איתן גרשוני
308350503

אלגברה לינארית - תרגיל בית 2

1) נראה בעזרת הכלה דו-כיוונית:

ראשית, בהינתן $v \in Ker(X^T)$, על הצדדים הבאים
מתקיים כי $X^T v = 0$, נכפיל כאן באלכסון X ונקבל:

$$X(X^T(v)) = X(0) = 0$$

ומכאן קיבלנו כי $v \in Ker(XX^T)$ על הצדדים הבאים.

עבור הכיוון השני, בהינתן $v \in Ker(XX^T)$ אזי על הצדדים
מתקיים $XX^T v = 0$ ומכאן:

$$(X^T v)^2 = (X^T v)^T (X^T v) = (v^T X)(X^T v) = v^T (XX^T v)$$

נשתמש בהנחה ונקבל:

$$v^T (XX^T v) = v^T \cdot 0 = \underline{0}$$

ולכן קיבלנו כי $v \in Ker(X^T)$ על הצדדים הבאים.
הראנו הכלה דו-כיוונית ולכן מתקיים:

$$Ker(X^T) = Ker(XX^T)$$

כנדרש.

ב) נוכיח כאן בקצרה הכלה דו-כיוונית:

ראשית, בהינתן $v \in V$ נראה ש- $A^T v \in Im(A^T)$

באמצעות פונקציה על $w \in Ker(A)$ ולכן:

$$< A^T v , w > = (A^T v)^T w = v^T (Aw) = v^T \cdot 0 = 0$$

ולכן לפי הגדרת הצמוד קיבלנו ש $A^T v \in Ker(A)^\perp$.

כעת, לפי הצד השני, נזכר בתשלומים שראינו בקורס הקודם:

(I) $$\dim V = \dim(Ker A) + \dim(Im A)$$

(II) $$\dim(Im A) = \dim(Im A^T)$$

$$\Downarrow$$

$$\dim V = \dim(Ker A) + \dim(Im A^T)$$

(III) $$\dim V = \dim(Ker A) + \dim(Ker A^\perp)$$

$$\Downarrow$$

$$\dim(Im A^T) = \dim(Ker A^\perp)$$

לכן, חסמנו שמלא ש $Im A^T \subseteq Ker A^\perp$ ולכן, בהמשך

לשוויון המימדים נסיק כי מתקיים:

$$Im A^T = Ker A^\perp$$

כנדרש.

3) נתבונן בשאלה הקודמת ונשים-לב כי בהינתן $y \perp \operatorname{Ker} X$

אזי $y \in \operatorname{Ker} X^{\perp}$ ולכן $y \in \operatorname{Im} X^T$ ולכן נוכל

כי בהינתן $y \in \operatorname{Im} X^T$ אזי מלעיל המשוואה אינה פריבילית:

קודם, נסמן את עמודות המטריצה X כלומר $x_1, \dots, x_n \in \mathbb{R}^n$.

כיוון ונתון כי X^T אינה הפיכה אזי קיימת שקולות

בין עמודות המטריצה וקיימים $c_1, \dots, c_k \in \mathbb{N}$ ו $a_1, \dots, a_k \neq 0$ המקיימים:

$$a_1 x_{c_1} + a_2 x_{c_2} + \dots + a_k x_{c_k} = 0$$

כלומר, בהינתן $y \in \operatorname{Im} X^T$ קיים ווקטור $v \in \mathbb{R}^n$ המקיים:

$$v_1 x_1 + v_2 x_2 + \dots + v_n x_n = y$$

נתבונן את המשוואה הבאה: עבור $t \in \mathbb{N}$

$$v_1 x_1 + v_2 x_2 + \dots + v_n x_n + t\underbrace{(a_1 x_{c_1} + a_2 x_{c_2} + \dots + a_k x_{c_k})}_{=0}$$

נפתח את המחוברים:

$$= \sum_{i=1}^{k} (v_i + t a_i) x_i + \sum_{j=k+1}^{n} v_j x_j$$

ונקבל כי המקיים:

$$X^T \cdot \begin{pmatrix} v_1 + t a_1 \\ \vdots \\ v_k + t a_k \\ v_{k+1} \\ \vdots \\ v_n \end{pmatrix} = y$$

כלומר, המשוואה $X^T w = y$ אינה פריבילית.

Scanned with CamScanner

4) ראשית, נשים לב כי כל המשוואות מעל קיים פתרון
שם אם $y \in \operatorname{Ker} X$ אזי $w=0$ מהווה פתרון וכן
אם $y \notin \operatorname{Ker} X$ אזי $y \in \operatorname{Ker} X^{\perp}$ ולכן משאלה 2 נקבל
כי $y \in \operatorname{Im} X^T$ כלומר, קיים וקטור $w \in \mathbb{R}^n$ כך ש- $X^T w = y$
ובהכפלה במטריצה X נקבל $XX^T w = Xy$.
כלומר, אם המטריצה XX^T הפיכה אזי לכל $b \in \mathbb{R}^m$
המשוואות $(XX^T)y = b$ קיים פתרון יחיד, כאשר
$y = (XX^T)^{-1} b$.
לעומת, אם המטריצה XX^T אינה הפיכה אזי נוכל כלומר
כי כן המטריצה X וכן המטריצה X^T אינן הפיכות
שכן אין יתרון כי מכפלת מטריצות הפיכות הינה הפיכה
וכן נוכל משאלה 1 שאם המטריצה X^T הפיכה אז"מ
המטריצה X הפיכה. כלומר, אם $X^T w = y$ אזי
$y \in \operatorname{Im} X^T$ ועל כל הפתרון משאלה 3 מתקיים שכל הפתרונות
המשוואות אינסוף פתרונות, כנדרש.

5) (a) בהינתן מטריצה P, נראה כי $P^T = P$ ומכן סימטרית:

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

לכן P סימטרית.

(b) בה"כ, נראה באופן מקרטיבי - v_j ונראה כי $Pv_j = v_j$:

$$Pv_j = \sum_{i=1}^{k} v_i v_i^T v_j = \sum_{i=1}^{k} v_i (v_i^T v_j) = \sum_{i=1}^{k} v_i \langle v_i, v_j \rangle =$$

$$= v_1 \langle v_1, v_j \rangle + v_2 \langle v_2, v_j \rangle + \dots + v_j \langle v_j, v_j \rangle + \dots + v_k \langle v_k, v_j \rangle$$

כלומר, מכיוון והבסיס הוא אורתונormלי ומקיים:

$$\langle v_i, v_j \rangle = \begin{cases} 0 & , j \neq i \\ 1 & , j = i \end{cases}$$

נקבל כי הביטוי שלמעלה:

$$= v_j \langle v_j, v_j \rangle = \underline{\underline{v_j}}$$

לכן $Pv_j = v_j$ ומכן מכיוון והבסיס אורתונormלי

הע"ע הוא 1, כנדרש.

(c) בהינתן $v = a_1 v_1 + \dots + a_k v_k \in V$ אזי:

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

לפי הסעיף הקודם

כנדרש.

Scanned with CamScanner

(ד) נתבונן בתנאי הכפלה כי אם P סימטרי אז מתקיים $P = P^T$ ולכן מתקיים:

$$P^2 = P \cdot P = P^T \cdot P = P \cdot P^T$$

כלומר, נראה כי $P^2 = P$ כלומר בהינתן EVD:

$$P^2 = (UDU^T)^2 = UDU^T \cdot UDU^T = UD^2U^T$$

כלומר, D^2 כלומר אלכסון שבו 0 או 1 ולכן מתקיים $D^2 = D$ ולכן:

$$= UDU = \underline{\underline{P}}$$

כנדרש.

(ה) בהתאם לנתון הקודם:

$$(I-P)P = I \cdot P - P^2 \underset{P^2 = P}{=} P - P = \underline{\underline{0}}$$

Scanned with CamScanner

6) נשים לב כי $X=U\Sigma V^T$ ולכן, מהגדרת הטרנספוז, נקבל:

$$X^T=(U\Sigma V^T)^T=V\Sigma^T U^T$$

כלומר,

$$XX^T=U\Sigma \underbrace{V^T V}_{\substack{\text{מאורתוגונלי}\\ =I}} \Sigma^T U^T = U\underbrace{\Sigma\Sigma^T}_{=D}U^T=UDU^T$$

לכן, מכיוון ש- U, U^T אורתוגונליים נקבל:

$$(XX^T)^{-1}=UD^{-1}U^T$$

נמשיך,

$$(XX^T)^{-1}X=UD^{-1}\underbrace{U^TU}_{=I}\Sigma V^T=UD^{-1}\Sigma V^T$$

כלומר, מכיוון ש- $D=\Sigma\Sigma^T$ ומן האפיון האלכסוני של Σ ושל Σ^T נתון, נקבל:

$$=U\begin{pmatrix}\sigma_1^{-2} & & & & \\ & \ddots & & & \\ & & \sigma_r^{-2} & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0\end{pmatrix}\begin{pmatrix}\sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0\end{pmatrix}V^T=U\Sigma^{-1}V^T=\underbrace{U(\Sigma^T)^+V^T}_{\Sigma^+ \text{ הגדרת}}$$

מצד שני,

$$X^{T+}=(X^+)^T=(V\Sigma^+U^T)^T=U(\Sigma^+)^TV^T=U(\Sigma^T)^+V^T$$

ולכן קיבלנו $(XX^T)^{-1}X=X^{T+}$, כנדרש.

ז) מתוך, לפי הגדרה:

$$XX^T = \begin{pmatrix} | & & | \\ x_1 & \cdots & x_m \\ | & & | \end{pmatrix} \begin{pmatrix} - & x_1^T & - \\ & \vdots & \\ - & x_m^T & - \end{pmatrix} = \begin{pmatrix} \sum x_i^1 x_i^1 & \cdots & \sum x_i^1 x_i^m \\ \vdots & & \\ \sum x_i^m x_i^1 & \cdots & \sum x_i^m x_i^m \end{pmatrix}$$

כלומר, אם XX^T הפיכה וגם XX^T מטריצה חיובית, מאחר שקיים קיטלאריות, $X^TX = (XX^T)^T$ הפיכה גם היא ולכן, מה קורה אם נ"ל:

$$\dim(\mathrm{Ker}(X)) = \dim(\mathrm{Ker}(X^TX)) = 0$$

$$\Updownarrow$$

$$\dim(\mathrm{Im}(X)) = \dim(\mathrm{Im}(XX^T)) = d$$

נשים לב כי XX^T הינה מטריצה $d \times d$ ולכן אם נפרק אותה בפירוק SVD נקבל מלבון:

$$\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_d \end{pmatrix}$$

מטריצה $d \times d$ ומהמשפט נקבל d ע"ע ולכן מימד התמונה שלה הוא d.

מה קורה אם מטריצה המימד של X הוא מימד הקטן מהפרמטר כלומר $\mathrm{span}\{x_1, \dots, x_m\} = \mathbb{R}^d$, כלומר.

8) לפי הגדרה,

$$XX^T w = Xy$$

$$U D V^T w = U \Sigma V^T y \quad / \cdot U$$

$$D V^T w = \Sigma V^T y$$

כלומר, בכתיב מלא יש לנו:

$$\begin{pmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix} V^T w = \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix} V^T y$$

לכן, נגדיר עבור $i \geq r+1$ $w_i = 0$ ונכפיל כל אגף ב-D^T ונקבל:

$$\underbrace{D^T D}_{\begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}} V^T w = \underbrace{D^T \Sigma}_{\begin{pmatrix} \sigma_1^{-1} & & & & \\ & \ddots & & & \\ & & \sigma_r^{-1} & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix} = \Sigma^+} V^T y$$

ולכן,

$$V^T w = \Sigma^+ V^T y \quad / \cdot V$$

אורתוגונלי

$$w = V \Sigma^+ V^T y = X^{+T} y$$

לפי הגדרה שהגדרנו $w_i = 0$ עבור $i \geq r+1$ לכל w אמיתי

לכל מלבד יוצא לכן כל $\bar{w}$ כזה ש- $\hat{w}$ בכל קואורדינטה

עבורה $i \leq r$ שם לא מתקיים מהגדרת מקרה לכן σ_i , u_i

שונים מאפס.

בנוסף, עבור $i \geq r+1$ לכל מספר שאר מאפס המכפלה

לכל סקלר $\bar{w}$ מתקיים $\hat{w}_i = \bar{w}_i$ עבור $i \geq r$.

נחשוב על נורמה 2:

$$\|\hat{w}\|_2 = \langle \hat{w}, \hat{w} \rangle = (\hat{w})^T \cdot \hat{w} = \sum_{i=1}^{d} \hat{w}_i \hat{w}_i \underset{\text{השאר אפסים}}{=} \sum_{i=1}^{r} \hat{w}_i \hat{w}_i$$

Scanned with CamScanner

באופן דומה, עבור הוקטור $\bar{w}$, נקבל:

$$\|\bar{w}\|_2 = \langle \bar{w}, \bar{w} \rangle = \sum_{i=1}^{d} \bar{w}_i \bar{w}_i = \sum_{i=1}^{r} \bar{w}_i \bar{w}_i + \underbrace{\sum_{i=r+1}^{d} \bar{w}_i \bar{w}_i}_{\geq 0}$$

ומכאן קיבלנו $\|\hat{w}\|_2 \leq \|\bar{w}\|_2$ וכן $\hat{w}$ בעל הנורמה המינימלית, כנדרש.

Scanned with CamScanner

לירון גרשוני
308350503

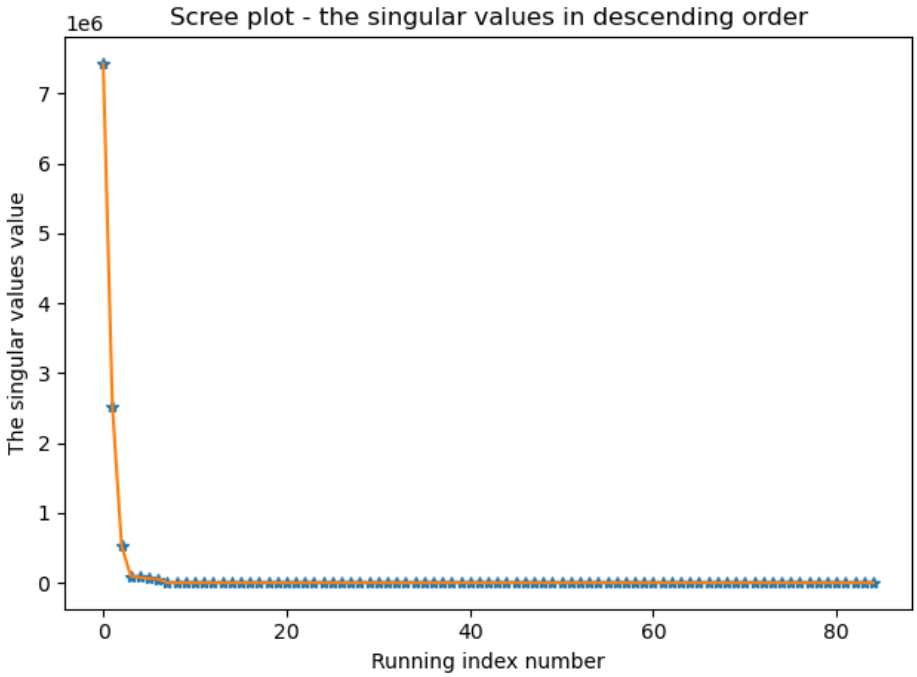
מערכות לומדות – המשך תרגיל בית 2

13) על-מנת לענות על השאלה הזו נכנסתי תחת הקישור שניתן לאתר Kaggle ומצאתי בדיוני הפורום הסבר קצר על כל feature וכך הבנתי יותר מה הכוונה בכל אחד מהם. בהמשך לכך בחרתי להשמיט מהמטריצה הסופית כמה מה-feature's כגון 'id', 'date', 'lat' ו-'long' כאשר כמובן גם הושמטה העמודה של מחירי הבתים שמהווה עבורי את וקטור ה-response. עבור העמודות שנותרו ביצעתי preprocessing שכלל התאמה של הערכים התקינים ע"פ אותו הסבר שמצאתי באתר כמו למשל: ערכי ה-condition שיכולים לנוע בין 1 ל-5 וכן הלאה.

את המידע לקחתי מהאתר: https://www.slideshare.net/PawanShivhare1/predicting-king-county-house-prices

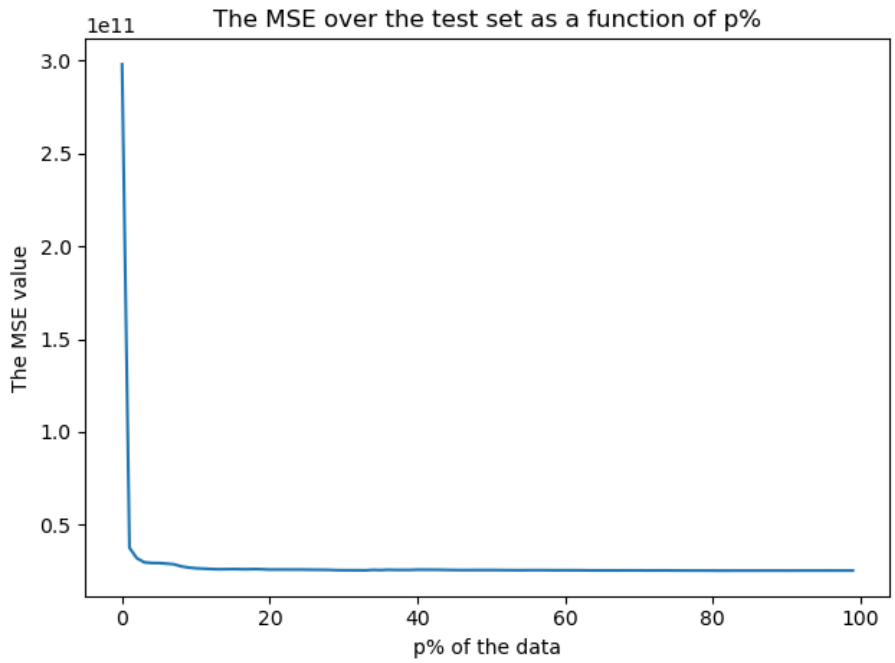
לבסוף, החלטתי שאת העמודה 'zipcode' יש להמיר למשתנה קטגורי שכן לאחר ההסבר שניתן מדובר בקבוצה חסומה של ערכים שאינם מקיימים יחס סדר וכן אין חשיבות לגודלו של המספר זאת למרות שלמיקוד זה השפעה על מחיר הבית.

15)



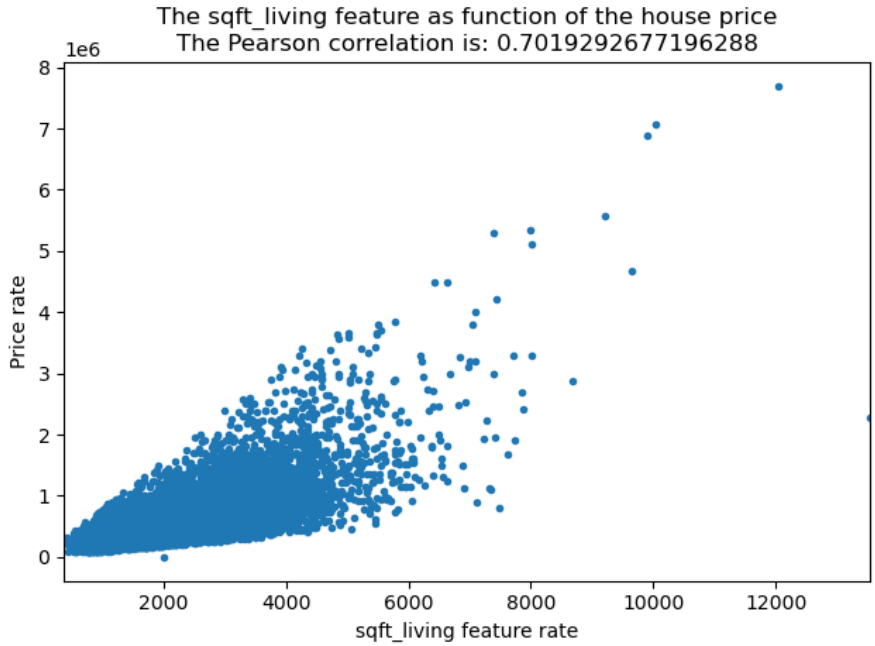
הגרף מעלה משקף את זה שישנם למטריצה ערכים סינגולריים מעטים שאינם אפס אך ערכים אלו מתכנסים מאוד מהר לאפס וכן ישנם לפחות שלושה ערכים שהם יחסית גבוהים לכן המטריצה ככל הנראה לא תהא סינגולרית אך מרבית הערכים הסינגולריים חיוביים ולכן ככל הנראה הפיכה.

(16



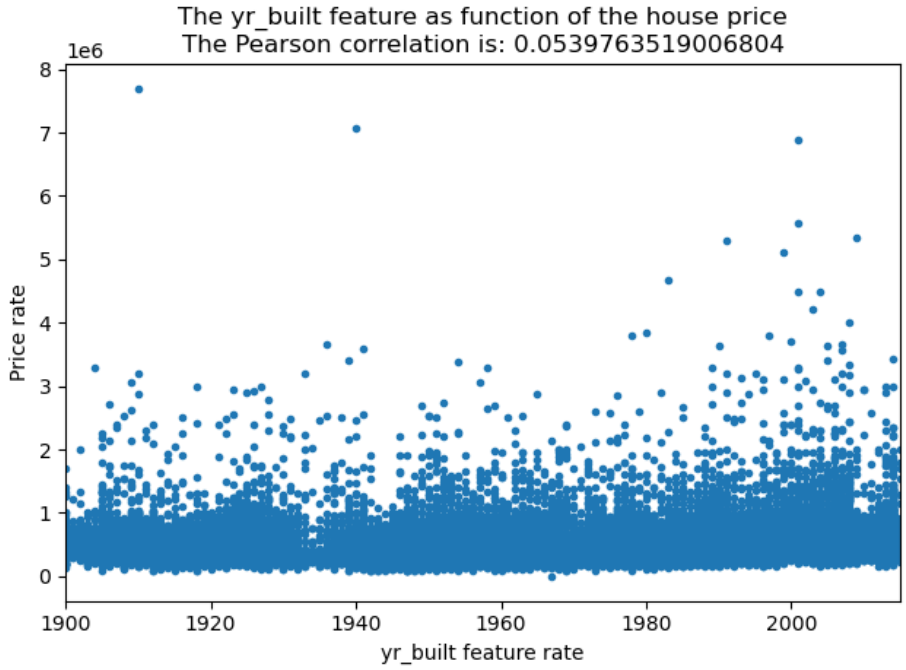
בהמשך לשאלה הקודמת, התקבלו ערכי MSE גבוהים כריבוע הערכים הסינגולריים. בנוסף, ניתן להסיק בקלות שככל ונגדיל את אחוז ה-data בו נשתמש לאימון כך ההערכה שלנו עבור w תהא מדויקת יותר וכן ערכי ה-MSE יתכנסו מהר יותר.

(17



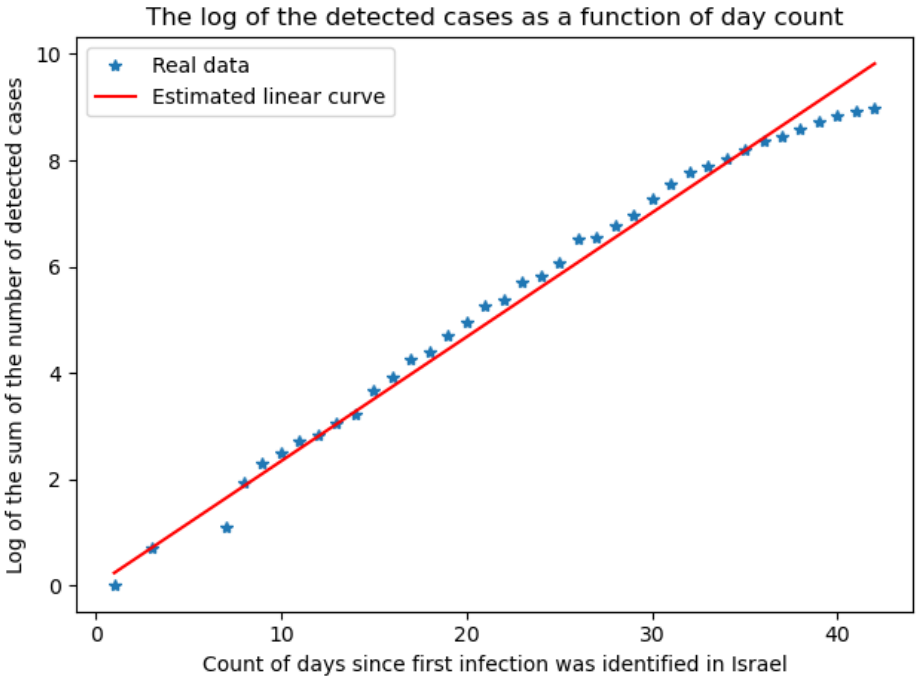
נשים-לב כי את הקשר ההדוק ביותר אל מחיר הבית נמצא בפרמטר המודד את גודל שטח המחייה בבית-ראשית, מכיוון וניתן לראות את ערך הקורלציה (כ-0.7) אשר הינו גבוה ביחס לשאר הפרמטרים וכן גם בהתבוננות פשוטה בגרף בו ניתן לראות עלייה במחיר ככל ששטח המחייה גדל (בנוסף, גם צפיפות הדגימות

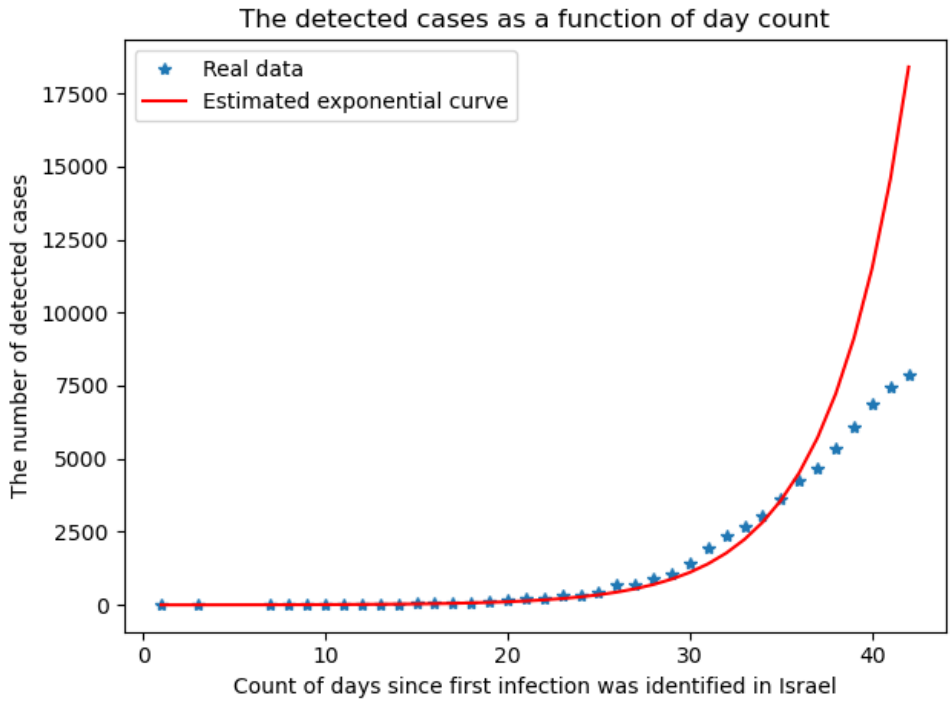
הולכת וקטנה מה שמלמד אותנו כי מספר הבתים בעלי שטח גדול מעטים ביחס לאלו עם שטח קטן- אבל זה כבר דיון אחר..)



לעומת הגרף הקודם, בגרף לעיל קיים קשר נמוך בין שנת בניית הבית ומחירו- הקורלציה נמוכה (כ-0.054) וכן גרפית לא ניתן להסיק מסקנות מיוחדות בהעדפה לשנת בנייה מסוימת לעומת אחרת.

(21





22) ראשית, נזכיר כי כאשר עסקנו ברגרסיה לינארית התקבל הוקטור w=[w_0, w_1] כך שהאומדן שלנו נראה y_hat = w_0 + w_1x. לכן, נוכל להגדיר את פונקציית ה-loss כסכום על ‖y_hat – y_real‖^2 עבור כל קואורדינטה של y. במקרה בו נפעיל את פונקציית הלוג על הערכים האמיתיים של y אנו למעשה מנסים למזער את הפונקציה ‖y_hat – log(y_real)‖^2. לעומת זאת, אם נרצה לשמור על ערכיה המקוריים של y_real ללא הפעלת פונקציית הלוג נוכל להגדיר את y_exp = exp(w_0 + w_1x). אם-כך, פונקציית ה-loss תהא כעת ‖y_exp – y_real‖^2 עבור כל קואורדינטה.

על-מנת למצוא את ה-ERM נגדיר תחילה את y_exp = exp(w_0 + w_1x) = exp(w_0) * exp(w_1x).
אנו מעוניינים למצוא את w_0 ו-w_1 הממזערים את פונקציית ה-loss כלומר, עלינו למצוא את:

$$\text{argmin}(y_exp - y_real)^2 = \text{argmin}([\exp(w_0) * \exp(w_1x)] - y_real)$$

כעת, על-מנת למצוא את ערכים אלו נוכל לחשב את נגזרת הפונקציה לעיל ולהשוות אותה לאפס למציאת נקודת מינימום אשר תמזער את w_0 ו-w_1.