IML, תרגיל 2
מגיש: אבי כוגן, ת.ז: 205417710

שאלה 1

תהי $X \in R^{m\times d}$ צ"ל: $Ker(X) = Ker(X^TX)$
$Ker(X) \subset Ker(X^TX)$: יהי $v \in Ker(X) \subseteq R^d$ לכן מתקיים $v \in ker(X^TX) \Leftarrow$
$X^TXv = X^T\underline{0} = \underline{0} \Leftarrow Xv = \underline{0}$.
$Ker(X) \supset Ker(X^TX)$: יהי $v \in Ker(X^TX) \subseteq R^d$ לכן מתקיים $X^TXv = \underline{0} \Rightarrow$
$v^TX^TXv = ||Xv|| = 0 \Leftrightarrow Xv = 0 \Leftrightarrow v \in Ker(x)$.

שאלה 2

תהי $A \in R^{n\times n}$.
$Im(A^T) \subseteq Ker(A)^\perp$:
יהי $v \in Im(A^T)$ אזי קיים $u \in R^n$ כך ש-$A^Tu = v$. יהי $w \in Ker(A)$.
נקבל ש $v \in Ker(A)^\perp \Leftarrow <v,w> = <A^Tu,w> = (A^Tu)^Tw = yAw = 0$.

כעת מתקיים

$$dim(Im(A^T)) \overset{\text{הדרגה שווה לדרגת המשוחלפת}}{=} dim(Im(A)) \overset{\text{משפט המימדים}}{=}$$

$$n - dim(Ker(A)) \overset{\text{משלים אורתוגונלי של הגרעין}}{=} dim(Ker(A)^\perp)$$

קיבלנו הכלה ושוויון מימדים לכן מתקיים $Im(A^T) = Ker(A)^\perp$.

שאלה 3

נתון שהמטריצה X אינה הפיכה, לכן יש לה אינסוף פתרונות או 0 פתרונות.
לכן נקבל שלממ"ל יש אינסוף פתרונות אם"ם $y \in Im(A)$,
מכיוון שהוכחנו בשאלה 2 שמתקיים $Im(X) = Ker(X^T)^\perp$, נקבל שמתקיים $y \in Im(X) \Leftrightarrow y \in Ker(X^T)^\perp$.

שאלה 4

נחלק למקרים:

X^TX הפיכה: נקבל שקיים פתרון יחיד $w=(X^TX)X^Ty$.

X^TX לא הפיכה: מתקיים X^TX ריבועית ולכן מהשאלות הקודמות:

למשוואות הנורמליות יש אינסוף פתרונות אם"ם $X^Ty\perp Ker((X^TX)^T)=Ker(X^TX)$.

ראינו משאלה 1 שמתקיים $Ker(X^TX)=Ker(X)$. לכן צ"ל $X^Ty\perp Ker(X)$. יהי $v\in Ker(X)$ אזי מתקיים

$<X^Ty,v>=y^TXv=y^Tv\Rightarrow X^Ty\perp Ker(X^TX)$.

שאלה 5

א. מהגדרה מתקיים $P\in R^{d\times d}$, ומתקיים $[P]_{ij}=\Sigma_{t=1}^k[v_t]_i[v_t]_j=\Sigma_{t=1}^k[v_t]_j[v_t]_i=[P]_{ji}$.
כאשר $[v_t]_i$ הוא האיבר במקום ה-i של הוקטור בסיס v_t.

ב. נתון $(v_1,..,v_k)$ בסיס אותנורמלי של V.
מלינארית 2, קיים $V^\perp$ כאשר $(v_{k+1},..,v_d)$ הבסיס האותונורמלי של המשלים האורתוגונלי.
מתקיים $V^\perp\oplus V=R^d$ כאשר $(v_1,..,v_k,v_{k+1},..,v_d)$ הוא בסיס אורתונורמלי של R^d.
נמצא את הע"ע:
נחלק למקרים:

1. לכל $V\supseteq(v_1,..,v_k)\ni v_j$ מתקיים $Pv_j=\Sigma_{i=1}^k(v_iv_i^T)v_j=\Sigma_{i=1}^kv_i(v_i^Tv_j)\overset{*}{=}\Sigma_{i=1}^kv_i\delta_{i,j}=v_j$

* בסיס אותנורמולי, כאשר $\delta_{i,j}=\begin{cases}1 & i=j\\ 0 & o.w\end{cases}$. קיבלנו שכל $v_j\in(v_1,..,v_k)$ הוא ו"ע עם ע"ע 1.

2.
1. לכל $V\supseteq(v_{k+1},..,v_d)\ni v_j$ מתקיים $Pv_j=\Sigma_{i=1}^k(v_iv_i^T)v_j=\Sigma_{i=1}^kv_i(v_i^Tv_j)\overset{*}{=}\Sigma_{i=1}^kv_i0=0v_j$

* כל הוקטורים השייכים למשלים הניצב מאונכים לוקטורי הבסיס. קיבלנו שכל $v_j\in(v_{k+1},..,v_d)$ הוא ו"ע עם ע"ע 0.

קיבלנו שמימדי המ"ע מקיימים $dim(V_0)\ \geq\ k,\ dim(V_1)\ \geq\ d-k$ אבל מצד שני מתקיים $dim(V_0)+dim(V_1)\leq dim(R^d)=d$
לכן נקבל $dim(V_0)=k,\ dim(V_1)=d-k$.
כעת מכיוון שקיבלנו $\Sigma_{\lambda\in\{0,1\}}dim(V_\lambda)=dim(R^d)$ נסיק שאין עוד ע"ע כי אחרת נקבל שסכום מימדי המ"ע של הע"ע גדול ממימד R^d.

ג. לכל $v\in V$ מתקיים $v=\Sigma_{i=1}^ka_iv_i$ עבור $a_i\in R$ לכל $i\in[k]$.
לכן נקבל

$$Pv=\Sigma_{i=1}^kv_iv_i^Tv=\Sigma_{i=1}^kv_iv_i^T\Sigma_{j=1}^ka_jv_j=\Sigma_{i=1}^kv_i\Sigma_{j=1}^ka_jv_i^Tv_j=$$
$$=\Sigma_{i=1}^kv_i\Sigma_{j=1}^ka_j\delta_{i,j}\overset{\Sigma_{j=1}^ka_j\delta_{i,j}=a_i}{=}\Sigma_{i=1}^kv_ia_i=v$$

ד.

$$P^2 = \Sigma_{i=1}^{k} v_i v_i^T \Sigma_{j=1}^{k} v_j v_j^T = \Sigma_{i=1}^{k} v_i \Sigma_{j=1}^{k} v_i^T v_j v_j^T = \\ = \Sigma_{i=1}^{k} v_i \Sigma_{j=1}^{k} \delta_{i,j} v_j^T = \Sigma_{i=1}^{k} v_i v_i^T = P$$

ה.

$$(I-P)P = I \cdot P - P^2 \overset{P^2=p}{=} P - P = 0$$

שאלה 6

נתון X^TX הפיכה, כאשר $X^\dagger = V\Sigma^\dagger U^T$, צ"ל: $X^\dagger y = (X^TX)^{-1}X^Ty$ $\overset{\text{שקול להוכיח}}{\Leftrightarrow}$ $X^\dagger = (X^TX)^{-1}X^T$.

הוכחה:

נתון X^TX הפיכה, לכן $\{0\} = Ker(X) = Ker(X^TX)$ $\Leftarrow$ X הפיכה לכן כל הערכים הסינגולרים שלה > 0.

נקבל שעבור פירוק SVD של X: $X = U\Sigma V^T$ מתקיים $\Sigma^\dagger = \Sigma^{-1}$ כי X הפיכה ולכן גם Σ הפיכה. מכך נקבל שמתקיים:

$$(X^TX)^{-1}X = ((U\Sigma V^T)^T U\Sigma V^T)^{-1}(U\Sigma V^T) = (V\Sigma^T U^T U\Sigma V^T)^{-1} V\Sigma^T U^T \overset{U^TU=\underline{I,}\ \Sigma^T=\Sigma}{=} \\ = (V\Sigma\Sigma V^T)^{-1} V\Sigma U^T = (V^T)^{-1}(\Sigma^2)^{-1}\cancel{(V)^{-1}V}\Sigma U^T \overset{*}{=} \\ = V\Sigma^{-1}U^T \overset{\Sigma^T=\Sigma}{=} V(\Sigma^T)^{-1}U^T \overset{(\Sigma^T)^{-1}=\Sigma^\dagger}{=} V\Sigma^\dagger U^T$$

שאלה 7

צ"ל: X^TX הפיכה $\Leftrightarrow span(x_1,..,x_m) = R^d$.

הוכחה:

$$\text{הפיכה } X^TX \in R^{d\times d} \Leftrightarrow Ker(X^TX) = \{0\} \Leftrightarrow Ker(X) = \{0\} \Leftrightarrow \\ \Leftrightarrow dim(Im(X)) \overset{d\leq m}{=} dim(R^d) \Leftrightarrow dim(Im(X)) = dim(Im(X^T)) \Leftrightarrow \\ Im(X^T) = span(x_1,..,x_m) = R^d$$

שאלה 8

כאשר X^TX לא הפיכה צ"ל: $||\hat{w}|| = ||X^\dagger y|| \leq ||w||$ $\forall \bar{w} \in \{$פתרונות למשוואות הנורמליות$\}$
הוכחה: X^TX לא הפיכה, לכן נניח $rank(X^TX) = k < d$.
ראינו בהרצאה ש-$\hat{w}$ הוא תמיד פתרון של המשוואה.
מכיוון ש-$rank(X^TX) = k$ נסיק שישנם k ערכים סינגולרים כאשר $\hat{w}_i$ מתאים לערך הסינגולרי ה-i של X לכל $i \in [k]$ וזהו הפתרון היחיד המתאים להם ונגדיר לכל $i \in \{k+1, .., d\}$ $\hat{w}_i = 0$ המתאים לערכי ה-0 על האלכסון של Σ.
נקבל שכל פתרון אחר $\bar{w} \in \{$פתרונות למשוואות הנורמליות$\}$ יקיים $\bar{w}_i = \hat{w}_i$ על מנת לקיים את הפתרון המתאים לערך הסינגולרי ה-i לכל $i \in [k]$ (כאשר אני מניח בה"כ שהערכים הסינגולרים מסודרים מהגדול לקטן על האלכסון, אחרת נתאים את הערכים בהתאם למיקום על האלכסון).
נקבל שהאיברים האחרים ב-$\bar{w}$ אינם תלויים בערכים הסינגולרים ולכן:

$$||\bar{w}|| = \Sigma_{i=1}^{d} \bar{w_i}^2 = \Sigma_{i=1}^{k} \bar{w_i}^2 + \Sigma_{i=k+1}^{d} \bar{w_i}^2 \leq \Sigma_{i=1}^{k} \bar{w_i}^2 = \Sigma_{i=1}^{k} \hat{w_i}^2 = ||\hat{w}||$$

שאלות 9-11

מימוש בקובץ קוד.

שאלה 12

מימוש בקובץ קוד, הסבר:
בשלב ה-$preproccesing$ הורדתי תצפיות עם ערכים חריגים:
$price \leq 0$ - מתוך הנחה שבית לא נמכר בחינם או במחיר שלילי.
$bedrooms \leq 0$ - מכיוון שלא יתכן בית ללא חדרים (זה לא בית) או מספר חדרים שלילי.
$bathrooms < 0$ - מכיוון שלא יתכן מספר חדרי שירותים שלילי.
$sqft\_living \leq 0$ - לא יתכן ששטח של בית יהיה 0 או שלילי.
$sqft\_lot \leq 0$ - לא יתכן ששטח האדמה של בית יהיה 0 או שלילי.
$sqft\_above \leq 0$ - לא יתכן שהשטח מעל האדמה יהיה 0 או שלילי.
$yr\_built \leq 0$ - לא יתכן ששנת הבנייה תהיה 0 או שלילית (אמריקה התגלת ב-1492).
$floors \leq 0$ - לא יתכן שמספר הקומות בבית יהיה 0 או שלילי, לכל הפחות ישנה קומה אחת.
$sqft\_living15 \leq 0$ - לא יתכן שהשטח של 15 השכנים הקרובים ביותר של הבית יהיה 0 או שלילי.
$sqft\_lot15 \leq 0$ - לא יתכן ששטח האדמה של 15 השכנים הקרובים ביותר של הבית יהיה 0 או שלילי.
$yr\_renovated < 0$ - לא יתכן ששנת השיפוץ תתהיה שלילית.
$sqft\_basement < 0$ - לא יתכן ששטח מרתף יהיה שלילי.
עבור $waterfront, view, condition, grade$ נבדק שהערכים בהתאם לטווח כפי שמוגדר בתיאור הדאטה.
לא בוצע שינוי ב-$lat, long$.
שיניתי את התאריכים כך שיתקיים ביניהם יחס סדר בעזרת שינוי התאריך למספר שהוא מספר היום מאז התאריך 1/1/01. התאריכים הושארו כי יתכן שקיימת מגמה במחירי הדירות לאורך זמן (על אף שבשאלה 17 ניתן היה לראות שהקורולציה בין התאריכים למחיר היא 0).

שאלה 13

מימוש בקובץ קוד, הסבר:

משתנים קטגורים - $.id, zipcode$

מחקתי את עמודת id מכיוון שאינה צריכה להשפיע על מחיר הדירה, מדובר רק במזהה בקובץ הדאטה.

פרקתי את עמודת $zipcode$ בהתאם לעקרון $hot - code$ כיוון שלמיקום יכולה להיות השפעה על מחיר הדירה.

שאלה 14

מימוש בקובץ הקוד.

שאלה 15

מהערכים הסינגולרים ניתן להעריך את דרגת המטריצה - כמספר הערכים הגדולים ממש מ-0.

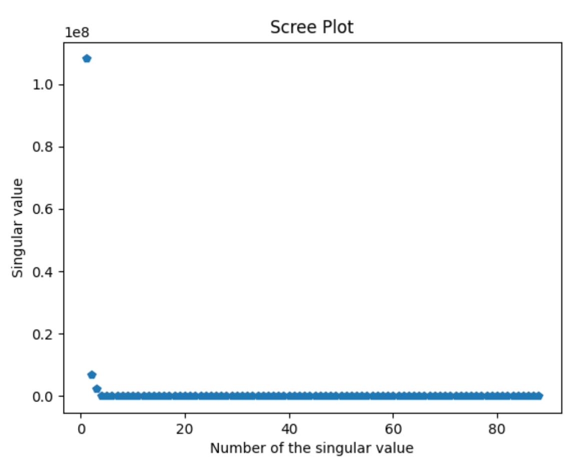
במקרה של המטריצה שמתקבלת מעיבוד קובץ הדאטה של הבתים, אין ערכים שהם 0 אבל הערך הקטן ביותר הוא $5.23189236e - 11 \approx 5.23 \cdot 10^{-11}$

ומכיוון שזהו ערך קטן מאוד "הדרגה האפקטיבית" של המטריצה אינה מלאה אך היא אכן קרובה להפיכות מכיוון שישנו רק ערך אחד קטן מאוד.

28 הערכים הסינגולרים הקטנים ביותר הם:

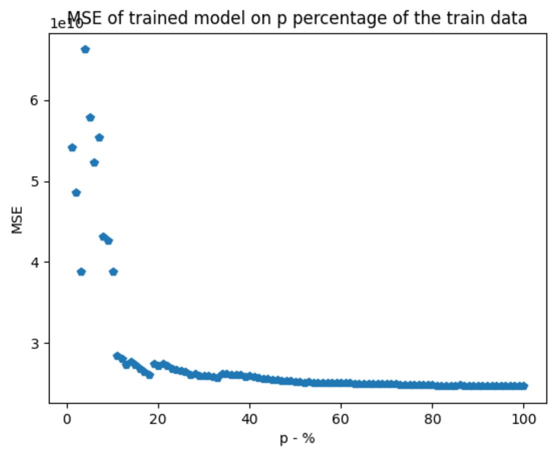
```
[1.57581084e+01 1.55764673e+01 1.52615574e+01 1.50825140e+01
 1.48432977e+01 1.42114746e+01 1.41323650e+01 1.39471565e+01
 1.37936692e+01 1.35991494e+01 1.32397139e+01 1.28220989e+01
 1.20223492e+01 1.19359739e+01 1.17203936e+01 1.12382281e+01
 1.11370887e+01 1.04916941e+01 1.02310405e+01 1.01429554e+01
 9.83180075e+00 8.99258741e+00 7.61791548e+00 7.04821007e+00
 2.80427518e+00 1.69948568e+00 2.04028842e-03 5.23189236e-11]
```

תרשים ה-$scree - plot$:



שאלה 16

ניתן לראות שכפי שניתן לצפות ה-MSE יורד ככל שהמודל מאומן על כמות גדולה יותר של תצפיות.
בנוסף יש חוסר עקביות כאשר המודל מאומן על כמות דאטה קטנה מכיוון שלכל תצפית יש השפעה גדולה על החיזוי.



שאלה 17

הפיצ'ר שלדעתי מועיל למודל הוא "$sqft\_living$".
באופן טבעי ניתן להסיק שגודל הבית ישפיע בצורה ישירה על מחיר הדירה וניתן לראות זאת בקורולציה הגבוהה ביניהם.
התרשים:

הפיצ'ר שלדעתי אינו מועיל הרבה למודל הוא "$date$".
ציר ה-x מייצג את מספר היום החל מתאריך $1/1/01$.

ניתן לראות בתרשים ששינוי בתאריך אינו מיצר מגמה כלשהי ביחס למחיר, כאשר ניתן לראות זאת גם בקורולציה 0 ביניהם, כלומר אין קורולציה ביניהם.

התרשים: