6 תהי A מטריצה הפיכה, האליפסואיד המוגדר ע"י A הוא הקבוצה:

$$\{y : y = Ax,\ \|x\| = 1\} = A\cdot\{x : \|x\| = 1\} = A\cdot B_1(0)$$

ומספיק שנראה כי:

$$U\cdot B_1(0) = B_1(0)$$

ראשית:

$$U\cdot B_1(0) \subseteq B_1(0)$$

יהי ולכל $y \in U\cdot B_1(0)$

$y = Ux$

$x \in B_1(0)$

ומכך ש-U אורתוגונלית (ולכן איזומטריה) נסיק כי

$$\|y\| = \|Ux\| = \|x\| = 1$$

בכיוון השני, יהי $x \in B_1(0)$ U אורתו' ולכן הפיכה וגם $U^{-1} = U^T$ אורתו'

ולכן איזומטריה ובפרט

$$x = UU^Tx$$

ו- $\|U^Tx\| = 1 \Rightarrow U^Tx \in B_1(0)$

לכן $x = U(U^Tx) \in U\cdot B_1(0)$ קיבלנו שוויון

נסיק מאסוציאטיביות

$$AU\cdot B_1(0) = A(U\cdot B_1(0)) = A\cdot B_1(0)$$

7 יהי H מישור המוגדר ע"י $\langle w,x\rangle = b$ ויהי x בכל המישור הזה

ההטלה של x על w היא:

$$P = \frac{\langle x,w\rangle}{\|w\|^2}w = \frac{b}{\|w\|^2}w$$

נראה כי זאת הנק' הכי קרובה בכל המישור ל-0 ולכן המרחק הוא $\|p\| = \frac{b}{\|w\|}$

יהי y בכל המישור. נרשום את y כ: $y = y - \frac{b}{\|w\|^2}w + \frac{b}{\|w\|^2}w$

ונראה זה כי:

$$\left\langle y - \frac{b}{\|w\|^2}w \,\middle|\, \frac{b}{\|w\|^2}w\right\rangle = \langle y|w\rangle - b\frac{\langle w|w\rangle}{\|w\|^2} = b - b = 0$$

לכן מפיתגורס:

$$\|y\| = \left\|y - \frac{b}{\|w\|^2}w\right\| + \left\|\frac{b}{\|w\|^2}w\right\| \ge \frac{b}{\|w\|^2}\|w\| = \frac{b}{\|w\|}$$

משמע, $\|p\| \le \|y\|$ בכלל, p בכל מישור ולכן $\|p\|$ הוא המרחק מהראשית.

9) $A = \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix} =$

$AA^T \Rightarrow \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix} \begin{pmatrix} 5 & -1 \\ 5 & 7 \end{pmatrix} = \begin{pmatrix} 50 & 30 \\ 30 & 50 \end{pmatrix}$

lets find the singular values of $\left(AA^T - \lambda I \right) =$

$$\begin{matrix} 50-\lambda & 30 \\ 30 & 50-\lambda \end{matrix} \Rightarrow (50-\lambda)^2 - 900 =$$

$$2500 - 100\lambda + \lambda^2 - 900 =$$

$$= \lambda^2 - 100\lambda + 2410 = (\lambda - 20)(\lambda - 80)$$

$$\Rightarrow 20, 80$$

[illegible] ע"ע

$\Rightarrow \Sigma = \begin{pmatrix} 80 & 0 \\ 0 & 20 \end{pmatrix} =$ how lets find the ker:

$\Rightarrow \begin{pmatrix} 50 & 30 \\ 30 & 50 \end{pmatrix} - \begin{pmatrix} 20 & 0 \\ 0 & 20 \end{pmatrix} = \begin{matrix} 30 & 30 \\ 30 & 30 \end{matrix}$ דירוג $\Rightarrow \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ ו"ע

$\Rightarrow \begin{matrix} 50 & 30 \\ 30 & 50 \end{matrix} - \begin{pmatrix} 80 & 0 \\ 0 & 80 \end{pmatrix} \Rightarrow \begin{matrix} -30 & 30 \\ 30 & -30 \end{matrix}$ דירוג $\Rightarrow \begin{matrix} -1 & 1 \\ 0 & 0 \end{matrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ו"ע

$A^T A = \begin{pmatrix} 26 & 18 \\ 18 & 74 \end{pmatrix} \Rightarrow$

$A^T A - \lambda I = \begin{matrix} 26-\lambda & 18 \\ 18 & 74-\lambda \end{matrix} \Rightarrow (26-\lambda)(74-\lambda) - 18^2 \Rightarrow$

ע"ע 20,80 [illegible]

$\begin{matrix} 26 & 18 \\ 18 & 74 \end{matrix} - \begin{matrix} 20 & 0 \\ 0 & 20 \end{matrix} \Rightarrow \begin{matrix} 6 & 18 \\ 18 & 54 \end{matrix}$ דירוג $\Rightarrow \begin{matrix} 1 & 3 \\ 0 & 0 \end{matrix} \Rightarrow$ ו"ע $\begin{pmatrix} -3 \\ 1 \end{pmatrix}$

$\begin{matrix} 26 & 18 \\ 18 & 74 \end{matrix} - \begin{matrix} 80 & 0 \\ 0 & 80 \end{matrix} = \begin{matrix} -54 & 18 \\ 18 & -6 \end{matrix}$ דירוג $\Rightarrow \begin{matrix} -3 & 1 \\ 0 & 0 \end{matrix} \Rightarrow$ ו"ע $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$

לפי קודם:

(4 המשך

$$AA^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 80 & 0 \\ 0 & 20 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

אלה הן כיוונים [illegible]
בהן שמירה של עוצמת הנ"ס

$$A^T A = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} 20 & 0 \\ 0 & 20 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \end{bmatrix}$$

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

$$\Rightarrow \Sigma = \begin{bmatrix} \sqrt{80} & 0 \\ 0 & \sqrt{20} \end{bmatrix}$$

(מטריצה זו
אלכסונית)

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

$$A = U \Sigma V$$

Scanned by CamScanner

14) $Im(X^T) = ker(X)^\perp$

יהי $V \subseteq \mathbb{R}^d$ תת מרחב, נגדיר את המשלים האורתוגונלי של V כ:

$V^\perp = \{x\in\mathbb{R}^d \mid \langle x,v\rangle = 0 \ \forall v\in V\}$

נוכיח כי מתקיים $(Im(X^T))^\perp = ker(X) \iff Im(X^T) = ker(X)^\perp$ כלומר מספיק להראות

כי מתקיים $(V^\perp)^\perp = V$

לכן נוכיח כי מתקיים $(Im(X^T))^\perp = ker(X)$

תחילה נראה כי מתקיים $ker(X) \subseteq Im(X^T))^\perp$

יהי $u\in ker(X)$ ו $v\in V$ אז:

$$Xu=0 \Rightarrow \langle Xu, v\rangle = 0 \Rightarrow \langle u, X^T v\rangle = 0 \Rightarrow \langle u, Im(X^T)\rangle$$

כלומר אנו רואים כי המכפלה הפנימית $\langle u, X^T u\rangle$ מתאפסת לכל

ומכאן מתקיים $u\in Im(X^T)^\perp$

כלומר $ker(X) \subseteq Im(X^T))^\perp$ כנדרש

$Im(X^T))^\perp \subseteq ker(X)$

יהי $u\in V$, $v\in Im(X^T)^\perp$ אזי מתקיים

$\langle v, X^T u\rangle = 0$

$\Rightarrow \langle Xv, u\rangle = 0$

בפרט, מכאן מתקיים

$$\forall u\in V: \langle Xv, u\rangle = 0$$

אז בהכרח מתקיים כי $Xv=0$ ולכן

$$Xv=0 \Rightarrow v\in ker(X) \Rightarrow (Im(X^T))^\perp \subseteq ker(X)$$

$$\Rightarrow (Im(X^T))^\perp = ker(X) \Rightarrow Im(X^T) = ker(X)^\perp$$

■

Scanned by CamScanner

15. נתבונן במערכת המשוואות $X^T w = y$ כאשר נניח כי X^T אינה הפיכה

טענה: למערכת המשוואות יש אינסוף פתרונות אמ"מ $y \perp \ker(X)$

<u>כיוון ראשון $\Leftarrow$:</u>

נניח כי למערכת יש אינסוף פתרונות. מכיוון שבהכרח מתקיים $\dim(\ker(X)) > 0$

לכן בהכרח קיים וקטור $v \in \ker(X)$

כעת יהי $v \in \ker(X)$. הוכחנו בשאלה 14 כי אז $v \in Im(X^T)^\perp$

בנוסף, מכיוון ומתקיים $X^T w = y$ אזי בהכרח $y \in Im(X^T)$

כעת מכיוון שמתקיים 1. $v \in Im(X^T)^\perp$

2. $y \in Im(X^T)$ אזי בהכרח מתקיים

$$\langle y, v \rangle = 0$$

כלומר סה"כ קיבלנו כי:

$$\forall v \in \ker(X): \langle y, v \rangle = 0 \Rightarrow y \perp \ker(X)$$

<u>כיוון שני $\Rightarrow$:</u>

נניח כי מתקיים $y \perp \ker(X)$ נזכור כי X^T אינה הפיכה ולכן בהכרח מתקיים $\dim(\ker(X^T)) > 0$. יהי $v \in \ker(X)$ אז משאלה 14 מתקיים $v \in Im(X^T)^\perp$

כעת, מכיוון ש $y \perp \ker(X)$ אזי בהכרח

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

$$\langle y, v \rangle = 0 \Rightarrow y \in Im(X^T)$$

כלומר מתקיים כי בהכרח $\exists u \in V$. כך שמתקיים:

$$X^T u = y$$

לבסוף, מכיוון שמתקיים $\dim(\ker(X^T)) > 0$ אזי בהכרח קיימים אינסוף וקטורים w המקיימים $X^T w = 0$

כעת ניקח וקטור $w' = u + w$ ונראה כי מתקיים

$$X^T w' = X^T(u+w) = X^T u + X^T w = y + 0 = y$$

מכאן קיימים אינסוף וקטורים w' המקיימים $X^T w' = y$

כלומר, למערכת יש אינסוף פתרונות, כנדרש ■

ט) עבור מע' המשוואות הנורמליות $XX^Tw=Xy$ נוכיח כי קבוצת משוואות הנ"ל יכול להיות פתרון יחיד אם X^TX הפיכה או אינסוף פתרונות אחרת.

במקרה ומתקיים כי XX^T הפיכה נקבל:

$$XX^Tw=Xy \Rightarrow (XX^T)^{-1}XX^Tw=(XX^T)^{-1}Xy \Rightarrow w=(XX^T)^{-1}Xy$$

כפתרון ל $(XX^T)^{-1}Xy$

מכיוון ש XX^T הפיכה מתקיים כי $0=\ker(XX^T) \Rightarrow \ker(X^T)=\ker(XX^T)=0$ *

ולכן $\ker(X)=0$ מכאן, הביטוי $(XX^T)^{-1}Xy$ מוגדר בצורה יחידה ולכן פתרון

w הינו פתרון יחיד כנדרש.

אם XX^T איננה הפיכה, אזי מס' 13 מתקיים כי ישנם למע' המשוואות אינסוף

פתרונות אמ"מ מתקיים

$$Xy \perp \ker((XX^T)^T) \Rightarrow Xy \perp \ker(XX^T)$$

היות ו-* לכן מע' משוואות יש אינסוף פתרונות אמ"מ מתקיים

$$Xy \perp \ker(X^T)$$

נשים לב כי אם מתקיים $Xy \perp \ker(X^T)$ אזי בפרט

$Xy \in \ker(X^T)^\perp$

משאלה 14 מתקיים

$$Xy \in \ker(X^T)^\perp \Rightarrow Xy \in Im(X^{TT}) \Rightarrow Xy \in Im(X)$$

וכמובן הביטוי $Xy \in Im(x)$ מתקיים באופן ריק שכן Xy שייך לתמונה

של X עבור כל y שניתן להכפיל ב X.

מכאן מתקיים תמיד כי יש אינסוף פתרונות לקבוצת המשוואות הנ"ל

כאשר XX^T איננה הפיכה ∎

Scanned by CamScanner

22

ש. יהי מלבן סגור הגבול R ידוע. אנו מעוניינים שלמדן
עם דיוק ε וביטחון δ. מהו המספר (m)
הדגימות ה-i.i.d. הנדרשות כדי להבטיח דיוק ε וביטחון δ

ע"י שימוש בחסם הופדינג עם הדגימות הללו שיבטיח דיוק ε
וביטחון δ:

$$m_1(\varepsilon,\delta) \geq \left\lceil \frac{1}{2\varepsilon^2}\log\left(\frac{2}{\delta}\right) \right\rceil$$

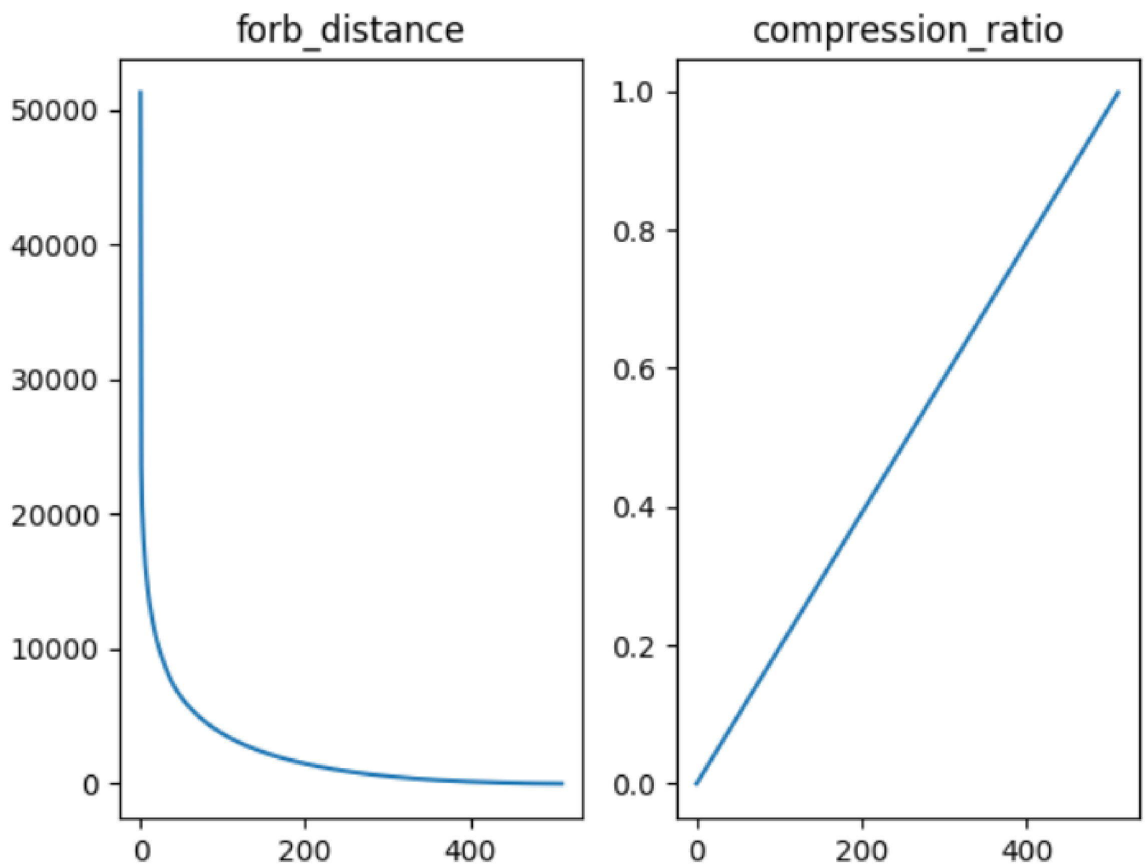
ע"י שימוש בחסם צ'בישב עם הדגימות הנ"ל
שיבטיח אותו הדבר

$$m_2(\varepsilon,\delta) \geq \left\lceil \frac{1}{4\delta\varepsilon^2} \right\rceil$$

לסיכום נקבל שיתקיים

$$m(\varepsilon,\delta) = \min\{m_1(\varepsilon,\delta), m_2(\varepsilon,\delta)\}$$

Question 10


forb_distance
50000
40000
30000
20000
10000
0
0
200
400
compression_ratio
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400

Question 11

0.comp: 0.0390625
norma: 10606.1853
k: 20

1.comp: 0.05859375
norma: 8707.5982
k: 30

2.comp: 0.1953125
norma: 3686.4259
k: 100

3.comp: 0.5859375
norma: 546.7496
k: 300

4.comp: 0.9765625
norma: 3.9496
k: 500

Question 12

Explanation for question 10:

The compression ratio has linear behavior (so in the lecture) as a function of the compression factor. Furthermore, we will expect to see the ratio getting bigger as the compression factor would grow. Since as 'k' (a.k.a the compression factor) grows, we lose less information.

Explanation for question 11:

Because the compression factor is representing the amount of the singular values of the original matrix M.

The different plotting of the picture according of the size off 'k' can be explained as such:

When k grows, the image become clearer. This happens since as 'k' grows, as less information we throw, (as 'k' grows) as more information is preserved in the compressed image (in particular – the compression is smaller) and the difference between the original and the compressed image becomes smaller and smaller. When performing the SVD, the singular values matrix Σ that is created, It upholds the following feature: $\forall i \in \sigma_{i,i} \geq \sigma_{i+1,i+1}$

Each singular value is equal or greater than its following singular value.
For here, we can explain the fact that difference between the different images decreases as 'k' grows (for example – the difference between the image with k=15 and k=50, seems much larger than the difference between an image with k=250 and k=450)

שאלה 17:

1. בדקתי האם ישנם נתונים שאינם הגיוניים ואם היו, ניקיתי את השורה.
2. בפרט, וידאתי שכל מה שמייצג מספר (מחיר, מספר שירותים, מספר חדרים וכו') הוא גדול מ0. כי אני מצפה שלא יכול להיות שלא יהיה כך.
3. אם ראיתי ששטח מחייה של דירה אינו יותר מ250 ראיתי זאת כערך לא הגיוני.
4. לבסוף אם ראיתי שורה שקיים בה ערך non אז הורדתי אותה גם כן.

שאלה 18:

1. בדקתי את מידע בטבלה ביצירת גרף עבור תכונה ספציפית והמחיר וניסיתי לראות האם קיימת קורלציה בין השניים.
2. גיליתי כי עבור ה ID, DATE, WATERFRONT, VIEW, LAT, LONG לא קיימת קורלציה בינם לבין המחיר.
3. יצרתי עבור ה ZIPCODE תכונות נוספות של dummy features בינאריים שנתנה לי עוד כ70 עמודות. (עוד סיבה שהורדנו את האורך והרוחב כי הייתה פה מעין כפילות.)
4. לבסוף, הוספתי עמודת w_0 של אחדות כפי שלמדנו.

.

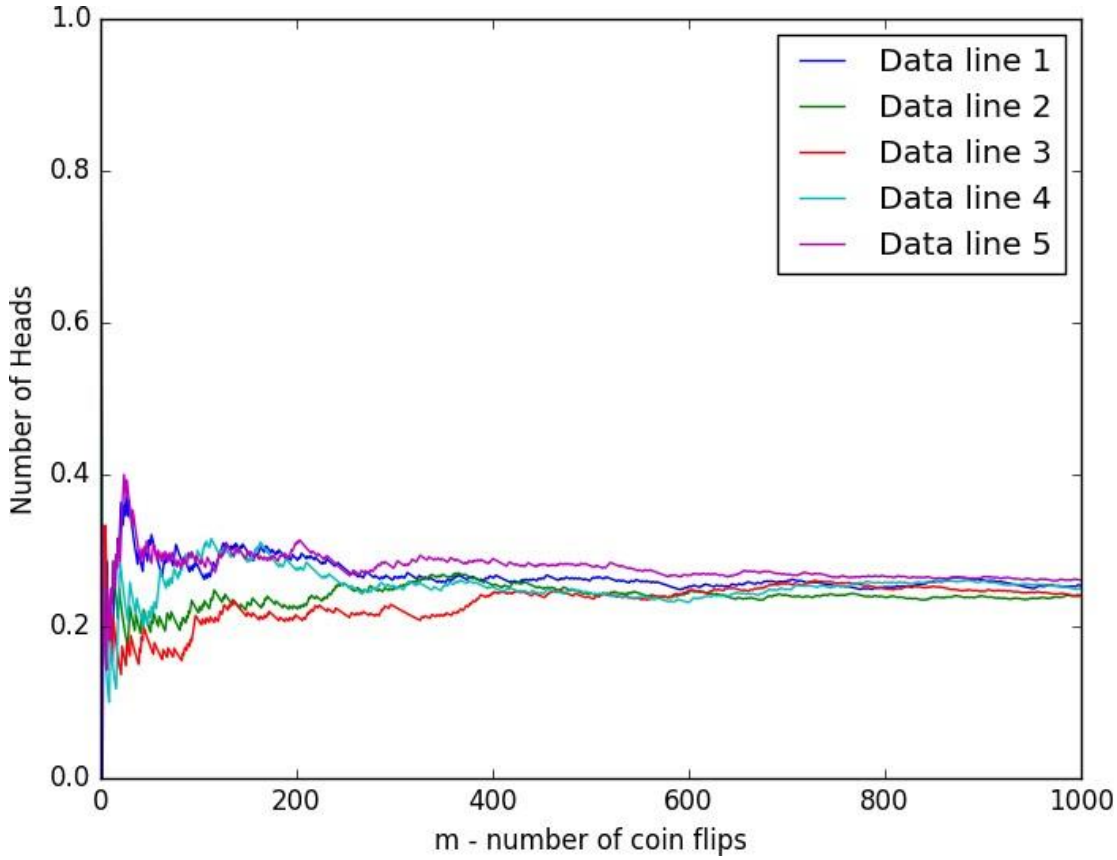
שאלה 21:

train error(g) vs test error(r)
1e11
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100

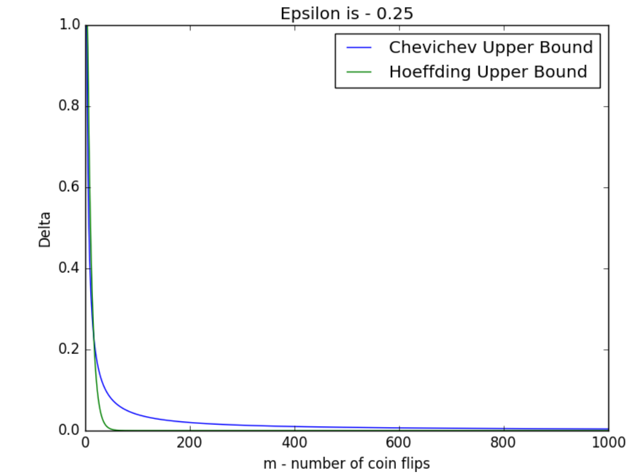
שאלה 23:

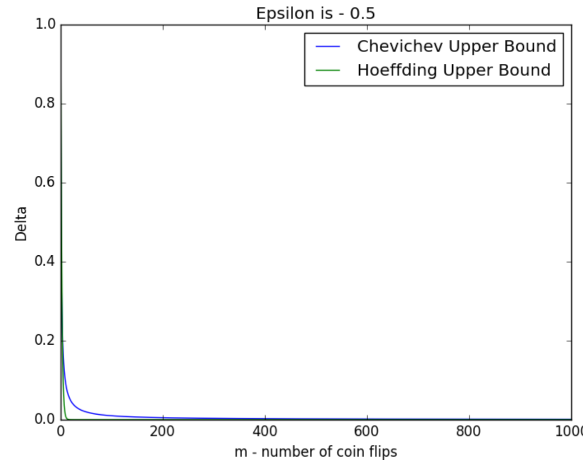
אני מצפה שככל ש m גדל (ככל שבוצעו יותר הטלות) כל בסדרת ההטלות השונות שורות (1-5) ההפרש בין הערכים המתקבלים עבורם יהיו קטנים יותר. כלומר, הגרף המייצג שלהם ישאף להתלכדות על גרף התוחלת שלהם y=0.25

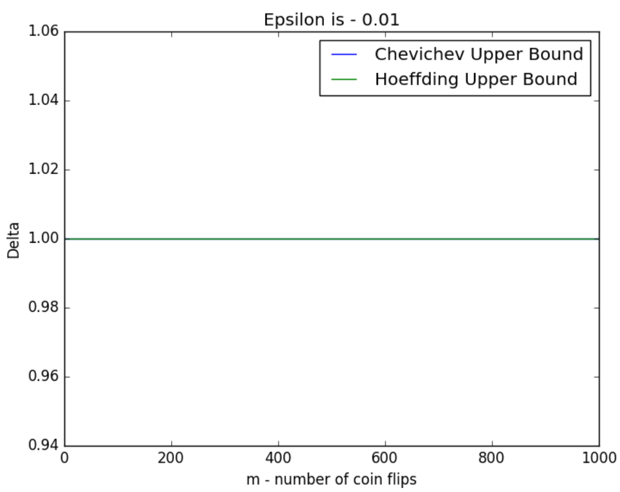
באופן דומה ככל ש m קטן כך הגרפים עשויים לקבל ערכים רחוקים יותר כפונקציה של m

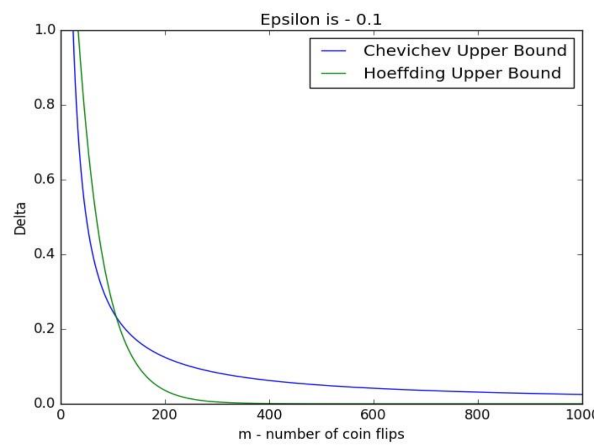


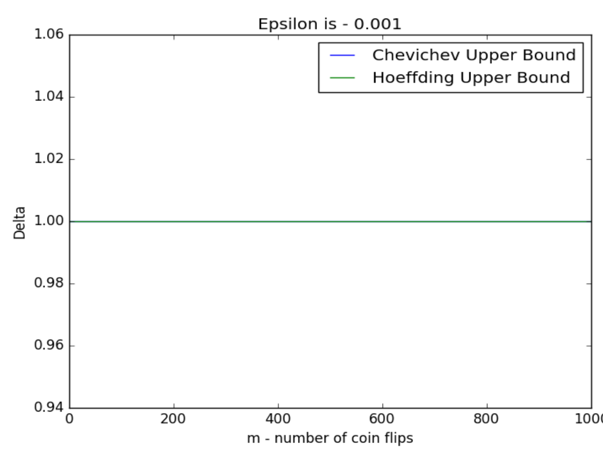
:B


Epsilon is - 0.25
Chevichev Upper Bound
Hoeffding Upper Bound
Delta
m - number of coin flips
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000


Epsilon is - 0.5
Chevichev Upper Bound
Hoeffding Upper Bound
Delta
m - number of coin flips
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000


Epsilon is - 0.01
Chevichev Upper Bound
Hoeffding Upper Bound
Delta
m - number of coin flips
0.94
0.96
0.98
1.00
1.02
1.04
1.06
0
200
400
600
800
1000


Epsilon is - 0.1
Chevichev Upper Bound
Hoeffding Upper Bound
Delta
m - number of coin flips
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000


Epsilon is - 0.001
Chevichev Upper Bound
Hoeffding Upper Bound
Delta
m - number of coin flips
0.94
0.96
0.98
1.00
1.02
1.04
1.06
0
200
400
600
800
1000

C

אני מצפה שלכל אפסילון ככך שנבצע יוותר הטלות כך X_m יתקרב לתוחלת שלו שהיא P ועל כן עבור אותם m'ים קטנים יותר יש יותר סיכוי לקבל שגיאה (לחרוג מאפסילון) וככל שה m גדל כל הסיכוי לחרוג מאפסילון קטן.

ההבדל בין האפסילונים הוא שערך האפיסלון ישפיע על מס ההטלות שדרוש עבור כל סדרת הטלות כדי שנחרוג פחות מאפסילון, כלומר ככל שאפסילון קטן כך נצטרך יותר הטלות (להגדיל את m) כדי להקטין את השגיאה.