בס"ד

# תרגיל 1 - IML

6. יהיו $U$ מט' אורתוגונלית ו- $A$ מטריצה.

נגדיר את האליפסואיד $B$ המוגדר ע"י $A$ להיות $B=\{Ax+v \mid \|x\|^2=1\}$

נשים לב כי האליפסואיד $C$ המוגדר ע"י $AU$ מקיים:

$$C=\{AUx+v \mid \|x\|^2=1\} = \{AUx+v \mid \|Ux\|^2=1\}$$

שכן הוכחנו בכיתה שמט' אורתו' הינה איזומטריה כלומר $\|Ux\|^2=\|x\|^2$

נסמן $Ux=w$ אזי $C=\{Aw+v \mid \|w\|^2=1\}=B$, כנדרש ∎

7. נחשב את המרחק של הראשית מהמישור $w^Tx=b$ (כלומר המישור הנפרש ע"י הק' כל ה-x-ים המקיימים את המשוואה הנ"ל).

נשים לב כי $\langle x,w\rangle=\langle w,x\rangle=w^Tx=b$

נבחר מהמישור הנ"ל (נגדיר אותו כמישור V) את הנק' $x_0=0_V$ ונחשב את המרחק לפי הכלל:

$$\|P_w(x-0_V)\| = \|P_w(x)\| = \left\|\frac{\langle x,w\rangle}{\|w\|}\cdot\frac{w}{\|w\|}\right\| =$$

$$= \left\|\frac{\langle x,w\rangle}{\|w\|}\right\|\cdot\left\|\frac{w}{\|w\|}\right\| = \frac{|b|}{\|w\|}\cdot\frac{\|w\|}{\|w\|} = \boxed{\frac{|b|}{\|w\|}}$$

∎

בס"ד

9) נתון מט' $A=\begin{pmatrix}5&5\\-1&7\end{pmatrix}\in\mathbb{R}^{2\times2}$ אזי $A^T=\begin{pmatrix}5&-1\\5&7\end{pmatrix}$

לכן $A\cdot A^T=\begin{pmatrix}5&5\\-1&7\end{pmatrix}\begin{pmatrix}5&-1\\5&7\end{pmatrix}=\begin{pmatrix}50&30\\30&50\end{pmatrix}$

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

נמצא את הע"ע של $AA^T$:

$$0=\det(AA^T-\lambda I)=\det\begin{pmatrix}50-\lambda&30\\30&50-\lambda\end{pmatrix}=(50-\lambda)^2-30^2=$$
$$=2500-100\lambda+\lambda^2-900=\lambda^2-100\lambda+1600=(\lambda-80)(\lambda-20)$$

כלומר הע"ע של $AA^T$ הן 80 ו-20. נמצא את הו"ע המתאימים להן:

עבור $\lambda=80$:
$$\begin{pmatrix}0\\0\end{pmatrix}=(AA^T-\lambda I)\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}-30&30\\30&-30\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}-30x_1+30x_2\\30x_1-30x_2\end{pmatrix}$$

כלומר $30x_1=30x_2 \iff x_1=x_2 \Rightarrow Sp\begin{pmatrix}1\\1\end{pmatrix}$

לכן $\frac{1}{\sqrt2}\begin{pmatrix}1\\1\end{pmatrix}$ הינו הו"ע של $AA^T$ המתאים לע"ע 80.

עבור $\lambda=20$:
$$\begin{pmatrix}0\\0\end{pmatrix}=\begin{pmatrix}30&30\\30&30\end{pmatrix}\begin{pmatrix}y_1\\y_2\end{pmatrix}=\begin{pmatrix}30x_1+30x_2\\30x_1+30x_2\end{pmatrix}$$

כלומר $30y_1=-30y_2 \iff -y_1=y_2 \Rightarrow Sp\begin{pmatrix}1\\-1\end{pmatrix}$

לכן $\frac{1}{\sqrt2}\begin{pmatrix}1\\-1\end{pmatrix}$ הינו הו"ע של $AA^T$ המתאים לע"ע 20.

לכן נוכל לכתוב EVD של $AA^T$, ונקבל הבא:

נסמן $U=\frac{1}{\sqrt2}\begin{pmatrix}1&1\\1&-1\end{pmatrix}$ אזי $U^T=\frac{1}{\sqrt2}\begin{pmatrix}1&1\\1&-1\end{pmatrix}$ ולכן

$$AA^T=U\Sigma^2U^T=\begin{pmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\\\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\end{pmatrix}\begin{pmatrix}80&0\\0&20\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\\\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\end{pmatrix}$$

כעת נחשב עבור $A^TA$ ונזכור כי למדנו בכיתה כי הע"ע של $AA^T$ ושל $A^TA$ זהים לכן הע"ע של $A^TA$ הינם גם 80 ו-20. נמצא את הו"ע המתאימים להן:

עבור $\lambda=80$:
$$\begin{pmatrix}0\\0\end{pmatrix}=(A^TA-80I)\begin{pmatrix}v_1\\v_2\end{pmatrix}=\begin{pmatrix}-54&18\\18&-6\end{pmatrix}\begin{pmatrix}v_1\\v_2\end{pmatrix}$$

כלומר
$$\begin{cases}\mathrm{I} & 0=-54v_1+18v_2\\ \mathrm{II} & 0=18v_1-6v_2 \quad /\cdot 3\\ \mathrm{II} & 0=54v_1-18v_2\end{cases}\Rightarrow \begin{matrix}54v_1=18v_2\\3v_1=v_2\end{matrix}$$

$\Rightarrow Sp\begin{pmatrix}1\\3\end{pmatrix} \Rightarrow \frac{1}{\sqrt{10}}\begin{pmatrix}1\\3\end{pmatrix}$ הינו ו"ע המתאים לע"ע 80.

המשך שאלה 9 -

עבור $\lambda = 20$:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = (A^TA - 20I)\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 6 & 18 \\ 18 & 54 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

$$\Leftarrow \begin{cases} \mathrm{I} & 6u_1 + 18u_2 = 0 \\ \mathrm{II} & 18u_1 + 54u_2 = 0 \end{cases} \Rightarrow u_1 = -3u_2 \Rightarrow Sp\begin{pmatrix} -3 \\ 1 \end{pmatrix}$$

$\Leftarrow$ $\frac{1}{\sqrt{10}}\begin{pmatrix} -3 \\ 1 \end{pmatrix}$ הוא וקטור המתאים לע"ע 20.

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

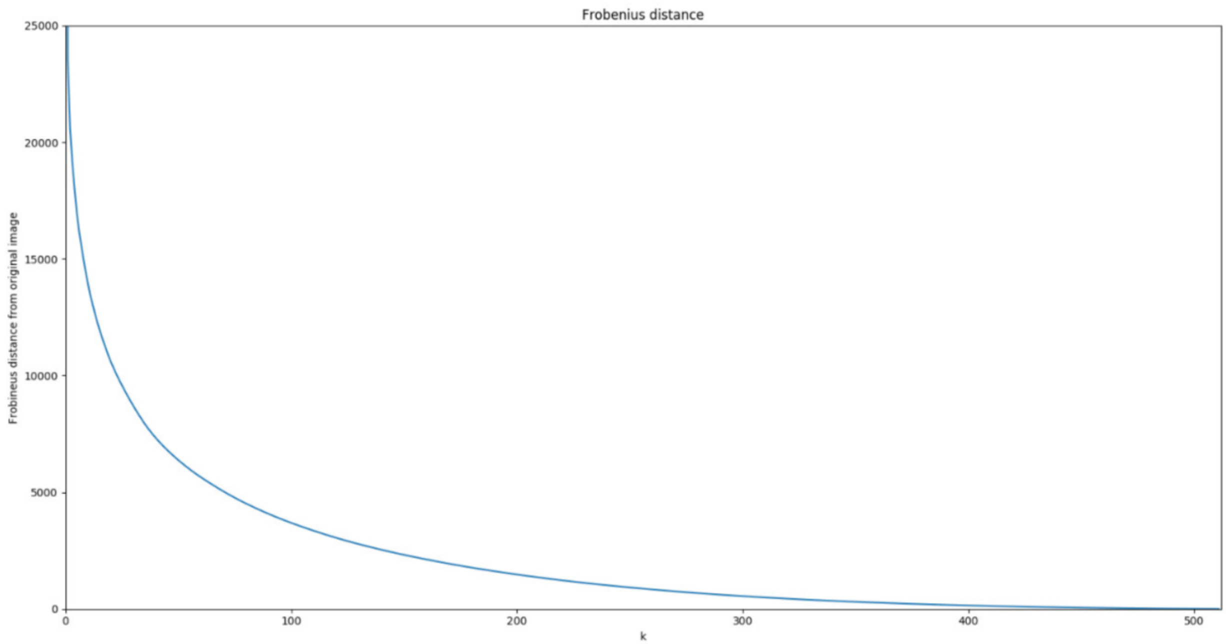
ולכן נוכל לכתוב את הפירוק של $A^TA$ באופן הבא:

$$A^TA = V\Sigma^2V^T = \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{-3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{pmatrix}\begin{pmatrix} 80 & 0 \\ 0 & 20 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{-3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{pmatrix}$$

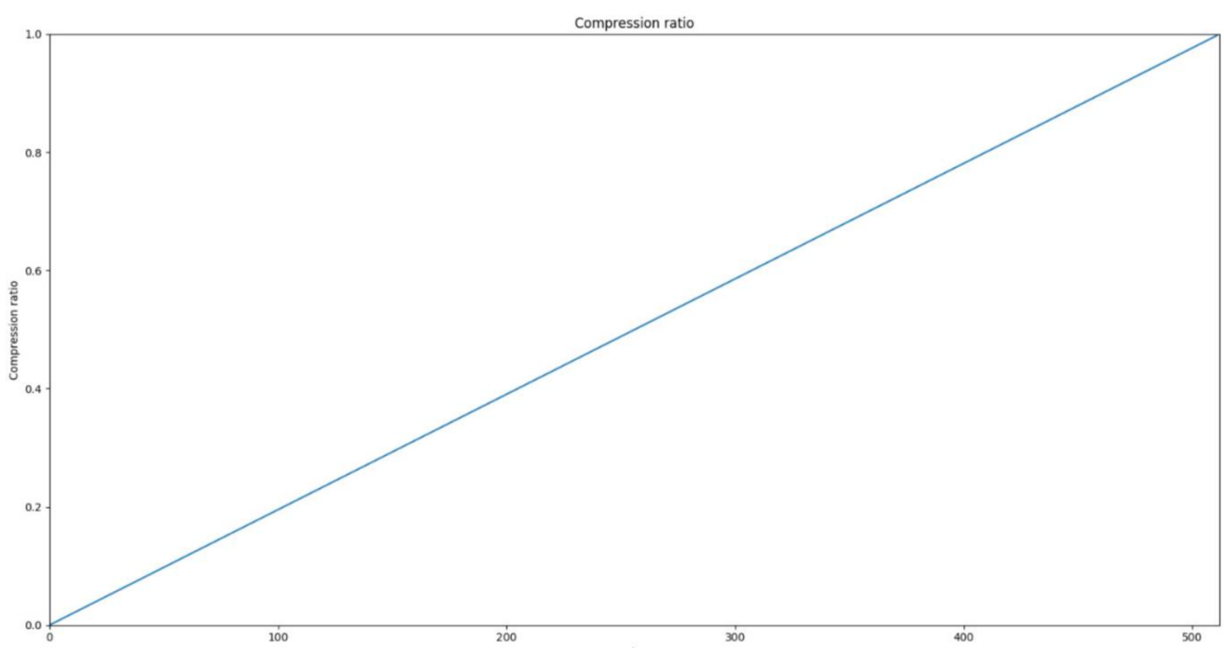
ולכן ע"פ שאלה 8 נוכל לכתוב את ה-SVD של A באופן הבא:

$$A = U\Sigma V^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} \sqrt{80} & 0 \\ 0 & \sqrt{20} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{-3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{pmatrix} \blacksquare$$

10. Frobenius distance as a function of k, for 512>=k>=0:

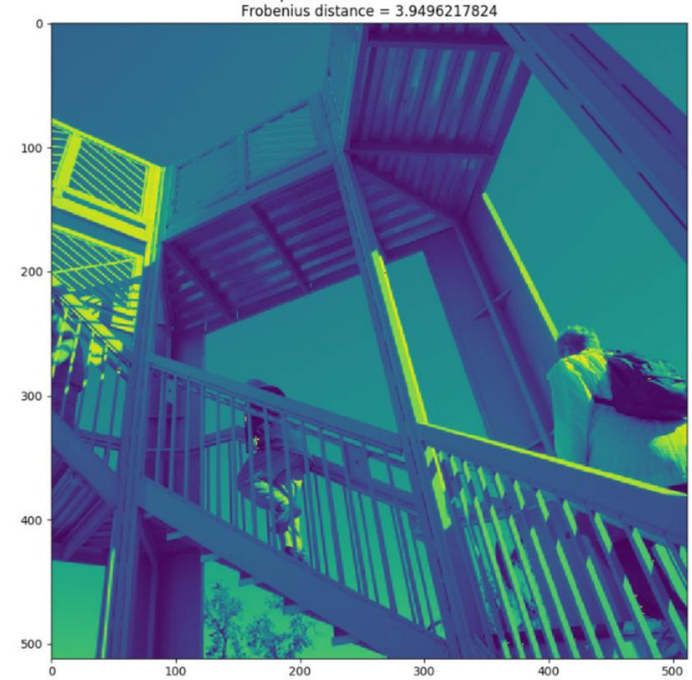



Compression ratio as a function of k, for 512>=k>=0:

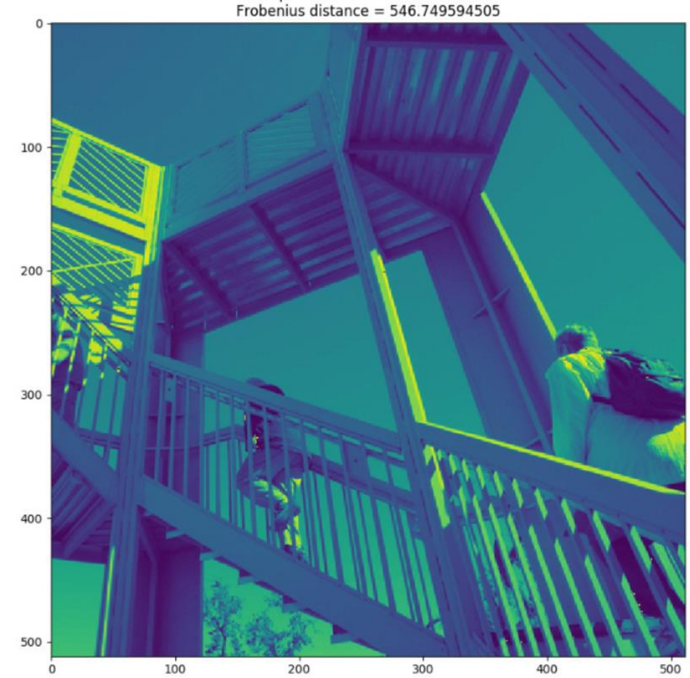


11.

k = 500
Compression ratio = 0.021484375
Frobenius distance = 3.9496217824

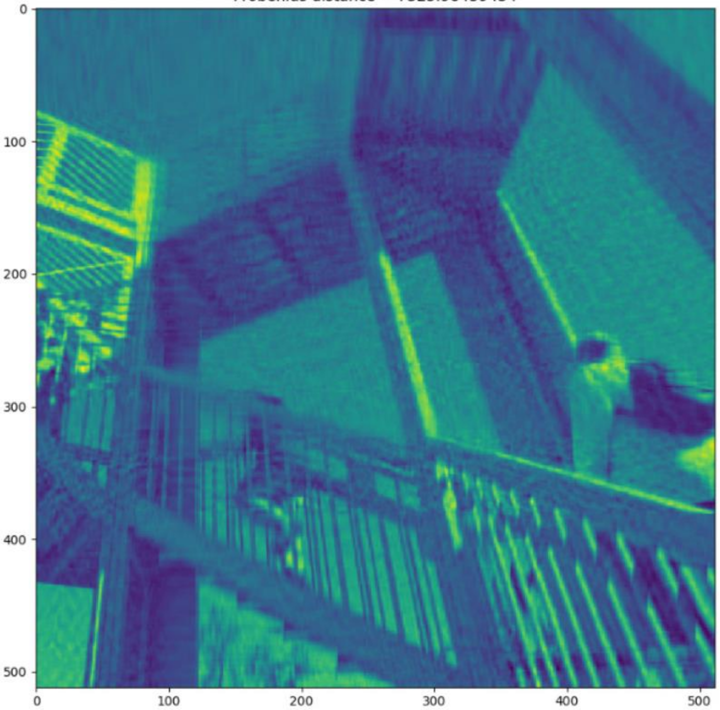


k = 300
Compression ratio = 0.412109375
Frobenius distance = 546.749594505

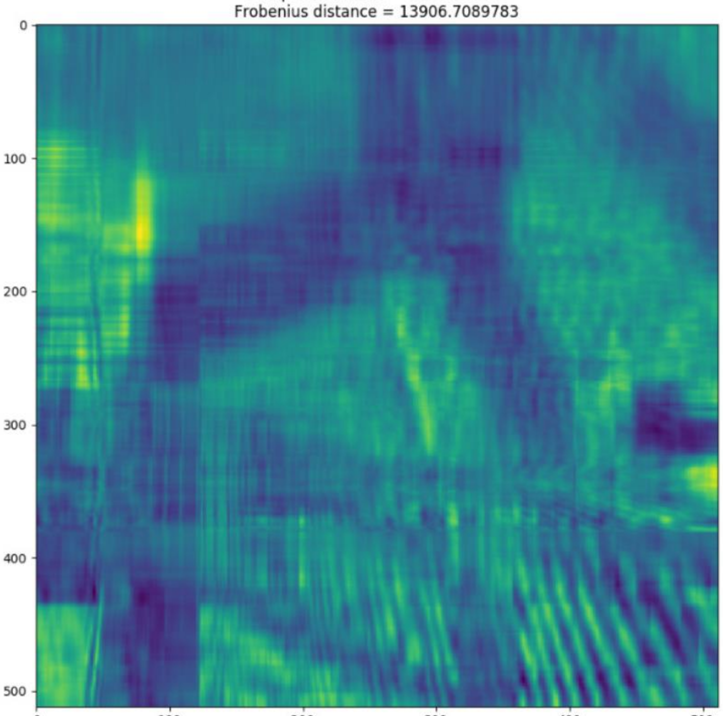


k = 100
Compression ratio = 0.802734375
Frobenius distance = 3686.42588797

k = 40
Compression ratio = 0.919921875
Frobenius distance = 7325.98439434



k = 10
Compression ratio = 0.978515625
Frobenius distance = 13906.7089783



12. First let’s notice that ‘k’ stands for the amount of singular values left, as such we see the the compression ratio is linear to the singular values, as expected. Interesting enough we see in the forbenius distance graph that there is an exponential drop within 100>=k>=0 due to this, we can see in the images that I chose that while k>=100 all though there are very big drops of data we see that the difference in sharpness between k=500 and k=300 is almost non existnat and when k=100 we can tell the difference but we see it is small and the picture is still clear. But for k<100, we see that the loss of sharpness is much more obvious all though the ‘leaps’ between k of 500,300,100 are bigger than the leaps between 100,40,10 yet the forbenius distance is MUCH bigger, and the sharpness much duller.

14) נראה כי $Im(X^T)=\ker(X)^\perp$: נראה הכלה דו כיוונית:

- נראה כי $Im(X^T)\subseteq\ker(X)^\perp$: יהי $v\in Im(X^T)$, כלומר עבור $X\in\mathbb{R}^{m\times n}$ קיים $u\in\mathbb{R}^m$ כך ש- $X^Tu=v$.

יהי $w\in\ker X$ כלומר $Xw=0$. נשים לב כי מתקיים:

$$\langle v,w\rangle = v^Tw=(X^Tu)^Tw=u^TXw=u^T(Xw)=0$$

$\Leftarrow$ $v\in\ker(X)^\perp$ $\Leftarrow$ $Im(X^T)\subseteq\ker(X)^\perp$.

- נראה כי $\ker(X)^\perp\subseteq Im(X^T)$: (נראה בשלילה:)

יהי $v\notin Im X^T$ ולכן בהכרח קיים $u\in Im(X^T)^\perp$ כך ש- $\langle v,u\rangle\neq 0$

אבל נשים לב כי $X^TXu\in Im X^T$ ולכן מתקיים:

$$0=\langle u,X^TXu\rangle=u^TX^TXu=(Xu)^TXu=\langle Xu,Xu\rangle$$

$\Leftarrow$ $\|Xu\|=0 \iff Xu=0$ (לפי הגדרת הנורמה)

$\Leftarrow$ $u\in\ker X$ ומכיוון שכבר ראינו כי $\langle v,u\rangle\neq 0$ אזי

$v\notin\ker(X)^\perp$. $\Leftarrow$ $\ker(X)^\perp\subseteq Im(X^T)$

$\Leftarrow$ $\ker(X^\perp)=Im(X^T)$, כנדרש. ■

15) נתונה המערכת $X^Tw=y$ ונניח כי $X^T$ איננה הפיכה אזי

נראה כי למערכת יש $\infty$ פתרונות $\iff$ $y\perp\ker(X)$:

"$\Leftarrow$": נניח כי למערכת יש $\infty$ פתרונות ולכן לכל פתרון מתקיים כי $y\in Im(X^T)$ ולכן משאלה 14 נקבל כי $y\in\ker(X)^\perp$ כלומר לכל $v\in\ker X$ $\langle y,v\rangle=0$ כלומר $y\perp\ker X$, כנדרש.

"$\Rightarrow$": נניח כי $y\perp\ker X$ אזי $y\in\ker(X)^\perp$ ולכן משאלה 14 $y\in Im(X^T)$ כלומר קיים $u\in\mathbb{R}^m$ כך ש- $X^Tu=y$ הינו פתרון לבעיה.

ומכיוון ש- $X^T$ איננה הפיכה אזי // למערכת יש 0 פתרונות // $\infty$ פתרונות (משפט מלינארית) ומכיוון שכבר ראינו פתרון אחד אזי בהכרח קיימים $\infty$ פתרונות, כנדרש. ■

16) נתונה המשוואה $XX^Tw = Xy$

- אם $XX^T$ הפיכה אזי לפי לינארית 1 יש פתרון יחיד

וניתן לראות זאת כי אם $XX^T$ הפיכה אז

$$\underbrace{(XX^T)^{-1}XX^T}_{I}w = (XX^T)^{-1}Xy$$

$w = (XX^T)^{-1}Xy$ , כלומר $w$ פתרון יחיד. ■

- נניח כי $XX^T$ אינה הפיכה. יהי $y \in Im(X^T)$

אזי משאלה 14 $ker(X)^\perp = Im(X^T)$ ולכן $y \in ker(X)^\perp$

ובתרגול ראינו כי $ker(X) = ker(XX^T)$. לכן $y \in ker(XX^T)$ כלומר

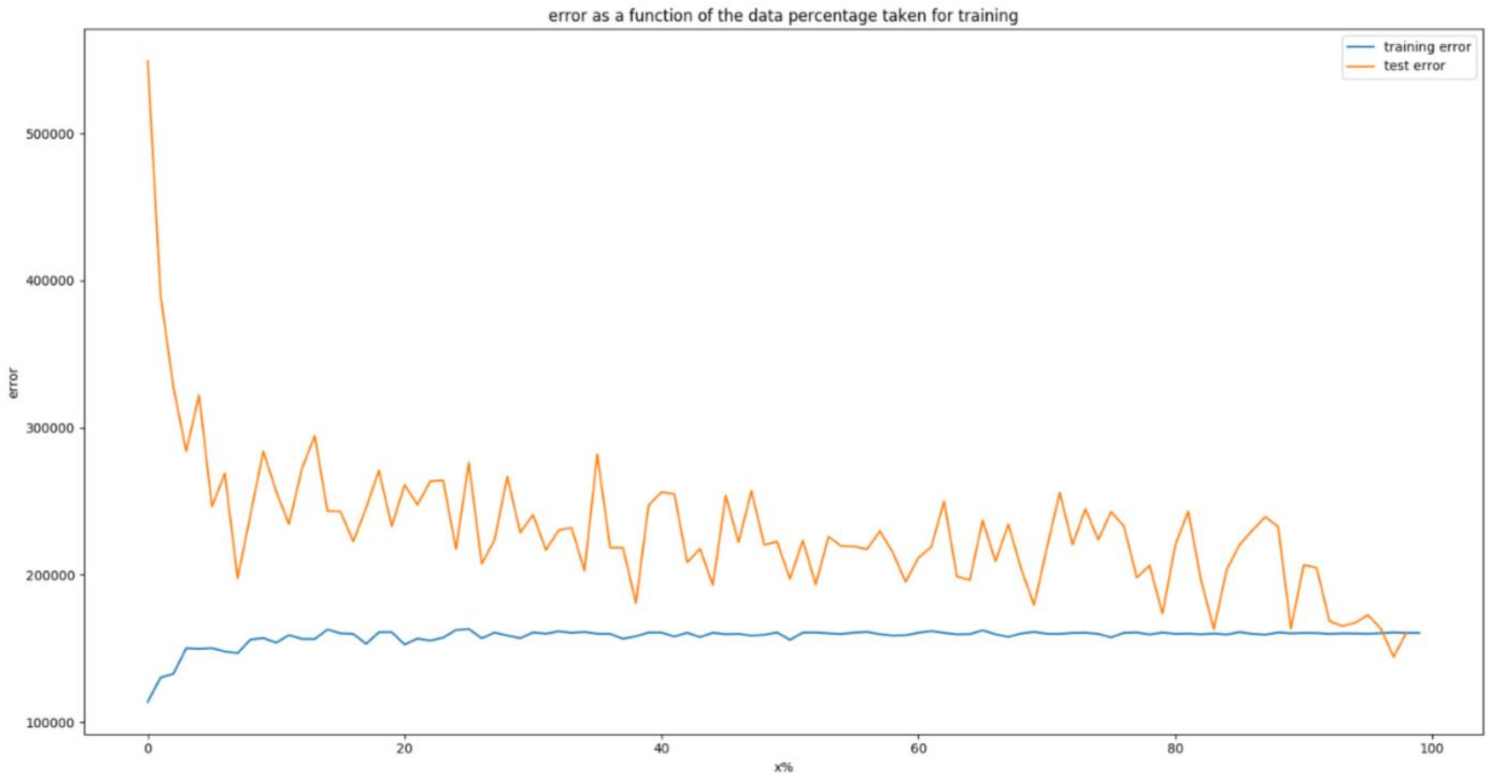
$y \perp ker(XX^T)$ ולכן משאלה 15, המערכת $X^Tw = y$ יש $\infty$ פתרונות ■

17. After looking in depth at the data I realized that there are irrelevant negative prices and ‘sqft lot 15’ and also houses with MUCH more bedrooms than the majority so I dropped all those. Next I dropped any house which had any NAN in one of its parameters.

18. I decided that ID and Date were irrelevant for the learning process so I dropped those entire columns. I also realized that the ‘sqft living’ colomn is linearly dependent with ‘sqft above’ and ‘sqft basement’ so I dropped ‘sqft living’. Finally, I realized that there is a limited amount of zipcodes and that it is hard to understand any numeric value from a zipcode so I used the ‘get_dummies’ function on it and changed it into a binary option of the different zipcode possibilities.

19+20: see attached code files.

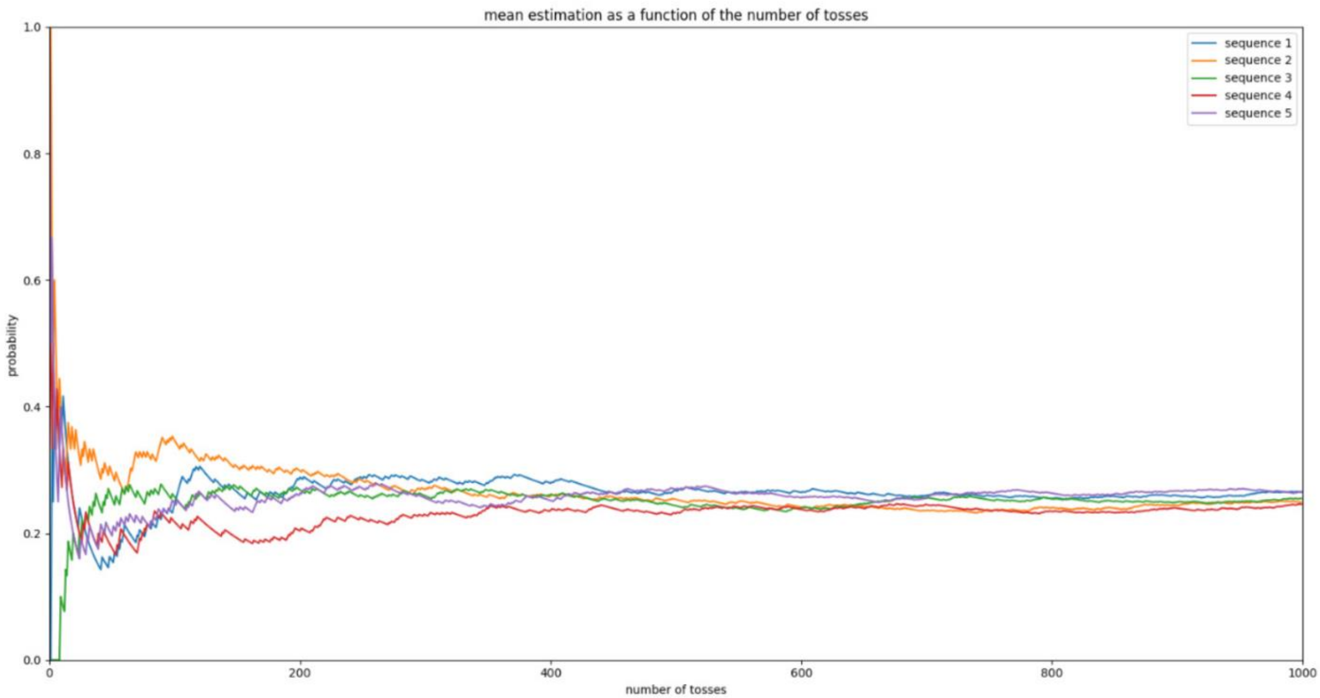
21. Train and test errors as functions of x (data percentage):



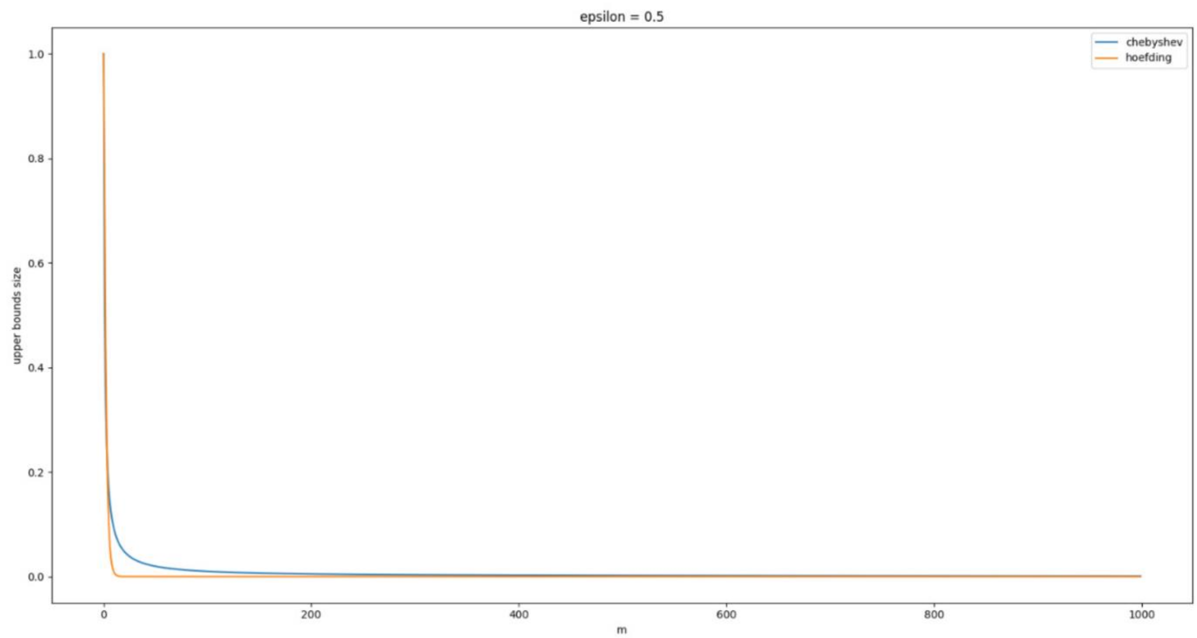

22) לפי מה שראינו בתרגול . $m \geq \left\lceil \frac{1}{2\varepsilon^2} \ln\left(\frac{2}{\delta}\right) \right\rceil$

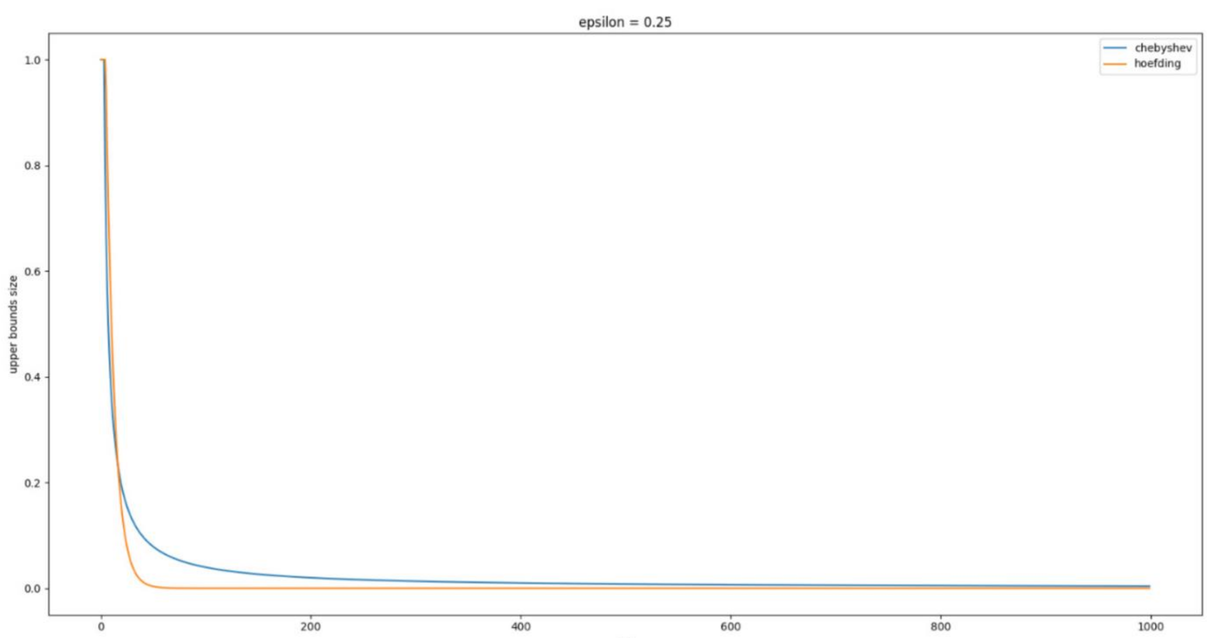
כלומר עבור כמות דוגמאות שקולה ל-m הנ"ל נקבל דיוק כרצוננו. ■

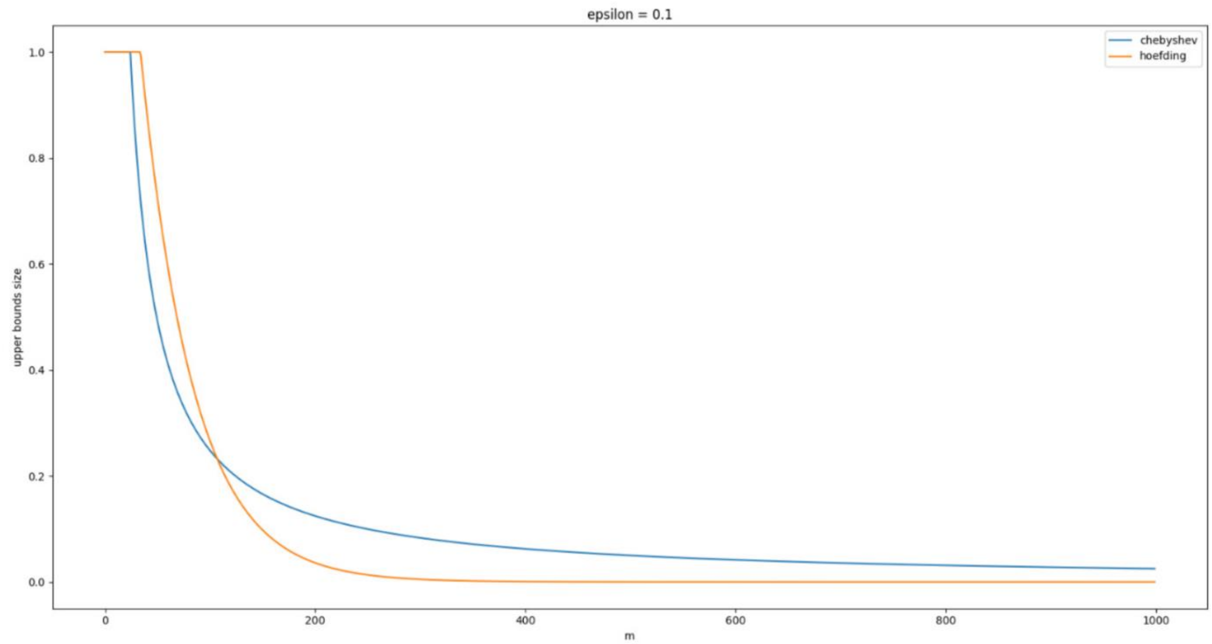
23. a) As m grows, I expect to see that the graph of each sequence should converge to the given statistical bias of the coins of p=0.25 (and this does happen as seen in the following graph):

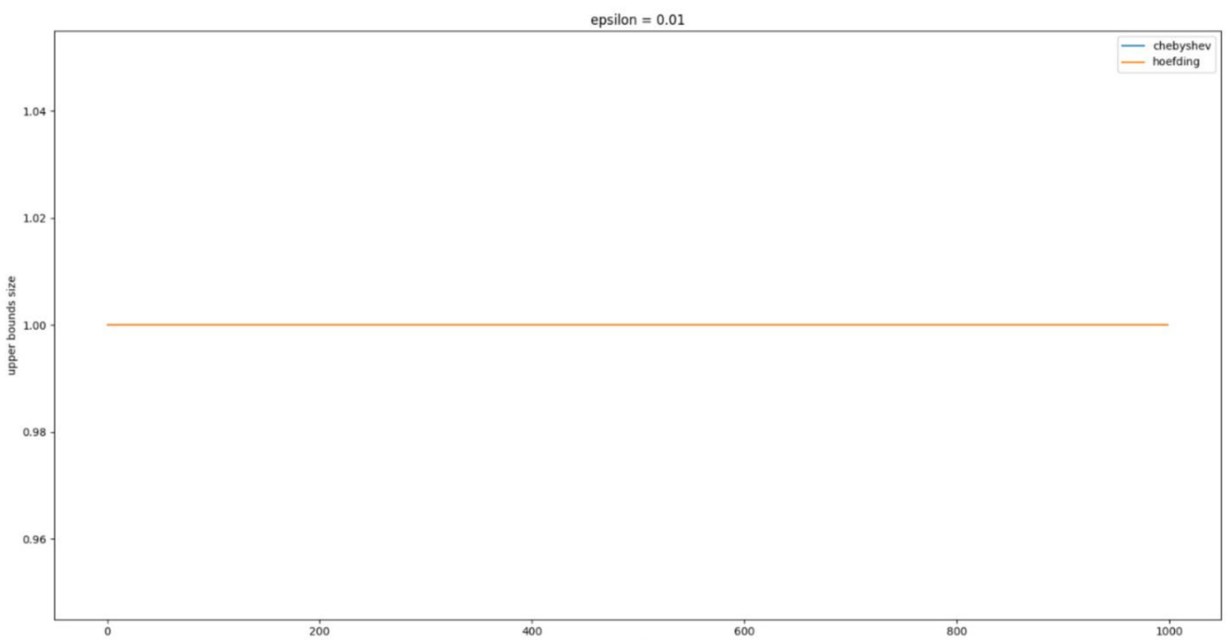


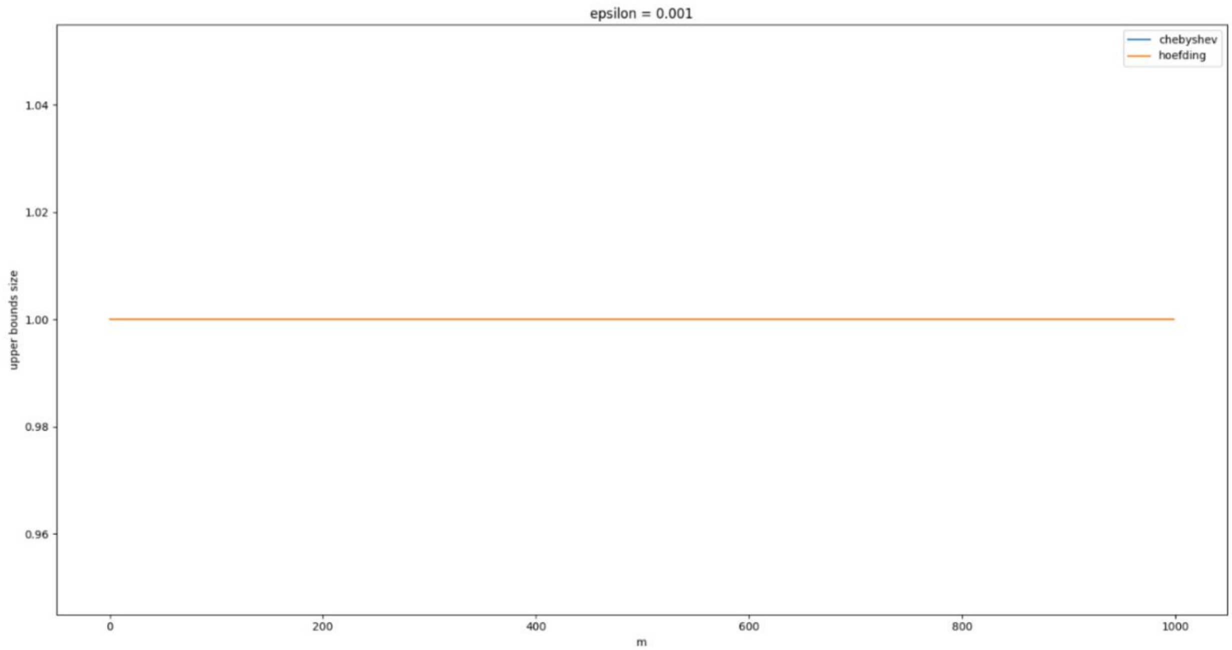
23 b) The next 5 figures show the upper bounds of the Hoeffding and and Chebyshev innequlities, as tought in class, per epsilon (while epsilon varies):



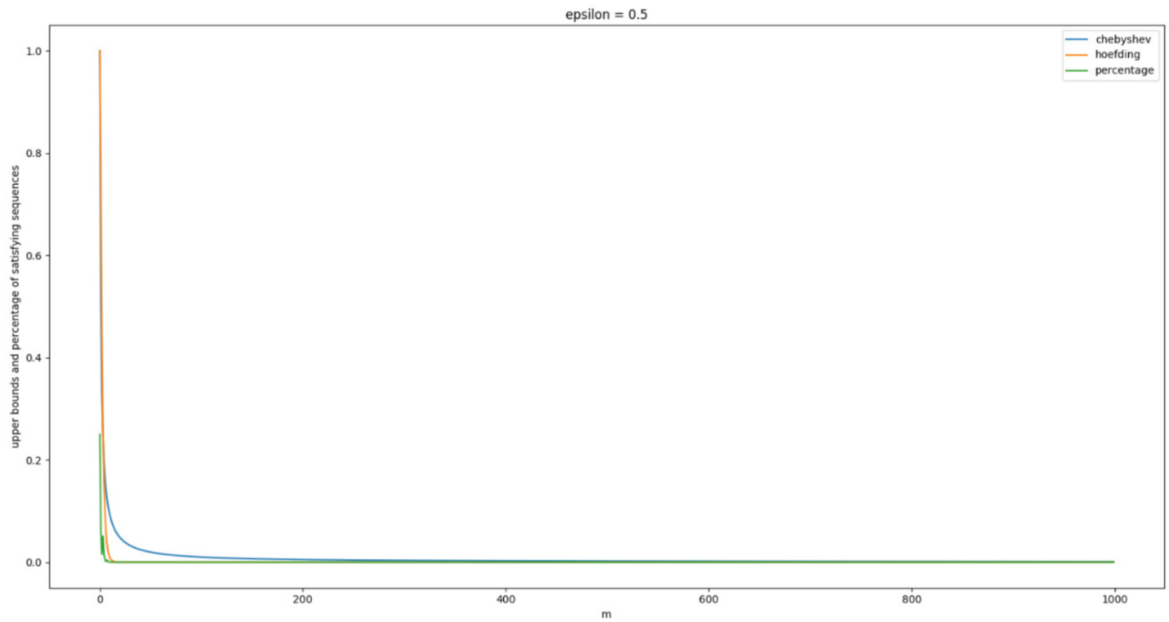

epsilon = 0.1
chebyshev
hoefding
1.0
0.8
0.6
0.4
0.2
0.0
upper bounds size
0
200
400
600
800
1000
m

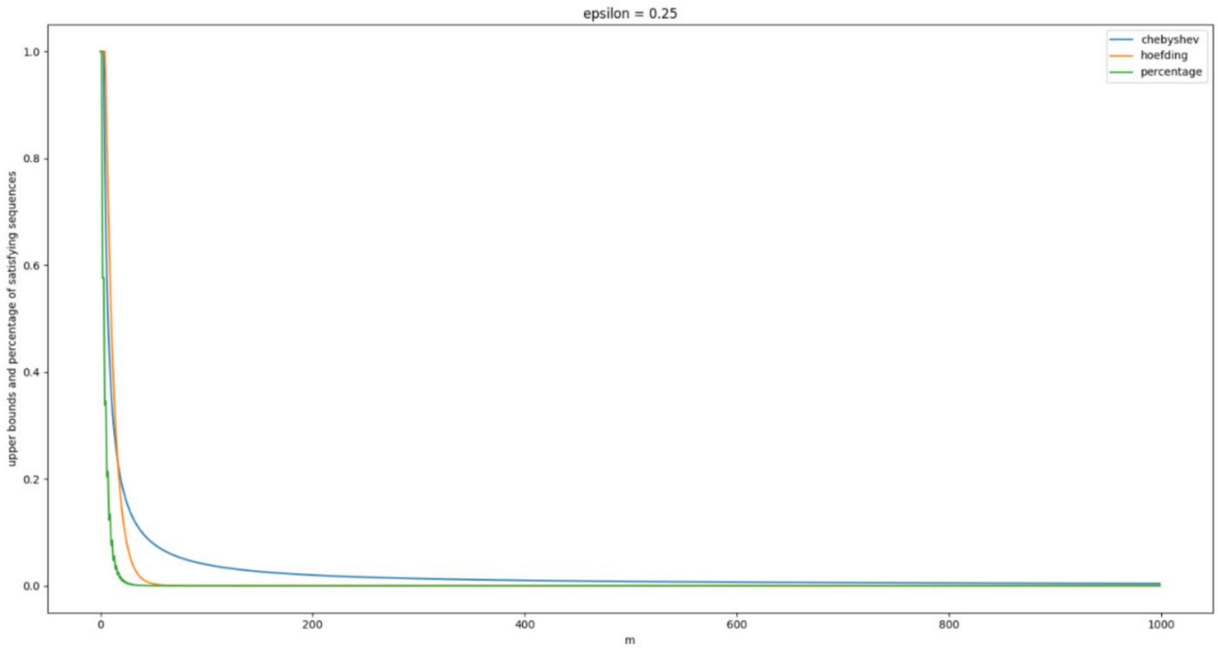

epsilon = 0.01
chebyshev
hoefding
1.04
1.02
1.00
0.98
0.96
upper bounds size
0
200
400
600
800
1000
m

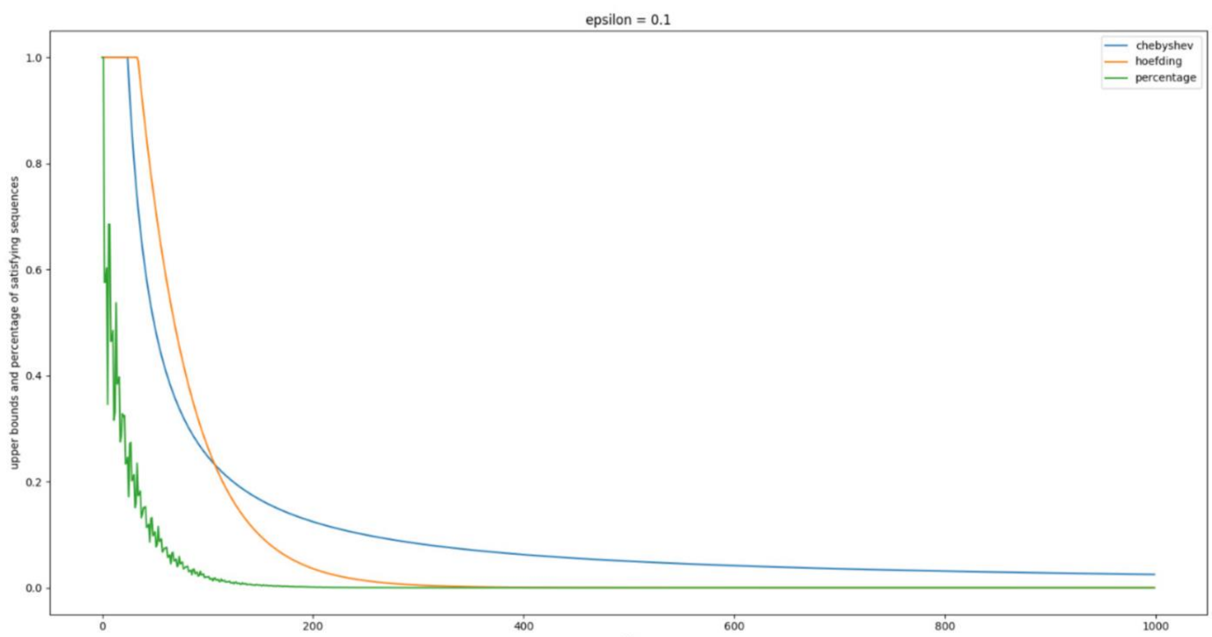


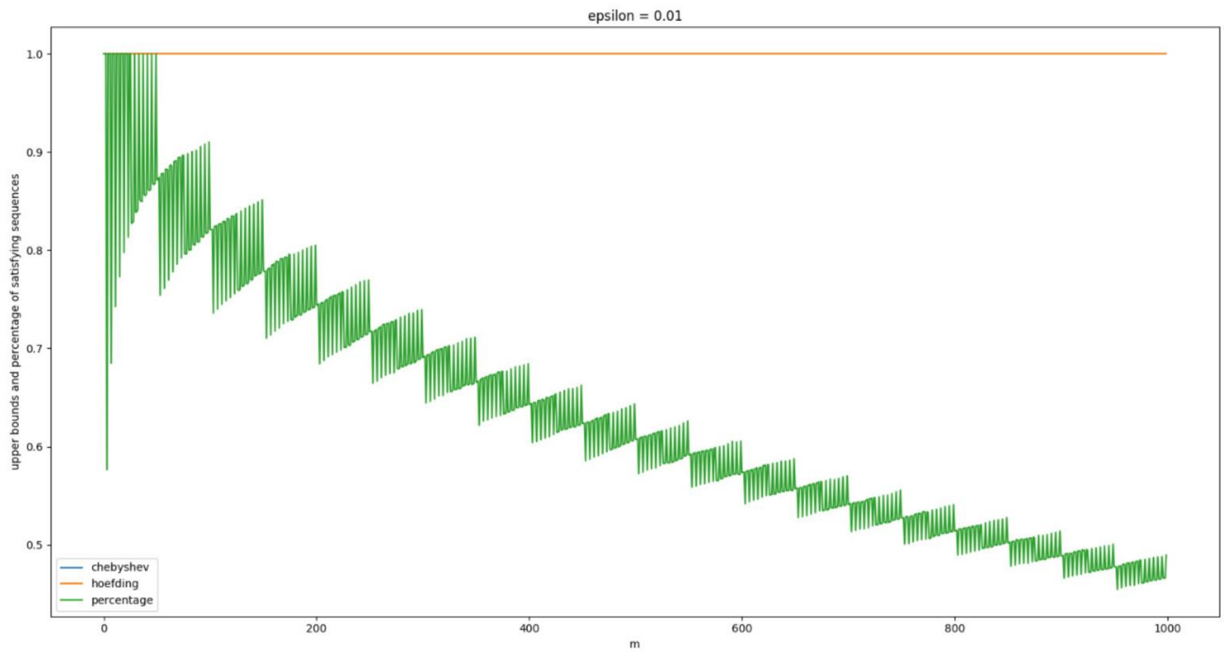

23. c) The next 5 figures show the percentage of sequences that satisfy $|\overline{X}_m - \mathbb{E}[X]| \geqslant \epsilon$ per epsilon (varies), beside the previous upper bound graphs, and while we now know that $\mathbb{E}[X] = p = 0.25$

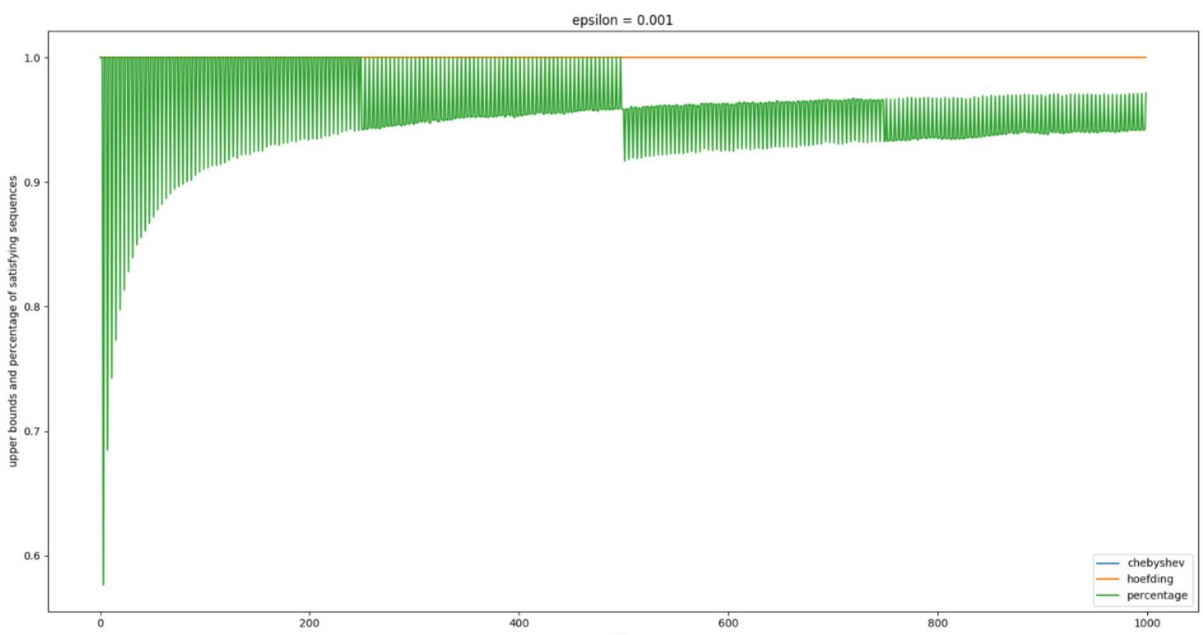
As such and based on section (a) I would expect to see that as m (the number of tosses) grows then the $|\overline{X}_m - \mathbb{E}[X]| \geqslant \epsilon$ difference grows smaller (because of the converging as seen in section a) and so the percent of sequences satisfying such should grow smaller. But as epsilon grows smaller, the tolerance is more strict and as such the percent of sequences should grow larger per m (in comparison to the same spot of m when calculated with larger epsilons):


epsilon = 0.5
chebyshev
hoefding
percentage
upper bounds and percentage of satisfying sequences
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
m


epsilon = 0.25
chebyshev
hoefding
percentage
upper bounds and percentage of satisfying sequences
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
m


epsilon = 0.1
chebyshev
hoefding
percentage
upper bounds and percentage of satisfying sequences
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
m


epsilon = 0.01
upper bounds and percentage of satisfying sequences
m
chebyshev
hoefding
percentage
1.0
0.9
0.8
0.7
0.6
0.5
0
200
400
600
800
1000


epsilon = 0.001
upper bounds and percentage of satisfying sequences
m
chebyshev
hoefding
percentage
1.0
0.9
0.8
0.7
0.6
0
200
400
600
800
1000