321123788

תרגיל 2 - IML

שמעון לובמן

1) נוכיח הכלה דו-כיוונית:

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

יהי: $\vec{v} \in \ker(X^T)$ s.t.

$$X^T\vec{v} = \vec{0}$$
$$\Downarrow$$
$$XX^T\vec{v} = X\vec{0}$$
$$\Downarrow$$
$$XX^T\vec{v} = \vec{0}$$
$$\Downarrow$$
$$\vec{v} \in \ker(XX^T)$$

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

יהי: $\vec{v} \in \ker(XX^T)$ s.t.

$$XX^T\vec{v} = \vec{0}$$
$$\Downarrow$$
$$\vec{v}^T XX^T\vec{v} = \vec{v}^T\vec{0}$$
$$\Downarrow$$
$$(X^T\vec{v})^T(X^T\vec{v}) = \vec{0}$$
$$\Downarrow$$
$$\|X^T\vec{v}\|^2 = \vec{0}$$
$$\Downarrow (*)$$
$$X^T\vec{v} = \vec{0}$$
$$\Downarrow$$
$$\vec{v} \in \ker X^T$$

(*) המעבר מתאפשר תודות להגדרת הנורמה: $\|X^T\vec{v}\|^2 = \sum_{i=1}^{n} \langle x_i, \vec{v} \rangle^2 \geq 0$

ב) נוכיח הכלה דו כיוונית:

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

יהי $\vec{v}\in \mathrm{Im}(A^T)$ ונסמן את A כך:

$$A=\begin{bmatrix} - & \vec{u}_1 & - \\ & \vdots & \\ - & \vec{u}_n & - \end{bmatrix} \Rightarrow A^T=\begin{bmatrix} | & & | \\ \vec{u}_1^T & \cdots & \vec{u}_n^T \\ | & & | \end{bmatrix}$$

לכן

$$\vec{v}=\sum_{i=1}^{n}\alpha_i \vec{u}_i^T$$

כעת יהי $\vec{w}\in\ker(A)$, כלומר:

$$A\vec{w}=\begin{bmatrix} - & \vec{u}_1 & - \\ & \vdots & \\ - & \vec{u}_n & - \end{bmatrix}\cdot\vec{w}=\begin{bmatrix} \vec{u}_1\vec{w} \\ \vdots \\ \vec{u}_n\vec{w} \end{bmatrix}=\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

כלומר $\forall i\in[n]\ \vec{u}_i\cdot\vec{w}=0$

כעת נראה ש- $\vec{v}\in\ker(A)^\perp$:

$$\langle\vec{v},\vec{w}\rangle=\overline{\langle\vec{w},\vec{v}\rangle}=\langle\vec{w},\sum_{i=1}^{n}\alpha_i\vec{u}_i^T\rangle=\sum_{i=1}^{n}\alpha_i\langle\vec{u}_i^T,w\rangle=$$

$$=\sum_{i=1}^{n}\alpha_i(\vec{u}_i\vec{w})=\sum_{i=1}^{n}\alpha_i\cdot 0=0$$

כלומר קיבלנו שהמכפלה הפנימית בין $\vec{v}$ ל-$\vec{w}$ שווה ל-0

ולכן $\vec{v}\in\ker(A)^\perp$

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

נוכיח שאם $\vec{v}\notin \mathrm{Im}(A^T)$ אז $\vec{v}\notin\ker(A)^\perp$.

נניח שאם $\vec{v}\notin \mathrm{Im}(A^T)$ ונשלול אזי צריך את הדימוי ש

$\vec{w}\in \mathrm{Im}(A^T)^\perp$ כך ש- $\langle\vec{v},\vec{w}\rangle\neq 0$. כי מתקיים מכיוון שאם

בשלילה: $\forall\vec{w}\in \mathrm{Im}(A^T)^\perp\ \langle\vec{v},\vec{w}\rangle=0$

אז $\vec{v}$ בהכרח אורתוגונלי ל-$\mathrm{Im}(A^T)^\perp$, כלומר:

$$\vec{v}\in \mathrm{Im}(A^T)^{\perp\perp}\underset{\text{משפט 3 סוגר}}{=}\mathrm{Im}(A^T)$$

בסתירה להנחה השלילה לפיה $\vec{v}\notin \mathrm{Im}(A^T)$

Сканировано с CamScanner

כעת נקבל ש- $\vec{w}\in Im(A^T)^\perp$ נותן ש:

$$\langle \vec{w}, A^T A\vec{w}\rangle = 0$$

מכיוון ש-$A^T(A\vec{w})\in Im(A^T)$. לכן!

$$0=\langle \vec{w}, A^T A\vec{w}\rangle = \langle A\vec{w}, A\vec{w}\rangle = \|A\vec{w}\|^2$$

$$\Downarrow$$

$$\vec{w}\in \ker(A)$$

כאמור מצאנו $\vec{w}\in\ker(A)$ כך ש- $\langle \vec{v},\vec{w}\rangle\neq 0$, לכן

$$\vec{v}\notin \ker(A)^\perp$$

לסיכום:

$$\vec{v}\notin Im(A^T) \Rightarrow \vec{v}\notin \ker(A)^\perp$$

$$\Downarrow$$

$$\vec{v}\in \ker(A)^\perp \Rightarrow \vec{v}\in Im(A^T)$$

$$\Downarrow$$

$$\ker(A)^\perp \subseteq Im(A^T)$$

3) אם הפיכות X^T אזי ניתן להסיק שלמערכת המשוואות

$X^T\vec{w}=\vec{y}$ או שקיימים ∞ פתרונות או שאין לה פתרון.

כמו כן ~~ב~~ נשים לב ש- $\vec{y}\in Im(X^T)$, ולכן משאלה 2

נובע ש- $\vec{y}\in ker(X)^{\perp}$.

לסיכום נציין שבמידה ו- $\vec{y}\in Im(X^T)$ אז חייב להיות צירוף ~~[illegible]~~

לינארי של וקטורי העמודה של X^T שנותנים את $\vec{y}$, כלומר

קיים לפחות פתרון אחד למערכת, ולכן מההבחנה הקודמת

נסיק שקיימים ∞ פתרונות. אמ"מ $\vec{y}\in Im(X^T)$ אמ"מ $\vec{y}\in ker(X)^{\perp}$.

4) נפצל למקרים

XX^T הפיכה

נשים לב שאם XX^T הפיכה אז $\det(XX^T)\neq 0$

ולכן $\det(X)\neq 0 \wedge \det(X^T)\neq 0$, כלומר X ו-X^T הפיכות.

לכן נוכל להציג את הפתרון היחיד הבא:

$$XX^Tw=Xy \Rightarrow w=(XX^T)^{-1}Xy$$

XX^T אינה הפיכה

נבחין תחילה ש- $(XX^T)^T=XX^T$, לכן:

$$XX^Tw=Xy \iff (XX^T)^Tw=Xy$$

ומשאלה 3 נסיק שלמערכת הנ"ל קיימים ∞ פתרונות אמ"מ $Xy\perp\ker(XX^T)$.

משאלה 1 נסיק שזה שקול ל- $Xy\perp\ker(X^T)$

יהי $\vec{v}\in\ker(X^T)$, מתקיים אם כן:

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

לכן Xy אכן אורתוגונלי ל- $\ker(XX^T)$ ולכן למערכת הנ"ל ∞ פתרונות.

5) (a) נוכיח תחילה ש:

$\vec{v}_i\vec{v}_i^T$ סימטרית $\forall i\in[n]$

הוכחה:

נסמן $v_i=\begin{bmatrix}a_1\\ \vdots\\ a_d\end{bmatrix}$ מתקיים:

$\forall m\in[k]$

$$(\vec{v}_m\vec{v}_m^T)_{ij}=a_i\cdot a_j=a_j\cdot a_i=(\vec{v}_m\vec{v}_m^T)_{ji}$$

לכן $\vec{v}_m\vec{v}_m^T$ סימטרית, נסמן אותה כ- A_m.

כעת מתקיים:

$$P=\sum_{i=1}^{k}A_i$$

לכן:

$$P_{ij} = \sum_{m=1}^{k} (A_m)_{ij} = \sum_{m=1}^{k} (A_m)_{ji} = P_{ji}$$

ולכן P סימטרית.

(ב) תחילה נשים לב שעבור v_i $i \in [k]$ מתקיים:

$$P\vec{v}_i = \left(\sum_{j=1}^{k} \vec{v}_j \vec{v}_j^T\right)\vec{v}_i = \sum_{j=1}^{k} \vec{v}_j (\vec{v}_j^T \vec{v}_i) = \sum_{j=1}^{k} \vec{v}_j \delta_{ij} = \vec{v}_i$$

כעת נרחיב את הבסיס $(v_1, \dots, v_k)$ לבסיס אורתונורמלי של $\mathbb{R}^d$ בעזרת בסיס שרירותי $(w_{k+1}, \dots, w_d)$ של $\mathbb{R}^d \setminus V$ ותהליך גרם שמידט. נסמן כתוצר בסיס $(v_1, \dots, v_d)$ אורתונורמלי לכל $\mathbb{R}^d$. נשים לב שעדיין מתקיים:

$$\forall i \in [k] \quad P\vec{v}_i = \vec{v}_i$$

ובנוסף ש:

$$\forall k+1 \le i \le d \quad P\vec{v}_i = \sum_{j=1}^{k} \vec{v}_j (\vec{v}_j^T \vec{v}_i) = 0$$

(c) יהי $\vec{v}\in V$, מתקיים: $\vec{v}=\sum_{i=1}^{k}\alpha_i\vec{v}_i$

לכן:

$$P\vec{v}=\left(\sum_{i=1}^{k}\vec{v}_i\vec{v}_i^{T}\right)\left(\sum_{j=1}^{k}\alpha_j\vec{v}_j\right)=$$

$$=\sum_{j=1}^{k}\left(\alpha_j\sum_{i=1}^{k}\vec{v}_i\vec{v}_i^{T}\vec{v}_j\right)=\sum_{j=1}^{k}\left(\sum_{i=1}^{k}\alpha_j\vec{v}_i\delta_{ij}\right)=\sum_{j=1}^{k}(\alpha_j\vec{v}_j)=$$

$$=\vec{v}$$

(d)

$$P^2=\left(\sum_{i=1}^{k}\vec{v}_i\vec{v}_i^{T}\right)\left(\sum_{j=1}^{k}\vec{v}_j\vec{v}_j^{T}\right)=\sum_{i=1}^{k}\sum_{j=1}^{k}\vec{v}_i\vec{v}_i^{T}\vec{v}_j\vec{v}_j^{T}=$$

$$=\sum_{i=1}^{k}\sum_{j=1}^{k}\vec{v}_i\delta_{ij}\vec{v}_j^{T}=\sum_{i=1}^{k}\vec{v}_i\vec{v}_i^{T}=P$$

(*) כעת ניתן לתת הוכחה פשוטה יותר עבור (ב):

יהי $\vec{v}\in V$ ו"ע של P, לכן:

$$P\vec{v}=\lambda\vec{v}$$

כמו כן מכיוון ש- $P^2=P$ מתקיים:

$$\lambda^2\vec{v}=P^2\vec{v}=P\vec{v}=\lambda\vec{v}$$

$$\Downarrow$$

$$(\lambda^2-\lambda)\vec{v}=0$$

$$\Downarrow$$

$$\lambda(\lambda-1)\vec{v}=0$$

$$\Downarrow$$

$$\lambda=0,1$$

(e) נראה כי P ומשלימתו הינן אורתוגונליות, לכן:

$$(I-P)P=IP-PP=P-P^2\underset{\substack{\uparrow\\ \text{לפי}\\ (d)}}{=}P-P=0_d$$

6) # פשוט נכפיל ונראה שהתוצאה אכן I_d:

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

$$= (U\Sigma \underbrace{V^TV}_{I}\Sigma^T \underbrace{U^TU}_{I} D^{-1}U^T) = U\underbrace{\Sigma\Sigma^T}_{D} D^{-1}U^T = UDD^{-1}U^T =$$

$$= UU^T = I_d$$

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

* נשים לב ש- $\Sigma\Sigma^T$ היא מטריצה אלכסונית, ואם XX^T הפיכה אז גם $\Sigma\Sigma^T$ הפיכה, ועל כן איברי האלכסון שלה הם:

$$(\Sigma\Sigma^T)_{ii} = \sigma_i^2$$

ומכאן נקבל:

$$(\Sigma\Sigma^T)^{-1}_{ii} = \frac{1}{\sigma_i^2}$$

לכן:

$$((\Sigma\Sigma^T)^{-1}\Sigma)_{ii} = \frac{1}{\sigma_i^2}\cdot\sigma_i = \frac{1}{\sigma_i}$$

ולסיכום

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

7) בשאלה 1 הוכחנו ש- $\ker(X^T)=\ker(XX^T)$, כלומר הדרגה

של X^T שווה לדרגה של XX^T (במימד סופי).

נזכור כמו כן ש- $X\in M_{d\times m}(\mathbb{R})$ ו- $XX^T\in M_{d\times d}$

$\{x_1,\dots,x_d\}$ הן השורות של X, לכן כמובן:

$$rk(X) = \dim(\mathrm{span}(\{x_1,\dots,x_m\}))$$

לכן:

$$rk(XX^T)=d \iff rk(X)=d \iff \dim(\mathrm{span}(\{x_1,\dots,x_m\}))=d$$

$$\Updownarrow$$

$$\mathrm{span}(\{x_1,\dots,x_m\})=\mathbb{R}^d$$

לסיכום:

$$rk(XX^T)=d \iff \mathrm{span}(\{x_1,\dots,x_m\})=\mathbb{R}^d$$

8) נשים לב שקיים פתרון יחיד לערכים בוקטור $\hat{w}$ מאחר והוא argmin, ושהערכים אלו מתאימים לערכים הסינגולרים במטריצה X.

מכאן נסיק שקיים $r\in[d]$ כך ש:

$$\begin{cases} \bar{w}_i = \hat{w}_i & 1\le i\le r \\ \hat{w}_i = 0 & r<i\le d \end{cases}$$

כאשר השורה השנייה נובעת מהעובדה שהערכים הסינגולריים מסודרים בסדר יורד ושעבור ערך סינגולרי 0 ערכי המקביל ב- X^{T+} הוא גם כן 0.

ולכן:

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

ולכן $\|\hat{w}\|$ מינימלי.

13) משתנים קטגוריים:

* 'zipcode': למיקוד אין יחס סדר, עצם העובדה שהמספר של מיקוד אחד גדול ממיקודו של מיקוד אחר לא מלמד אותנו שום דבר על הבתים עצמם, אבל המיקוד בעצמו כן מאוד חשוב. במקרה כזה ניתן להחליף את המיקודים בערכים סתמיים (dummy variables). המיקוד הוא כבר מספר לכן אפשר להשאיר אותו כפי שהוא, אבל יש בתכנית להחליף כל מיקוד לתוך מספר מ-0 ל-69 כדי לצמצם את סטיית התקן.
* 'id': המספר הסידורי של הבית לא מלמד אותנו שום דבר וגם לו אין יחס סדר. ערך זה הינו אמנם קטגורי אך לא נחוץ, לכן השמטתי אותו.
* 'view'/'waterfront': ערכים אלה גם חסרי יחס סדר וגם כבר מיוצגים בטבלה כערכים קטגוריים.

כמו כן הורדתי את העמודות הבאות:

* 'lat'/'long': ערכים אלה מציגים את מיקומו הגאוגרפי של הבית. הם משמשים בעצם את אותה המטרה כמו המיקוד. כמו כן בהערכת התכונות ראיתי שהשפעתם על המחיר נמוכה.
* 'date': בהערכת התכונות ראיתי שהשפעת תאריך המכירה נמוכה.
* 'yr_built': בהערכת התכונות ראיתי שהשפעת שנת בניית הבית נמוכה.

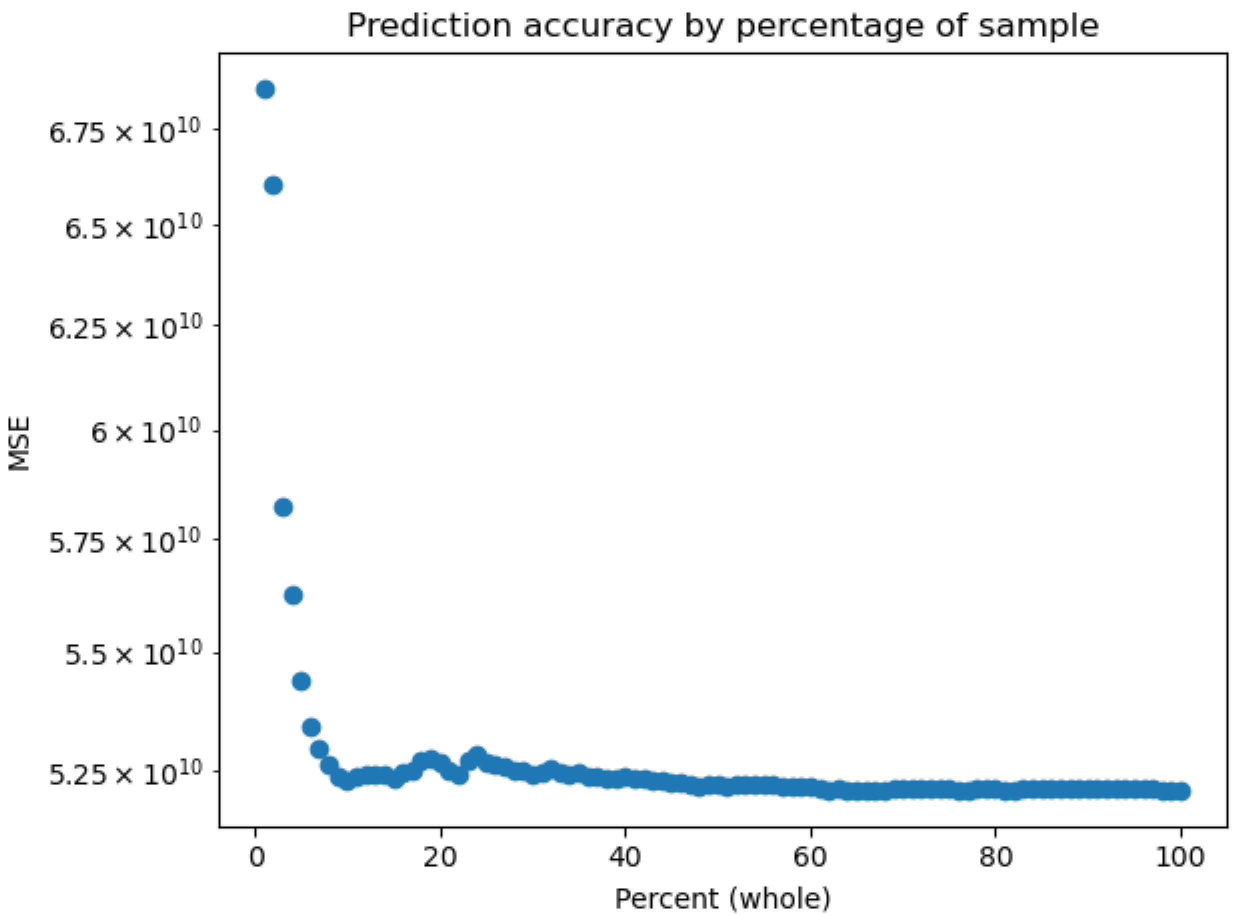
טו) בהינתן מטריצה Σ הערכים הסינגולריים שלה יכולים
ללמוד מהי מידת ההשפעה של כל הביטוי על הערך
של המטריצה.

16) בגלל שזמן ידוע שבכל ש' ם עוד רק הערכים שמפלים

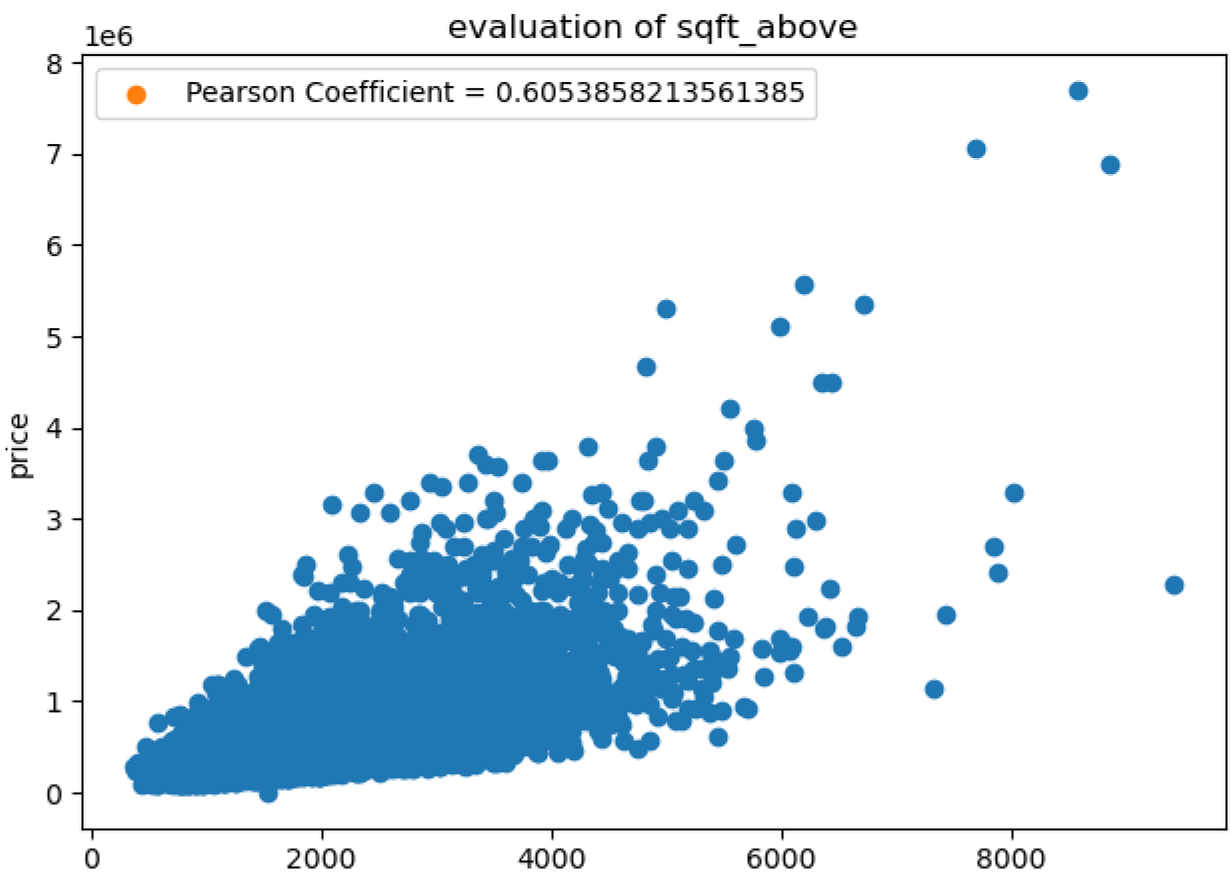
להתכנס סביב $4.4\cdot10^{11}$. ערך זה נמוך מכך שלא בוודאות

(בלי דיוק) היות ופונקציה המתאימה $(x_i, f(x_i))$, שלו

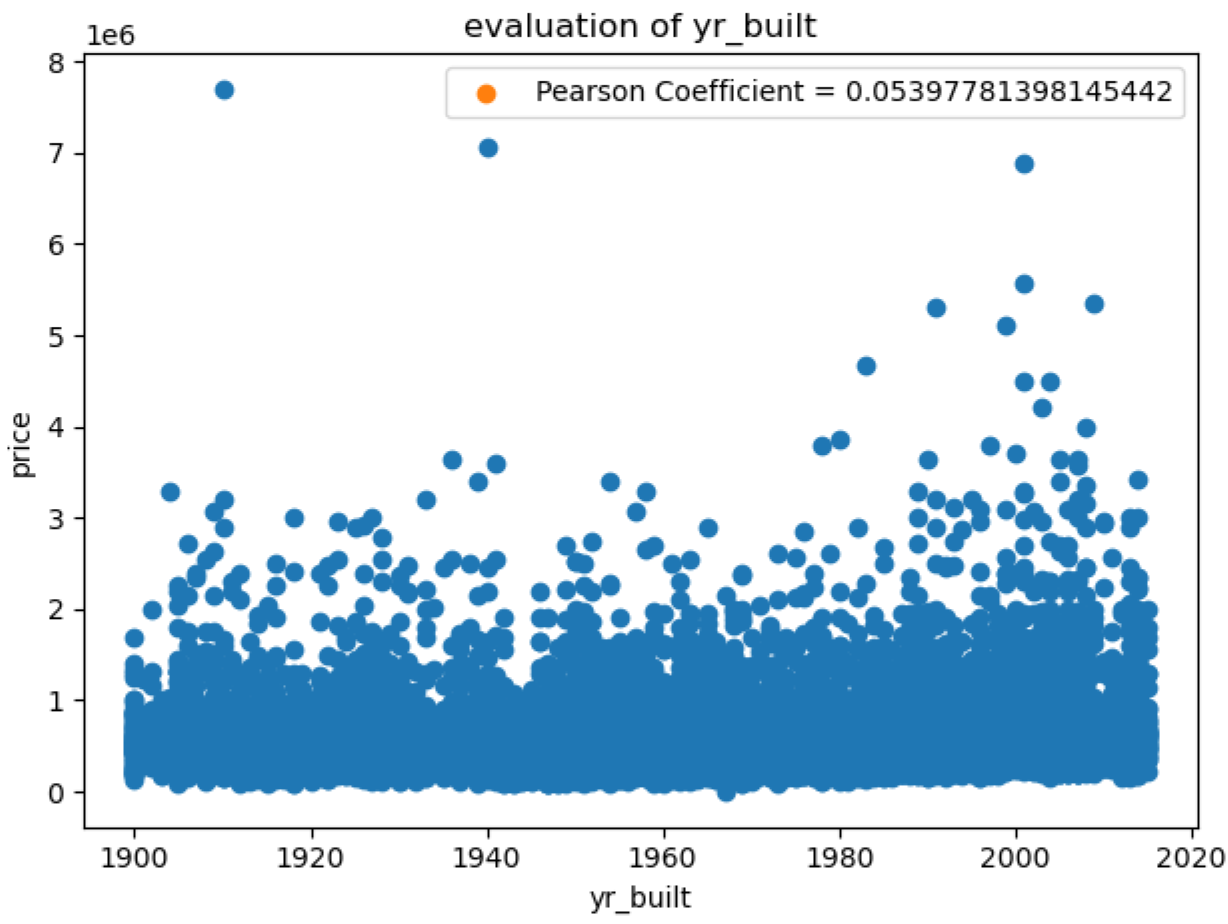
היום "רעש" (סביב $(0,\sigma)$) עד $z_1,\dots,z_m$. אם צאת הערכים בין מכנסים.


Prediction accuracy by percentage of sample
MSE
6.75×10^{10}
6.5×10^{10}
6.25×10^{10}
6×10^{10}
5.75×10^{10}
5.5×10^{10}
5.25×10^{10}
0
20
40
60
80
100
Percent (whole)

17) 'sqft_above' - בגרף הפיזור תכונה זו ניתן לראות
מגמת עליה בערכים של המחיר ככל שערך sqft_above עולה.
כמו כן חישבנו וגילינו שלשני משתנים אלו יש קורלציה
של 0.6503 מה שנחשב לגבוה לטעמנו. לכן נסיק
ש-sqft_above כן שימושי.
'yr_built' - לעומת זאת פה הקורלציה היא רק
0.0539 ונשים לב כל ערכי yr_built ערכי המחיר
מתפלגים באופן אקראי. לכן תכונה זו אינה שימושית.



evaluation of sqft_above
1e6
Pearson Coefficient = 0.6053858213561385
price
0
1
2
3
4
5
6
7
8
0
2000
4000
6000
8000



evaluation of yr_built
1e6
Pearson Coefficient = 0.05397781398145442
price
0
1
2
3
4
5
6
7
8
1900
1920
1940
1960
1980
2000
2020
yr_built

ב) שים לב שמה שאנחנו מחפשים הוא הלוג:

$$\log(y) = X^T W$$

ולכן כדי לקבל את y נפעיל $\exp$ משני הצדדים:

$$\Downarrow$$

$$\exp(\log(y)) = \exp(X^T w)$$

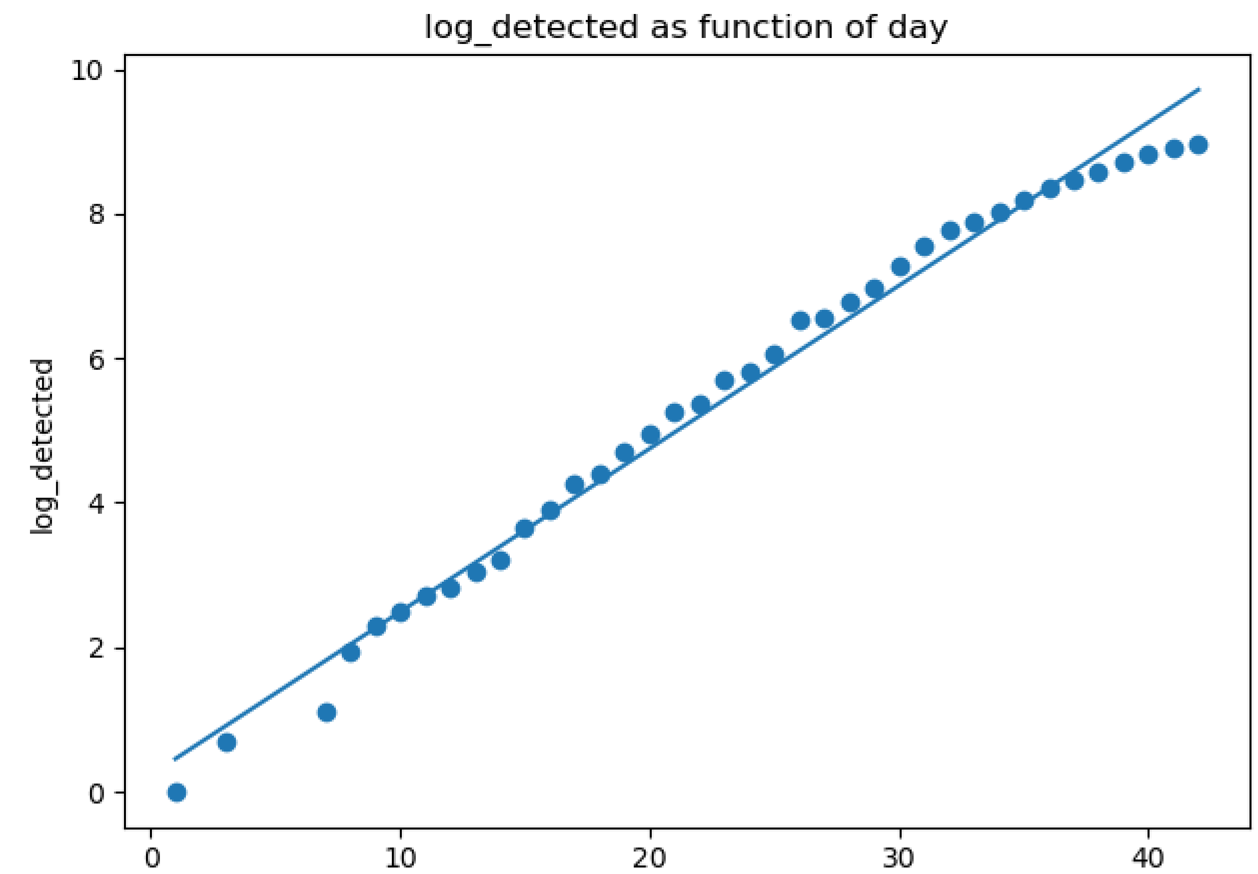
$$\Downarrow$$

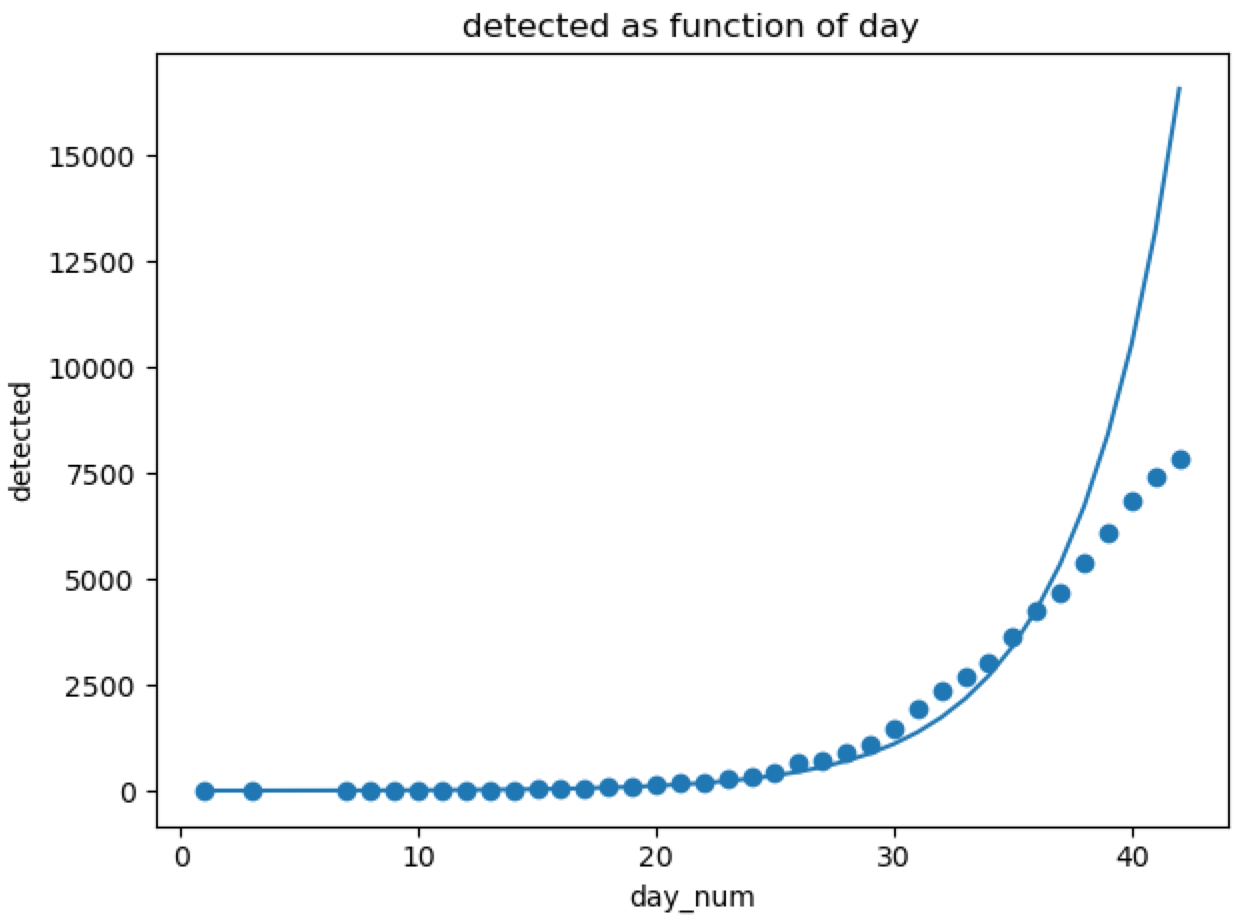
$$y = e^{\langle x, w \rangle}$$

ולכן המטרה ה- ERM תהיה:

$$ERM = \frac{1}{m}\sum_{i=1}^{m}\left(e^{\langle x_i, w \rangle} - y_i\right)^2$$

Сканировано с CamScanner


log_detected as function of day
log_detected
0
2
4
6
8
10
0
10
20
30
40


detected as function of day
detected
0
2500
5000
7500
10000
12500
15000
0
10
20
30
40
day_num