IML – Ex.2

Bar Aloni-[redacted]

חלק תאורטי:

<u>Solutions of the Normal Equations:</u>

1. נוכיח ש- $Ker(X^T) = Ker(XX^T)$ ע"י הכלה דו כיוונית:
<u>$\subseteq$:</u> יהא $v \in Ker(X^T) = \{X^Tv = 0\}$. מתקיים: $XX^Tv = X(X^Tv) = X \times 0 = 0$, כלומר: $v \in Ker(XX^T)$ כנדרש.
<u>$\supseteq$:</u> יהא $v \in Ker(XX^T) = \{XX^Tv = 0\}$. נניח כי $v \notin Ker(X^T)$. מתקיים ש:

$\langle X^Tv, X^T2v\rangle = 2\langle X^Tv, X^Tv\rangle \neq 0$ וכן: $\langle XX^Tv, 2v\rangle = 0$.

בנוסף נבחין ש: $\langle X^Tv, X^T2v\rangle = \langle XX^Tv, 2v\rangle$ מכיוון ש- $\langle Xv, u\rangle = \langle v, X^Tu\rangle$, וקיבלנו סתירה. לכן בהכרח $v \in Ker(X^T)$ כנדרש.

■

2. נראה ש: $Im(A^T) = \left(Ker(A)\right)^\perp$:
- $\subseteq$: יהא $x \in Im(A^T) = \{\, x \mid \exists y\ s.t:\ x = A^Ty\}$, אזי $x = A^Ty$ לאיזשהוא y.
נבחין ש:
$$\left(Ker(A)\right)^\perp = \{\, x \mid \langle x, v\rangle = 0\ \ \forall v \in Ker(A)\} = \{\, x \mid \langle x, v\rangle = 0\ \ \forall v\ s.t:\ Av = 0\}$$
אז יהא v המקיים: $Av = 0$. מתקיים:
$$\langle x, v\rangle = x^Tv = (A^Ty)^Tv = y^T(A^T)^Tv = y^TAv = y^T \cdot 0 = 0$$
כלומר $x \in \left(Ker(A)\right)^\perp$ כנדרש.
- $\supseteq$: נראה שאם $x \notin Im(A^T) = \{\, x \mid \exists y\ s.t:\ x = A^Ty\}$, כלומר אם $\forall y: x \neq A^Ty$, אז $x \notin \left(Ker(A)\right)^\perp$. יהא v המקיים $A^Tv = 0$. אזי:
$$\langle x, v\rangle = x^Tv \neq (A^Ty)^Tv = y^T(A^T)^Tv = y^TAv = y^T \cdot 0 = 0$$
ולכן $x \notin \left(Ker(A)\right)^\perp$, כנדרש.

■

3. תהא $y = X^Tw$ מערכת משוואות לינארית לא-הומוגנית. נניח כי $X^T \in \mathbb{R}^{n\times n}$ לא הפיכה.
- X^T לא הפיכה, כלומר מימד מרחב העמודות קטן מ-n, ולכן יש תלות לינארית ביניהן. נסמן את העמודות $X^T = \begin{bmatrix} | & & | \\ \varphi_1 & \cdots & \varphi_n \\ | & & | \end{bmatrix}$. נניח כי: $y \perp Ker(X)$. אזי מתקיים:
$y \in \left(Ker(X)\right)^\perp = Im(X^T)$, כלומר ניתן לתאר את y כצירוף לינארי של עמודות X^T.
מכיוון שעמודות X^T תלויות לינארית, יש ∞ דרכים לייצג את y וכל צורה כזו למעשה מהווה פתרון לבעיה ויש לנו ∞ פתרונות כנדרש.
- נניח כי יש ∞ פתרונות למשוואה $y = X^Tw$. אז קיים פתרון v, כלומר קיים v עבורו מתקיים: $X^Tv = y$. כלומר $y \in Im(X^T) = \left(Ker(X)\right)^\perp$, כנדרש.

■

4. תהא $XX^Tw = Xy$ מערכת משוואות נורמלית לינארית.
נראה ש:
- אם XX^T הפיכה אז יש למערכת פתרון יחד.
<u>הוכחה:</u> מתקיים: $(XX^T)^{-1}XX^Tw = (XX^T)^{-1}Xy$ כלומר $w = (XX^T)^{-1}Xy$ פתרון למערכת.
נניח בשלילה שיש פתרון נוסף למערכת: $\bar{w} \neq w$, כלומר מתקיים: $XX^T\bar{w} = Xy$. באותו אופן כמו מקודם, נקבל ש: $\bar{w} = (XX^T)^{-1}XX^T\bar{w} = (XX^T)^{-1}Xy$, ולכן: $\bar{w} = w$, בסתירה. לפיכך יש בדיוק

פתרון אחד למערכת המשוואות.

- אם XX^T אינה הפיכה אז יש למערכת אינסוף פתרונות.
<u>הוכחה:</u> נתבונן באגף הימני של מערכת המשוואות: $b \coloneqq Xy$. אז $b \in Im(X)$. משאלה (2): נקבל את הביטוי השקול הבא למה שהוכחנו: $Im(X) = \left(Ker(X^T)\right)^{\perp}$, ולכן: $b \perp Ker(X^T)$. משאלה (1) נקבל ש: $b \perp Ker(XX^T)$, ומשאלה (3), בתוספת ההנחה ש XX^T הפיכה, נקבל שלמערכת $(XX^T)^T w = b$ יש אינסוף פתרונות.
נבחין ש: $(XX^T)^T w = (X^T)^T X^T w = XX^T w$, ולכן למעשה למערכת שלנו, $b = Xy$, $XX^T w = Xy$ יש ∞ פתרונות כנדרש.

■

<u>Projection Matrices:</u>

5. הנחות: $V \subseteq \mathbb{R}^d$, $dim(V) = k \in \mathbb{N}$, $\{v_1, \dots, v_k\}$ בסיס אורתונורמלי ל-V.
נגדיר את מטריצת ההטלה האורתוגונלית: $P = \sum_{i=0}^{k} v_i v_i^T$, כאשר v_i נכפלים זה בזה ע"י המכפלה החיצונית $\otimes$.

a) <u>נראה ש-P סימטרית:</u>

$$P = \sum_{i=0}^{k} v_i v_i^T = \sum_{i=0}^{k} \begin{bmatrix} v_{i,1}\, v_{i,1} & \cdots & v_{i,1} v_{i,k} \\ \vdots & \ddots & \vdots \\ v_{i,k}\, v_{i,1} & \cdots & v_{i,k} v_{i,k} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{k} v_{i,1}\, v_{i,1} & \cdots & \sum_{i=1}^{k} v_{i,1}\, v_{i,k} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{k} v_{i,k}\, v_{i,1} & \cdots & \sum_{i=1}^{k} v_{i,k}\, v_{i,k} \end{bmatrix}$$

נבחין שכל איבר בסכום (לאחר השיוויון השני) הוא סימטרי מקומוטטיביות המכפלה, ולכן גם סכום האיברים הנ"ל הוא סימטרי, מקומוטטיביות פעולת החיבור. לכן P סימטרית.

b) <u>נראה שהע"ע של P הם 0 או 1, והו"ע המתאימים להם הם $\{v_1, \dots, v_k\}$:</u>
נקח $1 \le l \le k$, אז מתקיים:
$Pv_l = \sum_{i=0}^{k} (v_i v_i^T) v_l = \sum_{i=0}^{k} v_i (v_i^T v_l) = \sum_{i=0}^{k} v_i \langle v_i, v_l \rangle$.
נבחין שמכיוון שזהו בסיס אורתונורמלי מתקיים שלכל $i \neq l$: $\langle v_i, v_l \rangle = 0$, ואחרת: $\langle v_i, v_l \rangle = 1$.
אזי $Pv_l = v_l \cdot 1 = v_l$, כלומר לכל $1 \le l \le k$, v_l וקטור עצמי והע"ע המתאים לו הוא 1.

c) <u>$\forall v \in V\colon \mathrm{P}v = v$</u>
יהיה $v \in V$, אזי הוא ניתן לכתיבה כצירוף לינארי של של אברי הבסיס האורתונורמלי $\{v_1, \dots, v_k\}$: $v = \sum_{i=1}^{k} a_i v_i$, עבור $\{a_i\} \in \mathbb{R}$. מתקיים:

$$\begin{aligned} Pv &= \sum_{i=0}^{k} (v_i v_i^T) v \\ &= \sum_{i=0}^{k} v_i v_i^T \left(\sum_{j=1}^{k} a_j v_j \right) \\ &= \sum_{i=0}^{k} \left(\sum_{j=1}^{k} v_i v_i^T \left(a_j v_j\right) \right) = \sum_{j=0}^{k} a_j \left(\sum_{i=1}^{k} v_i (v_i^T v_j) \right) \end{aligned}$$

וכמו בסיף קודם נבחין ש: $\sum_{i=0}^{k} v_i (v_i^T v_j) = v_j$,
וקיבלנו: $Pv = \sum_{j=0}^{k} a_j \left(\sum_{i=1}^{k} v_i (v_i^T v_j) \right) = \sum_{j=0}^{k} a_j \left(v_j\right) = v$, כנדרש.

d) $\underline{P^2 = P}$:

$$P^2 = \left(\sum_{i=0}^{k} v_i v_i^T\right)\left(\sum_{i=0}^{k} v_i v_i^T\right) = \sum_{i=1}^{k}\left((v_i v_i^T)^2 + \sum_{i\neq j\in[k]} (v_i v_i^T)(v_j v_j^T)\right)$$
$$= \sum_{i=1}^{k}\left((v_i v_i^T)^2 + \underbrace{\sum_{i\neq j\in[k]} v_i \underbrace{\langle v_i, v_j\rangle}_{=0} v_j^T}_{=0}\right)$$
$$= \sum_{i=1}^{k} v_i (v_i^T v_i) v_i^T = \sum_{i=1}^{k} v_i \underbrace{(v_i^T v_i)}_{=1} v_i^T = \sum_{i=0}^{k} v_i v_i^T = P$$

e) $\underline{(I-P)P = 0}$:

מתקיים: $(I-P)P = I\cdot P - P^2 = P - P^2 = P - P = 0$, כאשר המעבר השלישי נובע מתכונה d.

<u>Least Squares:</u>

6. נסמן את פירוק ה-SVD של X: $X = U\Sigma V^T$.

- <u>נראה ש: $(XX^T)^{-1} = UD^{-1}U^T$, עבור $D \coloneqq \Sigma\Sigma^T$:</u>

מתקיים:

$$(XX^T)^{-1} = \left((U\Sigma V^T)\left((U\Sigma V^T)\right)^T\right)^{-1} = \left(U\Sigma \underbrace{V^T V}_{=I} \Sigma^T U^T\right)^{-1} = \left(U \underbrace{\Sigma\Sigma^T}_{\coloneqq D} U^T\right)^{-1}$$
$$= (U^T)^{-1} D^{-1} (U)^{-1} = (U^T)^T D^{-1} (U)^T = U D^{-1} U^T$$

כאשר $V^T V = I$ וכן $U^{-1} = U^T$ מכיוון ש-U,V מטריצות אורתונומליות.

- <u>נראה ש: $(XX^T)^{-1}X = X^{T\dagger}$:</u>

מתקיים: $(XX^T)^{-1}X = (UD^{-1}U^T)(U\Sigma V^T) = UD^{-1}\underbrace{U^T U}_{=I}\Sigma V^T = U(D^{-1}\Sigma)V^T$

נבחין ש: $X^{T\dagger} = (X^\dagger)^T = (V\Sigma^\dagger U^T)^T = U\Sigma^{T\dagger}V^T = U\Sigma^\dagger V^T$, כאשר המעבר האחרון נובע מכך ש- Σ אלכסונית ולכן לא משתנה תחת פעולת השחלוף. השמשתי גם בנתון שאומר ש: $(X^T)^\dagger = (X^\dagger)^T$.

לכן נותר להראות ש: $D^{-1}\Sigma = \Sigma^\dagger$.

נזכיר ש XX^T הפיכה, ולכן עמודותיה בת"ל, כלומר גם עמודות D בת"ל. בפרט זה אומר שאברי אלכסון D אינם 0 ולכן מתקיים:

$$D^{-1}\Sigma = (\Sigma\Sigma^T)^{-1}\Sigma = (\Sigma^2)^{-1}\Sigma = diag(\sigma_1^2, \dots, \sigma_d^2)^{-1} diag(\sigma_1, \dots \sigma_d) =$$
$$diag\left(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_d^2}\right) diag(\sigma_1, \dots \sigma_d) = diag\left(\frac{\sigma_1}{\sigma_1^2}, \dots, \frac{\sigma_d}{\sigma_d^2}\right) = diag\left(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_d}\right) = \Sigma^\dagger$$

השתמשתי בתכונה של מכפלת מטריצות אלכסוניות *A,B*: $C_{i,i} = A_{i,i}B_{i,i}$.

וקיבלנו כנדרש.

7. <u>נבחין שמתקיים:</u>

$\dim(Span\{x_1, \dots, x_m\}) = rank(X) = rank(XX^T)$ ולכן מכיוון ש$XX^T \in \mathbb{R}^{d\times d}$:

$$Span\{x_1, \dots, x_m\} = \mathbb{R}^d \Leftrightarrow rank(XX^T) = d \Leftrightarrow XX^T \ is\ invertible$$

8. נסמן את פירוק ה*SVD* של *X*: $X = U\Sigma V^T$. יהא *w* פתרון ל: $XX^T w = Xy$. אזי מתקיים:

$$XX^T w = Xy \Rightarrow U\Sigma \underbrace{V^T V}_{=I} \Sigma^T U^T w = U\Sigma V^T y \Rightarrow UDU^T w = U\Sigma V^T y \Rightarrow DU^T w = \Sigma V^T y$$

יהיו 2 פתרונות w_1, w_2. נבטא אותם בעזרת הבסיס האורתונורמלי ל- $\mathbb{R}^d$ שמגדירות עמודות המטריצה האורתונורמליות $U \in \mathbb{R}^{d\times d}$:

$$w_1 = \Sigma_{i=1}^d a_i u_i, \quad w_2 = \Sigma_{i=1}^d b_i u_i$$

מתקיים: $DU^T(w_1 - w_2) = DU^T(\Sigma V^T y - \Sigma V^T y) = DU^T \cdot 0 = 0$

נחשב את אגף שמאל:

המקטור המתקבל מכיל במקום ה-*j* את : $Dj\langle u_j, \Sigma_{i=1}^{d}(a_i-b_i)u_i\rangle$, כאשר D_j הוא *D* במקום ה-*j* (כלומר σ_j^2) והכפל מתאפשר מכיון ש-*D* מטריצה אלכסונית. נבחין שמתקיים:

$$\langle u_j, \Sigma_{i=1}^{d}(a_i-b_i)u_i\rangle = \Sigma_{i=1}^{d}(a_i-b_i)\langle u_j,u_i\rangle = (a_j-b_j)\underbrace{\langle u_j,u_j\rangle}_{=1} = (a_j-b_j),$$

מכיוון שלכל $i \neq j$: $\langle u_j,u_i\rangle = 0$. לכן התוצאה היא הוקטור $\begin{bmatrix}\sigma_1^2(a_1-b_1)\\ \vdots \\ \sigma_d^2(a_d-b_d)\end{bmatrix}$

נניח שיש ל-*X* *r* ערכים סינגולריים, אזי בה"כ הם מופיעים ב-*r* השורות הראשונות ב-Σ.

לכן:

* לכל $i > r$: $\sigma_i = 0$ כלומר אכן $\sigma_i^2(a_i-b_i) = 0$

* לכל $i \leq r$: $\sigma_i \neq 0$, ומהשוויון לאגף ימין צריך להתקיים: $\underbrace{\sigma_i^2}_{\neq 0}\ (a_i-b_i) = 0 \Rightarrow a_i = b_i$.

זאת אומרת שכל פתרון *w* מזדהה על *r* המקדמים ל-$u_1,\dots,u_r$, נסמנם: $\alpha_1,\dots,\alpha_d$.

יהא $\hat{w} = X^{T\dagger}y$ פתרון, ויהיה פתרון נוסף $\hat{w} \neq \bar{w}$.

נבחין ש:

$$\|\hat{w}\| = \|X^{T\dagger}y\| = \|U\Sigma^1V^Ty\| = \left\|U\cdot\begin{bmatrix}\frac{1}{\sigma_1}\langle v_1,y\rangle\\ \vdots\\ \frac{1}{\sigma_r}\langle v_d,y\rangle\\ 0\\ \vdots\\ 0\end{bmatrix}\right\| = \left\langle U\cdot\begin{bmatrix}\frac{1}{\sigma_1}\langle v_1,y\rangle\\ \vdots\\ \frac{1}{\sigma_r}\langle v_d,y\rangle\\ 0\\ \vdots\\ 0\end{bmatrix}, U\cdot\begin{bmatrix}\frac{1}{\sigma_1}\langle v_1,y\rangle\\ \vdots\\ \frac{1}{\sigma_r}\langle v_d,y\rangle\\ 0\\ \vdots\\ 0\end{bmatrix}\right\rangle$$

$$= \begin{bmatrix}\frac{1}{\sigma_1}\langle v_1,y\rangle\\ \vdots\\ \frac{1}{\sigma_r}\langle v_d,y\rangle\\ 0\\ \vdots\\ 0\end{bmatrix}\cdot\underbrace{U^TU}_{=I}\cdot\begin{bmatrix}\frac{1}{\sigma_1}\langle v_1,y\rangle\\ \vdots\\ \frac{1}{\sigma_r}\langle v_d,y\rangle\\ 0\\ \vdots\\ 0\end{bmatrix} = \left\|\begin{bmatrix}\frac{1}{\sigma_1}\langle v_1,y\rangle\\ \vdots\\ \frac{1}{\sigma_r}\langle v_d,y\rangle\\ 0\\ \vdots\\ 0\end{bmatrix}\right\| = \left\|\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_r\\ 0\\ \vdots\\ 0\end{bmatrix}\right\|$$

ומכיוון ש: $\bar{w} \neq \hat{w}$, אז קיים $r+1 \leq i \leq d$ כך ש: $\bar{w}_i \neq 0$, ולכן $\bar{w}_i^2 > 0$.

נבחין שמכך מתקיים: $\Sigma_{i=1}^{r}\bar{w}_i^2 \leq \Sigma_{i=1}^{d}\bar{w}_i^2$ ולכן:

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

ו- $\hat{w}$ הוא הפתרון מנורמה מינימלית למשוואה כנדרש.

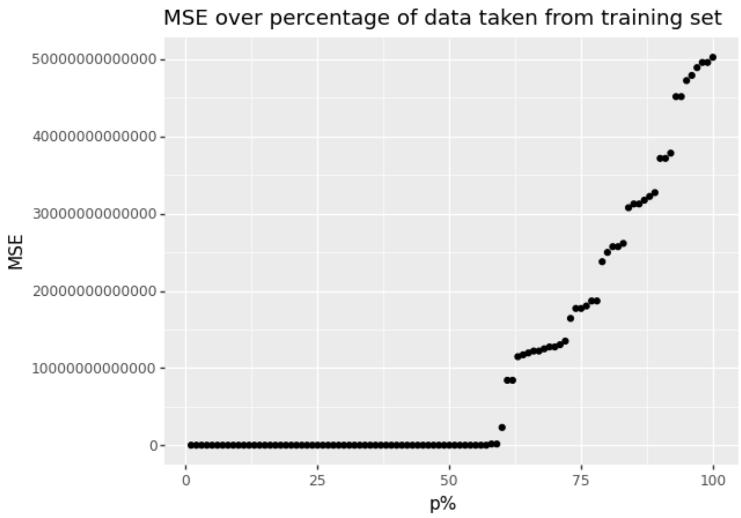
■

חלק מעשי:

<u>KC house data:</u>

9. מימוש בקובץ linear_model.py.
10. מימוש בקובץ linear_model.py.
11. מימוש בקובץ linear_model.py.
12. מימוש בקובץ linear_model.py.
13. המשתנה zipcode הוא קטגורי, מכיוון שאיו יחס סדר בין אזורי מיקוד שונה. מעבר לכך לא הגדרתי משתנים קטגוריים.
14. מימוש בקובץ linear_model.py.
15. ניתן ללמוד מהערכים הסינגולריים את מרחב האפס שהוא מרחב הפתרונות לבעיית הרגרסיה הלינארית ההומוגנית. הוקטורים הסינגולריים המתאימים לערכים הסינגולריים ששוים אפס, פורסים את מרחב הפתרונות. נבחין שהמטריצה X אכן קרובה להיות סינגולרית מכיוון שרוב הערכים הסינגולריים שלה קרובים ל-0.

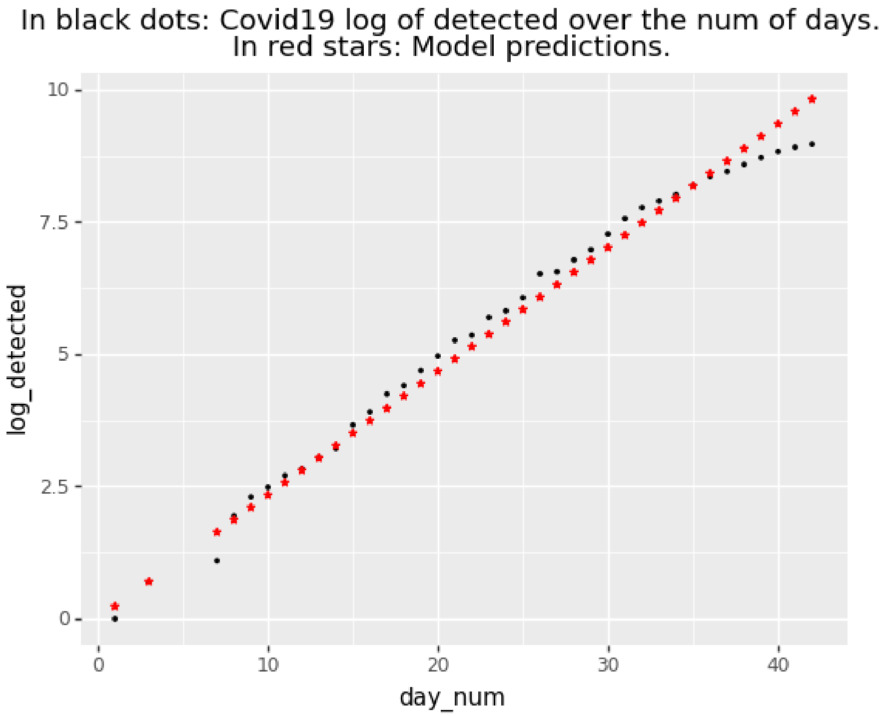
16.

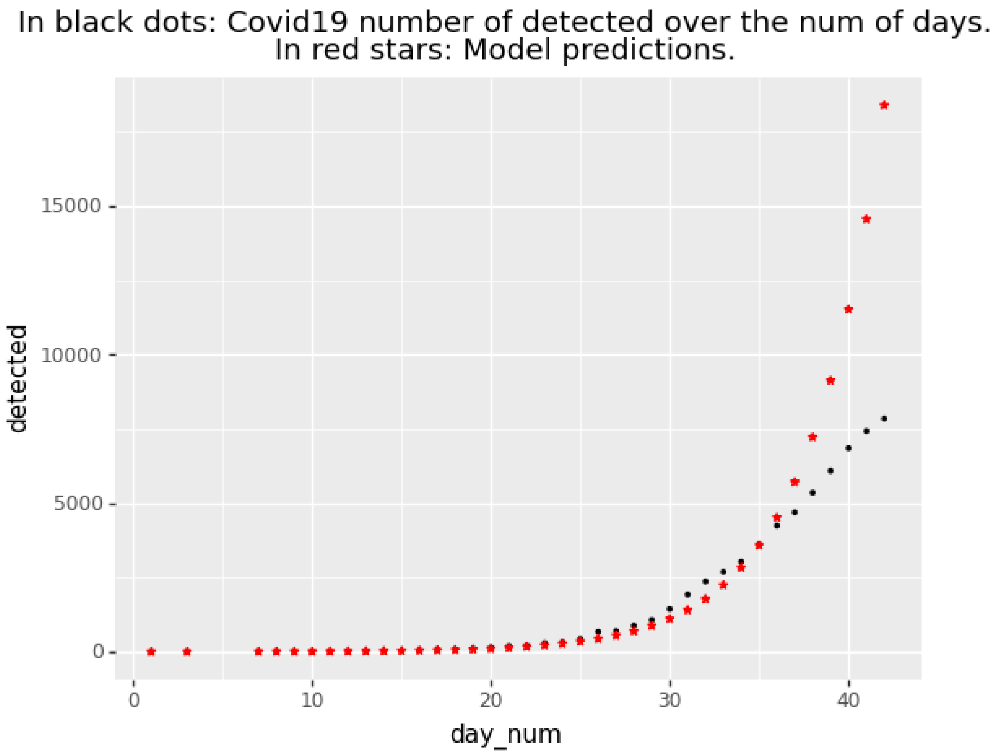


17.

<u>Covid19:</u>

18. מימוש בקובץ covid19.py
19. מימוש בקובץ covid19.py
20. מימוש בקובץ covid19.py
21.




In black dots: Covid19 number of detected over the num of days.
In red stars: Model predictions.
detected
15000
10000
5000
0
0
10
20
30
40
day_num