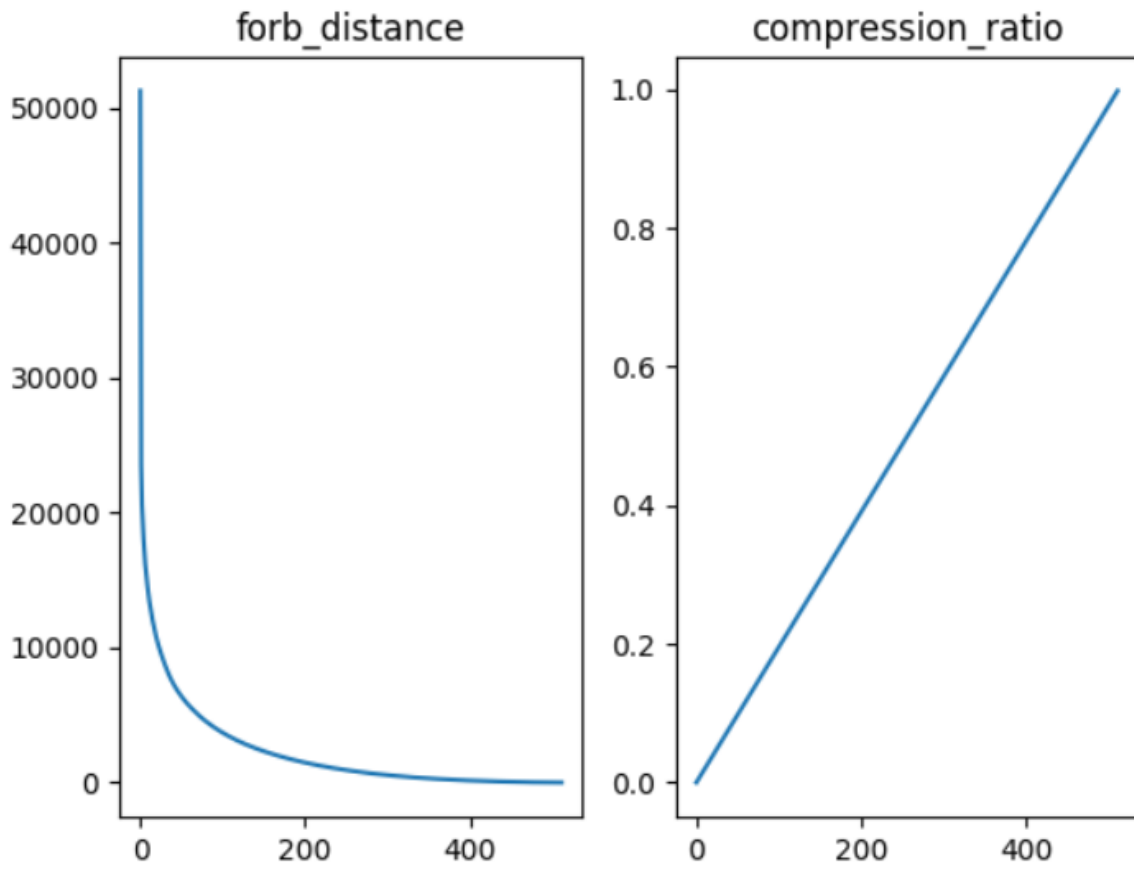


Question 10



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  $\Sigma$  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.

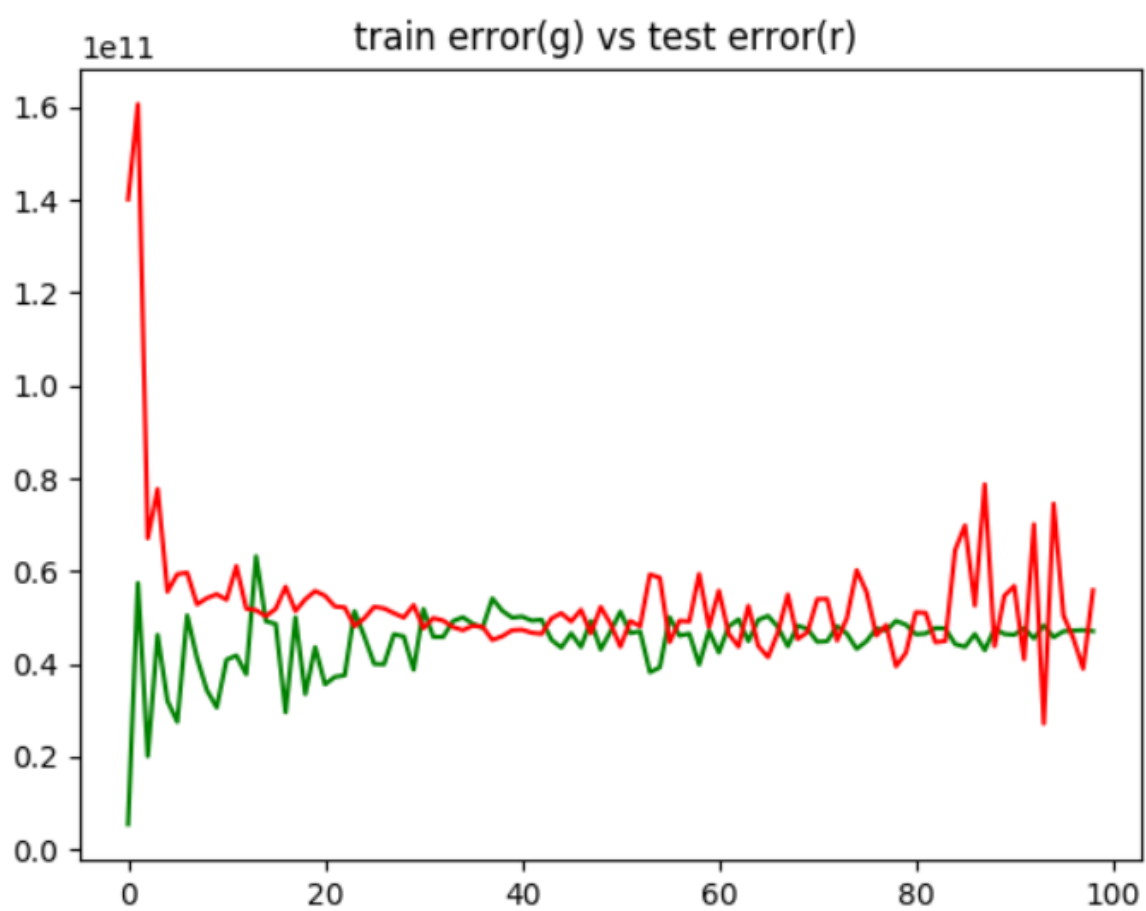
For here, we can explain the fact that difference between the different images decreases as 'k' grows (for example – the difference between the image with k=15 and k=50, seems much larger than the difference between an image with k=250 and k=450)

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:

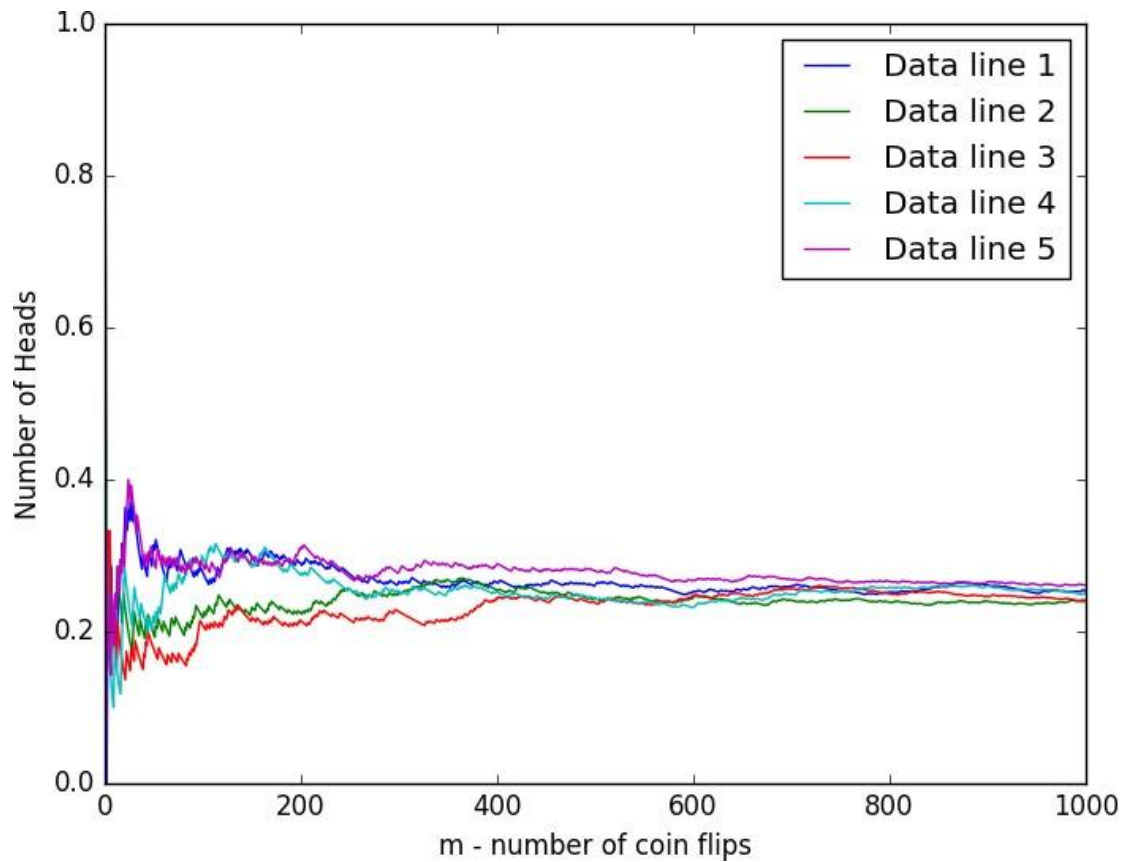
1. בדקתי את מידע בטבלה ביצירת גרף עבור תכונה ספציפית והמחיר וניסיתי לראות האם קיימת קורלציה בין השניים.
2. גיליתי כי עבור ה ID, DATE, WATERFRONT, VIEW, LAT, LONG לא קיימת קורלציה בינם לבין המחיר.
3. יצרתי עבור ה ZIPCODE תכונות נוספות של dummy features בינאריים שנתנה לי עוד כ70 עמודות. (עוד סיבה שהורדנו את האורך והרוחב כי הייתה פה מעין כפילות).
4. לבסוף, הוספתי עמודת w\_0 של אחדות כפי שלמדנו.



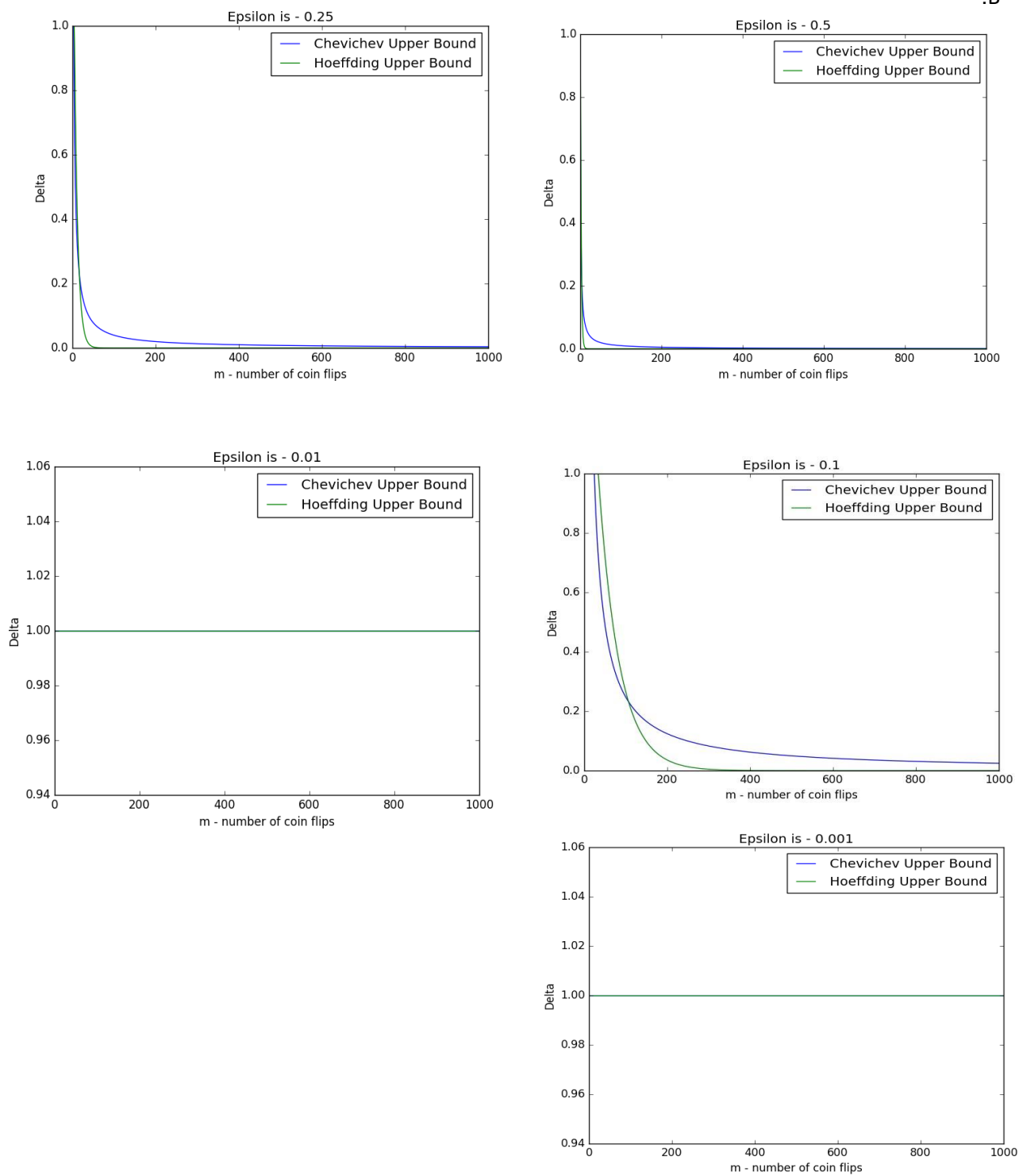
שאלה 23:

אני מצפה שכל  $m$  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  $\gamma=0.25$

באופן דומה כל  $m$  קטן כך הגרפים עשויים לקבל ערכים רחוקים יותר כפונקציה של  $m$



:B



C

אני מצפה שלכל אפסילון ככך שנבצע יותר 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 האפסילון 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.

