נועה בן דרור
316163260

IML - תרגיל 2

(1) נוכיח $ker(X) \subseteq ker(X^TX)$. יהי $0 \neq u \in ker(X)$, אזי $X^TXu = X^T0 = 0$
לכן $u \in ker(X^TX)$.

נוכיח $ker(X^TX) \subseteq ker(X)$. יהי $0 \neq u \in ker(X^TX)$, אזי $X^TXu = 0$. נכפיל
משמאל ב- u^T ונקבל:

$$u^TX^TXu = u^T0$$

$$(Xu)^TXu = 0 \Rightarrow \langle Xu, Xu \rangle = 0$$

ומתכונות בהגדרה של מכפלה פנימית מתקיים כי $\langle Xu, Xu \rangle = 0 \Rightarrow Xu = 0$
ולכן $u \in ker(X)$.

(2) יהי $v \in Im(A^T)$ ו- $w \in ker(A)$. נרצה להוכיח כי $\langle v, w \rangle = 0$. ידוע לנו כי $Aw = 0$
וכן שקיים u כך ש- $A^Tu = v$. מתקיים:

$$\langle v, w \rangle = \langle A^Tu, w \rangle = (A^Tu)^Tw = u^TAw = u^T \cdot 0 = 0$$

לכן $Im(A^T) \subseteq ker(A)^\perp$.

יהי $v \in ker(A)^\perp$. נרצה להוכיח כי קיים u כך ש- $A^Tu = v$.
נניח כי לא קיים u כך שלכל $w \in ker(A)$ מתקיים $\langle A^Tu, w \rangle \neq 0$, ומהגדרה מתקיים
$\langle A^Tu, w \rangle = u^TAw = 0$.

(3) X אינה הפיכה ולכן למערכת המשוואות יש או אינסוף פתרונות או 0 פתרונות.
כמו כן, למערכת יש לפחות פתרון אחד אם"ם $y \in Im(X)$. ומהסעיף הקודם
$y \in Im(X) \Leftrightarrow y \in ker(X^T)^\perp \Leftrightarrow y \perp ker(X^T)$ כלומר, קיים w כך ש- $Xw = y$

כלומר w הוא פתרון למשוואה, וזה אומר שיש ∞ פתרונות, כנדרש.

(4) אם X^TX הפיכה:

$$X^TXw = X^Ty \Rightarrow (X^TX)^{-1}X^TXw = (X^TX)^{-1}X^Ty \Rightarrow$$

$$w = (X^TX)^{-1}X^Ty$$

אם X^TX אינה הפיכה:

ראשית נראה $X^Ty \in \ker(X^TX)^{\perp}$. יהי $v \in \ker(X^TX)$*. מסעיף (1) מתקיים

ש- $v \in \ker(X)$, כלומר $Xv = 0$, ולכן:

$$\langle X^Ty | v \rangle = (X^Ty)^Tv = y^TXv = 0$$

קיבלנו ש- X^Ty מאונך לכל $v \in \ker(X^TX)$. מסעיף (3) ראינו שבמקרה כזה למשוואה יש אינסוף פתרונות.

(5) א) נשים לב כי P הוא סכום של מטריצות סימטריות, ולכן הוא סימטרי גם כן.

ב) מתקיים:

$$P v_j = \left(\sum_{i=1}^{k} v_i v_i^T\right) v_j = \sum_{i=1}^{k} v_i v_i^T v_j = \sum_{i=1}^{k} v_i \delta_{ij} = v_j$$

ג) יהי $v \in V$, אזי $v = \sum_{i=1}^{k} c_i v_i$ (לכל $1 \le i \le k$ $c_i \in \mathbb{R}$), ולכן

$$Pv = P\sum_{i=1}^{k} c_i v_i = \sum_{i=1}^{k} P c_i v_i = \sum_{i=1}^{k} c_i P v_i \underset{\text{(סעיף ב')}}{=} \sum_{i=1}^{k} c_i v_i = v$$

ד) נייצג את P בפירוק EVD כאשר D היא מטריצת היחידה משום שכפי שהוכחנו בסעיף ב' הע"ע הם 1.

$$P^2 = UDU^TUDU^T = UDDU^T = UDU^T = P$$

ה) $(I-P)P = IP - PP = P - P^2 = P - P = 0$

6) ג' שאם X^TX הפיכה אז $X^+y=(X^TX)^{-1}X^Ty$.

נראה כי $(X^TX)^{-1}X^T = X^+$. מתקיים:

$$(X^TX)^{-1}X^T = \left((U\Sigma V^T)^T(U\Sigma V^T)\right)^{-1}(U\Sigma V^T)^T =$$

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

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

$$V\underbrace{(\Sigma^T\Sigma)^{-1}\Sigma^T}_{\Sigma^+}U^T = V\Sigma^+U^T = X^+$$

7) הדרגה של X היא המימד של $Span\{x_1,\dots,x_m\}$. נשים לב כי הדרגה של X שווה לדרגה של X^TX. $\mathbb{R}^d = Span\{x_1,\dots,x_m\}$

אם"ם $rank(X^TX)=d$, ומכיוון ש- $X^TX \in M_{d\times d}$, X^TX הפיכה

אם"ם $rank(X^TX)=d$.

8) ה) w שנבחר הוא בהכרח פתרון עבור $Xw=y$. X^TX אינה הפיכה, ולכן נניח כי X מדרגה r כאשר $r<d$. נסתכל על פירוק SVD של $X=U\Sigma V^T$.
Σ מטריצה אלכסונית שעל האלכסון שלה הערכים הסינגולריים. מכיוון ש- X לא הפיכה ומדרגה r, אז קיימים r ערכים סינגולריים שונים מ-0.

ניכנס להיות הדרגה צירופים של U,V משמעותי: הן נראות כך:

$$U = \begin{bmatrix} | & | & & | & | & & | \\ u_1 & u_2 & \dots & u_r & u_{r+1} & \dots & u_d \\ | & | & & | & | & & | \end{bmatrix}$$

$$V = \begin{bmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_r^T & - \\ - & v_{r+1}^T & - \\ & \vdots & \\ - & v_m^T & - \end{bmatrix}$$

הוקטורים $u_1,\dots,u_r$ צריכים להיות שווים לכל אפשרויות ה-U

והוקטורים $v_1^T,\dots,v_r^T$ " " " " " ה-V.

הוקטורים $u_{r+1},\dots,u_d$, $v_{r+1}^T,\dots,v_m$ לא משפיעים על צורת ה-SVD

כי Σ תאפס אותם.

מה שמייחד את $\hat{w}$ הוא שהקורדינטות $r+1,\dots,d$ נמצאים אפסים, כלומר

$$\bar{w} = \begin{bmatrix} \bar{w}_1 \\ \vdots \\ \bar{w}_r \\ \bar{w}_{r+1} \\ \vdots \\ \bar{w}_d \end{bmatrix}, \qquad \hat{w} = \begin{bmatrix} \hat{w}_1 \\ \vdots \\ \hat{w}_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

ולכן

$$\|\hat{w}\| = \sum_{i=1}^{r} \hat{w}_i^2 + \sum_{i=r+1}^{d} 0^2 \leq \sum_{i=1}^{r} \bar{w}_i^2 + \sum_{i=r+1}^{d} \bar{w}_i^2 = \|\bar{w}\|$$

כנדרש.

316163260

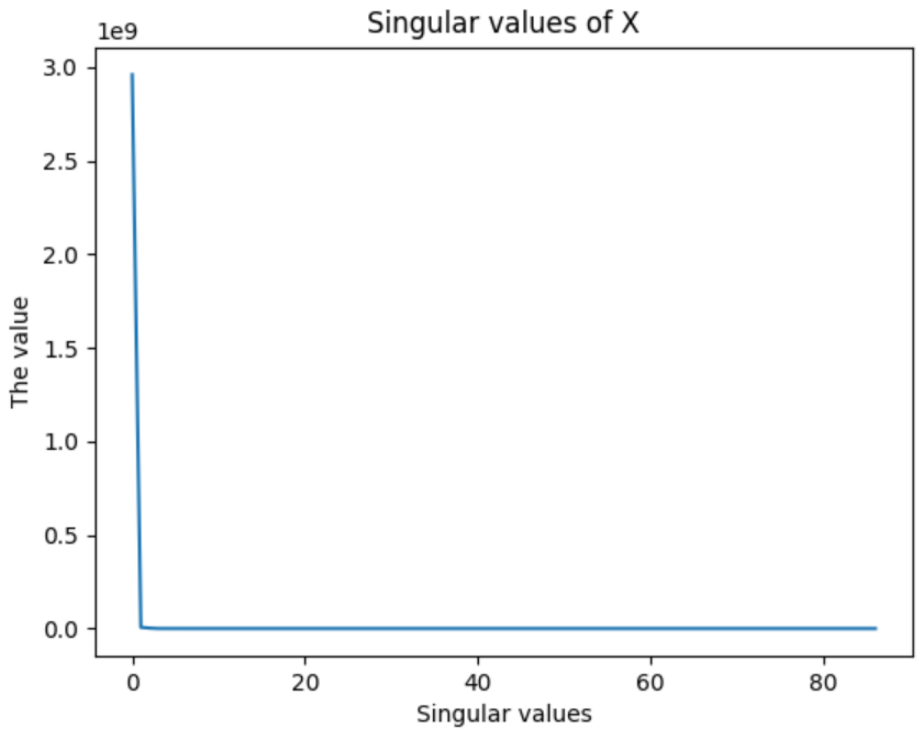
נועה בן דרור

IML – תרגיל 2 – חלק מעשי

13)

בחרתי להפוך את zipcode ל-feature קטגוריאלי, משום שזה feature שלא יכול להיות בעל השפעה על המחיר באופן לינארי. כלומר, זה feature שיכול להשפיע על המחיר, אבל אין משמעות למספר עצמו של ה- zipcode מבחינה לינארית.

15)

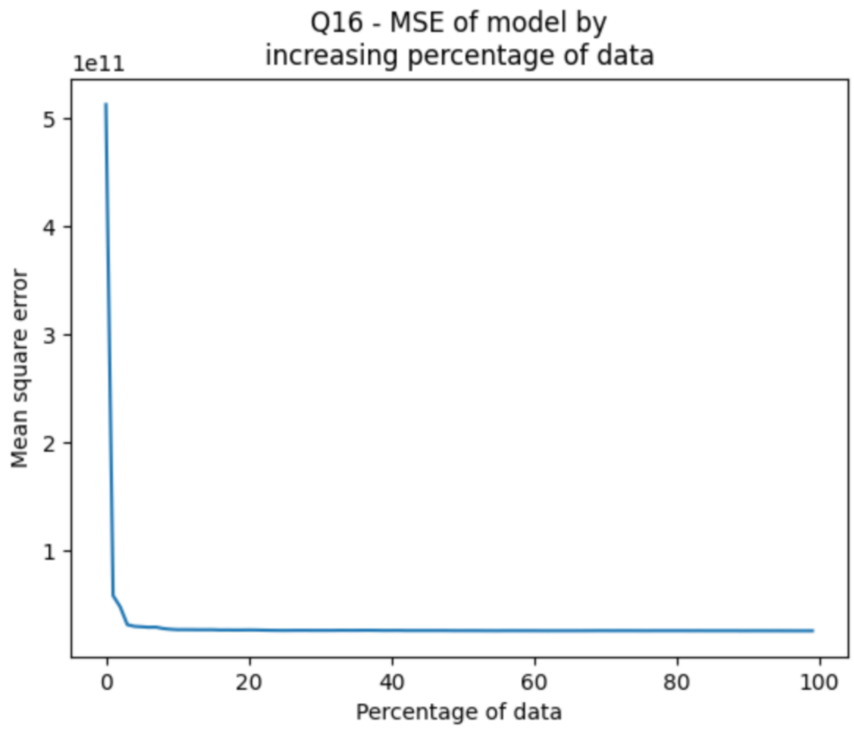


נשים לב בטבלה למטה כי 2 הערכים הסינגולריים הנמוכים ביותר הם 0, ואחרים 0.03.

84	85	86
0.03240	0.00000	0.00000

מכיוון שקיים ערך סינגולרי ששווה ל-0, המטריצה X היא לא הפיכה. כמו כן, קיים ערך סינגולרי שקרוב ל-0 (0.03), וניתן לפרש את זה כך שיש שני features שהזווית ביניהם חדה, ועל כן הם קירוב לינארי אחד של השני. זה גורר שגם אם לא היה ערך סינגולרי ששווה ל-0, המטריצה הייתה לכל הפחות "קרובה" ללהיות סינגולרית.

(16



ניתן לראות בגרף כי ככל שמאמנים את המודל על יותר דאטה, ה- mse קטן, כלומר הדיוק עולה. נשים לב כי השיפור ברמת הדיוק קורה די מהר.

(17

הקורלציה הכי גבוהה שנמצאה היא בין מחיר לבין sqft_living, ולכן ה- sqft_living הוא feature מועיל ומשמעותי למודל.

קורלציה נמוכה שנמצאה היא בין מחיר לבין מצב הדירה, ולכן מצב הדירה הוא feature לא מועיל ולא משמעותי למודל.