EX2 IML
Linear Regression

Let X be the design matrix of a linear regression problem with p rows (variables) and n columns (samples) in X.
Recall that for some vector space $V \subseteq \mathbb{R}^p$ the orthogonal complement of V is $V^\perp = \{x \in \mathbb{R}^p : <x,v> = 0 \ \forall v \in V\}$

1) Prove that $Ker(X^T) = Ker(XX^T)$

נתמוין מוכתרות

$Ker(X^T) = \{v \in V : X^Tv = 0\}$

$Ker(XX^T) = \{v \in V : XX^Tv = 0\}$

I) יהי $v \in V$ כך ש $v \in Ker(X^T)$ אזי $X^Tv = 0$
נכפיל כי $XX^Tv = 0$ (נכ' יפ' הירסות מ- X)
ולכן $v \in Ker(XX^T)$

II) יהי $v \in V$ כך ש $v \in Ker(XX^T)$ אזי $XX^Tv = 0$
נקבל כי $X^{-1}XX^Tv = 0$ ולכן $X^Tv = 0$.
כלומר $v \in Ker(X^T)$

$<XX^Tv, v> = 0 = <X^TXX^Tv, X^Tv> = <X^Tv, X^Tv> = \|X^Tv\|^2$

מ- I נסיק כי $Ker(X^T) \subseteq Ker(XX^T)$
II נסיק כי $Ker(XX^T) \subseteq Ker(X^T)$

מסקנה קיבלנו כי $Ker(X^T) = Ker(XX^T)$

2) Prove that for some square A : $Im(A^T) = Ker(A)^\perp$

נתמוין מוכתבת

$Im(A^T) = \{0 \neq w \in W ; w = A^Tv, v \in V\}$

$Ker(A)^\perp = \{x \in \mathbb{R}^p : <x,v> = 0 \ \forall v \in Ker(A)\}$

$Ker(A) = \{v \in V : Av = 0\}$

I) יהי $w \in Im(A^T)$ אזי $w = A^Tv$ עבור $v \in V$ וגם $w \neq 0$.
יהי $v \in Ker(A)$ אזי $Av = 0 \Leftarrow$
נראה כי $<w,v> = 0$

$<w,v> = <A^Tv, v> = v^T(A^Tv) = v^TA^Tv \underset{v \in Ker(A)}{=} (Av)^T \cdot v = 0 \cdot v = 0$

כלומר עבור $v \in Ker(A)$ מתקיים $<w,v> = 0$ ולכן $w \in Ker(A)^\perp$

לסיכום $w \in Im(A^T) \Rightarrow w \in Ker(A)^\perp$
כלומר $Im(A^T) \subseteq Ker(A)^\perp$

II) יהי $w \in Ker(A)^\perp$ ולכן עבור $v \in Ker(A)$ מתקיים $<w,v> = 0$

$\Rightarrow v^Tw = 0$ ונזכור $Av = 0 \Rightarrow (Av)^T = v^Tw$

$\Rightarrow v^TA^T = v^Tw \Rightarrow A^T = w \Rightarrow A^T \cdot \begin{pmatrix} \vdots \end{pmatrix} = w \Rightarrow w = A^Tv'$

$\Rightarrow w \in Im(A^T)$

לסיכום $Ker(A)^\perp \subseteq Im(A^T)$

ומסקנה הוכחנו הכלה משני הכיוונים ולכן השוויון מתקיים.

3) Let $y = X^T w$ be a non-homogeneous system of linear equation. Assume that X^T is square and not invertible. Show that the system has ∞ solutions $\iff y \perp \ker(X)$

נזכיר כי שמטריצה ריבועית למערכת המשוואות $X^T w = y$ יש אינסוף פתרונות אם"ם $\dim(\ker(X^T)) \neq 0$

$$\ker(X^T) = \{v \in V : X^T v = 0\}$$

נתון כי X^T היא מטריצה ריבועית ואין לה הופכית

I) נתון כי למערכת יש אינסוף פתרונות
ונזכיר כי $Im(X^T) = \{0 \neq w \in W : w = X^T v, v \in V\}$

ולכן $y = X^T w \in Im(X^T)$
כמו כן הוכחנו בסעיף קודם כי $Im(X^T) = \ker(X)^\perp$,
ולכן $y \in \ker(X)^\perp \Leftarrow y \perp \ker(X)$

II) נתון כי $y \perp \ker(X) \Leftarrow y \in \ker(X)^\perp$ ולכן $y \in Im(X^T)$
כמו כן הראינו כי $Im(X^T) = \ker(X)^\perp$
$\Leftarrow$ $Im(X^T)$ אינה ריקה

מכיוון ש-X^T אינה הפיכה $\Leftarrow$ למשוואה $y = X^T w$ קיים $w \in V$
יש 0 פתרונות או אינסוף פתרונות
$\Leftarrow$ למערכת המשוואות יש אינסוף פתרונות
שכן $Im(X^T)$ אינה ריקה

4) Consider the normal linear system $XX^Tw = Xy$. Using what you have proved above, prove that the normal equations can only have a unique solution if XX^T is invertible or infinitely many solutions.

Normal Equations are $Xy = XX^Tw$

I) אם ל-XX^T קיימת הופכית $(XX^T)^{-1}$ אזי $(XX^T)^{-1}Xy = w$
ומכיוון ש-XX^T היא הפיכה אזי לכל $v \in V$ קיים $w \in W$ כך ש-$XX^Tv = w$ ואזי קיים פתרון יחיד

II) אם XX^T אינה הפיכה, אזי משאלה 3 נסיק כי:
למערכת המשוואות $Xy = XX^Tw$ יש אינסוף פתרונות אם"ם $Xy \perp \ker(XX^T)$

נתון כי הוכחנו משאלה 1 כי $\ker(X) = \ker(XX^T)$
ולכן אם נוכיח כי $Xy \perp \ker(X)$ נוכל להסיק שיש אינסוף פתרונות

$Xy \perp \ker(X) \iff Xy \in \ker(X)^\perp \iff$ לכל $v \in \ker(X)$
מתקיים כי $\langle Xy, v\rangle = 0$

לכן צריך להראות שלכל $v \in \ker(X)$ מתקיים $\langle Xy, v\rangle = 0$

יהי $v \in \ker(X) \Rightarrow X^Tv = 0$

$$\langle Xy, v\rangle = y^T X^T v = y^T 0 = 0$$

לכן אם XX^T אינה הפיכה יש למשוואה אינסוף פתרונות

5) a) נוכיח כי P מטריצה סימטרית ע"י כך שנראה כי $P=P^T$

$$P = \sum_{i=1}^{k} v_i v_i^T = v_1 v_1^T + v_2 v_2^T + \dots + v_k v_k^T$$
$$= (v_1^T)^T v_1^T + (v_2^T)^T v_2^T + \dots + (v_k^T)^T v_k^T$$
$$= (v_1 v_1^T)^T + (v_2 v_2^T)^T + \dots + (v_k v_k^T)^T$$
$$= (v_1 v_1^T + v_2 v_2^T + \dots + v_k v_k^T)^T$$
$$= \left(\sum_{i=1}^{k} v_i v_i^T\right)^T = P^T$$

$\Rightarrow P = P^T \Rightarrow$ P סימטרית

(b) + (d) נוכיח כי $P=P^2$ ומכאן נוכיח כי הע"ע של P הם 0 או 1 ושהבסיס האורתונורמלי $(v_1,\dots,v_k)$ מכיל את הו"ע שמתאימים לערך העצמי 1

(d)
$$P^2 = \sum_{i=1}^{k} v_i v_i^T \sum_{j=1}^{k} v_j v_j^T = \sum_{i=1}^{k}\sum_{j=1}^{k} v_i v_i^T v_j v_j^T$$
$$= \sum_{i=1}^{k}\sum_{j=1}^{k} v_i \langle v_j, v_i\rangle v_j^T = \sum_{i=1}^{k}\sum_{j=1}^{k} v_i \delta_{ij} v_i^T$$
$$= \sum_{i=1}^{k} v_i v_i^T = P$$

שכן כל האיברים שעבורם $i\neq j$ נמחקים כי המכפלה הפנימית אפס ונותרים רק כאשר $i=j$

אנו משתמשים בכך ש- v_i, v_j הם וקטורים אורתונורמליים ולכן $\langle v_i, v_j\rangle = \begin{cases} 1 & i=j \\ 0 & else \end{cases} = \delta_{ij}$

כעת יהי x ו"ע שמתאים לע"ע λ אזי $Px = \lambda x$

נשים לב ש- $P^2 x = Px = \lambda x$

ולכן
$$P^2 x = P(Px) = P(\lambda x) = \lambda Px = \lambda(\lambda x)$$

ולכן $\lambda x = \lambda^2 x \Rightarrow \lambda x(1-\lambda) = 0$

$\Rightarrow \lambda_0 = 0, \quad \lambda_1 = 1$ עבור $x\neq 0$

וקיבלנו כי הע"ע של P הם $\lambda_1 = 1$, $\lambda_0 = 0$

יהי v_j שייך למטריצה P אזי
$$Pv_j = \sum_{i=1}^{k} v_i v_i^T v_j = \sum_{i=1}^{k} v_i \langle v_j, v_i\rangle = \sum_{i=1}^{k} v_i \delta_{ij} = \begin{cases} v_i & i=j \\ 0 & else \end{cases}$$
וקטורים אורתונורמליים

מכאן שהמטריצה האלכסונית של המערכת היא $M = \begin{bmatrix} I_{k\times k} & 0 \\ 0 & 0 \end{bmatrix}$

$I_{k\times k}$ מטריצת היחידה בגודל $k\times k$ שהעמודות שלה הן בסיס אורתונורמלי

מכילה הו"ע של M ו- P דומים

נשים לב כי בסיס אורתונורמלי $(v_1, \dots, v_k)$ ניתן להרחבה לבסיס אורתונורמלי $(v_1, \dots, v_k, \dots, v_n) \in \mathbb{R}^n$

הוקטור נשאר ללא שינוי ע"י P ומתאים לע"ע 1 ; הוקטור מתאפס ומתאים לע"ע 0

(c) נוכיח כי לכל $v \in V$ מתקיים $Pv = v$

יהי $v \in V$ לכן יש $v = \sum_{j=1}^{k} a_j v_j$ עבור סקלרים a_j לכל $1 \le j \le k$

$$Pv = \sum_{i=1}^{k} v_i v_i^T v = \sum_{i=1}^{k} v_i v_i^T \sum_{j=1}^{k} a_j v_j = \sum_{i=1}^{k} \sum_{j=1}^{n} a_j v_i v_i^T v_j$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{n} a_j v_i \langle v_j, v_i \rangle = \sum_{i=1}^{k} \sum_{j=1}^{n} a_j v_i \delta_{ij}$$

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

(d) הוכחה נוספת ש- $P = P^2$ בהסתמך על כך שהע"ע הם $0,1$

$\mathbb{R}^d = V_{\lambda_1} \oplus V_{\lambda_0} \Rightarrow$ קיימים $a \in V_{\lambda_1}$, $b \in V_{\lambda_2}$ כך ש- $v = a + b$

$$P^2 v = P^2(a+b) = P\,P(a+b) = P(Pa + Pb) = P(0 + b)$$

$$= Pb = Pb + 0 = Pb + Pa = P(a+b) = Pv$$

(e) $P^2 = P \iff P - P^2 = 0 \iff P(I - P) = 0 \iff (I - P)P = 0$

(6) Assume XX^T is invertible

Show that $(XX^T)^{-1} = UD^{-1}U^T$ where $D = \Sigma\Sigma^T$

Use this to show that $(XX^T)^{-1}X = X^{T+}$

נכון כי מטריצה ממשית ניתן לרשום כפירוק הבא:

$$X = U\,\Sigma\,V^T$$

מטריצה אורתוגונלית   מטריצה אלכסונית   מטריצה אורתוגונלית

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

$$= U\Sigma V^T V\Sigma^T U^T$$

אורתוגונלית V

$$= U\Sigma I \Sigma^T U^T$$

$$= UDU^T$$

נכון כי

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

נראה שזה אכן ההופכי כי

$$XX^T \cdot (XX^T)^{-1} = I$$

$$UDU^T \cdot UD^{-1}U^T = UDID^{-1}U^T$$

אורתוגונלית U

$$= UDD^{-1}U^T = UIU^T = UU^T = I$$

כנדרש

כעת נרצה להוכיח כי

$$(XX^T)^{-1}X = X^{T+}$$

נשים לב:

$$\Sigma^+_{ii} = \begin{cases} \frac{1}{\Sigma_{ii}} & \Sigma_{ii} \neq 0 \\ 0 & \text{else} \end{cases}, \quad X^{T+} = (V\Sigma U^T)^+ = (X^+)^T = (V\Sigma^+ U^T)^T = U\Sigma^{T+}V^T$$

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

U אורתוגונלית

$$= UD^{-1}\Sigma V^T$$

Σ מטריצה אלכסונית

$$\Sigma = \left(\begin{array}{ccc|c} \Sigma_{11} & & & \\ & \Sigma_{22} & & 0 \\ & & \Sigma_{kk} & \\ \hline & 0 & & 0 \end{array}\right) = \begin{pmatrix} \Sigma_{k\times k} & 0 \\ 0 & 0 \end{pmatrix}$$

$$\Sigma^T = \left(\begin{array}{ccc|c} \Sigma_{11} & & & \\ & \Sigma_{22} & & 0 \\ & & \Sigma_{kk} & \\ \hline & 0 & & 0 \end{array}\right) \Longrightarrow (\Sigma\Sigma^T)^{-1} = (\Sigma^2)^{-1}$$

הוכחה: XX^T הפיכה $\Longleftarrow$ $U\Sigma V^T V\Sigma^T U^T$ הפיכה

$\Longleftarrow$ $U\Sigma^2 U^T$ הפיכה

Σ אלכסונית ולכן $\Sigma\Sigma^T = \Sigma^2$

$\Longleftarrow$ D הפיכה

$\Longleftarrow$ ההופכית של D היא

$$D^{-1} = \begin{pmatrix} \frac{1}{\Sigma_{11}^2} & & \\ & \frac{1}{\Sigma_{22}^2} & \\ & & \ddots \\ & & & \frac{1}{\Sigma_{nn}^2} \end{pmatrix}$$

$$D^+ = (\Sigma\Sigma^T)^+ = \begin{pmatrix} \frac{1}{\Sigma_{11}^2} & & \\ & \ddots & \\ & & \frac{1}{\Sigma_{nn}^2} \end{pmatrix}$$

כמובן

כעת נבחין שזה נכון עבור $m>n$

$$\begin{array}{ccc} & D^{-1} & \Sigma \\ (XX^T)^{-1}X = U & \begin{pmatrix} \frac{1}{\Sigma_{11}^2} & & \\ & \frac{1}{\Sigma_{22}^2} & \\ & & \ddots \\ & & & \frac{1}{\Sigma_{nn}^2} \end{pmatrix} & \left(\begin{array}{cccc|c} \Sigma_{11} & & & 0 & \\ & \Sigma_{22} & & & 0_{n\times m-n} \\ & & \ddots & & \\ 0 & & & \Sigma_{nn} & \end{array}\right) \\ & & n\times m \end{array}$$

$$\begin{array}{cc} & n\times n \\ V^T = U & \left(\begin{array}{cccc|c} \frac{1}{\Sigma_{11}} & & & 0 & \\ & \frac{1}{\Sigma_{22}} & & & 0_{n\times m-n} \\ & & \ddots & & \\ 0 & & & \frac{1}{\Sigma_{nn}} & \underbrace{\qquad}_{m-n} \end{array}\right) V^T \end{array}$$

$$= U\Sigma^{T+}V^T$$

כל העמודות של האפסים
שווים ל-0 ולכן
$D^{-1} = D^+$
נשאר בכפל הזה
רק העמודות הלא אפסים בכל מקרה
הוכחנו

נשים לב שמתקיים

$$D^{-1}\Sigma = \Sigma^+$$

ולסיכום

$$(XX^T)^{-1}X = UD^{-1}\Sigma V^T$$

$$= U\Sigma^+V^T = X^+$$

7) $XX^T \in \mathbb{R}^{d\times d}$, המטריצה הפיכה $\iff rank(XX^T)=d$

הדרגה של מטריצה מוגדרת להיות המימד של התמונה

$$rank(XX^T) = dim(Im(XX^T))$$
$$rank(X) = dim(Im(X))$$

הדרגה של X זהה למימד של את המרחב הפורש $\{x_1,..x_m\}$

כלומר

$$Span\{x_1,..x_m\}=\mathbb{R}^d \iff rank(X)=d=dim(Im(X))$$

מסק"ה

$$XX^T \text{ הפיכה } \iff rank(XX^T)=d \iff dim(Im(XX^T))=d \iff$$
$$\iff dim(Ker(XX^T))=d-d=0 \overset{}{\underset{\text{לפי שאלה 1}}{\iff}} dim(Ker(X^T))=0$$
$$\iff dim(Im(X^T))=d-0=d \iff rank(X^T)=d \iff$$

מרחב השורות זהה למרחב העמודות

$$rank(X)=d \iff Span\{x_1,..x_m\}=\mathbb{R}^d$$

8) תהי $\bar{w}$ שמקיימת את ה- normal equation $XX^T\bar{w} = Xy$

I) אם XX^T הפיכה אזי $(XX^T)^{-1}XX^T\bar{w}=(XX^T)^{-1}Xy$

עפ שאלה 6 נסיק כי $\bar{w} = (XX^T)^{-1}Xy = X^{T\dagger}y=\hat{w}$

לכן $\|\bar{w}\|_2=\|\hat{w}\|_2$

II) אם XX^T לא הפיכה אזי

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

לכל $r+1\le i\le d$ $\hat{w}_i=0$

$$=\sum_{i=1}^r \bar{w}_i^2$$

תהי $\bar{w}$ שמקיימת $XX^T\bar{w}=Xy$ אזי לכל $1\le i\le r$ $\hat{w}_i=\bar{w}_i$

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

כמו מ שאלה 7 אם XX^T לא הפיכה אזי $r=rank(X)<d$

Shelly Gamlielly

Iml Ex2

13.

- In your submitted PDF explain what were the features you found to be categorical.

בתרגיל מוסבר כי משתנים קטגוריאלים הם פיצ'רים שאין סדר הגיוני לערכים שלהם במובן שאין משמעות לזה שאחד גדול מהשני. אנו בודקים את המשמעות של אותו פיצ'ר מול הresponse vector שהוא המחיר, כלומר כיצד ישפיע המיקוד על המחיר. אכן מיקום עשוי להשפיע על פרדיקציה של מחיר הדירה שכן יש מיקומים יוקרתיים, אולי קרובים לים או לפארק שמעלה את המחיר. אולם המיקוד מבחינת הערך המספרי שלו לא נותן אינפורמציה שתשפיע על המחיר. בית עם מיקוד n לא בהכרח יהיה זול יותר מבית במיקוד n+1 , כלומר הגודל של המספר לא משפיע. מכאן שאין קשר לינארי בין גודל במיקוד לבין המחיר, ולכן זהו משתנה שעל פי הגדרה קטגוריאלי.

14+15.

14. Implement a function called 'plot_singular_values' that receives a collection of singular values and plots them in descending order. That is: x-axis a running index number and y-axis the singular value's value. This kind of plot is called a scree-plot.

Describe what can be learned from the singular values.

נשים לב להסבר מתוך התרגול:

$A : \mathbb{R}^d \rightarrow \mathbb{R}^m$ (matrix $A \in M_{m\times d}(\mathbb{R})$) admits a singular value decomposition of the form:

$$A = U\Sigma V^\top$$

which means that $A(w)$ actually does the following:

1. rotation of w (multiply by orthogonal matrix $V^\top$).
2. stretch the rotated vector (multiply by diagonal matrix Σ).
3. apply another rotation to the stretched vector (multiply by orthogonal matrix U).

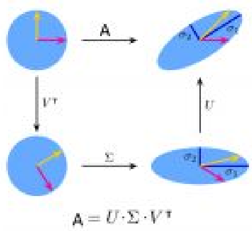
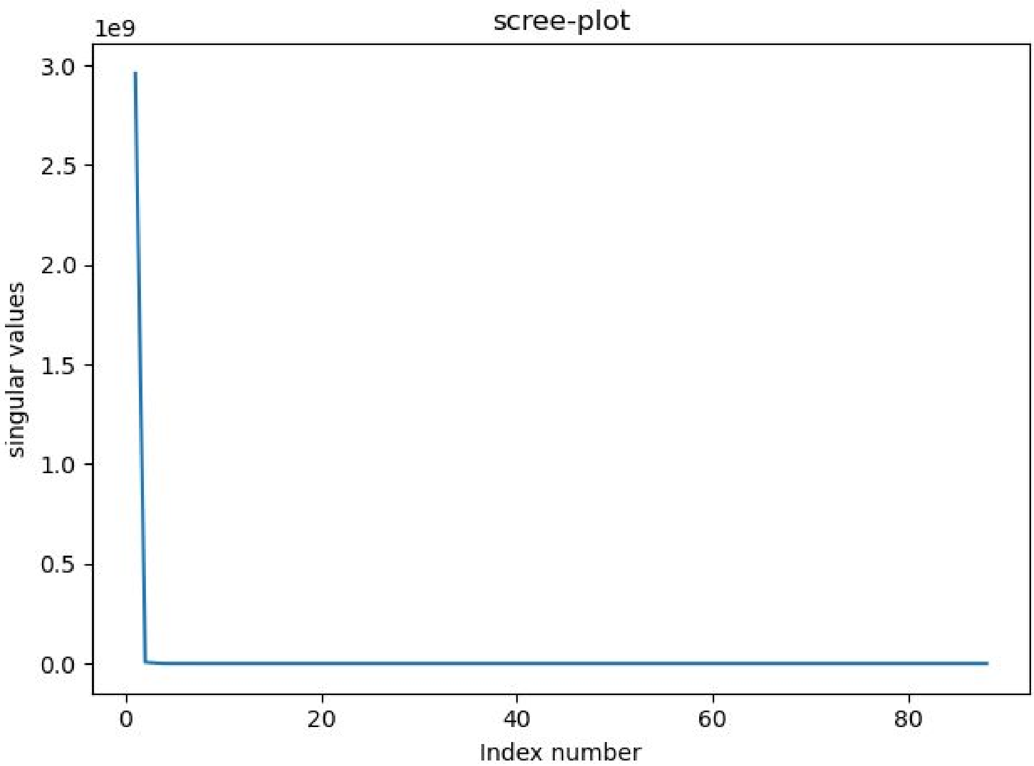


Figure 3: SVD intuition

כלומר המטריצה האלכסונית (על האלכסון מצויים הערכים הסנגולריים) אחראית על המתיחה.

מכאן שבפירוק SVD הערכים הסינגולריים ישפיעו על המתיחה ולמעשה ישפיעו על המשקל של כל פיצ'ר שחשבננו על מנת לחזות מהו המחיר בהינתן פיצ'רים.



למעשה ככל שהערך הסינגולרי גדול יותר כך מידת ההשפעה על המחיר של אותו הפיצ'ר ,שמתאים לערך הסינגולרי ,תהיה גדולה יותר.
כמו כן, ככל שהערכים הסינגולרים קטנים יותר קיימת תלות בין הפיצ'רים המתאימים להם.
יש לנו מעט ערכים סינגולריים גדולים, והשאר קרובים ל-0, כך שהמטריצה ניתנת לייצוג במימד נמוך יותר.
כמות החדרים וגודל הדירה משפיעים במידה רבה על המחיר, אך ההשפעה היא במימד אחר.

<u>Is the design matrix (X) close to being singular or not?</u>

נתבונן בערכים הסינגולרים:
ונבחין כי ישנו ערך ‎-13^10 * 2.78.
מכאן שהמטריצה כמעט סינגולרית (הערך אינו 0 אך מאוד קרוב ל-0).

Definition 6. *Let $A \in \mathbb{R}^{n\times n}$ be a square matrix. A is called invertible (or non-singular) if there exists a matrix $B \in \mathbb{R}^{n\times n}$ such that $AB = I_n = BA$. We denote A's inverse by A^{-1}.*

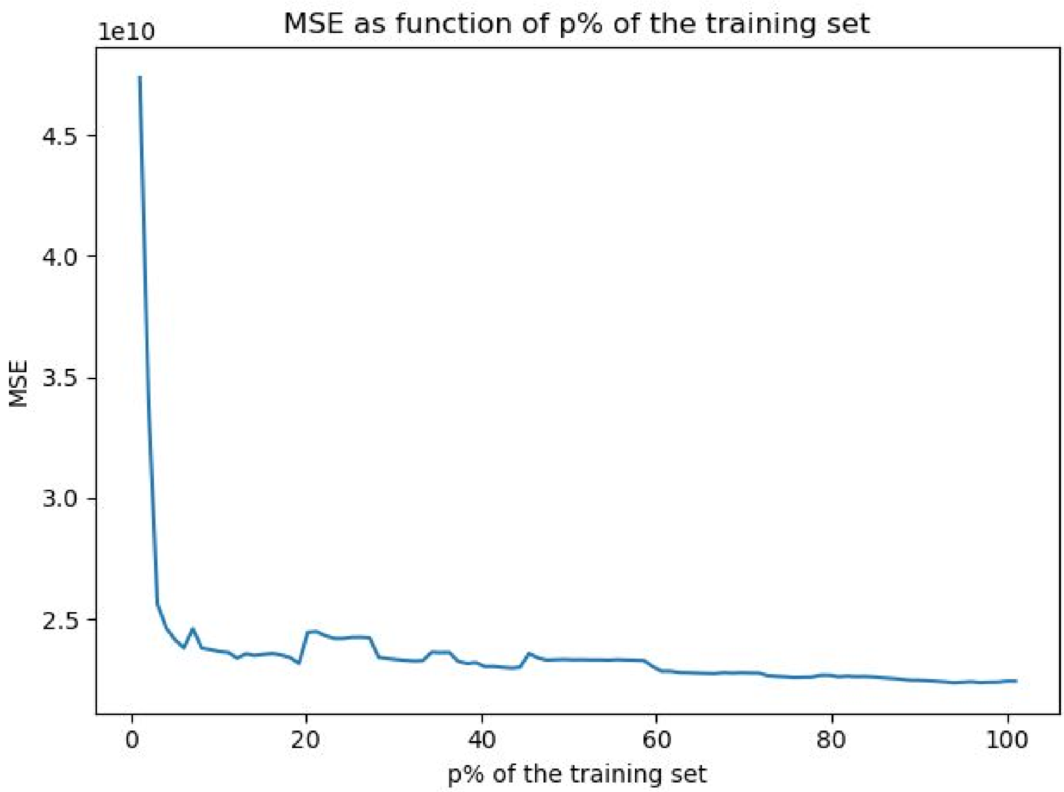
למדנו כי מטריצה סינגולרית היא מטריצה שאין לה הופכית, כאשר בפירוק SVD אם במטריצה האלכסונית קיים ערך סינגולרי שהוא 0 אזיי היא לא הפיכה ובהתאם המטריצה שלה בצענו פירוק.
הערכים הסינגולרים שהצגנו מופיעים על אלכסון המטריצה האלכסונית, כך שאם אחד מהם הוא 0, היא לא הפיכה ולכן סינגולרית.

[2.96002536e+09 6.84653975e+06 2.47573993e+06 1.90644171e+05
7.74035145e+04 5.81938652e+04 8.15367369e+03 3.72533296e+03
1.16840223e+02 1.01544634e+02 9.57995510e+01 8.77627322e+01
6.86982942e+01 5.20182591e+01 2.77800538e+01 2.41983198e+01
2.39471879e+01 2.38439049e+01 2.37372232e+01 2.32513301e+01

2.29386432e+01 2.28751631e+01 2.22151595e+01 2.21473020e+01
2.16477090e+01 2.16382584e+01 2.14344112e+01 2.11177092e+01
2.08274306e+01 2.06521048e+01 2.04891164e+01 2.00903785e+01
1.99575984e+01 1.98373026e+01 1.94706516e+01 1.87285912e+01
1.87001881e+01 1.86073001e+01 1.83535942e+01 1.82075088e+01
1.81254274e+01 1.77845556e+01 1.76432698e+01 1.74950469e+01
1.69762785e+01 1.68878070e+01 1.67942833e+01 1.67731572e+01
1.67286317e+01 1.66010025e+01 1.65414575e+01 1.64966113e+01
1.64091676e+01 1.63109383e+01 1.62343031e+01 1.62050557e+01
1.61337494e+01 1.60701726e+01 1.59652342e+01 1.56658969e+01
1.54764730e+01 1.50586254e+01 1.49139710e+01 1.44846981e+01
1.43829571e+01 1.39775714e+01 1.38092132e+01 1.37552995e+01
1.35000640e+01 1.32485493e+01 1.29037860e+01 1.20309997e+01
1.19297532e+01 1.17822418e+01 1.11992784e+01 1.09346425e+01
1.04323191e+01 1.01867896e+01 1.01016386e+01 9.83462855e+00
8.95093478e+00 7.91702189e+00 7.15769055e+00 6.79298100e+00
1.70861561e+00 5.33330065e-11 **2.78716840e-13**]

16.

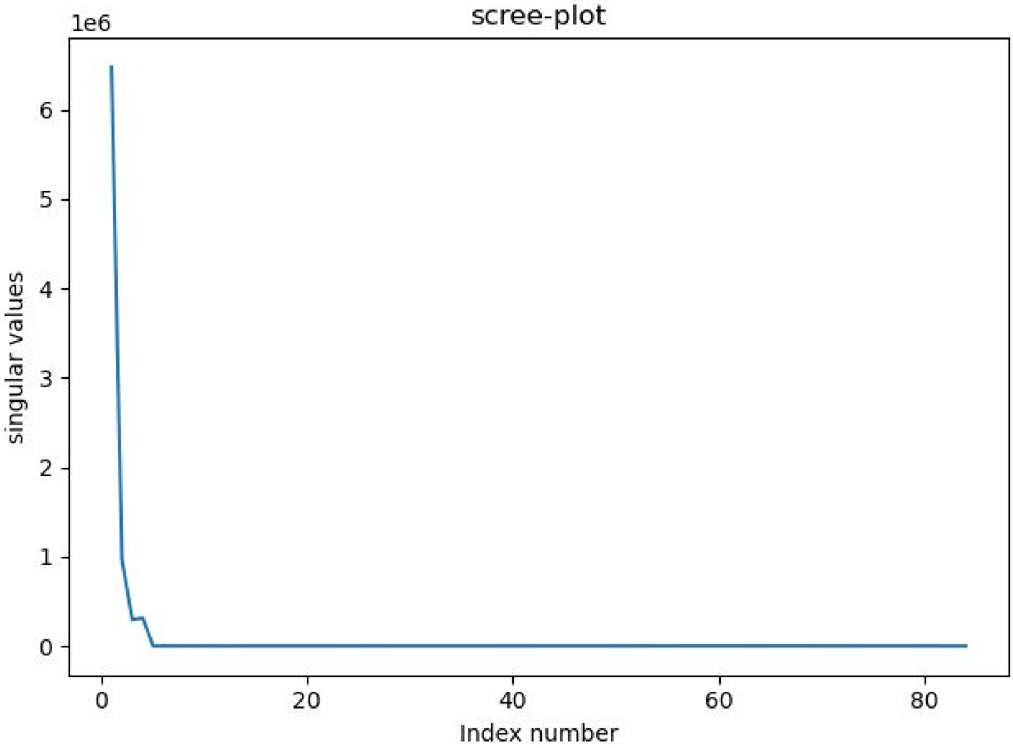
- Add to your answers PDF the plot of the MSE over the test set as a function of $p\%$.
- Explain the results you got.



אנו מחשבים את השגיאה הריבועית הממוצעת, ובודקים עד כמה טעינו בניחוש.

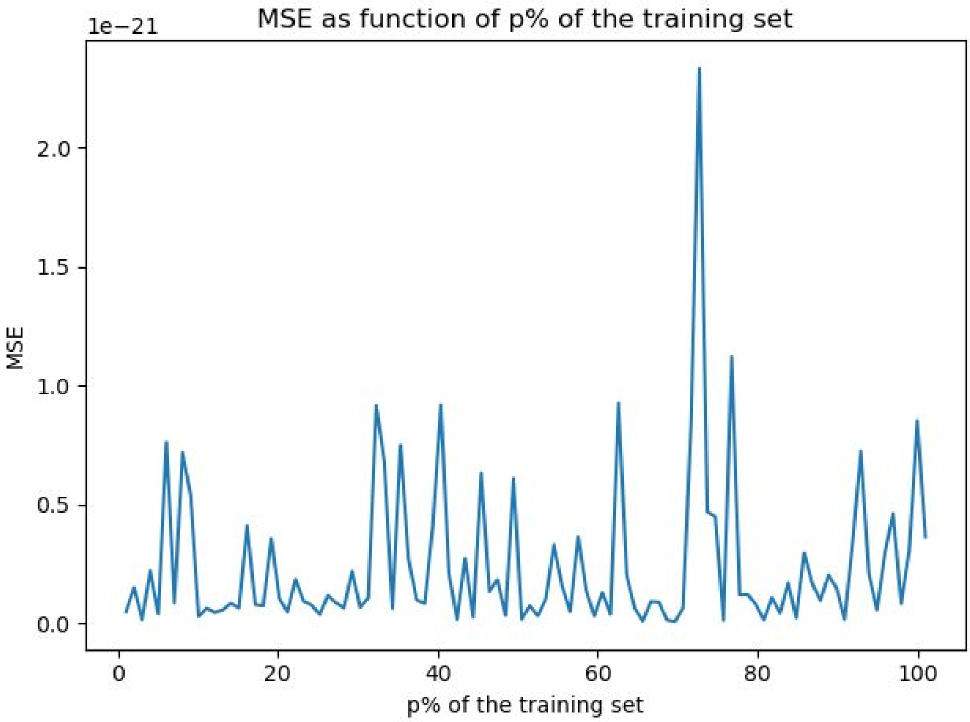
כלומר אנו מקבלים מטריצת פיצ'רים ובעזרתה מנסים לנחש מה המחיר של הבית, נרצה לבדוק עד כמה הניחוש שלנו היה טוב.
נעשה זאת על ידי כך שנשווה בין הפרדיקציה לבין המחיר הידוע האמיתי.
התבקשנו לבצע חיתוך של הדאטא לtraining set ולtest set, ובכל פעם עבור אחוז מסויים של הדאטא לחשב את המשקלים על הtraining set ולבדוק את השגיאה על הtest set.
בכל איטרציה האחוז של הדאטא שנדגם גדל כלומר בצענו חישוב של המשקלים על כמות יותר גדולה של דאטא. ככל שהדאטא גדול יותר כך נצפה לקבל תוצאה מדוייקת יותר וקרובה יותר למציאות.
רואים שהחל מ60% אנחנו נשארים די יציבים מבחינת השגיאה עד שמכסים את כל הtraining set.
ובאופן יחסי רואים ירידה בשגיאה ככל שהאחוז גדל.

- כמו כן, אני מוספיה את התוצאות שהתקבלו כאשר הורדתי לחלוטין את הפיצ'ר של הdate:
פונקציית הmse נראית שונה והסדר גודל שנה מאוד קטן - 21-^10



[6.47588522e+06 4.80970692e+05 9.70783222e+04 7.74043545e+04
1.22007344e+02 1.04505530e+02 1.02724317e+02 9.26438524e+01
7.34482709e+01 5.30185097e+01 2.95051535e+01 2.42264695e+01
2.39762866e+01 2.38854589e+01 2.37539135e+01 2.32846233e+01
2.31665312e+01 2.29866194e+01 2.23738408e+01 2.22343961e+01
2.19101935e+01 2.17605809e+01 2.14780562e+01 2.12187942e+01
2.08923521e+01 2.08348742e+01 2.05082716e+01 2.01674006e+01
2.01114179e+01 1.98534506e+01 1.95126852e+01 1.89176511e+01
1.87573700e+01 1.86834064e+01 1.84018201e+01 1.83365365e+01
1.81544323e+01 1.78340978e+01 1.76934510e+01 1.75348047e+01
1.70546445e+01 1.68922913e+01 1.68401130e+01 1.67808558e+01
1.67688174e+01 1.67047845e+01 1.66300495e+01 1.65204014e+01
1.64695914e+01 1.63873774e+01 1.63293966e+01 1.62180526e+01

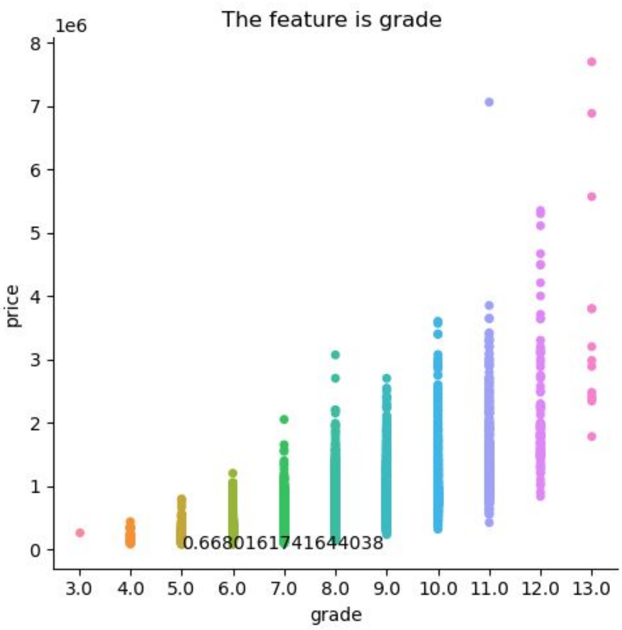
1.62095271e+01 1.60933275e+01 1.59741773e+01 1.58003262e+01
1.57364883e+01 1.56951754e+01 1.51591119e+01 1.49974323e+01
1.47021949e+01 1.41820308e+01 1.39672953e+01 1.38530351e+01
1.37709442e+01 1.35665796e+01 1.33932920e+01 1.28607010e+01
1.19842500e+01 1.18271294e+01 1.16783012e+01 1.11947717e+01
1.09841740e+01 1.04662501e+01 1.02190310e+01 1.00510142e+01
9.93735481e+00 8.87942877e+00 7.58459844e+00 7.04506238e+00
2.08924559e+00 1.28680969e+00 **1.19979102e-10**]

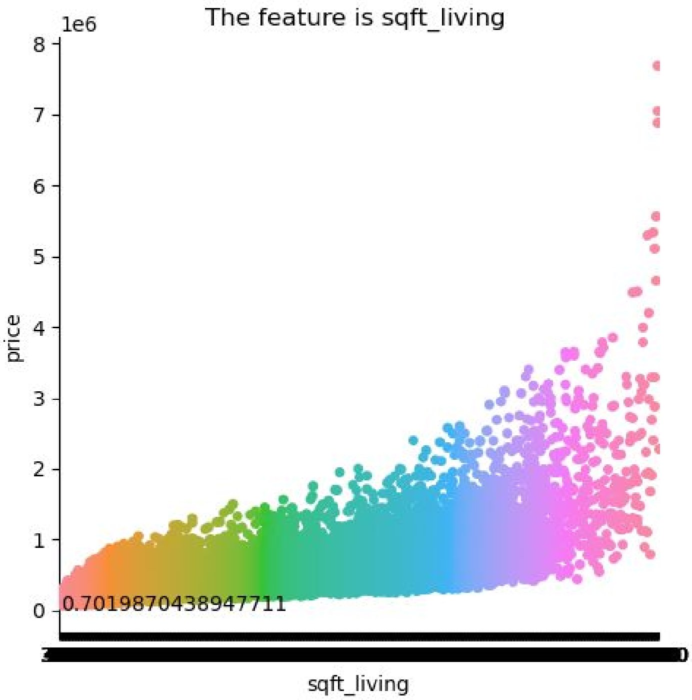


17.

Choose two features, one that seems to be beneficial for the model and one that does not. In your answers PDF add the graphs of these two chosen features and explain how do you conclude if they are beneficial or not.

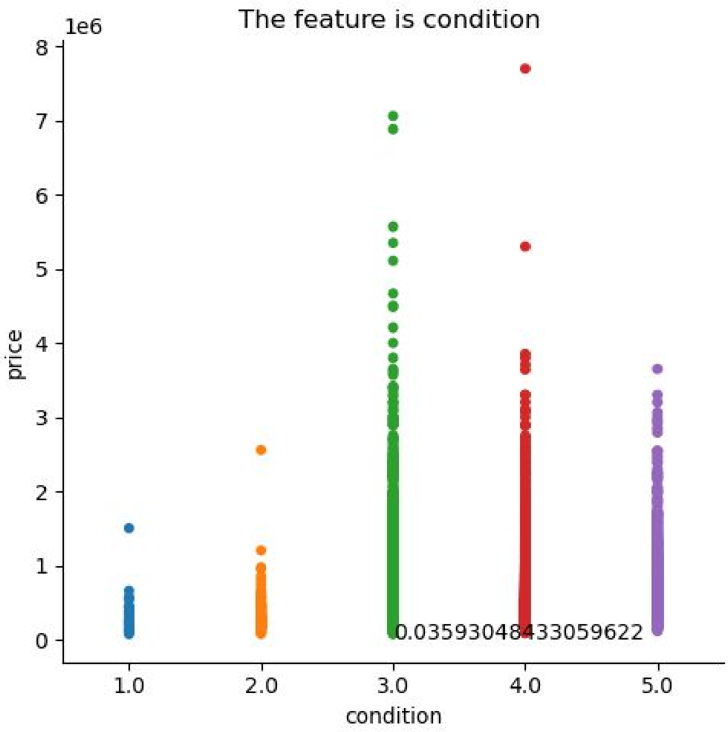
נתבונן כיצד הציון משפיע על המחיר :
נשים לב כי הקורלציה של פירסון הינה מדד לקשר לינארי בין משתנים ובמקרה זה היא בעלת ערך גבוה 0.67
ומכאן ניתן להסיק כי קיים קשר לינארי בין הציון למחיר הבית.
כמו כן ניתן לראות כי ככל שהציון גבוה יותר כך גם המחיר גבוה יותר.



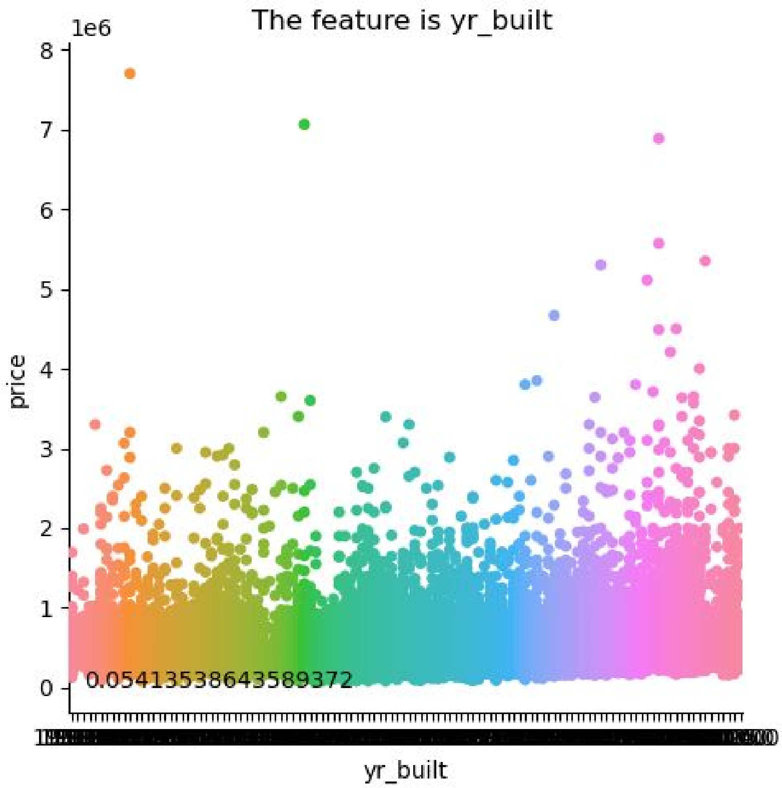


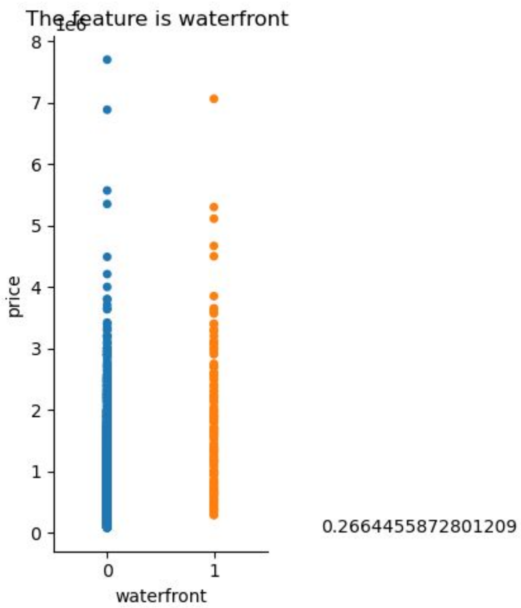
כאשר במקרה זה ניתן לראות כי מתאם
פירסון הינו גבוה - 0.7 ומצדיק את הקשר הלינארי בין גודל הדירה לבין המחיר. ככל שהדירה תהיה גדולה יותר כך המחיר עבורה יהיה גבוה יותר, דבר שתואם את המציאות.

התנאים הם בעצם אינדקסים מ-1 עד 5 על התנאים בדירה.
אנו רואים מקדם פירסון נמוך מאוד ולכן הקשר הלינארי בין התנאים בדירה למחיר הינו מזערי.
ניתן לראות שלא בהכרח ככל שהאינדקס גדול יותר כך המחיר גדול יותר.



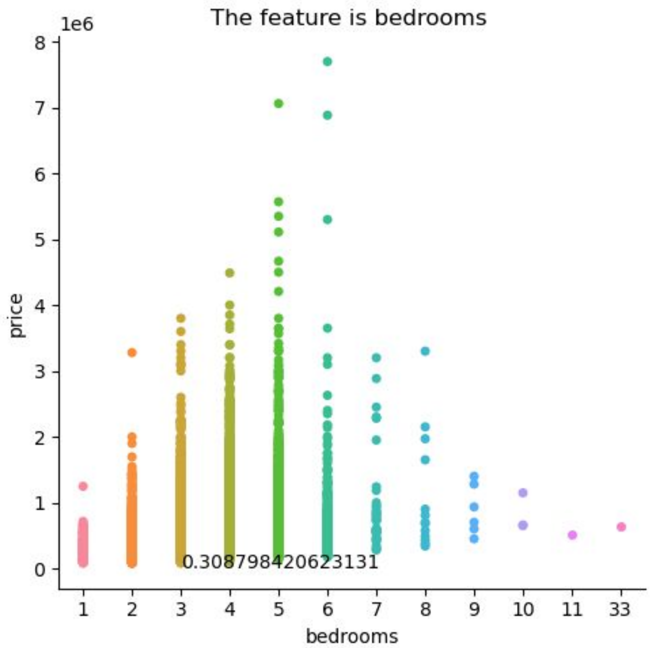
נתבונן כיצד משפיעה השנה שבה נבנה הבניין על המחיר:
אנו רואים כי על פי מתאם פרסון ההתאמה היא מזערית כלומר המדד קשר הלינארי בין שנת הבנייה למחיר הוא מועט.
כך שבתים שנבנו לפני מאה שנה לא בהכרח יותר זולים מבתים חדשים.

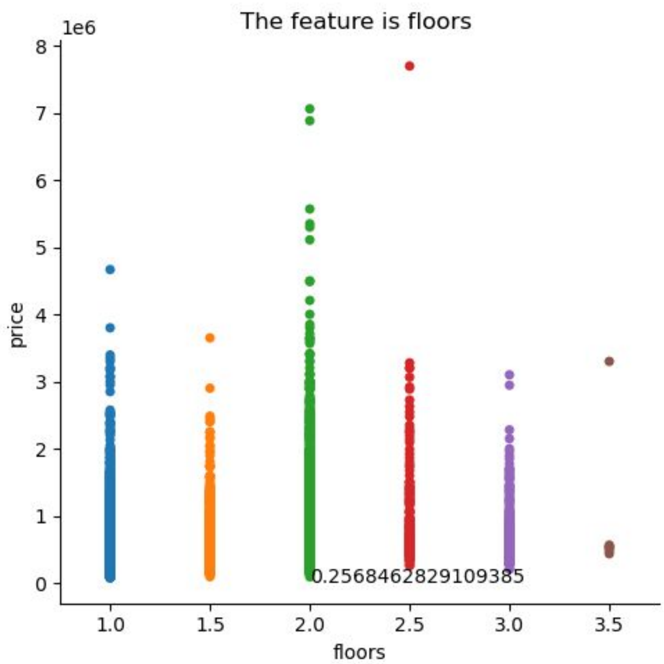


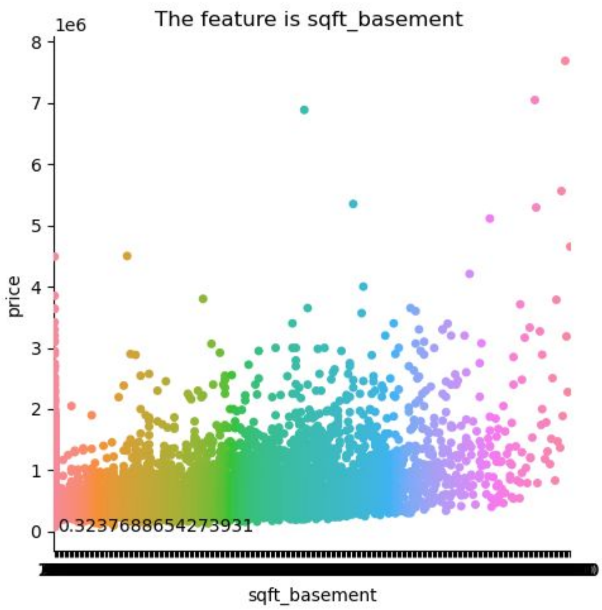


במקרה זה דווקא ניתן לראות כי מתאם פרסון נמוך יחסית כך שהמדד לקשר הלינארי בין מחיר לבין האם הדירה משקיפה לקו החוף אינו גבוה. נתון קצת מפתיע בעיניי שכן קרבה לחוף לפחות בישראל מקפיצה את מחירי הבתים.

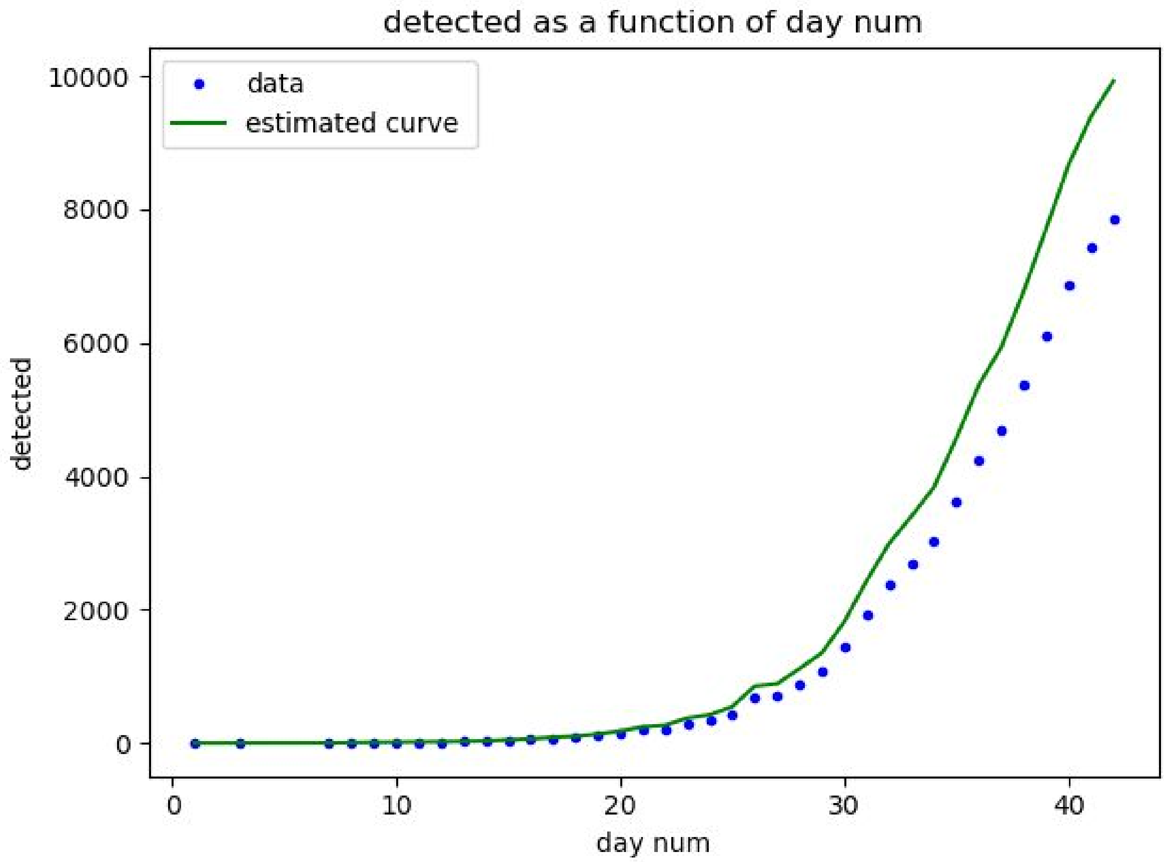
עוד דוגמאות מעניינות הן הקשר בין מספר חדרי השינה, גודל המרתף ומספר הקומות למחיר : שוב מדובר בקשר לינארי פחות חזק מדוגמאות שראינו כמו גודל הדירה והציון שלה.

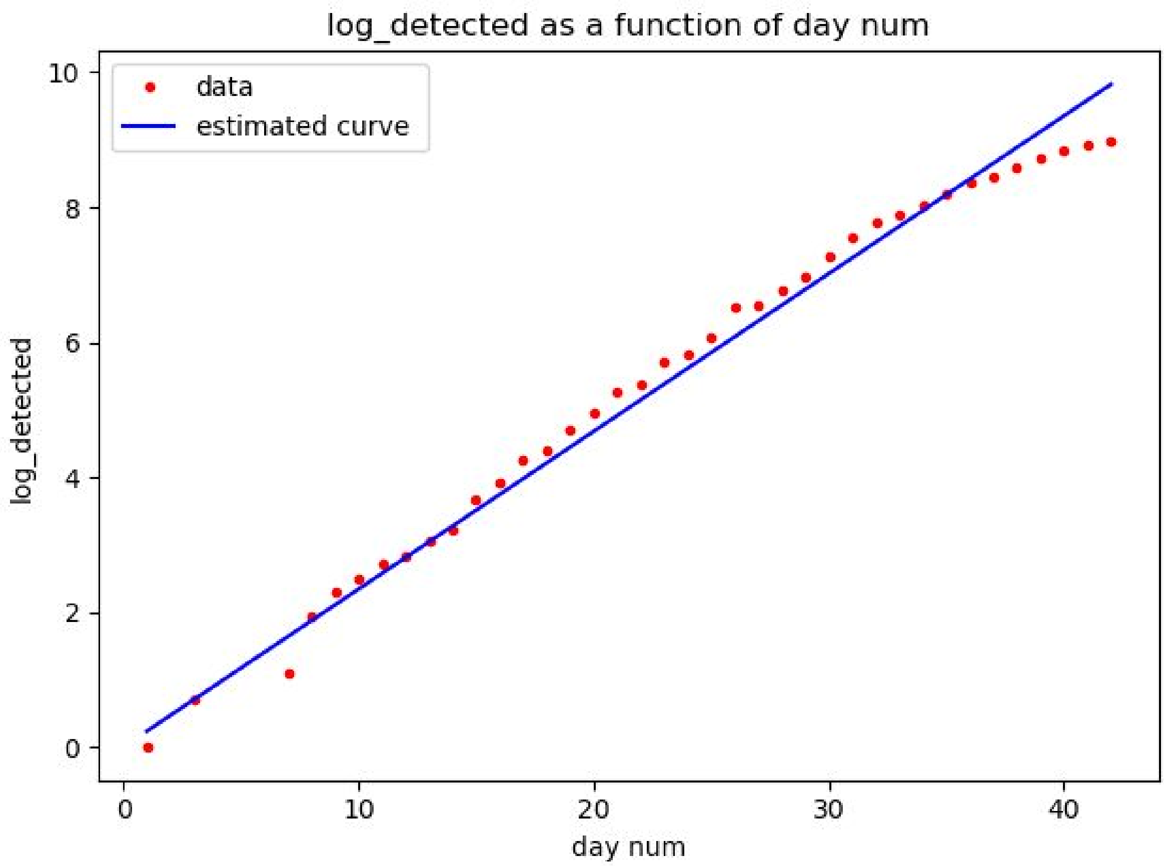






21. Plot two graphs: the first is the "log_detected" as a function of "day_num", and the second is "detected" as a function of "day_num". On each graph, add the data as single points, and add the fitted curve that you've estimated (think how to convert the linear result into an exponential result). Add those graphs to your answers PDF.


detected as a function of day num
data
estimated curve
10000
8000
6000
4000
2000
0
detected
0
10
20
30
40
day num


log_detected as a function of day num
data
estimated curve
10
8
6
4
2
0
log_detected
0
10
20
30
40
day num

22.

אז כמו שניתן לראות בשאלה 21 :
כאשר הצגנו את מספרים המקרים של קורונה שזוהו כפונקציה של מספר הימים שעברו מאז שזיהו את ההדבקה הראשונה בישראל קבלנו כי ה- fitted model הינו גרף אקספוננציאלי.
בהמשך נסביר כיצד עשינו זאת.
לעומת זאת כאשר הצגנו את log מספר המקרים של קורונה שאובחנו כפוקנציה של מספר הימים שעברו קבלנו כי ה- fitted model הינו גרף לינארי.

Exponential Growth is a mathematical function that can be used in several situations. The formula tells us the number of cases at a certain moment in time, in the case of Coronavirus, this is the number of infected people.

The formula of Exponential Growth

Exponential Growth is characterized by the following formula:

$$x(t) = x_0 * b^t$$

The Exponential Growth function

In which:

- *x(t)* is the number of cases at any given *time t*
- *x0* is the number of cases at the beginning, also called *initial value*
- *b* is the number of people infected by each sick person, the *growth factor*

Linear Regression allows us to estimate the best values for a and b in the following formula, given empirical observations for y and x. In this formula, y is the number of cases and x is the time. But we need to do some rewriting on the Exponential Growth function, because Linear Regression can only estimate formulas that look as below:

$$y = a + b * x$$

The type of formula that we need for Linear Regression

$$y = a + b * x$$

$$\log x(t) = \log(x_0) + \log(b) * t$$

- we use the **log of the number of infections** instead of the **number of infections**
- we use the **log of growth factor** instead of **growth factor**

ואז נעשה fit על המודל הלינארי נקבל את a,b ונחלץ את ה-growth factor, initial value.

22) $L_{exp}(f_w, (x,y)) = (\langle w, x\rangle - \log y)^2$

כעת נרצה להשתמש ב- least squares loss שראינו בהרצאה עבור המקרה של רגרסיה אקספוננציאלית: $(y - \langle w, x\rangle)^2$

נתבונן במודל regression model;

$$y = Ae^{BX} \longrightarrow \log y = \log A + BX$$

נרצה למצוא את A ו- B, ~~[illegible]~~

A ו- B הם הקבועים של המודל.

$$S_r = \sum_{i=1}^{n} (y_i - Ae^{BX_i})^2$$

נרצה למצוא את המינ׳

נקח את הנגזרות ביחס ל- A וביחס ל- B:

$$\frac{\partial S_r}{\partial A} = \sum_{i=1}^{n} 2(y_i - Ae^{BX_i})(-e^{BX_i}) = 0$$

$$\frac{\partial S_r}{\partial B} = \sum_{i=1}^{n} 2(y_i - Ae^{BX_i})(-\cancel{A}\, x_i e^{BX_i}) = 0$$

$$\Longrightarrow -\sum_{i=1}^{n} y_i e^{BX_i} + A\sum_{i=1}^{n} e^{2BX_i} = 0$$

$$\sum_{i=1}^{n} y_i x_i e^{BX_i} - A\sum_{i=1}^{n} X_i e^{2BX_i} = 0$$

וקיבלנו משוואות שהן nonlinear במונחים של A ו- B

אנו נרצה למצוא דרך למצוא את A ו- B ללא פתרון משוואות אלו.

לכן נפעיל log על שני האגפים:

$$\ln y = \ln A + \ln(e^{BX}) = \ln A + \cancel{\ln} BX$$

נסמן $a_0 = \ln A$, $Z = \ln y$, $a_1 = B$ כך ש $Z = a_0 + a_1 X$

שהוא מודל ליניארי, במקרה זה אנו יודעים למצוא את a_0, a_1.

נוכל לחזור אחורה ל- A ו- B בשימוש הבא:

$$A = e^{a_0}, \quad B = a_1$$

נשתמש בנוסחאות של linear model regression;

$$a_1 = \frac{n\sum_{i=1}^{n} x_i z_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} z_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \quad a_0 = \bar{Z} - a_1 \bar{X}$$

ERM problem: $\underset{w\in\mathbb{R}^d}{\arg\min} \|X^T w - y\|^2 = \arg\min \underbrace{\frac{1}{n}\sum_{i=1}^{n}(x_i^T w - y_i)^2}_{= S_r}$

ברגרסיה אקספוננציאלית $y = \alpha e^{\beta t}$ כאשר y הוא מספר המקרים המצטברים של המגיפה, t הם הימים שעברו מאז תחילת הדגימה,

נוכל להשתמש ב- least square ואז להתייחס למודל של גידול אקספוננציאלי עם לוג. הפעלת log על ה- response vector ולעשות fit למודל ליניארי

$\log y = \log\alpha + \beta t$ ואז log הוא פונקציה ליניארית של t עם intercept $\beta_0 = \log\alpha$

לכן כדי לעשות fit למודל עם גידול אקספוננציאלי לפי y נפעל כך:

1) נגדיר $Z = \log y$

2) נעשה fit למודל הליניארי עבור z ע״י שימוש ב- least squares

3) נקבל מהמודל האקספוננציאלי β ואת e^{α}: $y = \alpha e^{\beta t}$