מערכות לומדות תרגיל 2

שיר שבח 322407701

7 באפריל 2022

2. חלק תיאורטי

יהי X מטריצה (design matrix) של בעית רגרסיה לינארית עם m שורות (דגימות, samples) ו־d עמודות (משתנים / features).
יהי $y \in \mathbb{R}^m$ ה־ response vector התואם את הדגימות ב־ X. זכרו כי במרחב ווקטורי $V \subseteq \mathbb{R}^d$ המשלים האורתוגונלי של V הוא: $V^{\perp} = \left\{x \in \mathbb{R}^d \mid \langle x, v\rangle = 0 \;\; \forall v \in V\right\}$.

2.1. פתרונות של המשוואות הנורמליות

1. הוכיחו כי: $Ker\left(X\right) = Ker\left(X^TX\right)$.

נוכיח הכלה דו כיוונית.
$\underline{Ker\left(X\right) \subseteq Ker\left(X^TX\right)}:$
נניח כי $V \subseteq \overline{\{v_1, ..., v_k\} = Ker\left(X\right)}$. אזי לכל $i \in [k]$ מתקיים $Xv_i = 0$.
לכן לכל $i \in [k]$:

$$X^TXv = X^T0_V = 0_V$$

(כי תמיד ווקטור האפס מוביל לאפס).
$\underline{Ker\left(X\right) \supseteq Ker\left(X^TX\right)}:$
נניח כי $V \subseteq \overline{\{v_1, ..., v_k\} = Ker\left(X^TX\right)}$. אזי לכל $i \in [k]$ מתקיים $X^TXv_i = 0$.
נניח בשלילה כי קיים $Xv_i = u_i \neq 0$. אזי:

$$X^TXv_i = X^Tu_i = 0$$

נכפיל את שני הצדדים ב־ v_i^T ונפעיל $traspose$

$$\begin{aligned}\left(v_i^TX^Tu_i\right)^T &= \left(v_i^T \cdot 0\right)^T \\ u_i^T\left(v_i^TX^T\right)^T &= 0 \\ u_i^T\left(Xv_i\right) &= 0 \\ u_i^Tu_i &= 0\end{aligned}$$

הנחנו כי $u_i \neq 0$ לכן גם $u_i^Tu_i \neq 0$, סתירה.

2. הוכיחו כי עבור מטריצה ריבועית A: $Im\left(A^{T}\right)=Ker\left(A\right)^{\perp}$.

ראשית נוכיח את הטענה כי $Im\left(B\right)^{\perp}=Ker\left(B^{T}\right)$ ע"י הכלה דו כיוונית.

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

יהי $v\in Ker\left(B^{T}\right)$. נרצה להראות כי $v\in Im\left(B\right)^{\perp}$.

נסמן $Bw=u$. אזי:

$$\left\langle u,v\right\rangle =\left\langle Bw,v\right\rangle =\left\langle w,B^{T}v\right\rangle =\left\langle w,0\right\rangle =0$$

* שימוש בטענה מלינארית 2.

קיבלנו כי לכל $v\in Ker\left(B^{T}\right)$, v ניצב לתמונה של B. לכן $v\in Im\left(B\right)^{\perp}$.

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

יהי $v\in Im\left(B\right)^{\perp}$. מתקיים

$$\left\langle B^{T}v,B^{T}v\right\rangle =\left\langle v,BB^{T}v\right\rangle =0$$

משום שהווקטור v ניצב לתמונה של B.

מצד שני,

$$0=\left\langle B^{T}v,B^{T}v\right\rangle =\|B^{T}v\|$$

והנורמה שווה לאפס אמ"מ הווקטור בתוך הנורמה הוא 0. לכן $B^{T}v=0$ לכל $v\in Im\left(B\right)^{\perp}$.

לכן $v\in Ker\left(B^{T}\right)$.

נסמן $B=A^{T}$. לכן

$$Im\left(A^{T}\right)^{\perp}=Ker\left(A^{TT}\right)$$
$$Im\left(A^{T}\right)^{\perp}=Ker\left(A\right)$$

ניקח את המשלים האורתוגונלי ונקבל:

$$Im\left(A^{T}\right)=Ker\left(A\right)^{\perp}$$

מש"ל.

3. יהא $y = Xw$ מערכת לא הומוגנית של משוואות לינאריות. הניחו כי X ריבועית ולא הפיכה. הראו כי למערכת יש אינסוף פתרונות אמ"מ $y \perp Ker\left(X^T\right)$.

נסמן $A = X^T$, ולפי הסעיף הקודם נקבל כי:

$$Ker\left(X^T\right)^{\perp} = Ker\left(A\right)^{\perp} = Im\left(A^T\right) = Im\left(X\right)$$

משום ש $y \in Ker\left(X^T\right)^{\perp}$, קיבלנו כי $y \in Im\left(X\right)$.

X מטריצה לא הפיכה. לכן למשוואה $y = Xw$ אין פתרון או יש אינסוף פתרונות. משום שאנו יודעים לפי השיוויון $y \perp Ker\left(X^T\right)$ כי $y \in Im\left(X\right)$ אמ"מ יש פתרון ל Xw. לכן נשללה האפשרות כי אין פתרון למערכת הנ"ל וקיבלנו כי לXw יש אינסוף פתרונות.

4. חשבו את המערכת הלינארית (הנורמלית) $X^T X w = X^T y$. בשימוש במה שהוכחתם לעיל הוכיחו כי המשוואות הנורמליות יכולות שיהיה להן רק פתרון אחד (אם $X^T X$ הפיך) או אינסוף פתרונות (אחרת).

<u>אם $X^T X$ הפיך:</u>

אזי $Ker\left(X^T X\right)$ טריוויאלי לכן הפיך. לכן ניתן לחשב את $\left(X^T X\right)^{-1}$, ולכן:

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

וקיבלנו פתרון יחיד w למערכת המשוואות הנורמליות.

<u>אם $X^T X$ לא הפיך:</u>

נסמן $X^T X = X'$, $X^T y = y'$. אזי מתקיים $X' w = y'$. נשים לב כי X' ריבועית ולא הפיכה. לכן אנו עומדים בתנאים של הסעיף הקודם, אם נוכיח כי $y' \perp Ker\left(\left(X'\right)^T\right)$, אזי למערכת המשוואות הנורמליות יש אינסוף פתרונות.

נרצה להוכיח כי עבור $v \in Ker\left(\left(X'\right)^T\right) = Ker\left(\left(X^T X\right)^T\right)$ $\langle y', v\rangle = 0$.

$$\langle y', v\rangle = \left\langle X^T y, v\right\rangle = \left(X^T y\right)^T v = y^T X v = y^T \vec{0} = 0$$

קיבלנו את מה שרצינו לכן למערכת המשוואות הנורמליות יש אינסוף פתרונות.

5. בשאלה הזו אתם תוכיחו כמה תכונות של מטריצות הטלה אורתוגונליות שראינו בתרגול. יהי $V \subseteq \mathbb{R}^d$, $dim\left(V\right) = k$ ויהי $v_1, ..., v_k$ בסיס אורתוגונלי של V. נסמן את המטריצה ההטלה האורתוגונלית $P = \sum_{i=1}^{k} v_i v_i^T$ (שימו לב כי זה מכפלה חיצונית)

א. הראו כי P סימטרית.

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

קיבלנו כי P סימטרית.

ב. הוכיחו כי הערכים העצמיים של P הם 0 או 1 וה $v_1, ..., v_k$ הם הערכים העצמיים המתאימים לערך עצמי 1.

לכל v_i בבסיס האורתוגנלי מתקיים:

$$Pv_j = \left(\sum_{i=1}^{k} v_i v_i^T\right) \cdot v_j = \sum_{i=1}^{k} v_i v_i^T v_j$$
$$= \sum_{i=1}^{k} v_i \langle v_i, v_j \rangle$$

מתכונת הבסיס האורתוגנלי, $\langle v_i, v_j \rangle = 0$ ו־ $\langle v_j, v_j \rangle = 1$
לכן,

$$Pv_j = v_j$$

לכן קיבלנו כי לכל הבסיס האורתוגנלי של V, מתקיים כי אלו ווקטורים עצמיים עם ערך עצמי 1 והשאר זהו ערך עצמי 0 (ערך טרוויאלי).

ג. הראו כי לכל $\forall v \in V$, $Pv = v$.

נסמן $v = \alpha_1 v_1 + ... + \alpha_k v_k = \sum_{i=1}^{k} \alpha_i v_i$

$$v = \sum_{i=1}^{k} \alpha_i v_i$$
$$Pv = P \sum_{i=1}^{k} \alpha_i v_i$$
$$Pv = \sum_{i=1}^{k} \alpha_i P v_i$$

לכל i, הוכחנו v_i ווקטור עצמי עם ערך עצמי 1.

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

ג. הוכיחו כי $P^2 = P$.

על ידי פירוק EVD ניתן לכתוב את P (משום שהיא אלכסונית) כך: $P = UDU^T$. כך ש־U אותוגונלית שעמודותיה הם הווקטורים העצמיים של A ו־D אלכסונית שערכיה הם הערכים העצמיים של P. הוכחנו כי הערכים של 0 או 1 והווקטורים העצמיים מתאימים

לערך עצמי 1, לכן $D = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$ ו־ $U = \begin{bmatrix} \vdots & & \vdots \\ v_1 & \cdots & v_k \\ \vdots & & \vdots \end{bmatrix}$

$$P^2 = \left(UDU^T\right)^2 = UDU^TUDU^T = UD^2U^T$$

$D^2 = D$ משום שזוהי מטריצה אלכסונית עם ערכים של 1 על חלק מהאלכסון, לכן העלאה בריבוע לא משנה את ערך המטריצה.
לכן קיבלנו

$$P^2 = UDU^T = P$$

ד. הוכיחו כי $(I-P)\,P = 0$

מהסעיף הקודם

$$\begin{aligned} P^2 &= P \\ 0 &= P - P^2 \\ 0 &= (I-P)\,P \end{aligned}$$

2.3 Least Square

בהינתן דגימה $S = \{(x_i, y_i)\}_{i=1}^m$, כלל ERM לרגרסיה לינארית squared loss w.r.t הוא

$$\hat{w} \in argmin_{w\in\mathbb{R}^d} \|Xw - y\|^2$$

כאשר X היא מטריצת הפיצ'רים של הרגרסיה הלינארית עם שורות כדגימות וy הוא ווקטור responses. יהי $X = U\Sigma V^T$ ה־SVD של X, כאשר U הוא $m \times m$ מטריצה אורתוגונלית, Σ הוא σ_i־ות לא שוות לאפס הערכים הסינגולרים של X. היזכרו כי הפסאודו־אינברס של X מוגדר להיות $X^\dagger = V\Sigma^\dagger U^T$ כאשר $\Sigma^\dagger$ היא מטריצה אלכסונית $d \times m$, כך ש:

$$\Sigma^\dagger_{i,j} = \begin{cases} \sigma_i^{-1} & \sigma_i \neq 0 \\ 0 & \sigma_i = 0 \end{cases}$$

6. הראו שאם X^TX הפיך, הפתרון הכללי שהסקנו בתרגול שווה לפתרון שראינו בכיתה. לחלק הזה, הניחו כי X^TX הפיך.

צריך להוכיח שבמקרה בו X^TX הפיך, מתקיים:

$$X^{\dagger}=\left[X^TX\right]^{-1}X^T$$

(המטריצה שלמדנו בתרגול

מתקיים:

$$\begin{aligned}X^TX &\underset{SVD}{=}\left(U\Sigma V^T\right)\left(U\Sigma V^T\right)^T=U\Sigma V^TV\Sigma^TU^T\\ &\underset{\clubsuit}{=}U\Sigma\Sigma^TU^T\underset{\Sigma\Sigma^T=D}{=}UDU^T\end{aligned}$$

$\clubsuit$ - במטריצה אורתוגונלית $V^TV=I$.

משום שמתקיים:

$$X^TX\left(UD^{-1}U^T\right)=UDU^TUD^{-1}U^T=I$$

אזי $\left(X^TX\right)^{-1}=UD^{-1}U^T$.

ולכן:

$$\begin{aligned}\left(X^TX\right)^{-1}X &= UD^{-1}U^TU\Sigma V^T=UD^{-1}\Sigma V^T\\ &=U\Sigma^{\dagger T}V^T=\left(X^{\dagger}\right)^T\end{aligned}$$

לכן

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

7. הראו כי X^TX הפיך אמ"מ $\mathrm{span}\left\{x_1,...,x_n\right\}=\mathbb{R}^d$.

יהי X נפרש ע"י $\{x_1,...,x_n\}$. אנו יודעים כי $Ker\left(X\right)=Ker\left(X^TX\right)$. לפי פירוק SVD גם הדרגה של X שווה לדרגה של X^TX. לכן $\{x_1,...,x_n\}$ פורש את $\mathbb{R}^d$ אמ"מ

הדרגה של X^TX שווה ל-d. משום ש X^TX זוהי מטריצה $d\times d$ (כי X ריבועית מדרגה d) והפיכה, הדרגה שלה היא d ולכן $\mathrm{span}\left\{x_1,...,x_n\right\}=\mathbb{R}^d$.

8. היזכרו כי אם X^TX לא הפיך יש לו הרבה פתרונות. הראו כי $\hat{w} = X^\dagger y$ הוא פתרון כאשר נורמה L_2 היא מינימלית. אם כך, הראו שבשביל כל פתרון אחר $\overline{w}$, $\|\hat{w}\| \leq \|\overline{w}\|$.

לפי פירוק SVD, $X = U\Sigma V^T$. נסמן r התמונה של X. ונסמן:

$$V = \begin{bmatrix} V_r & V_2 \end{bmatrix}, U = \begin{bmatrix} U_r & U_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma' & 0 \\ 0 & 0 \end{bmatrix}$$

כאשר V_r ו־U_r מייצגים את r הטורים הראשונים, ז"א $1 \leq r \leq d$, ו־Σ' מייצג את האלכסון עם הערכים העצמיים החיוביים במטריצה.

מהגדרת המקרה הסינגולרי, $\hat{w} = X^\dagger y$ אומר כי $\dim(\ker(X)) \neq 0$ משום שהדרגות של X ו־X^TX שוות למספר הערכים הסינגולרים של המטריצות הללו.

כל הפתרונות האפשריים ל $X^\dagger y$ חייבים להתאים לערכים הסינגולרים הללו, לכן עבור $1 \leq i \leq r$ מתקיים ש$\hat{w} = \overline{w}$. אולם לכל $i > r$ נקבל כי $\hat{w} = 0$. כל פתרון אחר $\overline{w}$ בהכרח מקיים $0 = \hat{w} \leq \overline{w}$. מכאן

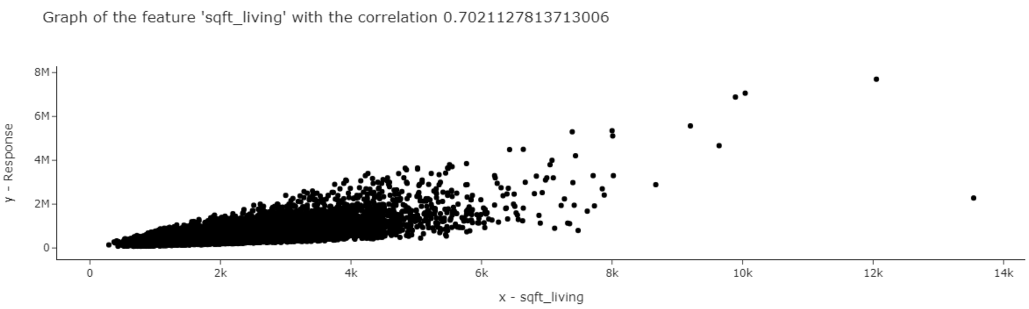
$$\begin{aligned} \|\hat{w}\|^2 &= \sum_{i=1}^{d} \hat{w}_i^2 = \sum_{i=1}^{r} \hat{w_i^2} + \sum_{i=r+1}^{d} \hat{w}_i^2 \\ &= \sum_{i=1}^{r} \overline{w_i}^2 + \sum_{i=r+1}^{d} 0 \leq \sum_{i=1}^{r} \overline{w_i}^2 \\ &= \|\overline{w}\|^2 \end{aligned}$$

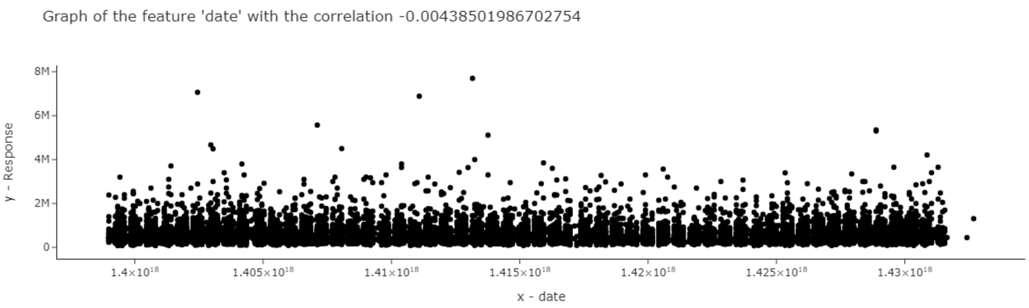
קיבלנו $\|\hat{w}\| \leq \|\overline{w}\|$.

חלק מעשי

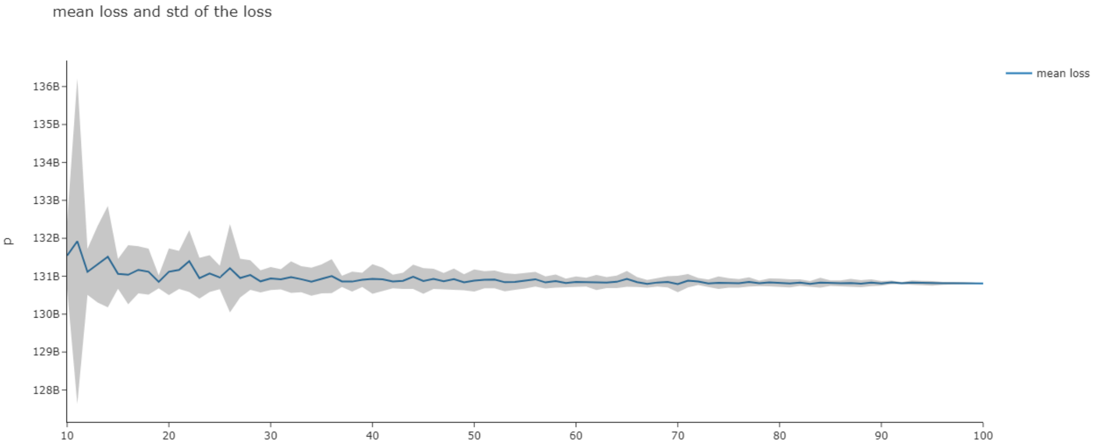
מודל רגרסיה לינארית

1. הורדתי את הפיצ'ר zip_code משום שהוא לא לינארי וחלוקה קטגורית לא תהיה יעילה (צריך לחלק כל מיקוד בנפרד).
שמתי את הפיצ'ר "גודל מרתף" כקטגורי משום שהוא לא לינארי. יש בתים בהם אין מרתף ולכן מסומן כגודל 0. לכן חילקתי עמודה זה לקטגוריה בינארית "האם יש מרתף" ולקטגוריות "מרתף בגודל עד 500", "מרתף בגודל 500 עד 1000", "מרתף בגודל יותר מ1000".
מחקתי מחירים שליליים הוא 0, משום שהם לא מייצגים ולא יועילו לחיזוי.
2.





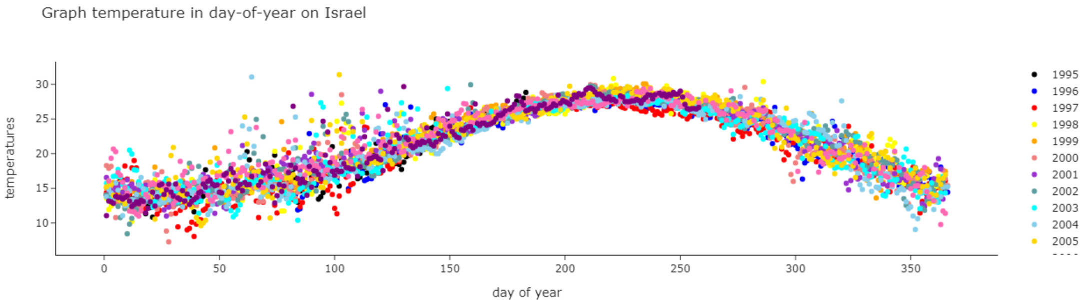
הגרף הראשון הוא גודל הבית והגרף השני הוא תאריך קניית הבית. ניתן לראות כי הגרף הראשון לינארי יותר, ולכן משמעותי יותר לחיזוי מחיר הבית, כי גודל הבית משפיע על מחיר הבית. אנו רואים כי הגרף של תאריך הקניה לא לינארי. ולכן ניתן להסיק כי תאריך קניית הבית אינו משמעותי לחיזוי מחיר הבית.



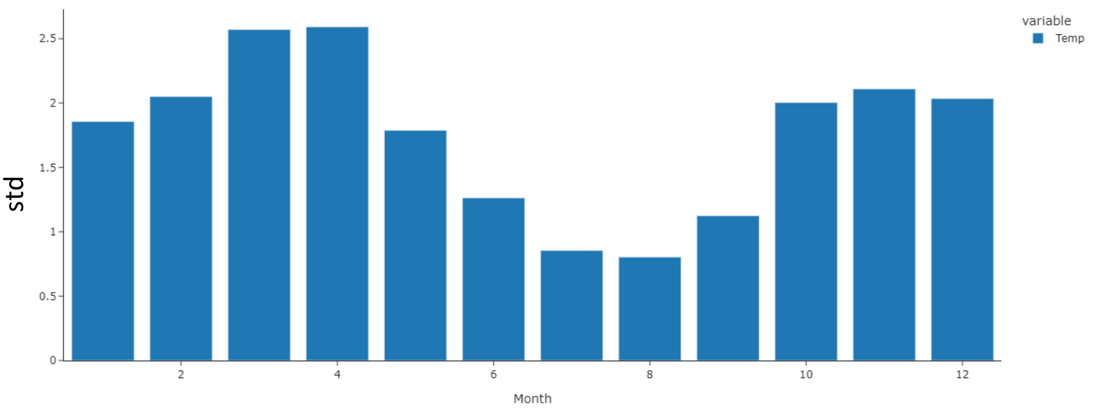
.4
אנו רואים לפי הגרף שככל שיש יותר דגימות, כך מסיקים יותר טוב וסטיית התקן קטנה יותר.

התאמה פולינומיאלית

. 2

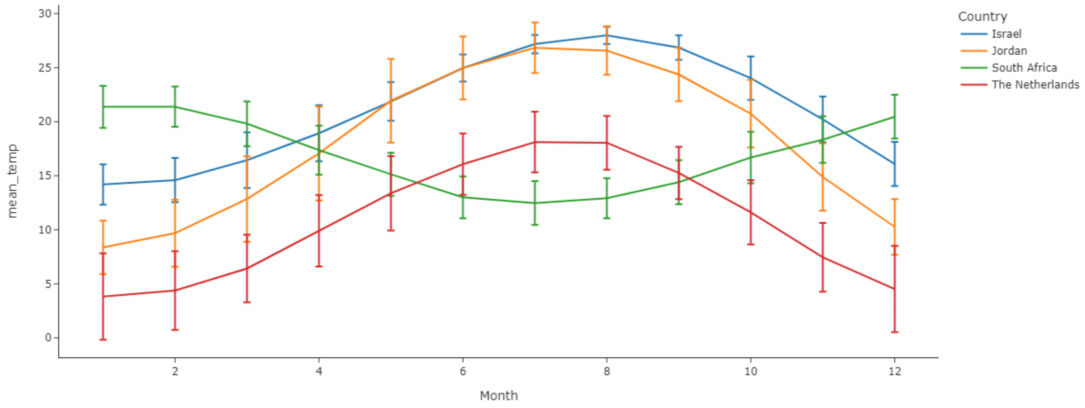


פולינום מדרגה 3 יכול להתאים למידע הנ"ל.



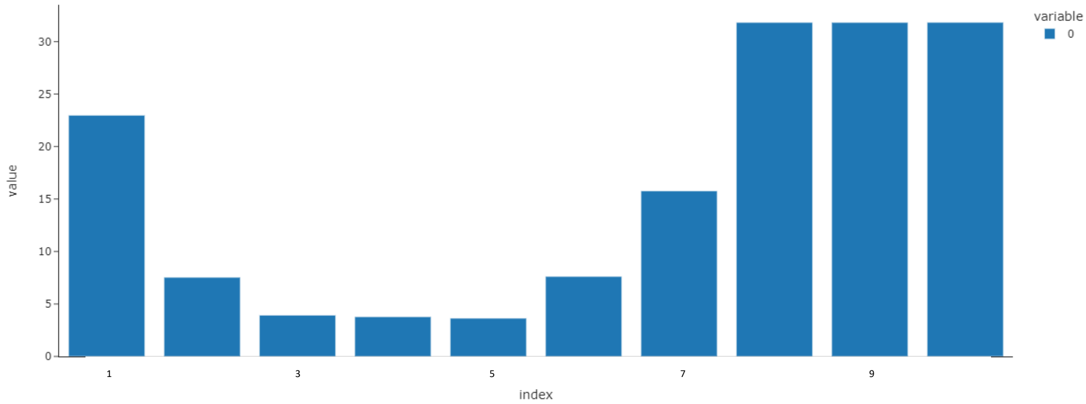
ניתן לשים לב כי בחודשים בהם יש עונות מעבר, החיזוי פחות טוב -הסטית תקן יותר גדולה, משום שהטמפרטורה בחודשים אלו פחות קבועה.

. 3



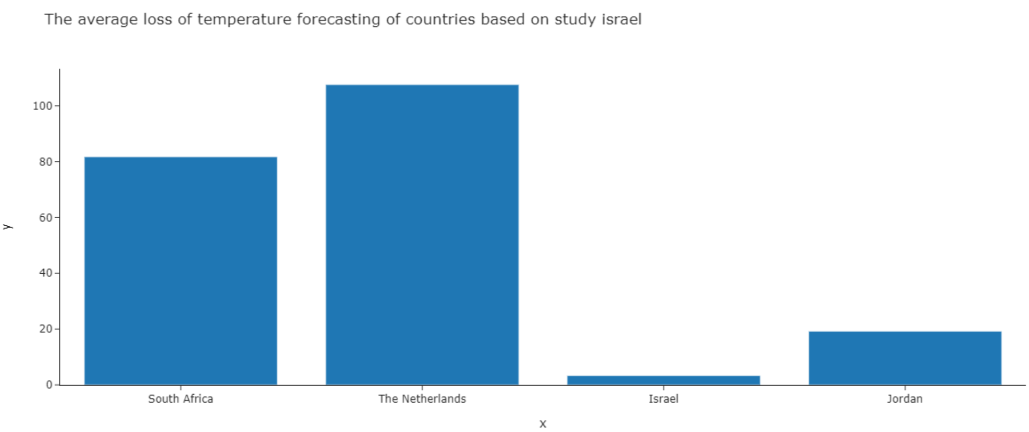
לא לכל המדינות יש את אותה תבנית. ניתן לשים לב כי לישראל וירדן יש דגימות קרובות, בשל כך שאלו מדינות קרובות, ולמדינות אחרות הדגימות שונות יותר.

. 4



נראה של-k=5 זהו ההתאמה הטובה יותר בשל כך שיש את השגיאה הכי קטנה.

. 5



ניתן לראות כי השגיאה הגדולה ביותר נמצאת ב Netherlands וב South Africa וניתן לראות בגרף בשאלה 3 שאכן גרפים אלו פחות קרובים ודומים לישראל. השגיאה בירדן קטנה משום שלפי גרף 3 הטמפרטורה שלה קרובה לטמפרטורה בישראל.