פתרון תרגיל מספר 2 ־ מערכות לומדות

שם: מיכאל גרינבאום, **ת.ז:** 211747639

23 באפריל 2020

1. **צ"ל:** $\ker\left(X^T\right)=\ker\left(X\cdot X^T\right)$
הוכחה:
תחילה נראה כי $\ker\left(X^T\right)\subseteq\ker\left(X\cdot X^T\right)$

$$v\in\ker\left(X^T\right)\Rightarrow X^T\cdot v=0\Rightarrow X\cdot X^T\cdot v=X\cdot 0=0\Rightarrow v\in\ker\left(X\cdot X^T\right)$$

כלומר $\boxed{\ker\left(X^T\right)\subseteq\ker\left(X\cdot X^T\right)}$, עתה נשים לב כי

$$\begin{aligned}&v\in\ker\left(X\cdot X^T\right)\Rightarrow X\cdot X^T\cdot v=0\Rightarrow v^T\cdot X\cdot X^T\cdot v=v^T\cdot 0=0\\&\Rightarrow\left(X^T\cdot v\right)^T\cdot\left(X^T\cdot v\right)=0\Rightarrow\left\langle X^T\cdot v,X^T\cdot v\right\rangle=0\Rightarrow\left\|X^T\cdot v\right\|=0\\&\Rightarrow X^T\cdot v=0\Rightarrow v\in\ker\left(X^T\right)\end{aligned}$$

כלומר הראנו כי $\boxed{\ker\left(X\cdot X^T\right)\subseteq\ker\left(X^T\right)}$, ולכן מתקיים $\boxed{\ker\left(X^T\right)=\ker\left(X\cdot X^T\right)}$

מ.ש.ל.☺

2. **צ"ל:** $\mathrm{Im}\left(A^T\right)=\left[\ker\left(A\right)\right]^\perp$
הוכחה:
יהי $y\in\mathrm{Im}\left(A^T\right)$, אזי $\exists x\in\mathbb{R}^p$ כך ש־ $Ax=y$ ו־ $v\in\ker\left(A\right)$, נשים לב כי

$$0=\langle x,0\rangle=\langle x,A\cdot v\rangle=\left\langle A^T\cdot x,v\right\rangle=\langle y,v\rangle$$

כלומר קיבלנו כי $\langle y,v\rangle=0$, לכל $v\in\ker\left(A\right)$ ולכן $y\in\left[\ker\left(A\right)\right]^\perp$, כלומר $\boxed{\mathrm{Im}\left(A^T\right)\subseteq\left[\ker\left(A\right)\right]^\perp}$
עתה נשים לב שממשפט המימדים מתקיים

$$\dim\left[\ker\left(A\right)\right]^\perp+\dim\ker\left(A\right)=\dim\mathbb{R}^p=p\Rightarrow\boxed{\dim\left[\ker\left(A\right)\right]^\perp=p-\dim\ker\left(A\right)}$$

עתה נזכר מלינארית 1 כי $\dim\mathrm{Im}\left(A\right)=\dim\mathrm{Im}\left(A^T\right)$, עתה ממשפט המימדים נקבל

$$\dim\mathrm{Im}\left(A\right)+\dim\ker\left(A\right)=\dim\mathbb{R}^p=p\Rightarrow\boxed{\dim\mathrm{Im}\left(A^T\right)=\dim\mathrm{Im}\left(A\right)=p-\dim\ker\left(A\right)}$$

כלומר קיבלנו כי $\dim\mathrm{Im}\left(A^T\right)=p-\dim\ker\left(A\right)=\dim\left[\ker\left(A\right)\right]^\perp$,
כלומר קיבלנו כי $\mathrm{Im}\left(A^T\right)\subseteq\left[\ker\left(A\right)\right]^\perp$ וגם $\dim\mathrm{Im}\left(A^T\right)=\dim\left[\ker\left(A\right)\right]^\perp$, לכן ממשפט מלינארית 1 נקבל

$$\boxed{\mathrm{Im}\left(A^T\right)=\left[\ker\left(A\right)\right]^\perp}$$

מ.ש.ל.☺

3. **צ"ל:** יש ∞ פתרונות אם"ם $y \perp \ker(X)$
הוכחה:
מהיות X^T לא הפיכה, מלינארית אנחנו יודעים שקיימים 0 פתרונות או ∞ פתרונות,
בנוסף לכך קיים פתרון למשוואה אם $y \in \mathrm{Im}\left(X^T\right)$ מההגדרה,
לכן אם $y \in \mathrm{Im}\left(X^T\right)$ אז קיימים ∞ פתרונות כי קיים לפחות 1 והאפשרויות הם $0, \infty$.
ואם קיימים ∞ פתרונות, אז $\exists y, w \in \mathbb{R}^p$ כך ש־ $y = X^T w$, כלומר $y \in \mathrm{Im}\left(X^T\right)$,
לכן $y \in \mathrm{Im}\left(X^T\right)$ אם"ם קיימים ∞ פתרונות למשוואה $y = X^T \cdot w$
בשאלה הקודמת הוכחנו כי $\mathrm{Im}\left(X^T\right) = [\ker(X)]^{\perp}$, נשתמש בכל אלה ונסיק כי:
$y \perp \ker(X)$ אם"ם $y \in [\ker(X)]^{\perp}$ אם"ם $y \in \mathrm{Im}\left(X^T\right)$ אם"ם קיימים ∞ פתרונות למשוואה $y = X^T \cdot w$,
כלומר קיבלנו כי יש ∞ פתרונות אם"ם $y \perp \ker(X)$

מ.ש.ל.☺

4. **צ"ל:** למשוואות הנורמליות יש פתרון אחד אם $X \cdot X^T$ הפיכה ו־ ∞ אם $X \cdot X^T$ לא הפיכה
הוכחה:
תחילה נשים לב שאם $X \cdot X^T$ הפיכה אז הפתרון היחיד המקיים את הנוסחא הוא $y = \left(X \cdot X^T\right)^{-1} \cdot X \cdot w$, כלומר יש פתרון יחיד למשוואות הנורמליות במקרה הזה.
אחרת $X \cdot X^T$ לא הפיכה, אזי מהשאלה הקודמת קיימים ∞ פתרונות אם"ם $X \cdot y \perp \ker\left(X \cdot X^T\right)$,
משאלה 1 נקבל כי $\ker\left(X \cdot X^T\right) = \ker\left(X^T\right)$, ולכן קיימים ∞ פתרונות אם"ם $X \cdot y \perp \ker\left(X^T\right)$,
נשים לב כי $\forall v \in \ker\left(X^T\right)$ מתקיים

$$\langle X \cdot y, v\rangle = \left\langle y, X^T \cdot v\right\rangle = \langle y, 0\rangle = 0$$

לכן $X \cdot y \perp \ker\left(X^T\right)$, כלומר קיימים ∞ פתרונות למשוואות הנורמליות, כנדרש

מ.ש.ל.☺

5. **פתרון:**

(א) **צ"ל:** $P = P^T$
הוכחה:
נסמן $v_i = (a_{i,1} \ldots, a_{i,n})$, לכן

$$[P]_{j,k} = \left[\sum_{i=1}^{n} v_i \cdot v_i^T\right]_{j,k} = \sum_{i=1}^{n} \left[v_i \cdot v_i^T\right]_{j,k} = \sum_{i=1}^{n} \left[v_i \cdot v_i^T\right]_{j,k} = \sum_{i=1}^{n} a_{i,j} \cdot a_{i,k}$$
$$= \sum_{i=1}^{n} a_{i,k} \cdot a_{i,j} = \sum_{i=1}^{n} \left[v_i \cdot v_i^T\right]_{k,j} = \sum_{i=1}^{n} \left[v_i \cdot v_i^T\right]_{k,j} = [P]_{k,j} = \left[P^T\right]_{j,k}$$

כלומר קיבלנו כי $[P]_{j,k} = \left[P^T\right]_{j,k}$ לכל איבר במטריצה ולכן $\boxed{P = P^T}$

מ.ש.ל.א.☺

(ב) **צ"ל:** P לכסינה עם ערכים עצמיים של 1 ו־ 0
הוכחה:
תחילה נשלים את $v_1, \ldots, v_k$ לבסיס אורתונורמלי של $\mathbb{R}^p$ (הערה: $v_1, \ldots, v_k$ הם כבר אורתונורמליים), ונסמן את הבסיס ב־ $v_1, \ldots, v_p$
עתה נשים לב כי אם $1 \leq j \leq k$ אזי מתקיים

$$P \cdot v_j = \sum_{i=1}^{k} v_i \cdot v_i^T \cdot v_j = \sum_{i=1}^{k} v_i \cdot \delta_{i,j} = v_j \Rightarrow P \cdot v_j = 1 \cdot v_j$$

כלומר v_j הוא וקטור עצמי עם ערך עצמי 1, אחרת נשים לב כי $j > k$ מתקיים

$$P \cdot v_j = \sum_{i=1}^{k} v_i \cdot v_i^T \cdot v_j = \sum_{i=1}^{k} v_i \cdot \delta_{i,j} = 0 \Rightarrow P \cdot v_j = 0 \cdot v_j = 0$$

כלומר v_j הוא וקטור עצמי עם ערך עצמי 0,
כלומר $v_1, \ldots, v_p$ הינו בסיס אורתונורמלי של וקטורים עצמיים,
נסמן $U = \left[\begin{array}{ccc} (v_1) & \cdots & (v_p) \end{array}\right]$, אזי מתקיים

$$P = U \cdot \left[\begin{array}{cc} I_k & \\ & 0_{p-k} \end{array}\right] \cdot U^T$$

מ.ש.ל.ב.☺

(ג) **צ"ל:** $Pv = v$
הוכחה:
יהי $v \in V$ אזי $\exists a_1, \ldots, a_k \in \mathbb{R}$ כך ש- $v = \sum_{j=1}^{k} a_i \cdot v_i$, נשים לב כי

$$P \cdot v = \sum_{i=1}^{k} v_i \cdot v_i^T \cdot v = \sum_{i=1}^{k} v_i \cdot \langle v_i, v \rangle = \sum_{i=1}^{k} v_i \cdot \left\langle v_i, \sum_{j=1}^{k} a_i \cdot v_i \right\rangle = \sum_{i=1}^{k} v_i \cdot a_i = v$$

כלומר קיבלנו כי $\boxed{Pv = v}$

מ.ש.ל.ג.☺

(ד) **צ"ל:** $P^2 = P$
הוכחה:
הראנו בסעיף ב כי $P = U \cdot D \cdot U^T$ כאשר U אורתוגונלית ו- D אלכסונית עם $1,0$ באלכסון,
נשים לב כי $D^2 = D$, לכן

$$\begin{aligned} P^2 &= P \cdot P = \left(U \cdot D \cdot U^T\right) \cdot \left(U \cdot D \cdot U^T\right) = U \cdot D \cdot \left(U^T \cdot U\right) \cdot D \cdot U^T \\ &= U \cdot D \cdot I_n \cdot D \cdot U^T = U \cdot D^2 \cdot U^T = U \cdot D \cdot U^T = P \end{aligned}$$

כלומר קיבלנו כי $\boxed{P^2 = P}$

מ.ש.ל.ד.☺

(ה) **צ"ל:** $(I - P) \cdot P = 0$
הוכחה:
בסעיף הקודם ראינו כי $P = P^2$, לכן נקבל כי $P - P^2 = 0$, נשים לב כי

$$(I - P) \cdot P = P - P^2 = 0$$

כלומר קיבלנו כי $\boxed{(I - P) \cdot P = 0}$

מ.ש.ל.ה.☺

6. **צ"ל:** $\left(X \cdot X^T\right)^{-1} \cdot X = X^{T^\dagger}$
הוכחה:
נעשה לפי המומלץ בתרגיל, נשים לב כי $X = U \cdot \Sigma \cdot V^T$ כאשר U, V אורתוגונליות ו- Σ אלכסונית, נשים לב כי

$$\begin{aligned} X \cdot X^T &= U \cdot \Sigma \cdot V^T \cdot \left(U \cdot \Sigma \cdot V^T\right)^T = U \cdot \Sigma \cdot V^T \cdot \left(V \cdot \Sigma^T \cdot U^T\right) \\ &= U \cdot \Sigma \cdot \left(V^T \cdot V\right) \cdot \Sigma^T \cdot U^T = U \cdot \left(\Sigma \cdot I_n \cdot \Sigma^T\right) \cdot U^T = U \cdot \left(\Sigma \cdot \Sigma^T\right) \cdot U^T \end{aligned}$$

נסמן $D=\Sigma\cdot\Sigma^T$, נשים לב כי נתון ש $X\cdot X^T$ הפיכה וגם U הפיכה, ולכן D הפיכה,
לכן מתקיים

$$\left(X\cdot X^T\right)^{-1}=\left(U\cdot D\cdot U^T\right)^{-1}=\left(U^T\right)^{-1}\cdot D^{-1}\cdot U^{-1}=U\cdot D^{-1}\cdot U^T$$

עתה נשים לב כי

$$\begin{aligned}\left(X\cdot X^T\right)^{-1}\cdot X&=\left(U\cdot D^{-1}\cdot U^T\right)\cdot X=\left(U\cdot D^{-1}\cdot U^T\right)\cdot\left(U\cdot\Sigma\cdot V^T\right)\\&=U\cdot D^{-1}\cdot\left(U^T\cdot U\right)\cdot\Sigma\cdot V^T=U\cdot D^{-1}\cdot I_n\cdot\Sigma\cdot V^T\\&=U\cdot D^{-1}\cdot I_n\cdot\Sigma\cdot V^T=U\cdot D^{-1}\cdot\Sigma\cdot V^T=U\cdot\Sigma^{T^\dagger}\cdot V^T\\&=\left(V\cdot\Sigma^\dagger\cdot U^T\right)^T=X^{T\dagger}\end{aligned}$$

כלומר קיבלנו כי

$$\boxed{\left(X\cdot X^T\right)^{-1}\cdot X=X^{T\dagger}}$$

מ.ש.ל.☺

7. **צ"ל:** $X\cdot X^T$ הפיכה אם"ם $\operatorname{span}\{x_1,\ldots,x_m\}=\mathbb{R}^d$
הוכחה:
$\operatorname{span}\{x_1,\ldots,x_m\}=\mathbb{R}^d$,
$\Leftrightarrow$ השורות של X^T פורשות את $\mathbb{R}^d$,
$\Leftrightarrow \ker\left(X^T\right)=\{0\}$
$\Leftrightarrow \ker\left(X\cdot X^T\right)=\{0\}=\ker\left(X^T\right)$, לפי שאלה 1 ראינו כי $\ker\left(X\cdot X^T\right)=\ker\left(X^T\right)$
$\Leftrightarrow \ker\left(X\cdot X^T\right)=\{0\}$
$\Leftrightarrow$ $X\cdot X^T$ הפיכה
כלומר קיבלנו כי $X\cdot X^T$ הפיכה אם"ם $\operatorname{span}\{x_1,\ldots,x_m\}=\mathbb{R}^d$

מ.ש.ל.☺

8. **צ"ל:** $\|\hat{w}\|_2\leq\|\overline{w}\|_2$
הוכחה:
יהי $v_1,\ldots,v_n$ בסיס של וקטורים עצמיים אורתונורמליים של $X\cdot X^T$, (קיימים מהמשפט הספקטרלי)
נסמן $\hat{w}=\sum_{i=1}^n a_i\cdot v_i$,$\overline{w}=\sum_{i=1}^n b_i\cdot v_i$,
עתה נשים לב כי לכל פתרון הממוזער את $X^T\cdot w=y$, מתקיים $X\cdot X^T\cdot w=X\cdot y$, לכן

$$X\cdot X^T\cdot\overline{w}=\sum_{i=1}^n\lambda_i\cdot b_i\cdot v_i=X\cdot y$$

$$X\cdot X^T\cdot\hat{w}=\sum_{i=1}^n\lambda_i\cdot a_i\cdot v_i=X\cdot y$$

$$0=X\cdot y-X\cdot y=\sum_{i=1}^n\lambda_i\cdot a_i\cdot v_i-\sum_{i=1}^n\lambda_i\cdot b_i\cdot v_i=\sum_{i=1}^n\lambda_i\cdot(a_i-b_i)\cdot v_i$$

לכן $\lambda_i=0$ או $a_i=b_i$,
לכן נסמן את האינדקסים של הערכים העצמיים שנותנים 0 ב־ $n,\ldots,l+1$, לכן נקבל כי עבור $1\leq i\leq l$ מתקיים $a_i=b_i$,
ולכן

$$\overline{w}=\sum_{i=1}^l a_i\cdot v_i+\sum_{i=l+1}^n b_i\cdot v_i$$

ונשים לב שמאיך שחישבנו את $\hat{w}$, מתקיים כי $(0,\dots,0) = a_{r+1}\dots,a_n$, לכן

$$\overline{w} = \sum_{i=1}^{l} a_i \cdot v_i + \sum_{i=l+1}^{n} b_i \cdot v_i = \hat{w} + \sum_{i=l+1}^{n} b_i \cdot v_i$$

לכן מאורתונרמליות הוקטורים וממשפט פיתגורס נקבל

$$\|\overline{w}\|^2 = \|\hat{w}\|^2 + \left\|\sum_{i=l+1}^{n} b_i \cdot v_i\right\|^2 \geq \|\hat{w}\|^2 \Rightarrow \boxed{\|\hat{w}\| \leq \|\overline{w}\|}$$

כלומר קיבלנו את מה שרצינו, שמתקיים $\|\hat{w}\|_2 \leq \|\overline{w}\|_2$, כנדרש

מ.ש.ל.☺

9. **בוצע**

10. **בוצע**

11. **בוצע**

12. **צ"ל:** הסבר על סינון מידע
הוכחה:
תחילה בדקתי מה אמור להיות חיובי ואי שלילי וסיננתי לפי זה.
לאחר מכן מצאתי קישור למה הטווחים של חלק מהעמודות וראיתי שרוב המידע שקיבלנו עומד בתנאים ולכן בדקתי שהעמודות מקיימות את דרישות הטווחים, הלינק מצורף בקוד.
לבסוף החלטתי למחוק שורות שהופיעו פעמיים כדי לעודד גיוון (זה היה רק 6 שורות מתוך 21 אלף).
בנוסף לזה החלטתי למחוק את העמודות $id, date$. הסיבה שבחרתי למחוק את id כי הוא מיוחד לכל קנייה ואינו מהווה תלות לינארית, ואם נמיר ל־ $one\ hot\ encoding$ נקבל התאמה אחד לאחד לכל המידע רק בגלל המשתנים הנ"ל.
לגבי $date$ בדקתי עם המרתו ל־ $one\ hot\ encoding$ וכשהוא לא קיים ולא היו שינויים משמעותיים בשגיאה ולכן בחרתי למחוק כי ככה החישובים יוצאים יותר מהירים. אין היגיון בלהשאיר את $date$ כתאריך כי אין תלות לינארית בתאריך עצמו (נבדק עם $pearson\ correlation$)

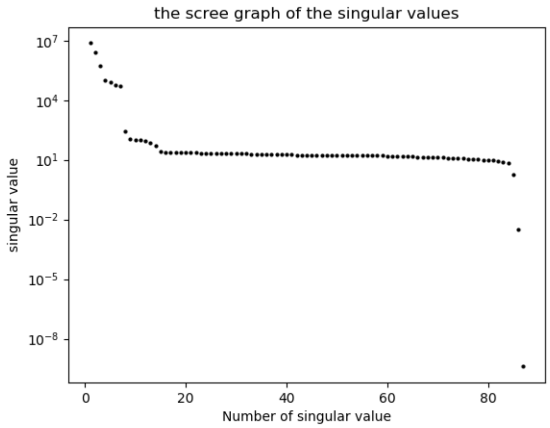
מ.ש.ל.☺

13. **צ"ל:** הסבר על משתנים קטגוריים
הוכחה:
בסעיף הקודם ציינתי שמחקתי את $id, date$ עם הנימוק שלי לכך.
המשתנה היחיד שהחלטתי שהוא קטגורי הוא $zipcode$ כי הוא מכיל מידע שאינו מהווה חלק מהתלות הלינארי (אם אני בזיפ 22, זה לא פי 3 פחות טוב מ זיפ 66 או הפוך).
אבל לעומת $date$, לאיזור שבו קניתי הייתה קורלציה לא קטנה עם המחיר, ולכן החלטתי להמיר עם $one\ hot\ encoding$ (בהתאם להסבר שניתן בפורום)

מ.ש.ל.☺

14. **בוצע**

15. **צ"ל:** תמונה של גרף ערכים סינגולריים
הוכחה:

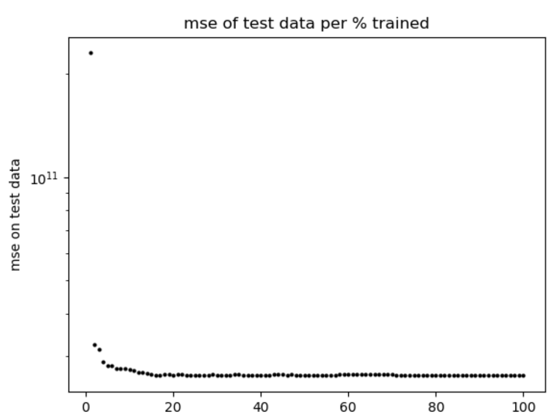



```
[7.42446895e+06 2.52633111e+06 5.33676731e+05 9.77649000e+04
 7.75964515e+04 5.94153840e+04 5.42400202e+04 2.65665202e+02
 1.18069595e+02 1.01543163e+02 9.57421407e+01 8.62961934e+01
 6.80045843e+01 5.21542844e+01 2.71660485e+01 2.42151900e+01
 2.38736930e+01 2.38456024e+01 2.36721066e+01 2.30507959e+01
 2.28873640e+01 2.27986634e+01 2.22624033e+01 2.18449986e+01
 2.16762858e+01 2.15341500e+01 2.13918228e+01 2.08911860e+01
 2.08305685e+01 2.05483509e+01 2.03920168e+01 2.01012420e+01
 1.98584723e+01 1.97459374e+01 1.90180231e+01 1.88174033e+01
 1.86568447e+01 1.83971730e+01 1.83670331e+01 1.81279957e+01
 1.79043875e+01 1.77260950e+01 1.76156764e+01 1.74851168e+01
 1.74364562e+01 1.69644906e+01 1.68884843e+01 1.67833278e+01
 1.67415004e+01 1.67048171e+01 1.66027975e+01 1.65496124e+01
 1.65124283e+01 1.63835845e+01 1.62684831e+01 1.62165847e+01
 1.61887093e+01 1.60865693e+01 1.60167215e+01 1.57608344e+01
 1.57181157e+01 1.52669836e+01 1.52002222e+01 1.50181856e+01
 1.45058205e+01 1.41903983e+01 1.40640401e+01 1.38834506e+01
 1.37573854e+01 1.35718668e+01 1.28883876e+01 1.23488518e+01
 1.19818688e+01 1.18312303e+01 1.17187293e+01 1.12379692e+01
 1.10118522e+01 1.03779861e+01 1.01754717e+01 1.00446734e+01
 9.57729943e+00 8.74399782e+00 7.61126283e+00 7.04720026e+00
 1.72328384e+00 3.12290185e-03 4.04954253e-10]
```

ניתן לראות שיש ערכים מאוד קטנים בסדר גודל של e^{-10} ולכן המטריצה X מאוד קרובה להיות סינוגלרית.

מ.ש.ל.☺

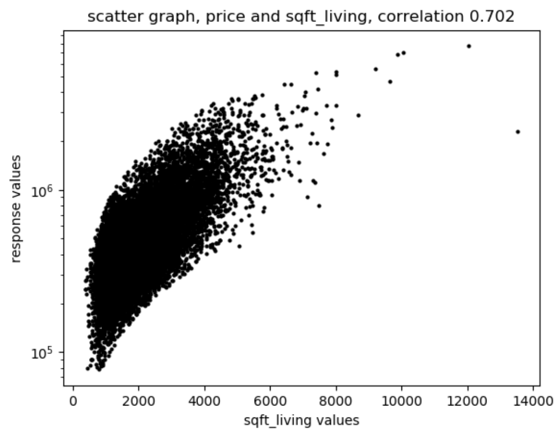
16. **צ"ל:** תמונה של למידה והסבר
הוכחה:

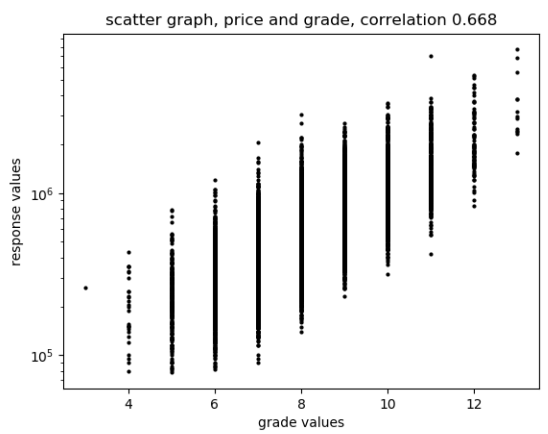


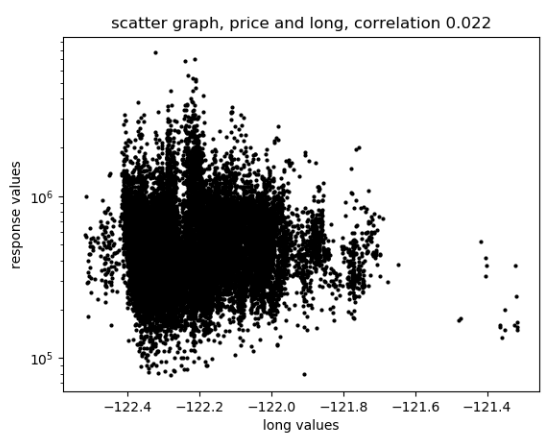
ניתן לראות שבהתחלה השגיאה גדולה כי למדנו רק חלק קטן מהמרחב והשגיאה יורדת ככל שלומדים יותר על המרחב.

מ.ש.ל.☺

17. **צ"ל:** תמונה של למידה והסבר
הוכחה:







ניתן להסיק מה תורם ומה לא לפי הקורלציה שחישבנו ושקיים יחס בתזוזה בx לתזוזה בy בגרף.
הגיונית, בתים שקיבלו ציון גבוה או עם הרבה שטח, ימכרו במכיר יותר גבוה כמו שניתן לראות מהקורלציה.
בנוסף לזה, לרוב בקניית בית אין חשיבות למיקום ה$longtitude$ וניתן לראות זאת בקורלציה נמוכה

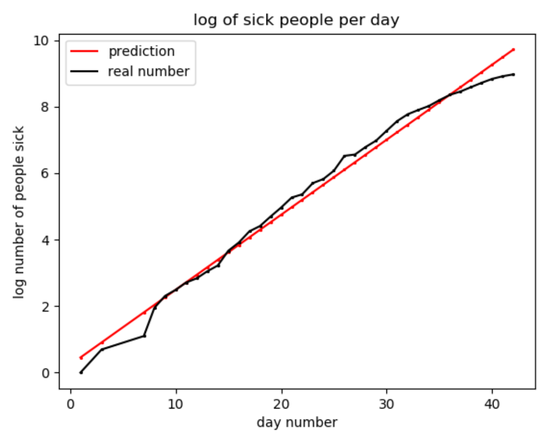
מ.ש.ל.☺

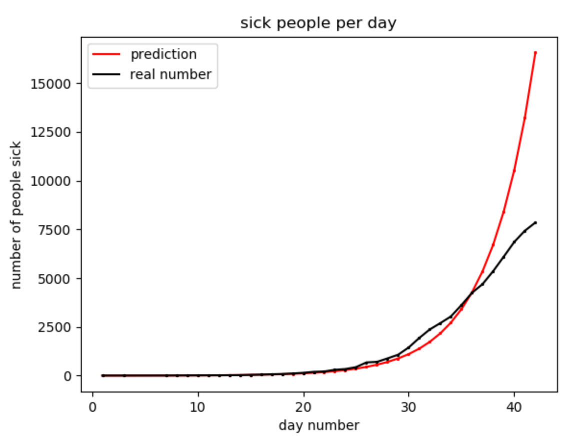
18. **בוצע**

19. **בוצע**

20. **בוצע**

21. **צ"ל:** תמונה של קוביד לוג וללא לוג
הוכחה:





מ.ש.ל.☺

22. **צ"ל:** לעשות $linear\ regresion$ על y ולא על $log\left(y\right)$
הוכחה:
אנחנו יודעים מהסעיף הקודם שהפונקציה שאנחנו מחפשים היא מהצורה $e^{\langle x,w\rangle}=\hat{f}\left(x\right)$,
בסעיף הקודם חישבנו את השגיאה על ידי $\left(\langle w,x\rangle-\log\left(y\right)\right)^{2}$,

עתה השגיאה האמיתית היא מהצורה $\hat{f}\left(x\right)-y$ (הניחוש - התשובה האמיתית), ומכיוון שאנחנו מעוניינים בשגיאה ריבועית, נגדיר

$$L\left(y,\hat{f}\left(x\right)\right)=\left(\hat{f}\left(x\right)-y\right)^{2}$$

מפה נקבל שהשגיאה הכללית היא

$$\sum_{i=1}^{m}L\left(y_{i},\hat{f}\left(x_{i}\right)\right)=\sum_{i=1}^{m}\left(\hat{f}\left(x_{i}\right)-y_{i}\right)^{2}=\sum_{i=1}^{m}\left(e^{\langle x_{i},w\rangle}-y_{i}\right)^{2}=\left\Vert e^{X^{T}\cdot w}-y\right\Vert ^{2}$$

נרצה למצוא וקטורים w הממזערים את הביטוי, לכן ננסה לגזור וננסה למצוא מקומות מינימום

$$\begin{aligned}J_{w}\left(\left\Vert e^{X^{T}\cdot w}-y\right\Vert ^{2}\right) & =J_{e^{X^{T}\cdot w}-y}\left(\left\Vert e^{X^{T}\cdot w}-y\right\Vert ^{2}\right)\cdot J_{w}\left(e^{X^{T}\cdot w}-y\right)=2\cdot\left(e^{X^{T}\cdot w}-y\right)^{T}\cdot J_{w}\left(e^{X^{T}\cdot w}-y\right)\\
 & =2\cdot\left(e^{X^{T}\cdot w}-y\right)^{T}\cdot J_{X^{T}\cdot w}\left(e^{X^{T}\cdot w}-y\right)\cdot J_{w}\left(X^{T}\cdot w\right)\\
 & =2\cdot\left(e^{X^{T}\cdot w}-y\right)^{T}\cdot e^{X^{T}\cdot w}\cdot X^{T}
\end{aligned}$$

ולכן

$$\nabla\left\Vert e^{X^{T}\cdot w}-y\right\Vert ^{2}=\left(J_{w}\left(\left\Vert e^{X^{T}\cdot w}-y\right\Vert ^{2}\right)\right)^{T}=2\cdot X\cdot\left(e^{X^{T}\cdot w}\right)^{T}\cdot\left(e^{X^{T}\cdot w}-y\right)$$

ננסה למצוא את הוקטורים המאפסים את המשוואה $\nabla\left\Vert e^{X^{T}\cdot w}-y\right\Vert ^{2}=0$ שהם מינימום גלובלי (על ידי חישוב ה־$Hessian$)

מ.ש.ל.☺