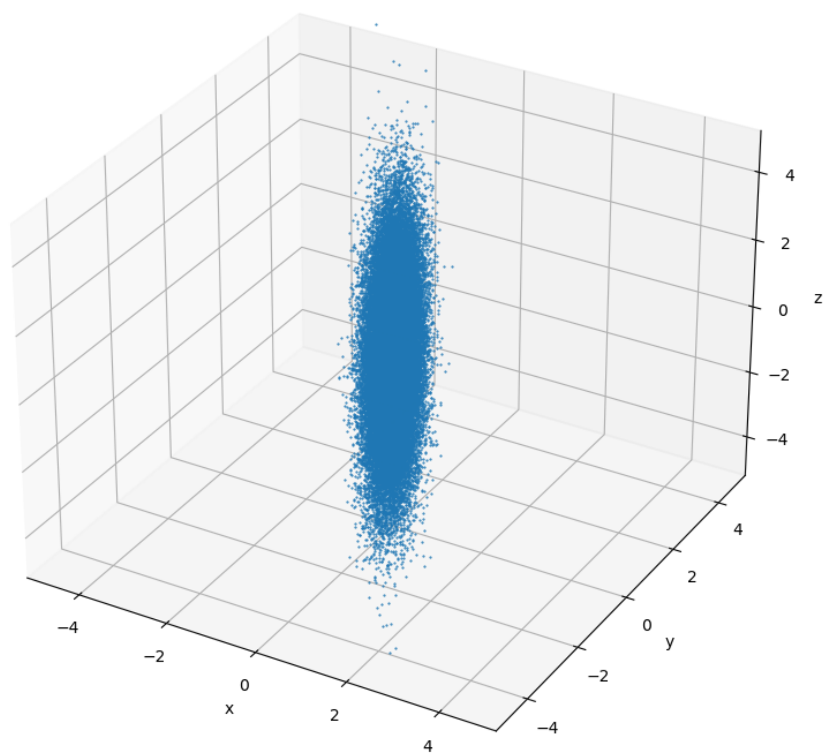
שאלה 11:

גרף המציג דגימות של הוקטור המקרי (X, Y, Z) אשר מטריצת השונות המשותפת שלו היא I_3



שאלה 12:

נכפיל את דגימות הוקטור הקודם במטריצת ה־scaling $\begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ ונציג:



מטריצת השונות המשותפת היא כעת (ע"פ חישוב נומרי באמצעות numpy.cov)

$$\begin{bmatrix} 0.00995 & 0.00021 & -0.00096 \\ 0.00021 & 0.25009 & 0.00158 \\ -0.000967 & 0.00158 & 4.00283 \end{bmatrix} \approx \begin{bmatrix} 0.1^2 & 0 & 0 \\ 0 & 0.5^2 & 0 \\ 0 & 0 & 2^2 \end{bmatrix}$$

קיבלנו בערך את ערכי מטריצת ה־scaling בריבוע וזה הגיוני, נראה זו בעזרת הנוסחא הבאה:

$$\text{cov}\left(AX\right) = A\text{cov}\left(X\right)A^T$$

כאשר A מטריצה כלשהי ו־X וקטור מקרי. במקרה שלנו A זו מטריצת ה־scaling ו־X זה הוקטור המקרי (x, y, z). בתרגיל 11 הנחנו כי

$$\text{cov}\left(X\right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

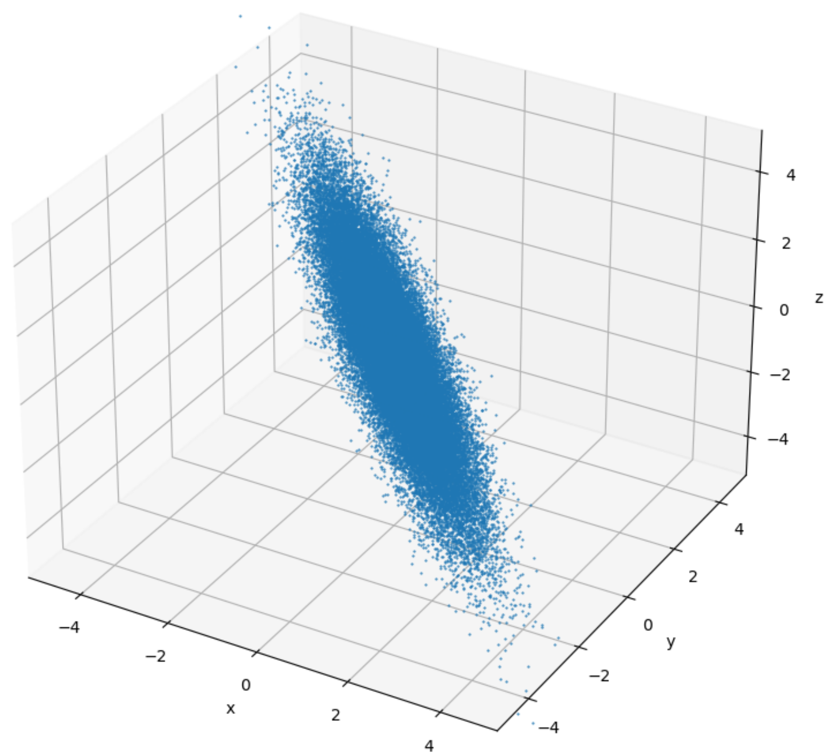
נרצה לחשב את $\text{cov}\left(AX\right)$. A מטריצה אלכסונית ולכן נקבל

$$\text{cov}\left(AX\right) = A\text{cov}\left(X\right)A^T = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 2 \end{bmatrix}^2 = \begin{bmatrix} 0.1^2 & 0 & 0 \\ 0 & 0.5^2 & 0 \\ 0 & 0 & 2^2 \end{bmatrix}$$

כמו שראינו שקיבלנו באמצעות numpy.cov.

שאלה 13:

כעת נכפיל את הנתונים במטריצה אורתוגונלית כלשהי. נקבל את גרף הדגימות הבא:

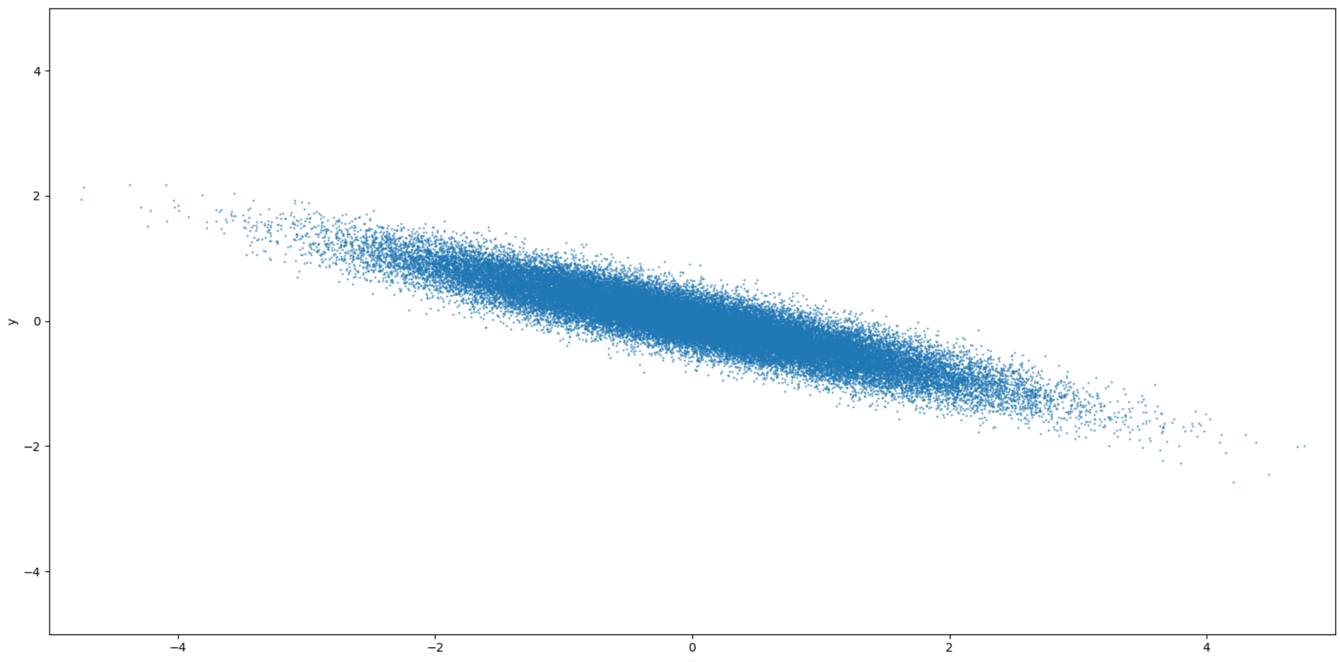


קיבלנו "הזזה" של הערכים. המטריצה בה הכפלנו ומטריצת השונות של המשותפת של הנתונים החדשים:

```
rand orthogonal matrix
[[-0.13902653  0.8487744  -0.51015061]
 [-0.96349463  0.00308801  0.26770983]
 [ 0.2288006   0.52874615  0.81736026]]
cov of 13
[[ 1.2205587  -0.54494605 -1.55624757]
 [-0.54494605  0.2966262   0.87491506]
 [-1.55624757  0.87491506  2.74570118]]
```

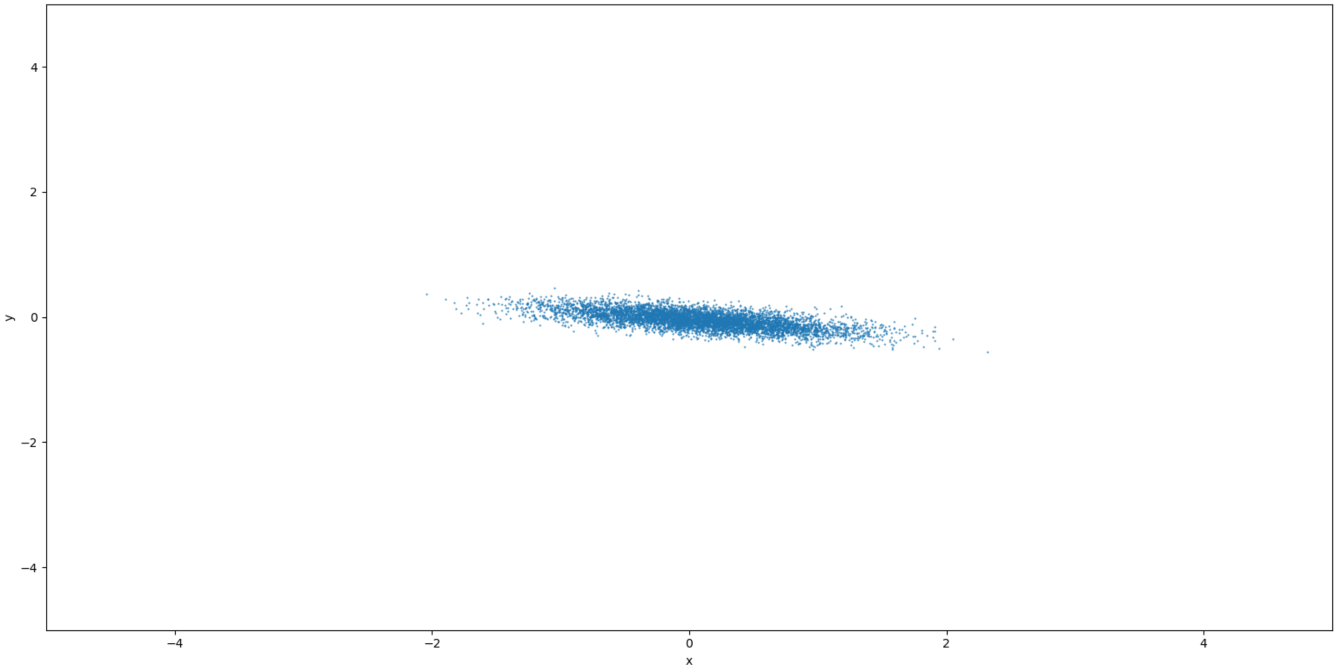
שאלה 14:

נטיל את הנתונים של שאלה 13 על המישור xy. (כלומר נציג את נתוני x, y בלבד, $z = 0$, הטלה אורתוגונלית)



שאלה 15:

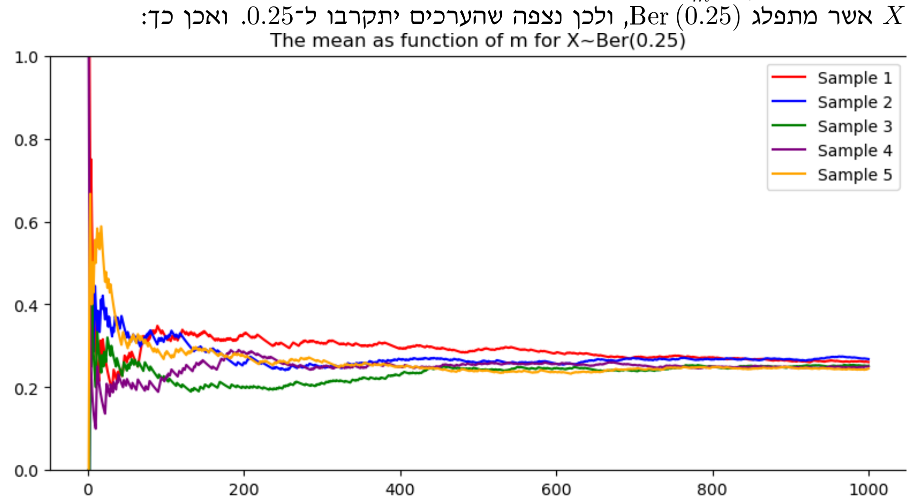
נטיל את כל הערכים בהם $0.1 > z > -0.4$ על מישור ה־xy.



שאלה 16:

<u>סעיף a</u>:

כיוון ש־ $\overline{X_m} = \frac{1}{m}\sum_{i=1}^{m} X_i$ הוא אומד חסר הטייה עבור התוחלת, נצפה לראות שככל ש־m עולה כך התוחלת תתקרב לתוחלת של X אשר מתפלג $\text{Ber}\,(0.25)$, ולכן נצפה שהערכים יתקרבו ל־0.25. ואכן כך:



<u>סעיף b</u>:

בסעיף זה נחשב את הדיוק של אומד חסר ההטייה עבור התוחלת. נבדוק עבור מספר ספים: $\varepsilon = 0.5, 0.25, 0.1, 0.01, 0.001$.
בחישוב בעזרת א"ש צ'בישב נשתמש בכך ש־

$$\text{Var}\,(X) = p\,(1-p) \leq \frac{1}{4}$$

ולכן לפי מסקנה 1.3.9 בספר א"ש צ'בישב נותן את החסם

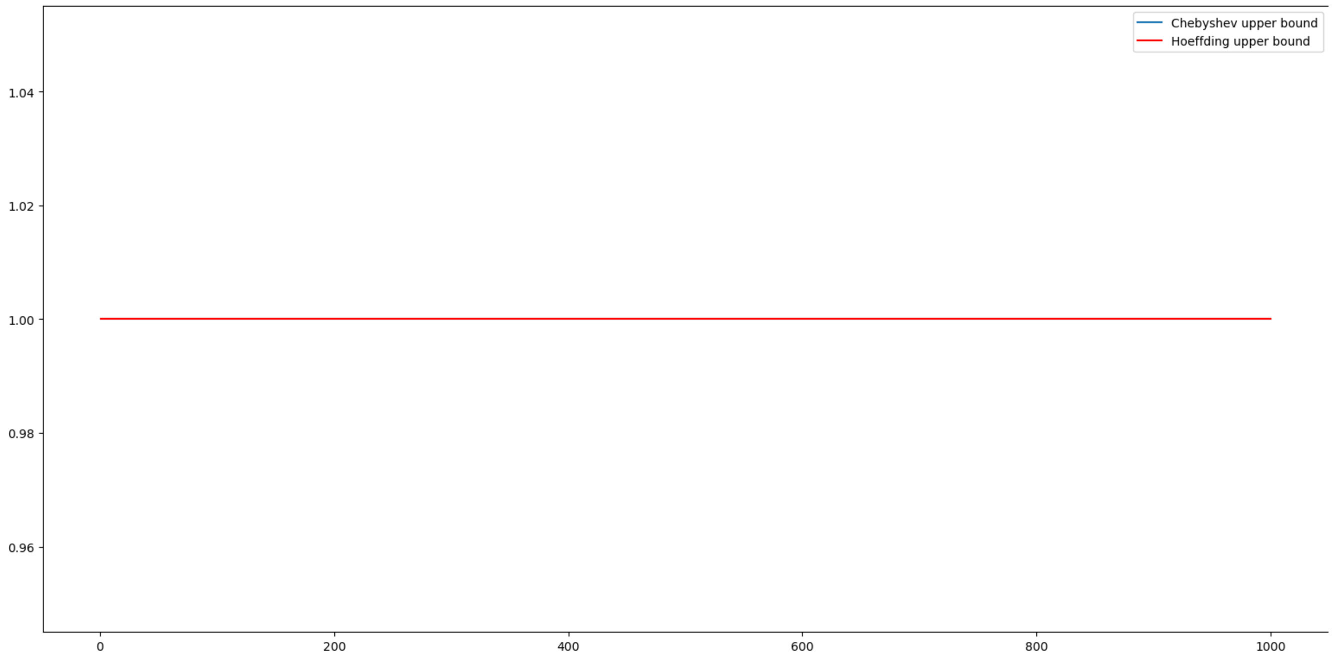
$$\mathbb{P}\left(\left|\overline{X_m} - \mathbb{E}\,(X)\right| \geq \varepsilon\right) \leq \frac{\text{Var}\,(X)}{m\varepsilon^2} \leq \frac{1}{4m\varepsilon^2}$$

ועבור א"ש הופדינג, לפי מסקנה 1.3.12 ועם החסמים $a=0,b=1$ עבור X_i (כי הערכים שלהם הם 1 או 0) נקבל את החסם:

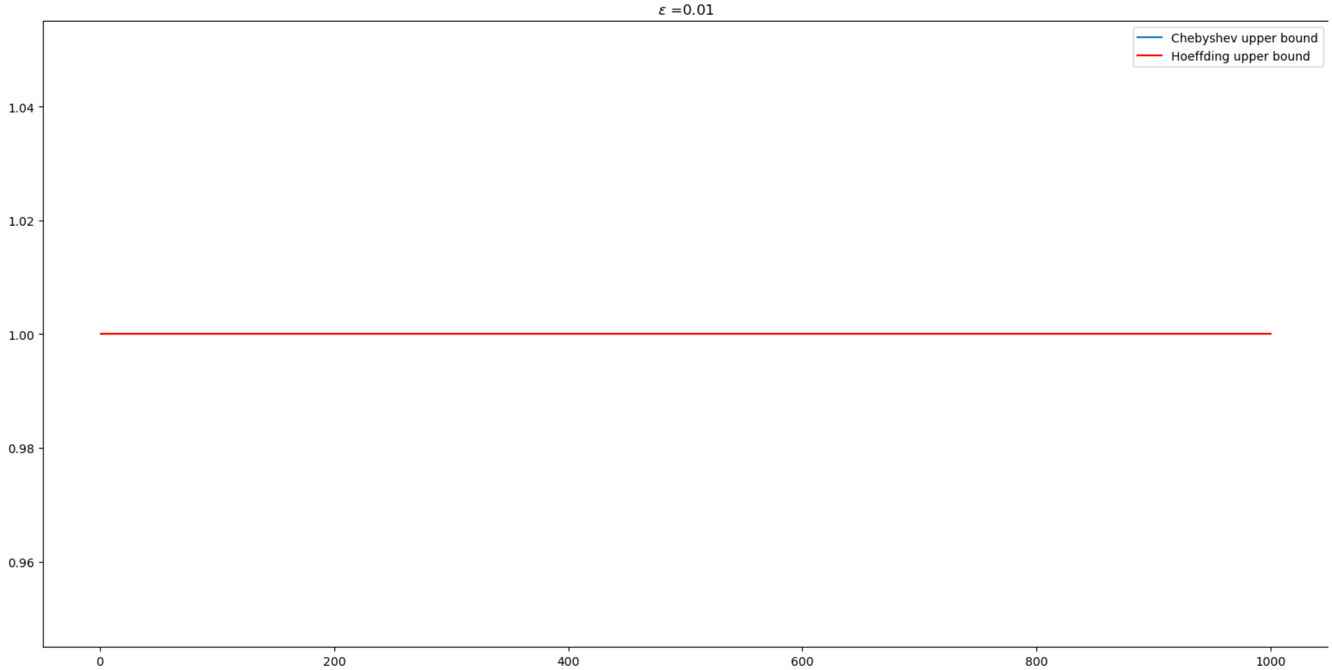
$$\mathbb{P}\left(\left|\overline{X_m}-\mathbb{E}\left(X\right)\right|\geq\varepsilon\right)\leq2\exp\left(\frac{-2m\varepsilon^{2}}{\left(b-a\right)^{2}}\right)=2\exp\left(-2m\varepsilon^{2}\right)$$

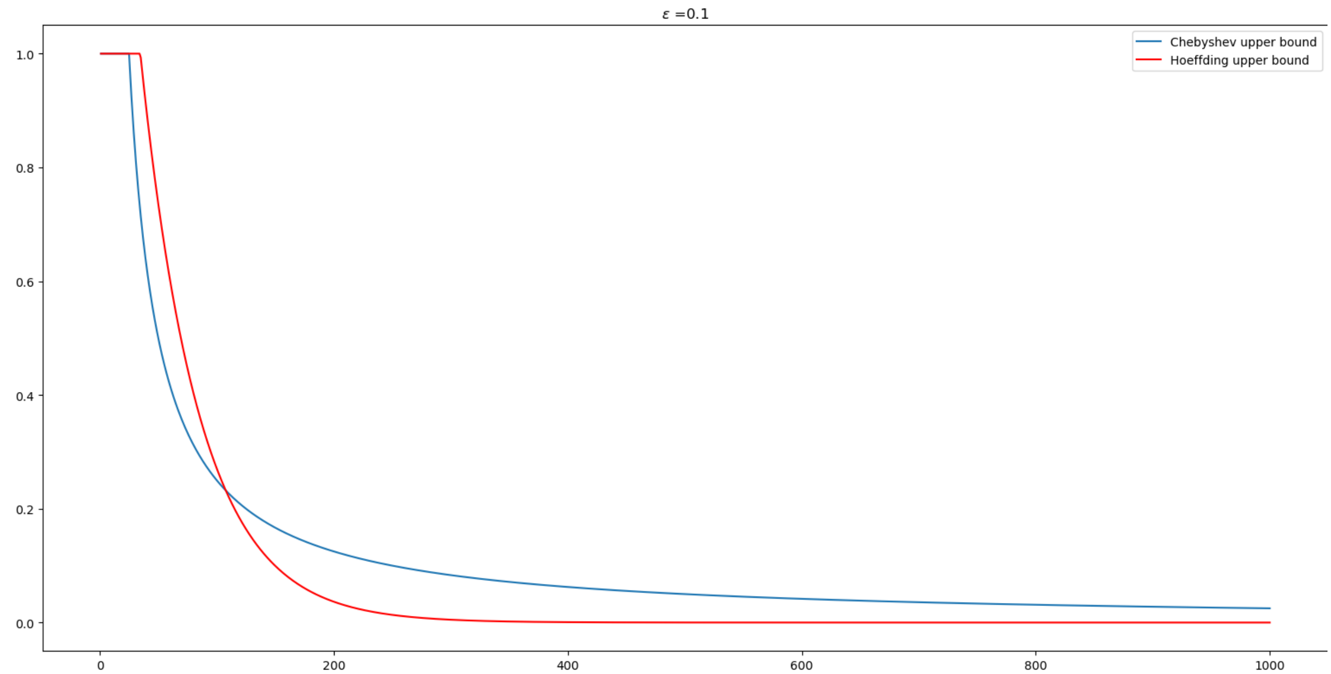
נציג את הגרפים:

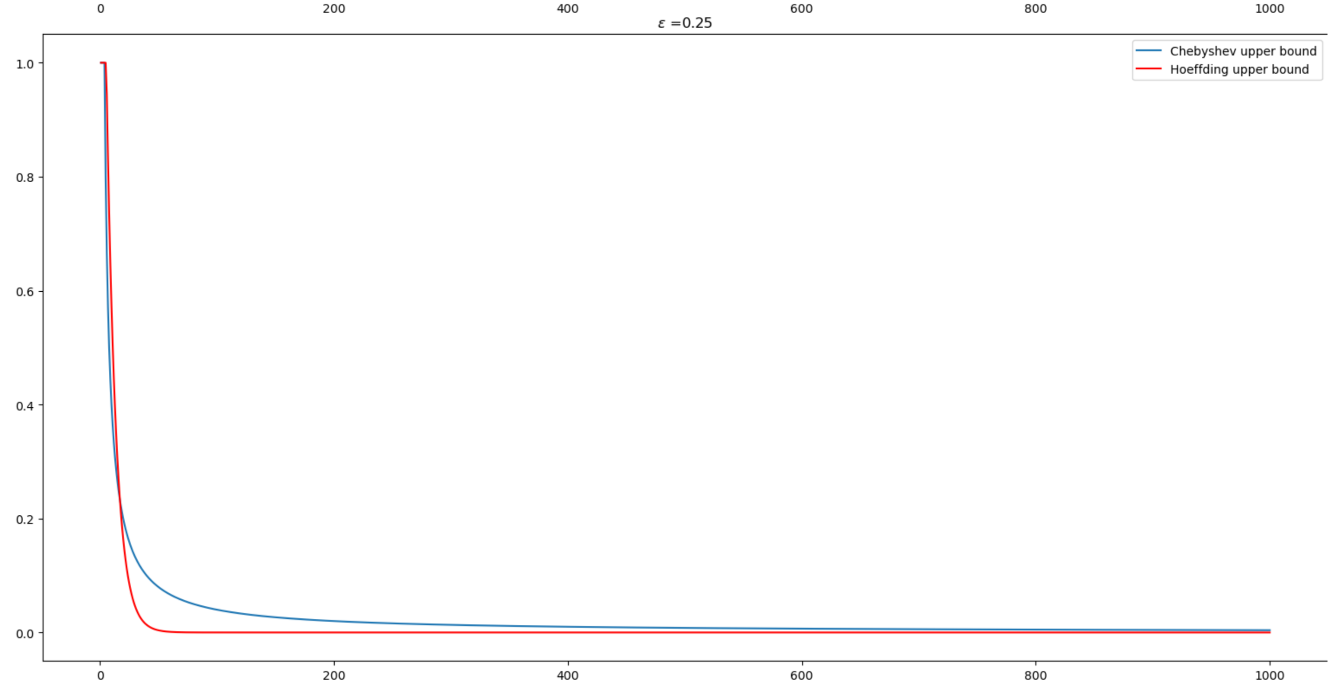
ε =0.001

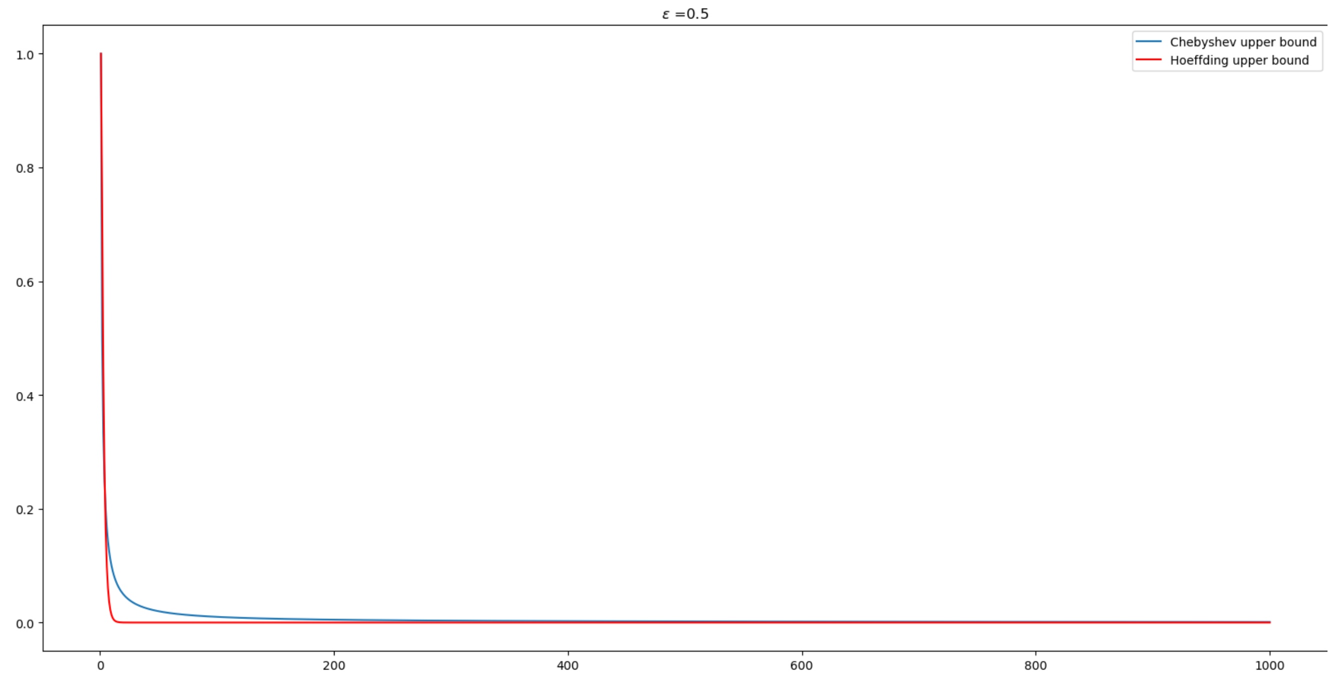


ε =0.01




ε =0.1
Chebyshev upper bound
Hoeffding upper bound
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000


ε =0.25
Chebyshev upper bound
Hoeffding upper bound
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000

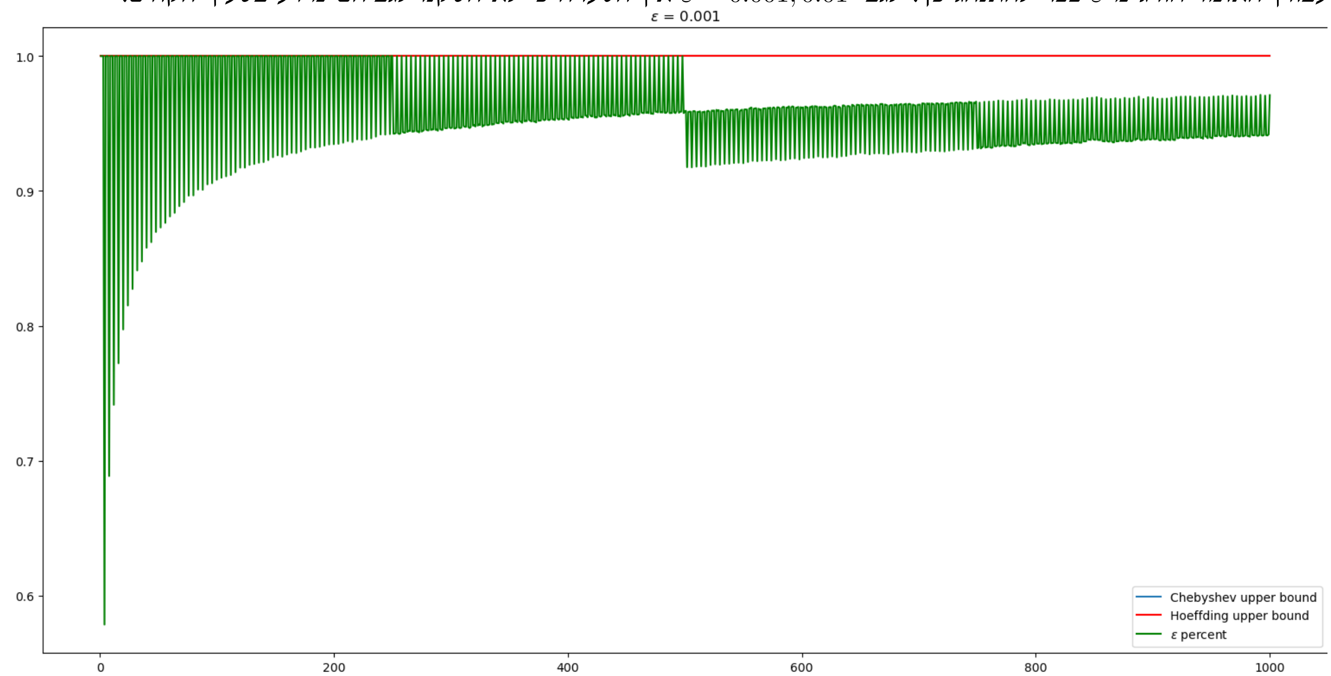


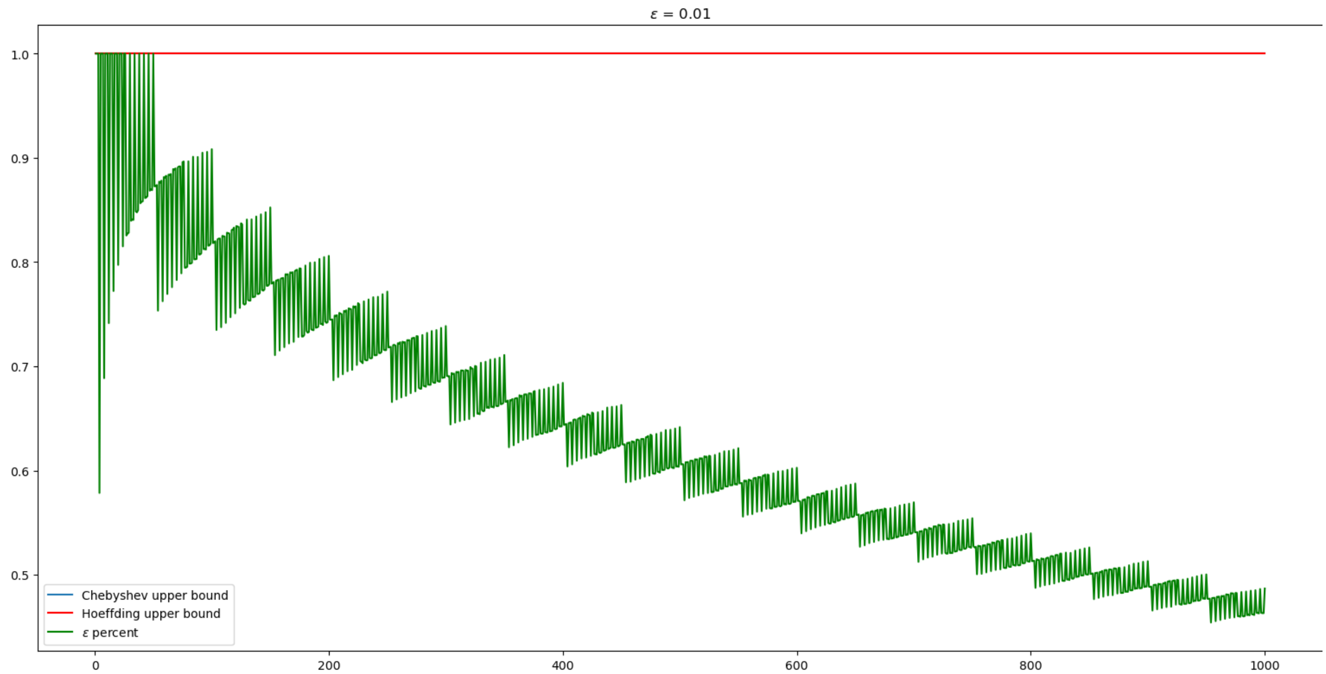
<u>מסקנות</u>:

1. עבור $\varepsilon = 0.001, 0.01$ החסמים לא עוזרים לנו כי אנו יודעים שהסתברות תמיד קטנה או שווה ל־1.
2. עבור $\varepsilon = 0.1, 0.25, 0.5$, נשים לב שככל ש־ε גדול יותר, כך ההסתברות שהאומד סוטה ביותר מ־ε שואפת לאפס מהר יותר וכן צריך פחות דגימות כדי שזה יקרה.
3. החסם העליון מא"ש הופדינג שואף מהר יותר לאפס מאשר החסם של א"ש צ'בישב ־ הוא נותן חסם הדוק יותר.

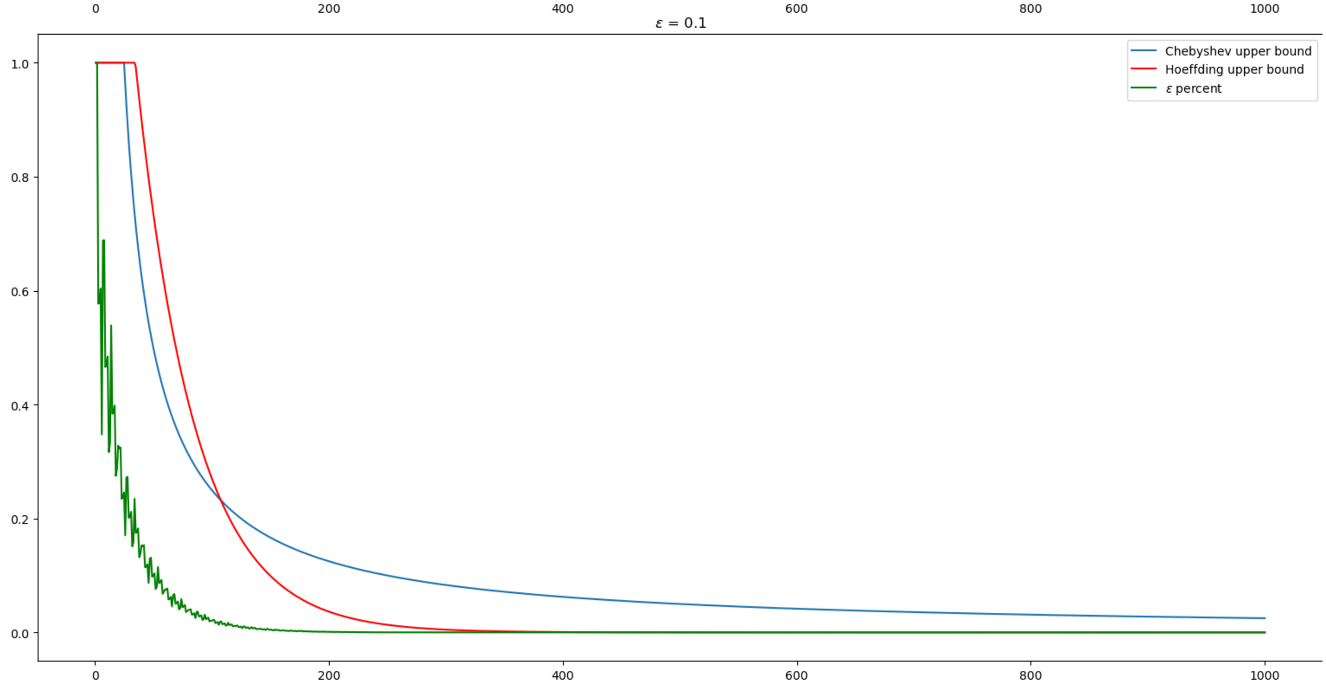
<u>סעיף c</u>:

בהנתן התוצאות של סעיף b, מה שמצופה לראות בגרפים זה שעבור $\varepsilon = 0.1, 0.25, 0.5$, ככל שמספר הדגימות ($m$) יעלה כך האחוזים יצנחו, כיוון שראינו שההסתברות של האומד לסטות מ־ε שואפת לאפס ככל שמספר הדגימות עולה, ולכן גם אחוז הדגימות עבורן האומד חורג מ־ε צפוי להתנהג כך. לגבי $\varepsilon = 0.001, 0.01$ אין השערה כי לא הסקנו לגביהם מידע בסעיף הקודם.

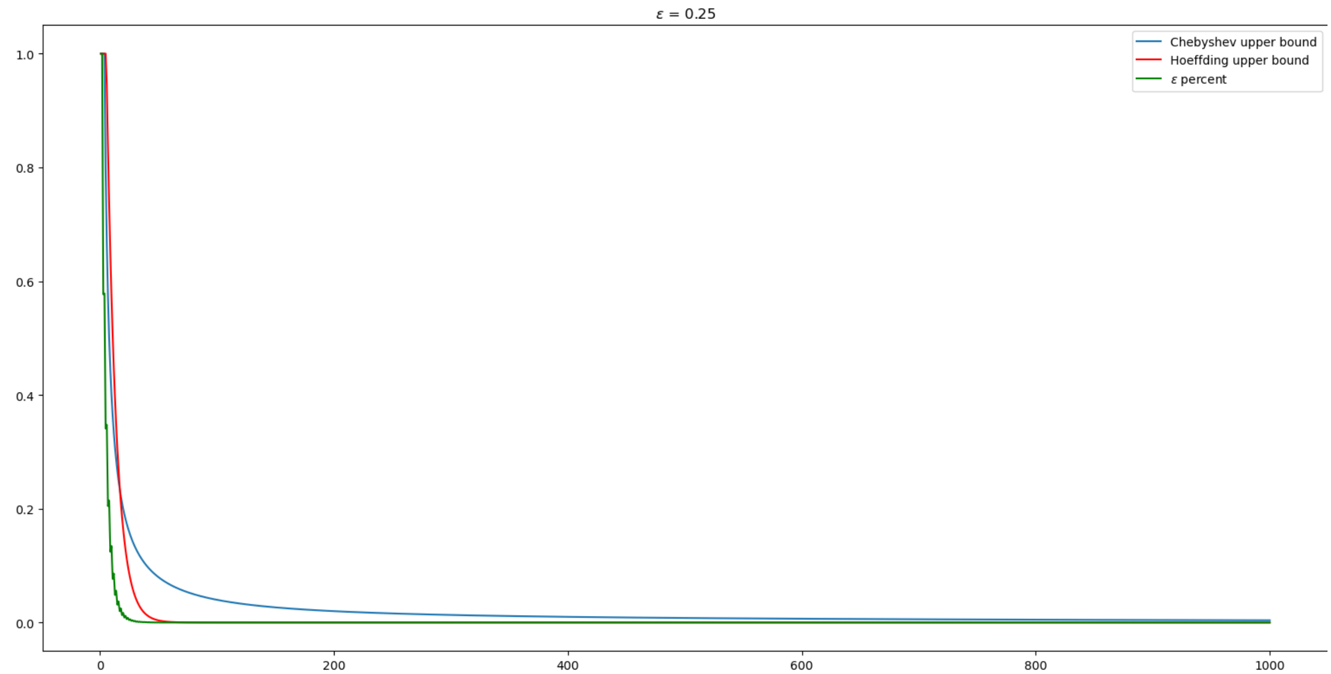


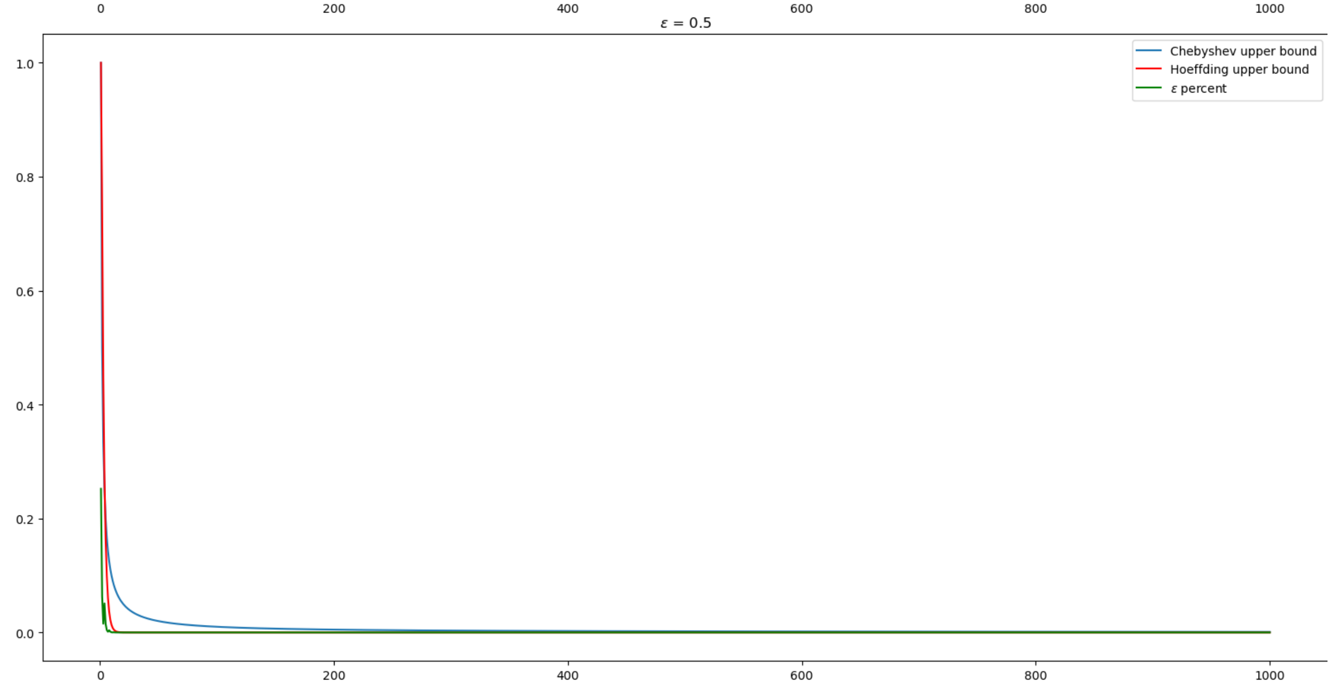


ε = 0.01
Chebyshev upper bound
Hoeffding upper bound
ε percent



ε = 0.1
Chebyshev upper bound
Hoeffding upper bound
ε percent





מסקנות:

1. צדקנו לגבי $\varepsilon = 0.1, 0.25, 0.5$
2. עבור $\varepsilon = 0.001, 0.01$, בסעיף הקודם לא הצלחנו לקבל הערכה לאומד עבור ערכים אלו. כאן, ניתן לראות שעבור $\varepsilon = 0.001$ האחוזים נשארים גבוהים - האומד חורג ברוב הדגימות. עבור $\varepsilon = 0.01$, רואים דווקא שהאחוזים יורדים כאשר מספר הדגימות המחושב עולה. הסקנו כאן מידע שלא הצלחנו להסיק בסעיף הקודם.