Iml - תרגיל 2

שם: שלום קצ'קו **ת"ז:** 313446320

שאלות תאורטיות

1.1 - פתרונות המשוואה הנורמלית

1. נוכיח כי מתקיים - $Ker(\boldsymbol{X}) = Ker(\boldsymbol{X}^T\boldsymbol{X})$ ע"י הכלה דו כיוונית.

$\Leftarrow$ נניח כי קיים וקטור $\vec{v} \in \mathbb{R}^d$ המקיים $\vec{v} \in Ker(\boldsymbol{X})$ אז בהכרח מתקיים $X\vec{v} = 0$ אז מתקיים:

$$(\boldsymbol{X}^T\boldsymbol{X})\vec{v} = \boldsymbol{X}^T(\boldsymbol{X}\vec{v}) = \boldsymbol{X}^T \cdot 0 = 0 \xrightarrow[\text{עפ הגדרה}]{} \vec{v} \in Ker(\boldsymbol{X}^T\boldsymbol{X})$$

$\Rightarrow$ נניח כי מתקיים $\vec{v} \in Ker(\boldsymbol{X}^T\boldsymbol{X})$ אז ע"פ הגדרה מתקיים:

$$(\boldsymbol{X}^T\boldsymbol{X})\vec{v} = 0$$

נכפול את שני האגפים ב - $\vec{v}^T$:

$$\vec{v}^T(\boldsymbol{X}^T\boldsymbol{X})\vec{v} = (\boldsymbol{X}\vec{v})^T(\boldsymbol{X}\vec{v}) = ||\boldsymbol{X}\vec{v}||^2 = 0 \rightarrow \boldsymbol{X}\vec{v} = 0 \rightarrow \vec{v} \in Ker(\boldsymbol{X})$$

הראנו כי מתקיים $Ker(\boldsymbol{X}) \subseteq Ker(\boldsymbol{X}^T\boldsymbol{X})$ וגם $Ker(\boldsymbol{X}^T\boldsymbol{X}) \subseteq Ker(\boldsymbol{X})$ ולכן $Ker(\boldsymbol{X}) = Ker(\boldsymbol{X}^T\boldsymbol{X})$**.**

2. נוכיח כי לכל מטריצה ריבועית A מתקיים:

$$Im(A^T) = Ker(A)^{\perp}$$

$\Leftarrow$ נניח כי קיימים $y \in Im(A^T)$ ו - $x \in Ker(A)^{\perp}$. אז בהכרח קיים וקטור u המקיים $A^Tu = y$ וגם:

$$\langle x, y\rangle = x^Ty = x^T(A^Tu) = (Ax)^Tu = 0$$

ולכן מתקיים $y \in Ker(A)^{\perp}$ ונקבל $Im(A^T) \subseteq Ker(A)^{\perp}$.

$\Rightarrow$ נרצה להוכיח כי מתקיים $\forall v \in Ker(A)^{\perp} \rightarrow \ v \in Im(A^T)$. נראה זאת ע"י הוכחת השלילה של הפסוק. השלילה:
נרצה להראות כי מתקיים:

$$v \notin Im(A^T) \rightarrow v \notin Ker(A)^{\perp}$$

בכדי להראות שהשלילה מתקיימת מספיק למצוא וקטור $z \in Ker(A)$ כך שמתקיים $\langle v, z\rangle \neq 0$. אז תחת ההנחה כי $v \notin Im(A^T)$ אז בהכרח קיים ל - v רכיב ב - $Im(A^T)^{\perp}$ אז נסתכל על וקטור $z \in Im(A^T)^{\perp}$ כך שמתקיים $\langle v, z\rangle \neq 0$. כעת מכיוון שמתקיים $z \in Im(A^T)^{\perp}$ ולכן הוא מאונך לכל וקטור ב - $Im(A^T)$ ולכן מתקיים:

$$||Az||^2 = \langle Az, Az\rangle = (Az)^T(Az) = z^TA^TAz = \langle z, A^TAz\rangle = 0$$

כאשר המעבר האחרון נכון מכיוון שנוכל להגדיר וקטור $x = Ac$ והוקטור $A^T x \in Im(A^T)$ ע"פ הגדרה ולכן מאונך לוקטור z, וסיימנו. הראה הכלה דו כיוונית ולכן מתקיים $Im(A^T) = Ker(A)^\perp$.

3. נוכיח את שני הכיוונים:

נניח כי מתקיים $y \perp Ker(X^T)$ ונרצה להראות כי למערכת המשוואת יש אינסוף פתרונות.

$$y \perp Ker(X^T) \Leftrightarrow y \in Ker(X^T)^\perp \Leftrightarrow y \in Im(X)$$

כאשר את המעבר האחרון ניתן להוכיח בצורה דומה להוכחה של סעיף 2 ע"י הגדרת $A = X^T$. כלומר קיים פתרון יחיד. מכיוון שהמטריצה X איננה הפיכה אז למערכת המשוואות יש אפס פתרונות או אינסוף. מכיוון שמצאנו פתרון קיים אז בהכרח יש אינסוף פתרונות.

4. נתונה מערכת המשוואות הנורמאלית:

$$X^T X w = X^T y$$

נרצה להוכיח כי למערכת משוואות נורמאלית יש פתרון יחיד או אינסוף פתרונות (בהנחה כי $X^T X$ הפיכה).
נניח כי המטריצה $X^T X$ איננה הפיכה, משאלה 3 נובע כי קיימים למערכת אינסוף פתרונות. ולכן:

$$X^T y \perp Ker(X^T X) \Leftrightarrow X^T y \perp Ker(X)$$

כאשר המעבר נובע מההוכחה בשאלה 1.

אז נסתכל על איבר $u \in Ker(X)$ והמכפלה הפנימית שלו עם $X^T y$:

$$\langle X^T y, u \rangle = (X^T y)^T u = y^T X u = 0$$

ולכן יש למערכת אינסוף פתרונות. במידה ו - $X^T X$ הפיכה אז נוכל להכפיל את המשוואות הנורמליות משמאל:

$$w = (X^T X)^{-1} X y$$

וקיבלנו פתרון יחיד עבור מערכת המשוואות הנתונה.

Projection Matrices – 1.2

5. **(א)** נסתכל על מטריצת ההטלה P ועל שני וקטורים כלשהם v, u. נוכיח תחילה כי מתקיים:

$$v \cdot (Pu) = (Pv) u$$

נפרק כל אחד מהוקטורים לרכיב מקביל ומאונך למישור אליו P מטיל את הוקטורים. נסמן את הרכיב המקביל ב - u_p ואת המאונך ב- u_n. אז מתקיים:

$$v \cdot (Pu) = v \cdot \Big(P(u_p + u_n) \Big) = v \cdot u_p = u_p (v_p + v_n) = u_p v_p$$

$$(Pv) u = P(v_P + v_n) u = v_p u = v_p (u_p + u_n) = v_p u_p$$

כעת הראנו שהשוויון מתקיים, אז נשתמש בכך שאפשר לכתוב מכפלה פנימית בצורה - $x \cdot y = x^\top y$ ונקבל:

$$v \cdot (Pu) = (Pv)u \Rightarrow v^\top \cdot (Pu) = (Pv)^\top u \Rightarrow v^\top Pu = v^\top P^\top u$$

מכיוון שלא הנחנו דבר על הווקטורים בהתחלה אז השוויון צריך להתקיים לכל u, v ולכן נקבל כי $P = P^\top$ כלומר המטריצה P סימטרית.

(ב) ע"פ הגדרת ערך עצמי מתקיים:

$$\lambda v = Pv = P^2 v = P(Pv) = \lambda Pv = \lambda^2 v$$

המעבר השני נכון ע"פ תכונה (ד) והמשוואה מתקיימת רק עבור $\lambda = 1{,}0$. כעת נניח כי $v \in \{v_i\}_1^k$:

$$Pv = \lambda v$$

נקח את השחלוף של המשוואה שקיבלנו:

$$v^\top P^\top = \lambda v^\top$$

נכפיל את המשוואת זו בזו ונקבל:

$$v^\top P^\top Pv = \lambda^2 v^\top v \to v^\top v = \lambda^2 v^\top v$$

מכיוון ש - v וקטור בבסיס אז הוא אורתונרמלי ולכן $v^\top v = \left|\left|v\right|\right|^2 \neq 0$ ולכן $\lambda = 1$.

(ג) נוכיח כי לכל וקטור $v \in V$ מתקיים $Pv = v$. מכיוון ש - $v \in V$ אז הוקטור v נפרש ע"י וקטורי הבסיס של תת המרחב ולכן ניתן לכתוב אותו כצירוף ליניארי שלהם:

$$v = \sum_{i=1}^{k} a_i v_i$$

אז נפעיל את מטריצת ההטלה על הוקטור ונקבל:

$$Pv = P \cdot \sum_{i=1}^{k} a_i v_i = P \cdot (a_1 v_1 + \cdots + a_k v_k) = a_1 v_1 + a_2 v_2 + \cdots + a_k v_k = v$$

כאשר המעבר השני נובע מסעיף (ב), הוכחנו כי עבור וקטורי הבסיס הערך העצמי הוא 1 ולכן קיבלנו שוב את v.

(ד) נרצה להוכיח כי מתקיים $P^2 = P$. נפתח לפי הגדרה:

$$P^2 = \sum_{i=1}^{k} v_i v_i^\top \cdot \sum_{j=1}^{k} v_j v_j^\top = \sum_{i=1}^{k} \sum_{j=1}^{k} v_i v_i^\top v_j v_j^\top \underset{(1)}{=} \sum_{i=1}^{k} \sum_{j=1}^{k} v_i \langle v_i^\top, v_j \rangle v_j^\top \underset{(2)}{=} \sum_{i=1}^{k} \sum_{j=1}^{k} v_i \delta_{ij} v_j^\top \underset{(3)}{=}$$

$$\sum_{i=1}^{k} v_i v_i^\top = P$$

(1) - מהגדרת מכפלה פנימית. (2) הבסיס אורתוגונלי ולכן המכפלה הפנימית תתאפס לכל $i \neq j$

(3) - הדלתא של קרוניקר מתאפסת לכל $i \neq j$ ונותנת 1 אחרת.

(ה) נוכיח כי מתקיים $(I-P)P = 0$:

$$(I-P)P = P - P^2 = P - P = 0$$

כאשר המעבר השני נובע מסעיף ד'.

1.3 - Least Squares

6. נרצה להוכיח כי אם $X^\top X$ הפיכה אז מתקיים $(X^TX)^{-1}X^\top = X^\dagger$. נסתכל על פירוק SVD:

$$X = U\Sigma V^\top \rightarrow X^\top = V\Sigma U^\top$$

$$(V\Sigma^\top U^\top U\Sigma V^\top)^{-1}(V\Sigma^\top U^\top) \underset{U^TU=I}{=} (V\Sigma^\top \Sigma V^\top)^{-1}(V\Sigma^\top U^\top) = V(\Sigma^\top\Sigma)^{-1}V^\top(V\Sigma^\top U^\top) =$$

$$V\Sigma^{-1}(\Sigma^\top)^{-1}\Sigma^\top U^\top = V\Sigma^{-1}U^\top$$

ומכיוון שהמטריצה Σ אלכסונית (אך לא ריבועית) אז מתקיים $\Sigma^\dagger = \Sigma^{-1}$ וקיבלנו - $(X^TX)^{-1}X^\top =$ $X^\dagger$ כנדרש.

7. נרצה להוכיח כי X^TX הפיכה אמ"מ $span\{x_1, \dots, x_m\} = \mathbb{R}^d$. ע"פ הגדרה אנו יודעים כי מתקיים:

$$X^TX \ is \ invertible \Leftrightarrow X^TX \ is \ full \ rank \Leftrightarrow X \ is \ full \ rank$$

כאשר המעבר השני מתקבל מתוך תרגול 1, ראינו כי מתקיים $rank(X) = rank(X^\top X)$ ובנוסף מתקיים $rank(X) = \dim(span\{x_1, \dots, x_m\})$. ונובע ישירות כי $span\{x_1, \dots x_m\} = \mathbb{R}^d \Leftrightarrow$ $rank(X) = d$ ומכיוון שהדרגות שוות נקבל:

$$X^\top X \ is \ invertible \Leftrightarrow span\{x_1, .. x_m\} = \mathbb{R}^d$$

8. מכיוון שפירוק SVD של X הוא יחיד, אז הערכים של $\hat{w}$ נקבעים בצורה יחידה כתלות בערכים הסינגולריים של של הפירוק. כאשר הגדרנו את המטריצה כך שאם יש r ערכים סינגולריים אז לכל $i \leq r$ מתקיים $\Sigma_{ii} = \sigma_i$ ולכל $i > r$ קבענו כי $\sigma_i = 0$. אז מכיוון שהגדרנו $\hat{w} = X^\dagger y$ אז מתקיים $\forall i > r \ \hat{w}_i = 0$. עבור כל פתרון אחר $\bar{w}$, r הערכים הראשונים חייבים להיות זהים לאלו של $\bar{w}$ בכדי לקיים את המשוואה, ושאר הערכים חופשיים אך בהכרח אי שליליים, ולכן:

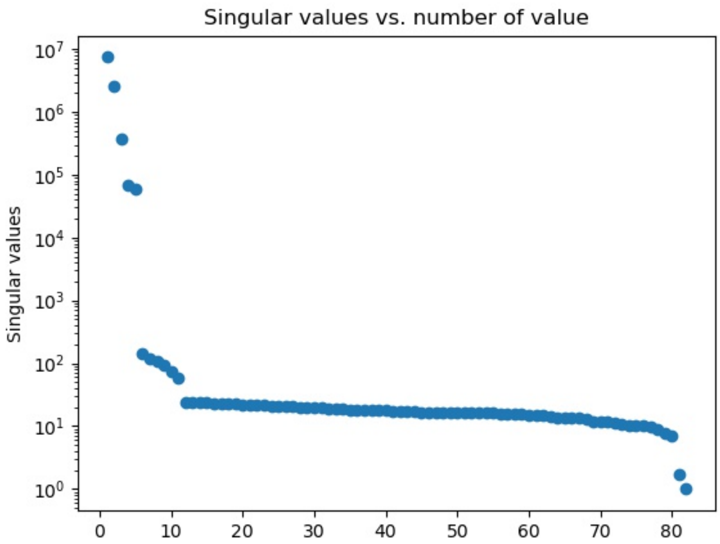
$$||\overline{w}||^2=\sum_{i=1}^{d}||\overline{w_i}||^2=\sum_{i=1}^{r}||\hat{w}_i||^2+\sum_{i=r+1}^{d}||\overline{w}_i||^2\leq\sum_{i=1}^{r}||\hat{w}_i||^2=\sum_{i=1}^{d}||\hat{w}_i||^2=||\hat{w}||^2$$

כלומר בהכרח הפתרון שמקיים $\hat{w}=X^{\dagger}y$ הוא בעל נורמת L_2 המינימלית מכל הפתרונות.

Practical Questions

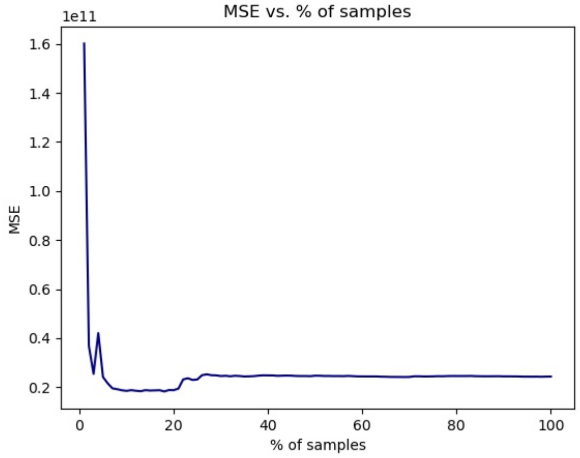
13. המשתנים שהחלטתי להתייחס אליהם כקטגוריים הם - zipcode, lat, long, waterfront. מכיוון שלא ניתן להגדיר יחס סדר בין קווי גובה/ רוחב וכנ"ל בין מיקוד אחד לאחר מכיוון שכן נצפה שיתקיים קשר בין המיקום למחיר אז הסרתי את קווי האורך והרוחב במהלך העיבוד המידע. החזית למים נתונה מראש בצורה בינארית ולכן התייחסתי אליה כאל משתנה קטגורי.

14. הגרף נתון בסקאלה לוגריתמית למטרת קריאות בלבד:



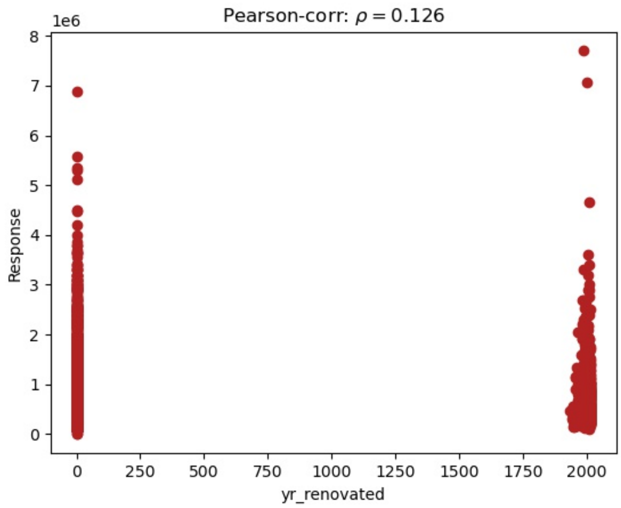
15. פירוק לפי ערכים סינגולריים מאפשר לנו להבין איזה מאפיינים משפיעים יותר על המחיר. בדומה לוקטור עצמי בעל ערך עצמי גדול בבסיס המלכסן עבור מטריצה מסוימת הוא בעל השפעה "גדולה" יותר על הטרנספורמציה הליניארית אז יש לנו קשר ישיר בין ערכים עצמיים לסינגולריים כאשר הוקטורים במקרה שלנו הם עמודות המאפיינים, כלומר מאפיין בעל ערך סינגולרי גדול יותר התואם לו, ישפיע יותר על חיזוי המודל. לפי שרטוט הערכים הסינגולריים, אין ערך סינגולרי קרוב ל-0 ולכן המטריצה איננה קרובה להיות סינגולרית.

.16



באופן כללי קיימת מגמה ברורה של ירידה בשגיאה כתלות בכמות הדגימות שעליהן אנו מאמנים את המודל, השגיאה מתחילה גדולה מאוד ויורדת בחדות ככל שכמות הדגימות עולה, בפחות מ- 20% מהדגימות השגיאה אינה יציבה, כנראה כי מדובר בכמות יחסית נמוכה של דגימות ולכן גם בכמה הרצות שונות היה ניתן לראות הבדלים בשגיאה בכמות נמוכה של דגימות התלויה כנראה בתנאי ההתחלה המוגרלים. מעל 20% מהדגימות נראה כי השגיאה מתייצבת על שגיאה של $\sim 0.2 \times 10^{11}$ ולא משתפרת, כנראה מכיוון שהגענו להגבלת הדיוק של המודל שלנו.

.17 הגרף הראשון הוא ה - response vector כתלות בשנת השיפוץ, קודם כל קורולציית פירסון נמוכה בינהם,לדעתי הפיצ׳ר לא תורם למודל הרבה מכיוון שיש בתים רבים שלא שופצו מעולם, וייתכן שהם חדשים ולכן אין צורך בשיפוץ. הפיצ׳ר בצורתו הבסיסית לא מספק את המידע הזה. ניתן לראות ויזואלית גם כי אין קשר ברור בין שנת השיפוץ לתוצאות המודל.



הגרף השני הוא ה - response כתלות בשטח הסלון. לדעתי הפיצ'ר תורם למודל, קודם כל מכיוון שיש קורלציה גבוהה יחסית בין הוקטורים, בנוסף ניתן לראות ויזואלית כי קיימת מגמה כלשהי בין גודל הסלון ל response: