# מבוא ללמידת מכונה - תרגיל 2

**מגיש:** יואב לוי 314963257

## Solutions of the Normal Equations

**שאלה 1**

יש להוכיח $Ker\left(X^T\right)=Ker\left(XX^T\right)$, אראה זאת בעזרת הכלה דו־כיוונית.

- $Ker\left(X^T\right)\subseteq Ker\left(XX^T\right)$:

יהי וקטור $w\in Ker\left(X^T\right)$ מתקיים,

$$XX^Tw=X\vec{0}=\vec{0}\implies w\in Ker\left(XX^T\right)$$

- $Ker\left(X^T\right)\supseteq Ker\left(XX^T\right)$:

יהי וקטור : $w\in Ker\left(XX^T\right)$ (נסמן ב־☺), נשים לב כי בעזרת פירוק ה־SVD נוכל לרשום את המטריצות הבאות כך:

קיימות מטריצות $U,V$ אורתוגונליות ומטריציה $\Sigma$ **ריבועית אלכסונית**

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

$$X^T=V\Sigma^TU^T\overset{\Sigma\text{ is squared mat}}{=}V\Sigma U^T$$

$$XX^T=U\Sigma^2U^T$$

כעת, נשים לב שמתקיים

$$XX^Tw=U\Sigma^2U^Tw\overset{☺}{=}\vec{0}$$

**למה** $0.1$ **- בהינתן מטריצה אורתוגונלית** $U\in M_{n\times n}\left(\mathbb{R}\right)$ **הפיכה (תכונה של מטריצות אורתוגונליות)**
**אז** $dim\left(Ker\left(U^T\right)\right)=0$

**<u>הוכחת למה</u>** $0.1$**:**

ראשית מהגדרת ההפיכות נובע כי $dim\left(Ker\left(U\right)\right)=0$, מלינארית אנו יודעים שמימד מרחב השורות שווה למימד מרחב העמודות של מטריצה.

בנוסף אנחנו יודעים כי מימד מרחב השורות של $U$ הוא $dim\left(Im\left(U\right)\right)$
והוא כאמור שווה למימד מרחב השורות של $U^T$. נשים לב שלפי משפט המימדים נקבל:

$$dim\left(Im\left(U\right)\right)+dim\left(Ker\left(U\right)\right)=n$$

ולכן גם:

$$dim\left(Im\left(U^T\right)\right)+dim\left(Ker\left(U^T\right)\right)=n\iff dim\left(Ker\left(U^T\right)\right)=n-dim\left(Im\left(U\right)\right)=dim\left(Ker\left(U\right)\right)=0$$

■

כעת נחזור למשוואה $XX^Tw=U\Sigma^2U^Tw=\vec{0}$ ונשים לב שמהלמה 0.1 נקבל ש־$U^Tw\neq 0$ ואם $\Sigma^2U^Tw\neq 0$ אז שוב מלמה 0.1 $U\Sigma^2U^Tw\neq\vec{0}$ בסתירה למשוואה שהגענו אליה, ולכן $\Sigma^2U^Tw=\vec{0}$ בהכרח.
מכאן שגם $\Sigma U^Tw=\vec{0}$, זאת מכיוון שהמטריצה $\Sigma$ ריבועית ואלכסונית ולכן להכפיל ווקטור בה זה כמו להכפיל כל קורדינטה של הווקטור
פי סקלר (הסקלרים שעל האלכסון), אם כך שלכפול פי עוד סיגמה זה כמו לכפול פי הסקלר בריבוע פשוט, לכן אם הסקלר אינו אפס מלכתחילה
הוא לא יאפס את הווקטור (אלא אם הווקטור הוא אפס מלכתיחלה), ומכאן נוכל לקבל בקלות שמתקיים:

$$X^Tw\overset{S.V.D}{=}V\Sigma U^Tw=\vec{0}$$

כלומר $w\in Ker\left(X^T\right)$ כנדרש.

■

**שאלה 2**
נוכיח את הטענה בעזרת הכלה דו־כיוונית.

- $Im\left(A^T\right)\subseteq Ker\left(A\right)^{\perp}$:

יהי: $w\in Im\left(A^T\right)$ כלומר קיים ווקטור $v\in\mathbb{R}^n$ כך שמתקיים $A^Tv=w$.
אזי לכל ווקטור $\mu\in Ker\left(A\right)$ מתקיים:

$$\left\langle w,\mu\right\rangle=\left\langle A^Tv,\mu\right\rangle\overset{\text{Hermitian adjoint over }\mathbb{R}}{=}\left\langle v,A\mu\right\rangle\overset{\mu\in Ker(A)}{=}\left\langle v,0\right\rangle=0$$

ולכן $w\in Ker\left(A\right)$.

- $Im\left(A^T\right)\supseteq Ker\left(A\right)^{\perp}$:

נניח בשלילה שעבור ווקטור $w\notin Im\left(A^T\right)$ ועלינו להראות כי $w\notin Ker(A)^{\perp}$.
כלומר עלינו להראות כי קיים $v\in Ker\left(A\right)$ כך שמתקיים $\left\langle w,v\right\rangle\neq 0$. נגדיר את $v\in Im\left(A^T\right)^{\perp}$ ונראה כי
מכיוון ש־$w\notin Im\left(A^T\right)$ אזי $w\in Im\left(A^T\right)^{\perp}$ כלומר $\left\langle w,v\right\rangle\neq 0$ וגם

$$\left\langle v,\overbrace{A^TAv}^{\in Im\left(A^T\right)}\right\rangle=0$$

ולכן,

$$\left\|Av\right\|^2=\left\langle Av,Av\right\rangle=\left\langle v,A^TAv\right\rangle=0$$

כלומר $Av=0$ א"ז $v\in Ker\left(A\right)$ ולכן $Im\left(A^T\right)\supseteq Ker\left(A\right)^{\perp}$ כנדרש.

■

**שאלה 3**

בהינתן שהמטרציה $X^T$ סינגולרית צריך להוכיח $y\perp Ker\left(X\right) \iff X^Tw=y \text{ has } \infty \text{ solutions}$.
($\Leftarrow$) אנחנו יודעים שיהיו למערכת המשוואות $X^Tw=y$ או 0 או $\infty$ פתרונות.
כעת, אם נראה ש-$y\in Im\left(X^T\right)$ אז זה אומר שקיים פתרון, ולכן יש $\infty$ פתרונות.
נשים לב כי אנחנו מניחים שמתקיים $y\perp Ker\left(X\right)$, כלומר $y\in Ker\left(A\right)^{\perp}$.
כעת נעזר בשאלה הקודמת ובה הראינו את השקילות $y\in Ker\left(A\right)^{\perp}=Im\left(X^T\right)$, ולכן $y$ אכן בתמונה ולמערכת יש $\infty$ פתרונות.
($\Rightarrow$) כעת נניח שלמערכת יש $\infty$ פתרונות. נסמן פתרון אחד ב-$w$ ונשים לב כי $y=X^Tw$, כעת לכל ווקטור $v\in Ker\left(X\right)$

$$\left\langle y,v\right\rangle =\left\langle X^Tw,v\right\rangle =\left\langle w,Xv\right\rangle \overset{v\in Ker(X)}{=}\left\langle w,0\right\rangle =0\implies y\perp Ker\left(X\right)$$

כנדרש.

■

**שאלה 4**

- אם $XX^T$ היא הפיכה (Invertible) אזי מתקיים,

$$XX^Tw=Xy\iff\left(XX^T\right)^{-1}XX^Tw=\left(XX^T\right)^{-1}Xy\iff w=\left(X^T\right)^{-1}y$$

- אחרת אם $XX^T$ סינגולרית (אינה הפיכה) אזי או שיש 0 פתרונות או שיש $\infty$, אם אצליח להראות כי $Xy\in Im\left(XX^T\right)$

  הרי שיש $\infty$ פתרונות למערכת המשוואות.

  נשים לב שניתן להמיר את הבעיה לכך שצריך להראות כי $Xy\in Ker\left(XX^T\right)^{\perp}$ מתוך הלמה שהוכחנו בשאלה השנייה.

  כעת, מהלמה שהוכחנו בשאלה הראשונה ניתן להמיר שוב את הבעיה לכך שצריך להראות כי $Xy\in Ker\left(X^T\right)^{\perp}$

  ואת זה קל לנו כבר לפתור, יהי $v\in Ker\left(X^T\right)$ נשים לב כי מתקיים

$$\left\langle Xy,v\right\rangle =\left\langle y,X^Tv\right\rangle =\left\langle y,0\right\rangle =0$$

  לכן $Xy\in Ker\left(X^T\right)^{\perp}=Ker\left(XX^T\right)^{\perp}=Im\left(XX^T\right)$ כלומר $Xy\in Im\left(XX^T\right)$ כנדרש.

  ולכן באם המטריצה $XX^T$ סינגולרית אז למערכת המשוואות יש $\infty$ פתרונות.

■

## Projection Matrices

**שאלה 5**

($a$) מטריצת ההטלה על $V\subseteq\mathbb{R}^d$ מוגדרת ע"י המכפלה החיצונית הבאה, $P=\sum_{i=1}^{k}v_i\cdot v_i^T$ (כאשר $v_1,\ldots,v_k$ בסיס אורתונורמלי).
מכאן שכל איבר בסכום הוא מטריצה בגודל $d\times d$, אם אראה שכל מטריצה כזו היא סימטרית הרי ש-$P$ עצמה סימטרית, כתוצאה של חיבור
של מטריצות סימטריות.
לכן, נתבונן במחובר כללי בסכום (למשל כאשר $i=j$)

$$\vec{v}_j \coloneqq \begin{bmatrix} v_{j,1} \\ v_{j,2} \\ \vdots \\ v_{j,d} \end{bmatrix} \in \mathbb{R}^d$$

$$\vec{v}_j \cdot \vec{v}_j^T = \begin{bmatrix} v_{j,1} \\ v_{j,2} \\ \vdots \\ v_{j,d} \end{bmatrix} \cdot \begin{bmatrix} v_{j,1} & v_{j,2} & \dots & v_{j,d} \end{bmatrix} = \begin{pmatrix} (v_{j,1})^2 & (v_{j,1} \cdot v_{j,2}) & \cdots & (v_{j,1} \cdot v_{j,d}) \\ (v_{j,2} \cdot v_{j,1}) & (v_{j,2})^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ (v_{j,d} \cdot v_{j,1}) & (v_{j,d} \cdot v_{j,2}) & \cdots & (v_{j,d})^2 \end{pmatrix}$$

נסמן את המטריצה שלעיל ב־$A$ וכך נוכל לכתוב בכתיב קומפקטי:

$$A_{\mathrm{x,y}} = (v_{j,x} \cdot v_{j,y})$$

לכן מכיוון שהכפל קומוטטיבי אז המטריצה הזאת היא סימטרית, וזה גורר שהמטריצה המקורית $P$ סימטרית גם כן.

$(b)$ ראשית אוכיח שהערכים העצמיים היחידים של מטריצת ההטלה הם 0,1.
נשים לב שמתקיים $P^2 = P$ כי

$$P^2 = \left(\sum_{i=1}^{k} v_i \cdot v_i^T\right)^2 = \left(\left(v_1 \cdot v_1^T\right) + \cdots + \left(v_k \cdot v_k^T\right)\right)^2 = \sum_{i=j} \left(v_i \cdot v_i^T\right) \cdot \left(v_j \cdot v_j^T\right) + 2 \sum_{i \neq j} \left(v_i \cdot v_i^T\right) \left(v_j \cdot v_j^T\right) =$$

$$\overset{(v_1, \dots, v_k \text{ orthonormal basis})}{=} \sum_{i=1}^{k} v_i \cdot v_i^T = P$$

כעת, עבור ווקטור עצמי $0 \neq v \in V$ נקבל,

$$\exists \lambda \in \mathbb{R} \ s.t. \ Pv = \lambda v$$

$$P^2 v = Pv = \lambda v$$

$$P^2 v = P(\lambda v) = \lambda P v = \lambda^2 v$$

$$\Downarrow$$

$$\lambda^2 v = \lambda v \iff \lambda^2 v - \lambda v = 0 \iff \lambda v (\lambda - 1) = 0$$

כלומר $\lambda = 0 \vee \lambda = 1$, כנדרש.
כעת, נשים לב שעבורים הווקטורים $v_1, \dots, v_k$ הערכים העצמיים המשוייכים להם הם 1:

$$\forall j \in [k]: \ P(v_j) = \sum_{i=1}^{k} v_i \cdot v_i^T \cdot (v_j) \overset{(v_1,\dots,v_k \text{ orthonormal basis})}{=} \sum_{i=1}^{k} v_i \overbrace{\delta_{i,j}}^{\text{Kronecker delta}} = 1 \cdot v_j$$

כנדרש.

$(c)$ יהי $v \in V$, כלומר ניתן לייצגו כסופרפוזיציה של הבסיס האורתונורמלי של $V$ הנתון לנו $B = (v_1, \dots, v_k)$.
בהינתן סקלרים $c_1, \dots, c_k \in \mathbb{R}$ נקבל:

$$v = c_1 \cdot v_1 + c_2 \cdot v_2 + \dots + c_k \cdot v_k$$

כעת,

$$Pv = P\left(c_1 \cdot v_1 + c_2 \cdot v_2 + \dots + c_k \cdot v_k\right) = \left(\sum_{i=1}^{k} v_i \cdot v_i^T\right) \cdot \left(\sum_{i=1}^{k} c_i \cdot v_i\right) =$$

$$= \sum_{j=1}^{k} \sum_{i=1}^{k} v_j \cdot \overbrace{v_j^T \cdot v_i}^{\delta_{i,j}} \cdot c_i = \sum_{i=1}^{k} c_i \cdot v_i = v$$

כנדרש.

$(d)$ הוכח כבר בסעיף ב' בשאלה זו (🤦‍♂️)

$(e)$ $(I - P)\,P = P - P^2 \overset{(d)}{=} P - P = \mathbf{0}$

## Least Squares

### שאלה 6

•

$$\left(XX^T\right)^{-1} = \left(U\Sigma \underbrace{V^T V}_{=0 \text{ because orthonormal matrix}} \Sigma^T U^T\right)^{-1} = \left(UDU^T\right)^{-1} = UD^{-1}U^T$$

כאשר השיוויון האחרון נובע מהעובדה שההופכי של מטריצות אורתוגונליות הוא המטריצה המשוחלפת.

•

$$\left(XX^T\right)^{-1} X = UD^{-1}U^T X = UD^{-1}U^T U \Sigma V^T = UD^{-1}\Sigma V^T = \left(V \overbrace{\Sigma^T \Sigma^T \left(\Sigma\Sigma^T\right)^{-1}}^{(\text{פ})} U^T\right)^T$$

כעת נעבור לנוטצית אינדקסים, ואגדיר את האופרטור † על סקלרים באופן הבא: $\left([\Sigma]_{i,j}\right)^{\dagger} := \begin{cases} \frac{1}{[\Sigma]_{i,j}} & \text{if } [\Sigma]_{i,j} \neq 0 \\ 0 & otherwise \end{cases}$

ונשים לב שמתקיים: (פ)

$$\Sigma^T \Sigma^T \left(\Sigma \Sigma^T\right)^{-1} \overset{\text{Index Notation}}{\Longrightarrow} [\Sigma]_{j,i} \cdot \left([\Sigma]_{i,j} \cdot [\Sigma]_{j,i}\right)^{-1} = \left([\Sigma]_{i,j}\right)^{\dagger}$$

ולכן,

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

כנדרש.

**שאלה 7**

ראשית נשים לב שהדרגה של המטריצה $X^T \in M_{m\times d}(\mathbb{R})$ היא המימד הנפרש ע"י הווקטורים $x_1, \ldots, x_m$.
כעת, מהשאלה הראשונה בתרגול נובע כי $Ker\left(X^T\right) = Ker\left(XX^T\right)$ ומכיוון ש־$XX^T$ הפיכה אז $Ker\left(X^T\right) = Ker\left(XX^T\right) = 0$ ולפי משפט המימדים נובע,

$$Im\left(X^T\right) + Ker\left(X^T\right) = d \iff Im\left(X^T\right) = d$$

כלומר, מימד מרחב השורות של $X^T$ שווה ל־$d$ כלומר $\text{span}\left\{x_1, \ldots, x_m\right\} = \mathbb{R}^d$, וזה קורה אם ורק אם המטריצה $XX^T$ הפיכה.

**שאלה 8**

מכיוון שכל קורדינטה ב־$\hat{w}$ היא ייחודית לכל ערך סינגולרי של $X$ אז לכל פתרון $\bar{w}$ אחר נקבל, $\bar{w}_i = \hat{w}_i$ עבור כל $i = 1, \ldots, r$, אבל $\forall i \in r+1, \ldots, d : \hat{w}_i = 0$ ולכן:

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

**שאלה 9+10+11+12**

אלו שאלות קוד שמומשו בקובץ "linear_model.py".

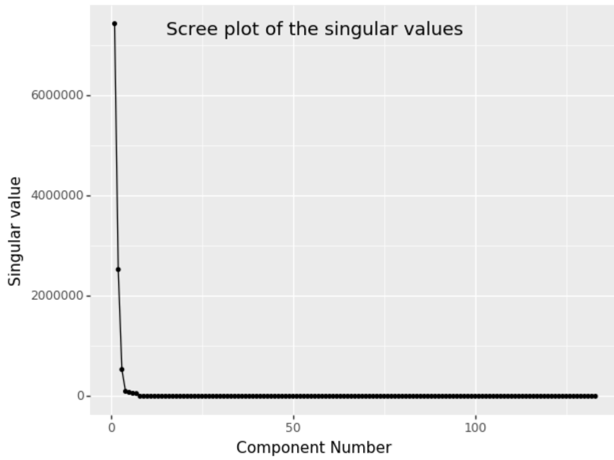
**שאלה 13**

מצאתי שה־features הבאים הם **categoricals**:

- $zipcode$ ־ המיקוד, כמו שנוסח בשאלה, אין קורלציה ישירה בין הגדלים של זיפ קוד אחד לאחר אך סביר להניח שהמיקוד יכול להועיל לרגרסיה הלינארית, מאחר וסביר שתהיה קורלציה בין המחיר של הבית למיקום שלו (שבא לידי ביטוי במיקוד).
- $date$ ־ זהו התאריך של מכירת הבית, נשים לב כי קיבלנו אותו בפורמט קצת מוזר yymmddT000000, התמודדתי איתו על ידי עריכה של העמודה ופיצולה ל־3 עמודות שונות (שנה, חודש, יום) ועבור כל אחד מהעמודות האלה, יצרתי dummy variables.

**שאלה 14**

<u>הגרף שמתאר את הערכים הסינגולרים:</u>

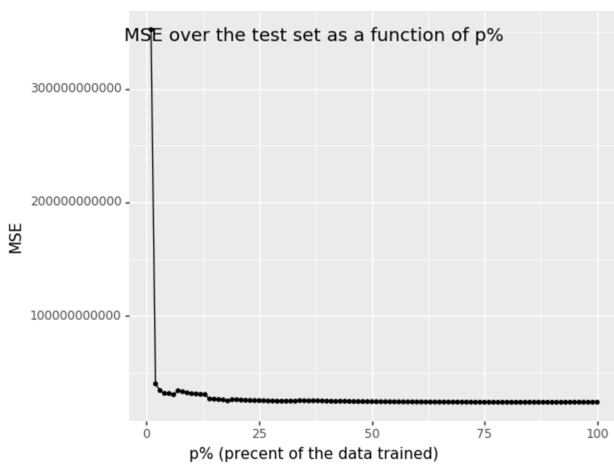



**שאלה 15**

מתוך ההרצאה "LectureHandout_2_Linear_Regression.pdf": (תרגמתי לעברית)
"לעיתים $XX^T$ היא פורמלית הפיכה אך קרובה ללא הפכיה (=סינגולרית). הדבר קורה כאשר העמודות של $X^T$ הן
כמעט תלויות לינארית, או כאשר אחת מהעמודות של $X^T$ הוא כמעט נפרש על־ידי שאר העמודות.
במקרה כזה, מספר הערכים הסינגולרים של $X$ שהם לא אפסים יהיה קטן".

כעת, בגרף שמופיע בשאלה 14 יש את התוצאה שלי,
נשים לב כי רוב הערכים הסינגולרים הם דווקא אפס, כלומר מספר הערכים הסינגולרים של $X$ שהם לא אפס קטן,
ולכן סביר להניח שהיא **אכן** קרובה להיות סינגולרית.

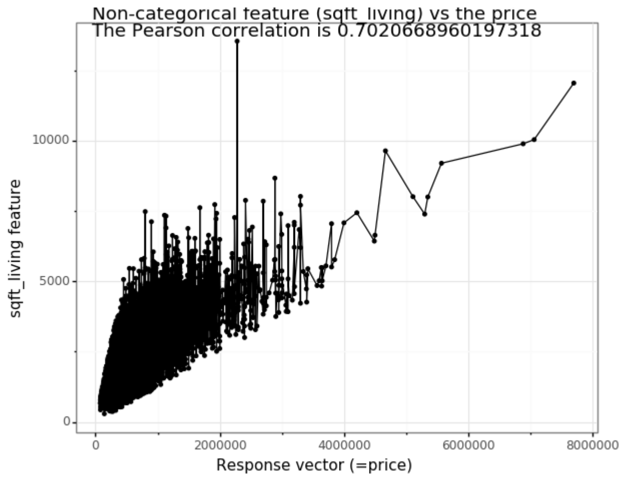
**שאלה 16**



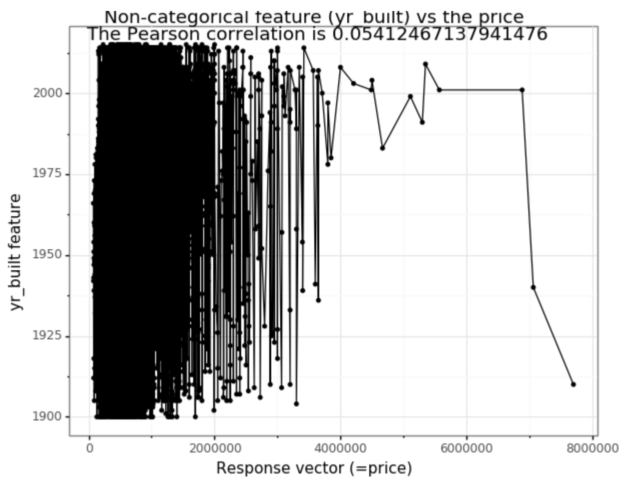

נשים לב כי אנחנו מתכנסים מהר מאוד ל־~$MSE = 24,000,000,000$ זאת מכיוון שאנחנו מאמנים על עוד נתונים.
בנוסף שמתי לב שכשאני לא מעבד את הדאטא שלי ולא מחליף משתנים קטגורים אלא מוחק אותם לוקח לי יותר זמן להתכנס,
פה אני מתכנס די מהר.

**שאלה 17**

בגרף הבא, ניתן לראות קורלציה גבוהה (יחסית) בין מחיר הדירה לשטחה ב־square feet. ניתן לראות זאת בכך
שה־Pearson correlation קרוב ל־1, בנוסף בגרף ניתן לראות התנהגות יחסית לינארית ככל שהמחיר עולה.
לכן ניתן להסיק שהשטח של הדירה יכול להועיל לרגרסיה הלינארית. (חוץ מזה אינטואיטיבית לא מפתיע שהמחיר של הדירה
קשור בקשר הדוק לגודל הדירה)




בגרף הבא, ניתן לראות קורלציה נמוכה (יחסית) בין מחיר הדירה לשנה שבה היא נבנתה. ניתן לראות זאת בכך שה־Pearson correlation קרוב ל־0, בנוסף בגרף ניתן לראות התנהגות שאינה לינארית לכל אורך הגרף. לכן ניתן להסיק שהשטח של הדירה יכול להועיל לרגרסיה הלינארית.

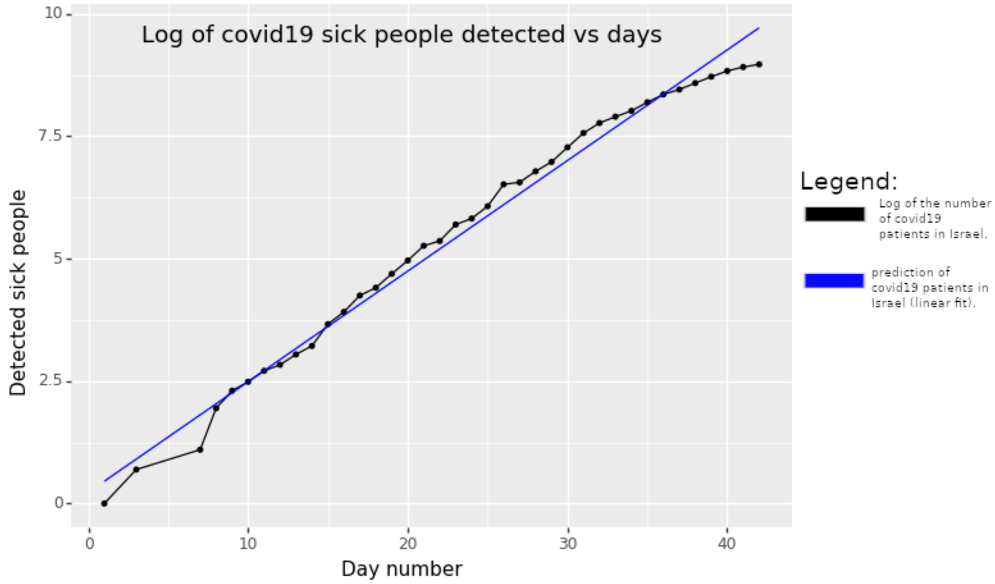


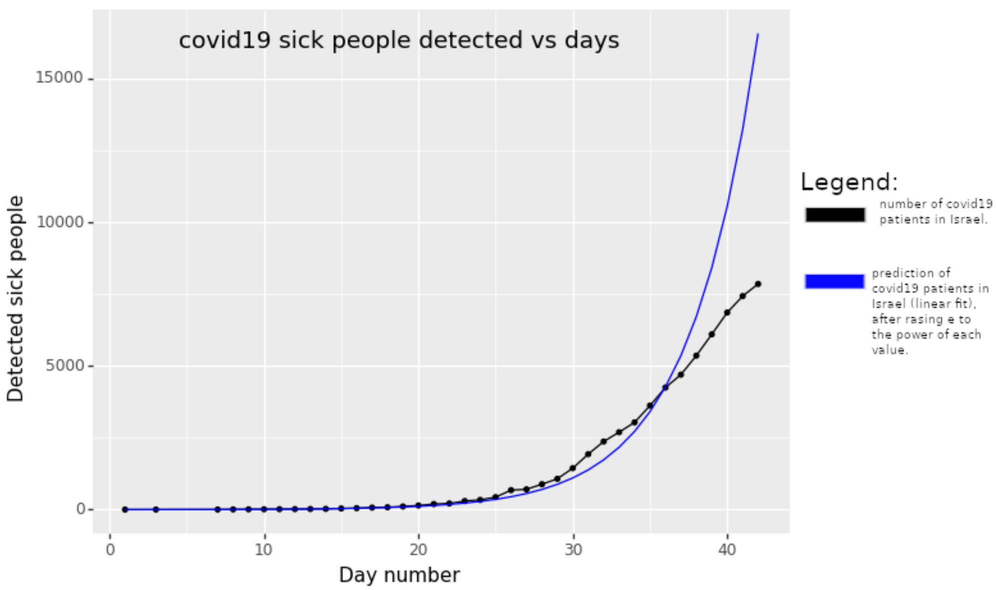

**שאלות 18+19+20**

אלו שאלות קוד הממומשות בקובץ "covid19.py".

**שאלה 21**

הגרפים הבאים מראים את המספר חולי הקורונה בישראל אל מול מספר הימים מאז שהתגלה החולה הראשון. הגרף הימני מתאר את מספר החולים באופן לוגריתמי והגרף השני לא. בנוסף בשני הגרפים יש את הקו הכחול שהוא השיערוך שלי, כאשר בגרף השמאלי זה השיערוך של החולים מועלים בחזקת $e$.




**שאלה 22**

- פונקצית ה־loss צריכה להיות כמו $L_{exp}$ כדי שנקבל תוצאות נכונות.
  כלומר אנחנו נשלח לפונקצית ה־fit_linear_regression את הלוגריתם הטבעי של $y$.
  כי אנחנו מנסים לעשות שערוך לפונקציה לינארית ולא אקספוננציאלית, ומכיוון שהלוגריתם של פונקציה אקספוננציאלית הוא לינארי, נבחר ב־loss function הזה.
- אם היינו משתמשים בפונקציית ה־least square loss, אז הייתי מנסה לעשות Gradient Descent שזו שיטת אופטימיזציה איטרטיבית מסדר ראשון למציאת מינימום מקומי של פונקציה.
  בשיטה זו, נעשה צעד נגדי לגרדיאנט ביחס לנקודה הנוכחית.