IML
תרגיל 2

1) יהי $x \in \ker(X^T)$ כלומר:

$$XX^Tx = X(X^Tx) = X \cdot 0 = 0$$

ולכן $x \in \ker(XX^T)$

$$\ker(X^T) \subseteq \ker(XX^T)$$

• יהי $x \in \ker(XX^T)$ כלומר

$$XX^Tx = 0 \iff x^t XX^t x = 0$$
$$\iff (X^t x)^t X^t x = 0$$
$$\iff \|X^t x\|_2^2 = 0$$
$$\iff X^t x = 0$$
$$\Rightarrow x \in \ker(X^t)$$

$$\ker(XX^t) \subseteq \ker(X^t)$$

ולכן מהכלה דו כיוונית $\ker(XX^t) = \ker(X^t)$.

2) יודעים שמתקיים $(A^T)^T = A$ ולכן
$$Im(A) = \ker(A^T)^\perp \iff Im(A^T) = \ker(A)^\perp$$

• יהי $x \in Im(A^T)$ ולכן ק"ם $v \in \mathbb{V}$ כך ש- $x = A^T v$

ויהי $y \in \ker(A)$ ולכן $A \cdot y = 0 \Leftarrow$

$$\langle x, y \rangle = x^T y = (A^T v)^T y = v^T A y$$

$$= v^T \cdot 0 = 0 \Rightarrow \langle x, y \rangle = 0$$
$$\Rightarrow x \in \ker(A^T)$$
$$Im(A^T) \subseteq \ker(A)^\perp.$$

$x \in Im(A^T)^{\perp}$. אזי $x^T(A^T)v$ לכל v

$\Leftarrow$ $(Ax)^T v = 0$ ונבחר $v = Ax$ כאשר $x \in \ker A$

ולכן $Im(A^T) \subseteq \ker(A)$ ומכאן $\ker(A)^{\perp} \subseteq Im(A)^{\perp\perp}$

ולכן $Im(A^T) \subseteq \ker(A)^{\perp} \subseteq Im(A^T)^{\perp\perp} = Im(A^T)$

מתקבלה השוויונית $\ker(A)^{\perp} = Im(A^T)$

3) X^T לא הפיכה ולכן יש אין סוף פתרונות או שאין פתרונות בכלל. (כאשר $y = X^t w$)

∞ פתרונות $\Leftrightarrow$ $y \in Im(X^t)$ $\leftarrow$ (ולכן יש לפחות פתרון אחד ולכן ∞ פתרונות.)

(סעיף 2) $\Leftrightarrow$ $y \in \ker(X)^{\perp}$

$\Leftrightarrow$ $y \perp \ker(A)$

4) נרצה להוכיח שיש פתרון יחיד למערכת $XX^t w = Xy$ אחרת XX^t הפיכה ולכן יש אינסוף פתרונות אחרת.

$$XX^t w = Xy \underset{X\text{-הכפלה ב}}{\overset{\text{שאלה 3.}}{\Leftrightarrow}} X^t w = y$$

נחלק לשני מקרים

(*) (XX^t) הפיכה $\Rightarrow$ יש פתרון יחיד והפתרון הוא. $w = (XX^t)^{-1}Xy$

(*) (XX^t) לא הפיכה

$$X^t y \in \ker(X^t X) \overset{\text{שאלה 2}}{=} Im(X)^{\perp} \Leftrightarrow$$

שאלה 3. $\Leftarrow$ ∞ פתרונות

5)

א) $P^T = \sum_{i=1}^{k} (v_i v_i^T)^T \underset{\uparrow}{=} \sum_{i=1}^{k} v_i v_i^T = P$

$(v_i v_i)^T = (v_i^T)^T v_i^T = v_i v_i^T$

ב) נשתמש בפירוק EVD כאשר $D = \begin{pmatrix} I_{k\times k} & 0 \\ 0 \cdots & 0 \end{pmatrix}$

ולכן פירוק של P הוא $P = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_d \\ | & & | \end{pmatrix} \underbrace{\begin{pmatrix} I_{k\times k} & 0 \\ 0 & 0 \end{pmatrix}}_{D} \begin{pmatrix} -v_1- \\ \vdots \\ -v_d- \end{pmatrix}$

ונקבל רק 0 או 1.

יהי j כך ש- $v_j \in \{v_1, \ldots, v_k\}$ ולכן

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

ידוע ש- $\langle v_i, v_j \rangle = \delta_{ij}$ מ אורתונורמליות.

$\Rightarrow P v_j = \sum_{i=1}^{k} v_i \langle v_i, v_j \rangle = \sum_{i=1}^{k} v_i \delta_{ij} = v_j.$

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

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

ד) $P^2 = \left(\sum_{i=1}^{k} v_i v_i^T\right)^2 = \sum_{i=1}^{k} (v_i v_i^T)^2$

$= \sum_{i=1}^{k} (v_i \underbrace{v_i^T)(v_i}_{I} v_i^T) = \sum_{i=1}^{k} (v_i v_i^T) = P.$

ה) $(I-P)P = P - P^2 \underset{\substack{\uparrow \\ \text{לפי} \\ \text{הקודם}}}{=} P - P = 0.$

Scanné avec CamScanner

6) נשתמש ב $U\Sigma U^T$ להיות את ה SVD של X.

(1)

$$XX^t = U\Sigma\Sigma^T U^T$$
$$\Leftrightarrow (XX^t)^{-1} = (U^T)^{-1}(\Sigma^T)^{-1}(\Sigma)^{-1}U^{-1}$$
$$= UD^{-1}U^T$$

(2)

ידוע לנו מהכיתה/תרגול

$$X^{\perp t} = X^{t\perp} = U\Sigma^{t+}V^T$$
$$(XX^t)^{-1}X = UD^{-1}U^t X = U\underbrace{D^{-1}\Sigma}_{\text{נשים לב:}} V^T$$
$$D^{-1}\Sigma = (\Sigma\Sigma^t)^{-1}\Sigma = (\Sigma^t)^{-1}\Sigma^{-1}\Sigma = (\Sigma^t)^{-1}$$
$$\Rightarrow (XX^t)^{-1}X = X^{t\perp} = U$$

7) XX^t הפיכה $\Leftrightarrow$ $\text{Rank}(XX^t) = d$

$\ker(X^t) = \{0\}$ $\Leftrightarrow$ $\text{Rank}(X^t) = d$

$\Leftrightarrow$ $\dim(\text{Im}(X^t)) = d$

$\Leftrightarrow$ $\text{Span}\{x_1, \dots, x_m\} = \mathbb{R}^d$

8) נרצה להוכיח ש: $\|\bar{\omega}\|_2 \geq \|\hat{\omega}\|$

ונניח שהדרגה של X היא R<d ולכן R ערכים הראשונים של $\bar{\omega}$ ו $\hat{\omega}$ הם שווים ונשים לב i > R

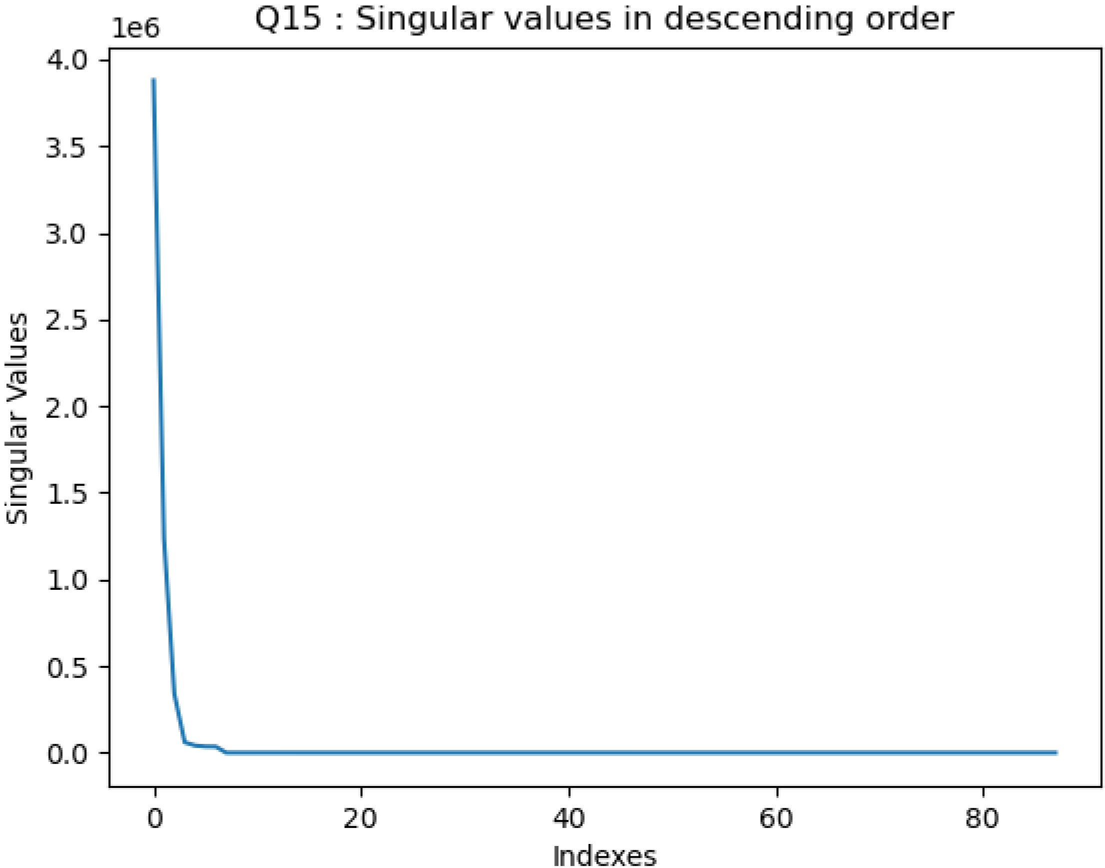
$\hat{\omega}_i = 0$ כלומר נקבל:

$$\|\bar{\omega}\|_2 = \sum_{i=1}^{d} \bar{\omega}_i^2$$
$$= \sum_{i=1}^{R} \bar{\omega}_i^2 + \sum_{i=R+1}^{d}$$
$$= \sum_{i=1}^{R} \hat{\omega}_i^2 + \sum_{i=R+1}^{d} \bar{\omega}_i^2 \geq \sum_{i=1}^{R} \hat{\omega}_i^2 = \|\hat{\omega}\|_2$$

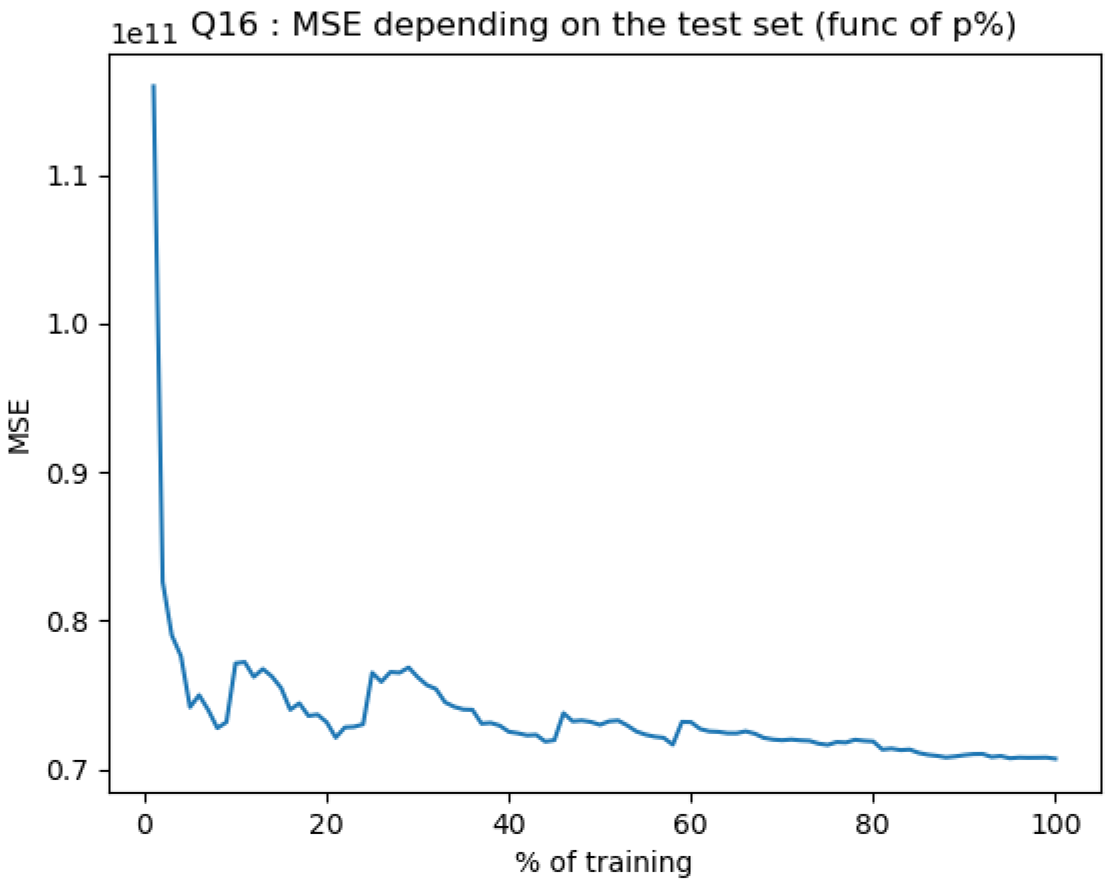
13) החלטתי לא לקחת בחשבון גם את ה-id
וגם את התאריך של הבית (date)
id לא רלוונטי למחיר של הבית. ובנוסף, כשפתחתי
את הקובץ אפשר לשים לב שכל שנה שמופיעה היא 2014
או 2015 ולכן אין לה הכרע גדול.

בנוסף לכך, ה zipcode הוא "categorical feature"
ולכן השתמשתי ב- get-dummies.

15) ל design matrix יש singular values ששווים
לאפס. <= מטריצה לא הפיכה.


Q15 : Singular values in descending order
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Singular Values
0
20
40
60
80
Indexes

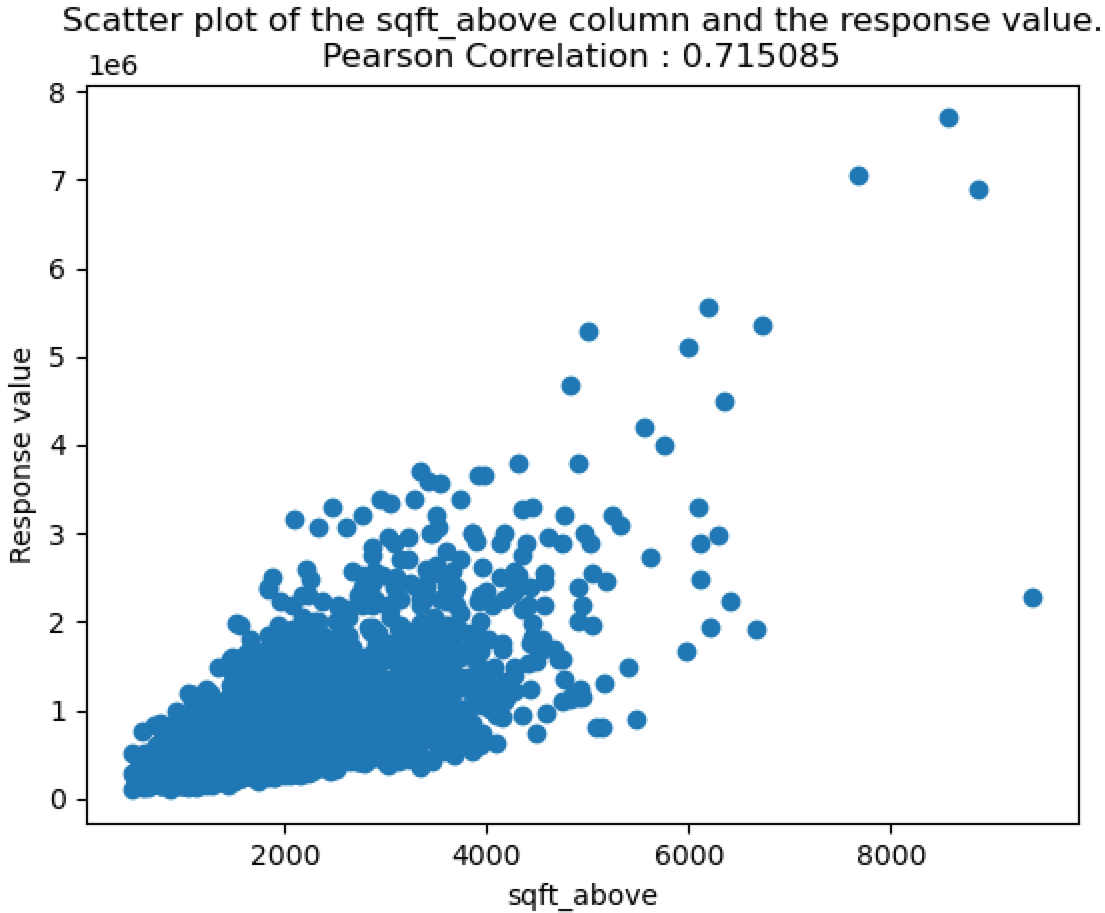
16) לכל אחוז, שהוא אחוז נמוך אפשר לשים לב
שכל סוג שלוקחים הרבה מה training set, ה MSE
מקטין.
בסוף שגדל לערך ש 0.7 ולא מקטין מהר והרבה
כמו לפני וכמו שרואים בגרף.

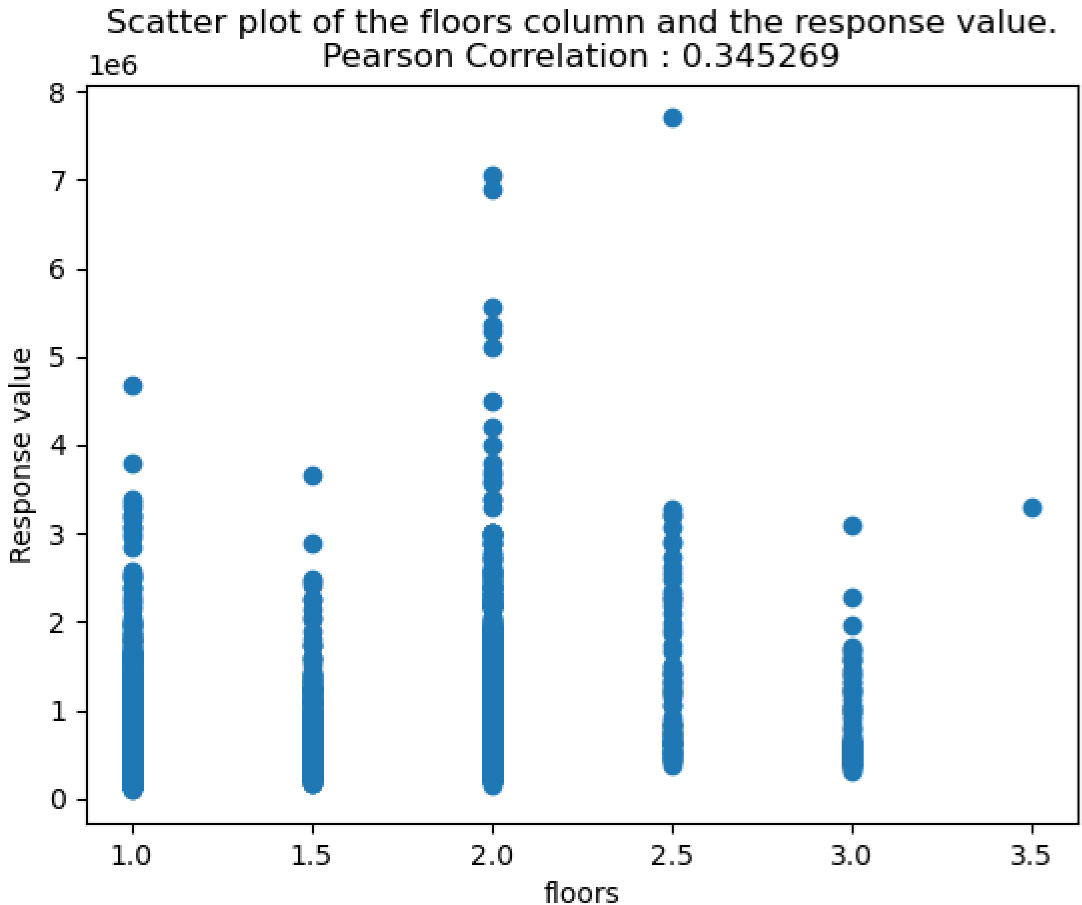

Q16 : MSE depending on the test set (func of p%)
1e11
1.1
1.0
0.9
0.8
0.7
MSE
0
20
40
60
80
100
% of training

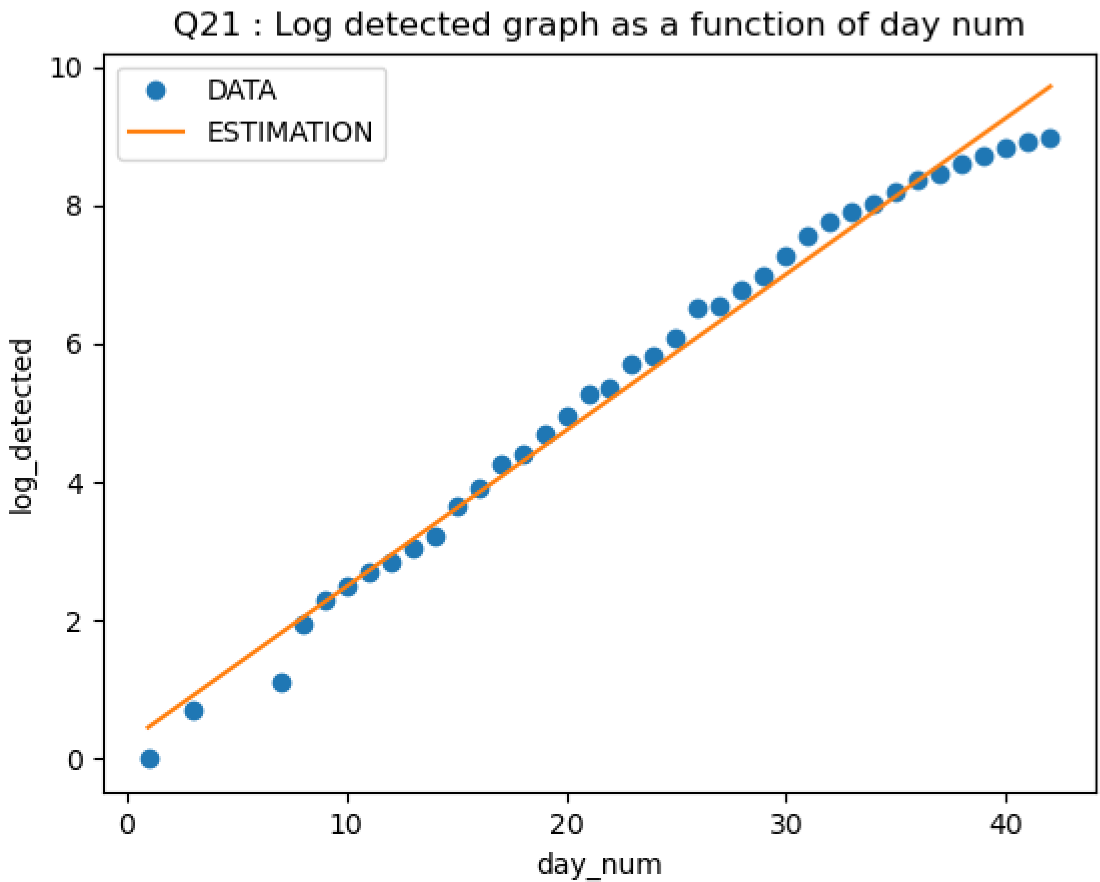
17) הגרפים מצורפים ובחרתי כי:

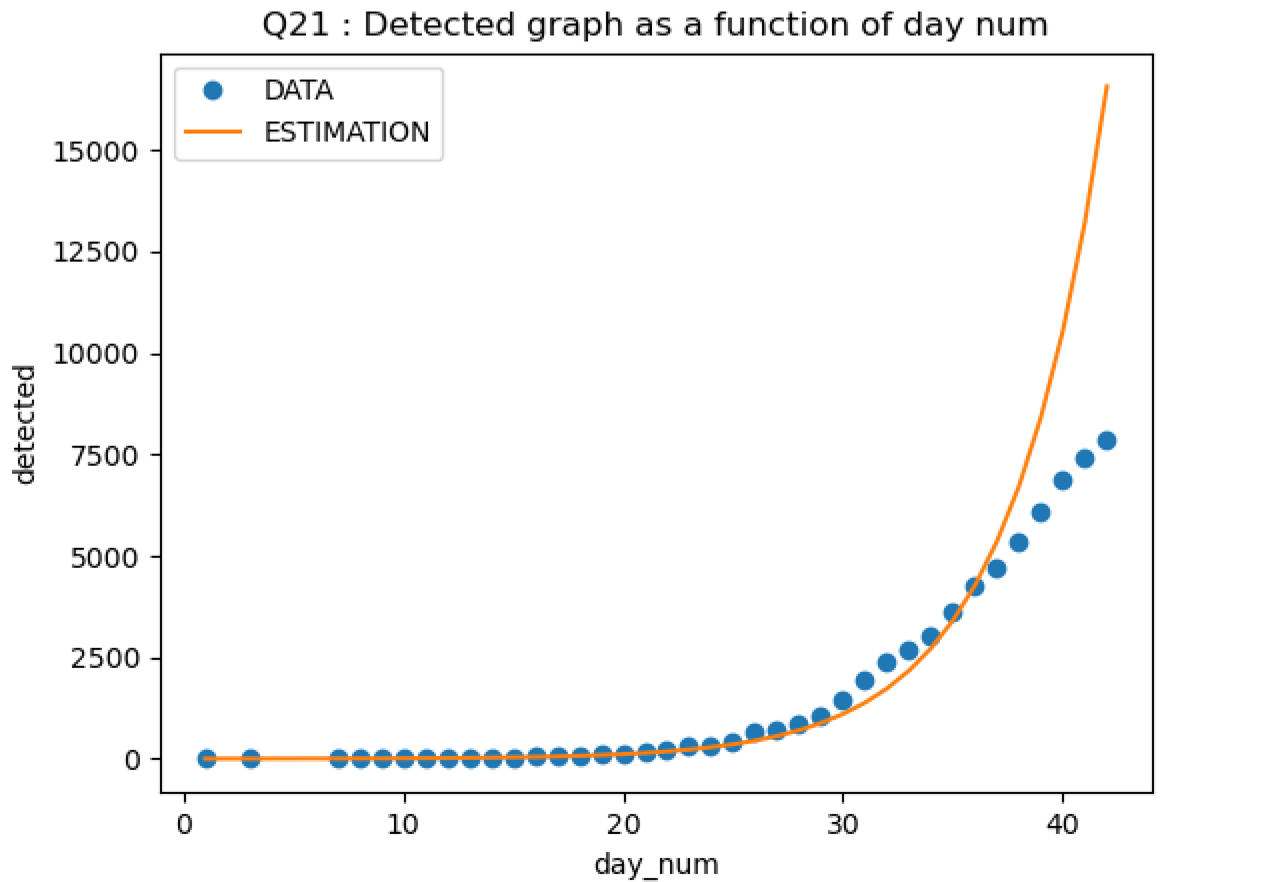
1) sqft_above: נלמד שהשטח מגדל אזי גם המחיר מגדל. 17-11
קורלציה גדולה ← משפיע הרבה.

2) floors: לא משפיע על המחיר לפי הגרף. 17-6


Scatter plot of the sqft_above column and the response value.
Pearson Correlation : 0.715085
1e6
8
7
6
5
4
3
2
1
0
Response value
2000
4000
6000
8000
sqft_above


Scatter plot of the floors column and the response value.
Pearson Correlation : 0.345269
1e6
Response value
floors
0
1
2
3
4
5
6
7
8
1.0
1.5
2.0
2.5
3.0
3.5


Q21 : Log detected graph as a function of day num
DATA
ESTIMATION
log_detected
day_num
0
2
4
6
8
10
0
10
20
30
40


Q21 : Detected graph as a function of day num
DATA
ESTIMATION
detected
15000
12500
10000
7500
5000
2500
0
0
10
20
30
40
day_num

(22

1. נצור design matrix חדשה שמכילה את log של ה features המקוריים

$y \rightarrow \log(y)$

2. נפעיל רגרסיה לינארית על ה designmatrix של 1

$L = (\langle w,x \rangle - \log(y))^2$.

3. נחשב את האקספוננט של ה- prediction שלנו כדי לקבל רגרסיה אקספוננציאלית.

1,2,3 השלבים כדי למצוא את ה- ERM.

ובנוסף ה-"כנס" היא פונק' פולינומיאלית (עם הרצאה)