תרגיל 2 – תומר ניסים 313232845

# חלק תיאורטי

## פתרונות למשוואות הנורמליות

1. נוכיח ש-:
   1. – יהי . אם ה- היחיד ששייך לגרעין הוא וקטור האפס אז מההגדרה הוא גם שייך לגרעין של . עבור – מההגדרה של וקטור בגרעין נקבל ש- ולכן  
      ומכאן נקבל ש-.
   2. – יהי . אם ה- היחיד ששייך לגרעין הוא וקטור האפס אז מההגדרה הוא גם שייך לגרעין של . עבור – מההגדרה של וקטור בגרעין נקבל ש- ולכן  
       מכאן נקבל ש- ולכן .
2. נוכיח שעבור מטריצה ריבועית מתקיים .
   1. - יהי כלומר קיים כך ש-. יהי כלומר . נסתכל על המכפלה הפנימית:  
      1 – הגדרת המכפלה הפנימית. 2 – שחלוף על מספר איברים. 3 - .  
      לכן קיבלנו ש-.
   2. – נסתכל על . יהי לכן מתקיים  
      מכיוון שזה נכון לכל זה נכון גם עבור ולכן נקבל ש:  
      לכן קיבלנו ש-. מהוכחה שעשינו בלינארית 2, כאשר נסתכל על המשלים האורתוגונלי נקבל:  
      1 – עבור מ"ו המשלים האורתוגונלי של המשלים האורתוגונלי של הוא .  
      לכן סה"כ .
3. תהי מערכות משוואות לינאריות לא הומוגניות. נתון ש- ריבועית ולא הפיכה. נוכיח שלמערכת המשוואות יש אינסוף פתרונות אם ורק אם .   
   בלינארית 1 הראנו שמכיוון ש- היא מטריצה לא הפיכה, למערכת המשוואות ישנם אינסוף פתרונות או שאין פתרונות כלל. כמו כן, הראינו שישנם אינסוף פתרונות אם ורק אם . בסעיף 2 הראנו ש- ובאותו אופן . לכן אם ורק אם . וסה"כ קיבלנו:
4. תהי מערכת המשוואות הנורמלית . נראה שלמערכת יש פתרון יחיד כאשר הפיכה ושלמערכת יש אינסוף פתרונות כאשר לא הפיכה.
   1. הפיכה – מכיוון שהמטריצה הפיכה נוכל לפתור את המשוואה ולקבל פתרון יחיד:
   2. לא הפיכה – נשים לב ש- היא ריבועית ולא הפיכה ולכן נסתכל על ההוכחה מסעיף 3. אנחנו צריכים להראות ש- על מנת לקבל שיש אינסוף פתרונות למערכת. בסעיף 1 הוכחנו ש- ולכן ניקח ונסתכל על המכפלה הפנימית:  
      לכן מסעיף 3 למערכת יש אינסוף פתרונות.

## מטריצות הטלה

1. תהיי המטריצה כאשר הינם בסיס אורתונורמלי ל-.
   1. נראה ש- היא סימטרית:  
      מכיוון שמדובר על כפל סקלרים נקבל ש-  
      ולכן היא מטריצה סימטרית.
   2. ראו סעיף קודם - נוכיח אותו ונסתמך על ההוכחה שם על מנת להוכיח את .  
      נניח בשלילה שקיים ערך עצמי . יהי (המרחב העצמי של הערך העצמי ). נסתכל על השוויון שהוכחנו בסעיף ונכפול את כל אחד מהאגפים על :  
      מכיוון ש-נקבל ש- ולכן קיבלנו סתירה לסעיף . מכאן נסיק שהערכים העצמיים היחידים הינם . מכיוון שהווקטורים הינם בסיס אורתונורמלי כל מכפלה . לכן נסתכל על המכפלה :  
      ולכן הערך העצמי שמתאים לווקטורים האלו הוא 1.
   3. נראה ש-. ניתן לכתוב את כך - . מכיוון ש- מתאימים לערך העצמי 1 של נקבל:
   4. מכיוון שנתון ש- הם בסיס אורתונורמלי, כל מכפלה ולכן כאשר נסתכל על נקבל:  
      1 – הנורמה של וקטור בבסיס אורתונורמלי היא 0.
   5. נשתמש בסעיף ונקבל:

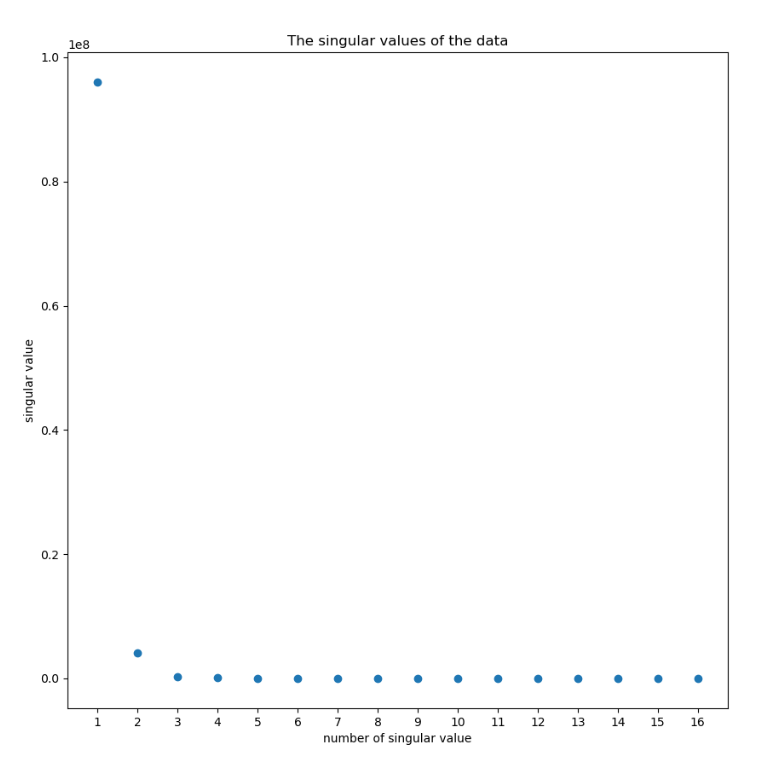
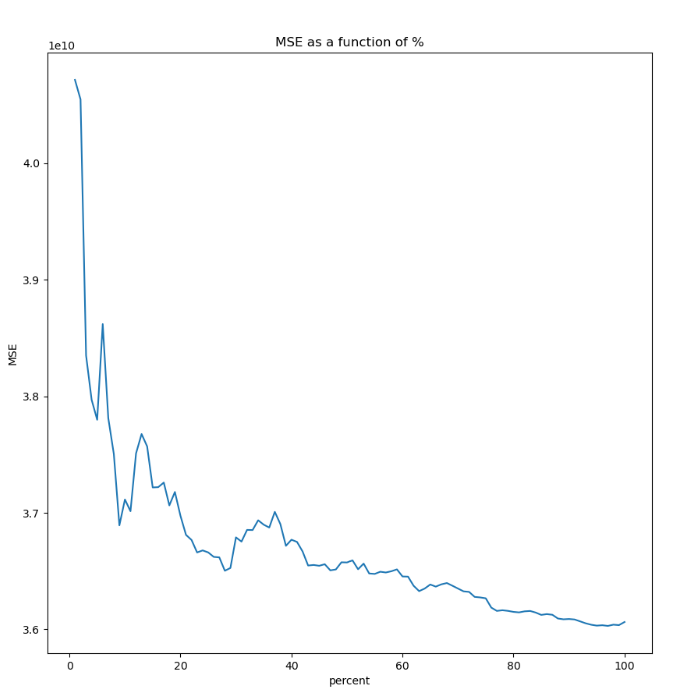
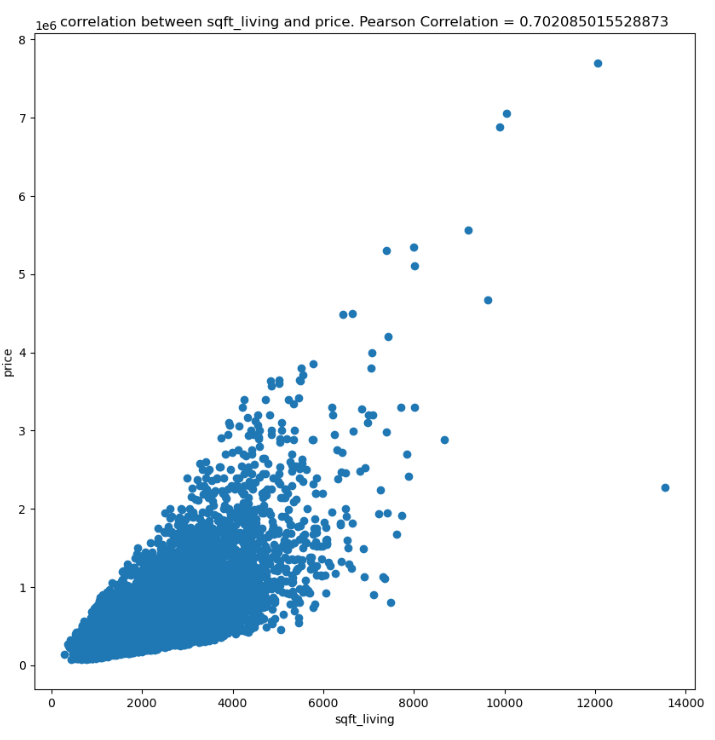
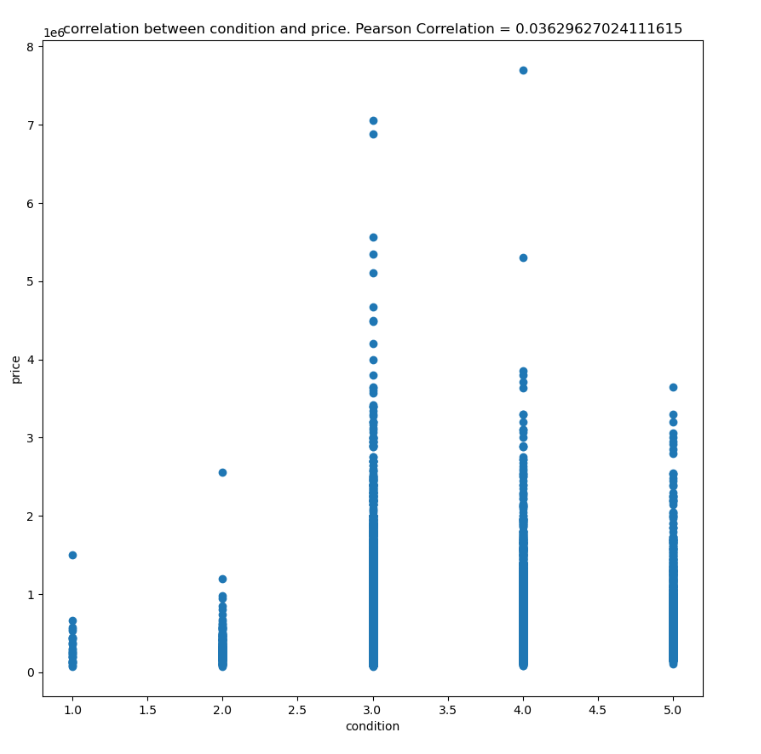
## הריבועים הפחותים

1. נראה שאם הפיכה מתקיים:  
   ניזכר בהגדרה של :  
   כעת, מכיוון ש- הפיכה הגרעין שלה הוא טריוויאלי ולכן גם הגרעין של הוא טריוויאלי כמו שהוכחנו בשאלה 1. לכן נסיק ש- כלומר .  
   כמו כן, לפי הפירוק של :  
   כעת נכפול בפירוק ה- של :  
   1 – אורתוגונלית ולכן ההופכית היא המשוחלפת. 2 – כפל של מטריצות אלכסוניות.  
   לכן סה"כ הראינו:
2. נתון ש- הפיכה ונוכיח ש-. ראשית ניזכר כי הראינו ש-. כמו כן, נשים לב ש- לכן מכיוון שהמטריצה הפיכה אנו יודעים שהגרעין שלה טריוויאלי, כלומר . לכן נסיק ש- גם הוא ולכן .  
    נתון ש- נוכיח ש- הפיכה. מהנתון נסיק ש- לכן גם . מכיוון ש- אנחנו יודעים שהגרעין שלה חייב להיות טריוויאלי ולכן היא הפיכה.
3. נראה שעבור הנורמה *מסוג היא מינימלית, כלומר לכל פתרון אחר מתקיים .  
   יהי פתרון . מהגדרת המשוואות הנורמליות אנו יודעים שעבור הקואורדינטות אשר מתאימות לערכים הסינגולריים של מתקיים (כי לכל פתרון מתקיים וזה נותן לנו פתרון יחיד ל- הקואורדינטות הראשונות).  
   כעת לפי הגדרת , שאר הערכים על האלכסון של הם אפסים ולכן הקואדינטות הנוספות של הינן אפס. כעת נחשב את הנורמה של :  
   1 – כפי שהראינו בתחילת הסעיף, הקואורדינטות הראשונות זהות בין כל פתרון.*

# חלק פרקטי

3. הפיצ'רים אשר אני חשבתי שהם קטגוריאליים הם – מיקוד, תאריך מכירה, תאריך שיפוץ, תאריך בנייה, קו גובה וקו אורך.
   1. **מיקוד** – לאחר שקראתי באינטרנט ולאחר שבחנתי את הנתונים, ראיתי כי יש כמעט אפס קורלציה בין המיקוד לבין המחיר ולכן בחרתי להוריד את העמודה הזו.
   2. **קו גובה וקו אורך** – מכיוון ששדות אלו הם יחסית גסים, יכולים להיות מצבים בהם יש בית שנמכר במחיר גבוה מאוד ובית שכן שנמכר במחיר נמוך מאוד ולכן שדות אלו לבדם יתקשו לחזות מחירים באופן מדויק. מכיוון שחששתי מרעש ששדות אלו יכולים ליצור החלטתי לבדוק את הקורלציה ביניהם לבין המחיר – הקורלציה בין קו הרוחב למחיר הוא 0.3 והקורלציה בין קו הגובה למחיר הוא . לכן החלטתי להוריד את עמודת קו הגובה והשארתי את עמודת קו הרוחב.
   3. **תאריך מכירה** – בדומה לקו גובה וקו אורך, יתכן ובאותו יום נמכרו בתים יקרים מאוד וזולים מאוד ולכן יכול להיווצר רעש משמעותי כתוצאה מבחירת יום ספציפי ולכן החלטתי להוריד את העמודה הזו.
   4. **תאריך שיפוץ** – מכיוון שלפי דעתי בית משופץ יכול להיות בעל ערך רב יותר מאשר בית שאינו משופץ החלטתי להפוך את העמודה הזו לעמודה עם ערכים של 0 ו-1 – 0 אם לא שופץ ו-1 אם כן שופץ.
   5. **תאריך בנייה** – טווח הבנייה של הבתים הוא גדול מאוד ולאו דווקא בית חדש שנבנה הוא בעל תשתיות טובות יותר מבית ישן. כמו כן, הקורלציה היא מאוד נמוכה (0.05) ולכן החלטתי להוריד את העמודה הזו.

השדות הם אמנם שדות בדידים אך יש עליהם יחס סדר ולכן השארתי אותם כמו שהם.  
כמו כן, כמו שפירטתי בקוד ובשאלה 15, מכיוון ש-  
החלטתי להוריד את העמודה של על מנת שהעמודות הללו לא תהיינה תלויות לינארית.

2. להלן הערכים הסינגולריים של הנתונים:  
     
   כפי שניתן לראות בגרף מעלה, המטריצה היא כמעט סינגולרית – יש 14 פיצ'רים שהערך הסינגולרי שלהם קרוב או ממש 0. מכיוון שהמטריצה היא כמעט סינגולרית ניתן ללמוד שחלק מעמודות מטריצת הנתונים שלנו כמעט תלויות לינארית, או ממש תלויות לינארית – למשל העמודה היא הסכום של שתי העמודות ו-, כלומר שלוש עמודות אלו ממש תלויות לינארית.
3. בגרף המצורף ניתן לראות שככל שאחוז הדגימות עליהם אנחנו מבססים את המודל שלנו גדל, כך ה- קטן. זה בדיוק מה שהייתי מצפה לראות מכיוון שיותר דגימות גורר יותר מקרים ללמידה גורר מודל יותר מדיוק.  
   
4. בחרתי את שני הפיצ'רים הבאים – (מ"ר של מרחב המחייה הפנימי) ו- (המצב של הדירה). להלן הגרפים:  
   הדרך שבה הסקתי האם הפיצ'ר הוא מועיל או לא לפי שתי סיבות עיקריות:
   1. **לוגית** – מצב הדירה הוא סובייקטיבי ולא ברור מי קבע אותו לכן הדירוג לא בהכרח מעיד על איכות הדירה ומחירה. לעומת זאת, גודל מרחב המחייה הוא כנראה הפרט הראשון אשר נבדק כאשר קונים דירה ולכן אני מצפה שהוא ישפיע מאוד על המחיר.
   2. **מקדם פירסון** – מקדם פירסון מודד את הקשר בין שני משתנים מקריים (1 – תלויים ביחס ישר, 1- - תלויים ביחס הפוך, 0 – בלתי תלויים). ניתן לראות שמקדם פירסון עבור המחיר וגודל מרחב המחייב הוא 0.702, כלומר היחס הוא כמעט ישר (ניתן לראות זאת גם בגרף). לעומת זאת, מקדם פירסון עבור מצב הדירה והמחיר הינו 0.036, כלומר המשתנים כמעט בלתי תלויים (ניתן לראות זאת בגרף – אין קשר בין דירוג גבוה יותר של מצב הדירה למחיר גבוה יותר).