מבוא למערכות לומדות (67577) | תרגיל 2

שם: אלון בינר | ת"ז: 318418118
April 29, 2023

שאלה 1

(1) <u>פיתרון:</u>

יהא $X \in Mat_{m\times d}$.
צריך להראות ש-$Ker\left(X\right) = Ker\left(X^TX\right)$.
יהא $v \in Ker\left(X\right) = \{x\in\mathbb{R}^m \mid Xv = 0_{\mathbb{R}^m}\}$. אזי $Xv = 0_{\mathbb{R}^m} \in \mathbb{R}^m$.
$X^T \in Mat_{d\times m}$. אזי $\left(X^TX\right)v = X^T\left(Xv\right) = X^T0_{\mathbb{R}^m} = 0_{\mathbb{R}^d}$. לכן $v \in Ker\left(X^TX\right)$.

מצד שני, אם $v \in Ker\left(X^TX\right) \subset \mathbb{R}^d$ אז $X^TXv = 0_{\mathbb{R}^m}$. נראה ש-$v \in Ker\left(X\right) \subset \mathbb{R}^m$. כלומר $Xv = 0_{\mathbb{R}^d}$. כלומר $\left\langle X^Tv, X^Tv\right\rangle = 0$.
מתקיים כי

$$X^TXv = 0_{\mathbb{R}^m}, v\in\mathbb{R}^d \Rightarrow v^TX^TXv = 0_{\mathbb{R}^d} \Rightarrow \left(Xv\right)^T\left(Xv\right) = 0_{\mathbb{R}^d} \Rightarrow \langle Xv, Xv\rangle = 0 \Rightarrow Xv = 0 \Rightarrow$$

$$\Rightarrow v \in Ker\left(X\right)$$

(2) <u>פיתרון:</u>

תהא $A \in \mathbb{R}^{n\times n}$. נראה ש-$Im\left(A^T\right) = Ker\left(A\right)^{\perp}$.
$Im\left(A^T\right) = \left\{x\in\mathbb{R}^n \mid \exists w\in\mathbb{R}^n \text{ s.t. } x = A^Tw\right\}$.
$Ker\left(A\right)^{\perp} = \{x\in\mathbb{R}^n \mid \langle x,v\rangle = 0, \forall v \in Ker\left(A\right)\} = \{x\in\mathbb{R}^n \mid \forall v\in V, Av = 0, \langle x,v\rangle = 0\}$
יהא $x \in Im\left(A^T\right)$. יהא $v \in Ker\left(A\right)$. אזי $Av = 0$. לכן קיים $w\in\mathbb{R}^n$ כך ש:

$$\langle x,v\rangle = \left\langle A^Tw, v\right\rangle = \left(A^Tw\right)^Tv = w^TAv = w^T\cdot 0 = 0$$

לכן $x \in Ker\left(A\right)^{\perp}$.
מצד שני, נניח ש-$x \notin Im\left(A\right)$. נראה ש-$x \notin Ker\left(A\right)$.
קיים $v \in Im\left(A\right)^T$ כך ש-v אורתוגונלי לכל וקטור ב-$Im\left(A\right)$.
בפרט $\left\langle v, AA^Tv\right\rangle = 0$ לכן:

$$0 = \left\langle v, AA^Tv\right\rangle = v^TAA^Tv = \left(A^Tv\right)^TA^Tv = \left\langle A^Tv, A^Tv\right\rangle = \parallel A^Tv \parallel$$

אם ורק אם $A^Tv = 0$ לכן $v \in Ker\left(A^T\right)$ לכן $x \notin Ker\left(A^T\right)$.

(3) <u>פיתרון:</u>

תהא $y = Xw$ מערכת משוואות לא הומוגנית עם X לא הפיכה.
אזי יש למערכת המשוואות 0 או אינסוף פתרונות.
אם יש למערכת אינסוף פתרונות אז $y \in Im\left(X\right)$. אחרת, למערכת יש אפס פיתרונות, ואז לא קיים $w\in\mathbb{R}^n$ עבורו $y = Xw$. כלומר $y \notin Im\left(X\right)$.

לכן למערכת יש אינסוף פתרונות אם"ם $y \in Im\left(X\right)$.
משאלה 2 נובע ש-$y \in Im\left(X\right)$ אם"ם $y \in Ker\left(X\right)^{\perp}$.
כלומר $y \perp Ker\left(X\right)$.

(4) <u>פיתרון:</u>

תהא $X^{T}Xw = X^{T}y$ מערכת משוואות נורמלית.
נניח ש-$X^{T}X$ הפיכה. אזי:

$$\left(X^{T}X\right)^{-1}X^{T}Xw = \left(X^{T}X\right)^{-1}Xy$$

$$\Updownarrow$$

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

ומיחידות מטריצה הפוכה, נובע שזה הפיתרון היחיד של המערכת.
כעת נניח ש-$X^{T}X$ לא הפיכה.
מהסעיף הקודם נובע שלמערכת יש אינסוף פתרונות אם ורק אם $X^{T}y \perp Ker\left(X^{T}X\right)$.
לפי סעיף (1), $Ker\left(X^{T}X\right) = Ker\left(X\right)$.
נובע שצריך להראות למערכת יש אינסוף פתרונות אם ורק אם $X^{T}y \perp Ker\left(X\right)$ (למה צריך להראות את זה וזה לא מתקיים מיד?!)
אם $u \in Ker\left(X\right)$ אז $\left\langle u, X^{T}y\right\rangle = u^{T}X^{T}y = \left(Xu\right)^{T}y = \left\langle Xu, y\right\rangle = 0$.
לכן למערכת יש אינסוף פתרונות.

(5) <u>פיתרון:</u>

יהא $V \subset \mathbb{R}^{n}$ ממימד k.
יהא $\mathcal{B} = \left(v_{i}\right)_{i=1}^{k}$ בסיס א"נ של V.
תהא $P = \sum_{i=1}^{k} v_{i}v_{i}^{T}$ מטריצת הטלה אורתוגונלית.

(1) צריך להראות שלכל $j,l \in [n]$ מתקיים ש-$P_{j,l} = P_{l,j}$.
יהיו $j,l \in [n]$. אזי ע"פ הגדרת P:

$$P = \sum_{i=1}^{k} v_{i}v_{i}^{T} = \sum_{i=1}^{k}\left(\left[\begin{array}{c} v_{1,i}\\ \dots\\ v_{n,i}\end{array}\right]\left[\begin{array}{ccc} v_{1,i} & \dots & v_{n,i}\end{array}\right]\right) = \sum_{i=1}^{k}\left(\left[\begin{array}{ccc} v_{1,i}\cdot v_{1,i} & \dots & v_{1,i}\cdot v_{n,i}\\ \dots & \dots & \dots\\ v_{n,i}\cdot v_{1,i} & \dots & v_{n,i}\cdot v_{n,i}\end{array}\right]\right)$$

מכאן ש-$\left(P\right)_{j,l} = \sum_{i=1}^{k} v_{j,i}\cdot v_{l,i} = \sum_{i=1}^{k} v_{l,i}\cdot v_{j,i} = \left(P\right)_{l,j}$.
לכן P סימטרית.
(2) יהא $m \in [k]$.
נחשב את Pv_{m}.

$$Pv_{m} = \sum_{i=1}^{k} v_{i}v_{i}^{T}v_{m}$$

מאורתונורמליות $\mathcal{B}$ נובע שלכל $j,l \in [n]$, אם $j \neq l$ אז $\left\langle v_{j}, v_{l}\right\rangle = 0$ כלומר $v_{j}^{T}v_{l} = 0$.
אחרת, $\left\langle v_{j}, v_{l}\right\rangle = 1$ כלומר $v_{j}^{T}v_{l} = 1$. לכן $P_{j,l} = \delta_{j,l}$ כש-δ מוגדרת להיות הדלתא של קרונקר.
לכן:

$$\sum_{i=1}^{k} v_{i}v_{i}^{T}v_{m} = \sum_{i=1}^{k} v_{i}\delta_{i,m} = v_{m}$$

כשהשוויון האחרון נובע מהגדרת $\delta_{i,m}$.

(3) יהא $v \in V$. אזי קיימים קבועים $a_1, a_2, ..., a_k \in \mathbb{R}$ כך ש-$v = \sum_{i=1}^{k} a_i v_i$ לכן:

$$Pv = P\left(\sum_{i=1}^{k} a_i v_i\right) \overset{(1)}{=} \sum_{i=1}^{k} P\left(a_i v_i\right) = \sum_{i=1}^{k} a_i P v_i \overset{(2)}{=} \sum_{i=1}^{k} a_i v_i = v$$

שוויון (1) נובע מדיסטריבוטיביות בכפל מטריצות.
שוויון (2) נובע מהטענה שהוכחנו בסעיף (2).

(4) צריך להראות ש-$P^2 = P$.
מתקיים כי:

$$P^2 = \left(\sum_{i=1}^{k} v_i v_i^T\right)^2 = \left(\sum_{i=1}^{k} v_i v_i^T\right)\left(\sum_{i=1}^{k} v_i v_i^T\right) = \sum_{i=1}^{k}\left(\sum_{j=1}^{k} v_j v_j^T\right) v_i v_i^T = \sum_{i=1}^{k}\left(\sum_{j=1}^{k} v_j v_j^T v_i\right) v_i^T =$$

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

(5) צריך להראות ש-$(I-P)\,P = 0$.
מתקיים:

$$(I-P)\,P = P - P^2 \overset{(1)}{=} P^2 - P^2 = 0$$

שוויון (1) נובע מהסעיף הקודם.

(6) <u>פיתרון:</u>

תהא $X \in \mathbb{R}^{m\times(d+1)}$ ויהא $y \in \mathbb{R}^m$.
נניח ש-$X^T X$ הפיכה.
צריך להראות ש-$\left(X^T X\right)^{-1} X^T y = X^\dagger y$.
יהא $X = U\Sigma V^T$ פירוק SVD קומפקטי של X.
אזי:

$$\left(X^T X\right)^{-1} X^T y = \left(\left(U\Sigma V^T\right)^T \left(U\Sigma V^T\right)\right)^{-1} \left(U\Sigma V^T\right)^T y =$$

$$= \left(\left((U\Sigma)\,V^T\right)^T \left(U\Sigma V^T\right)\right)^{-1} \left((U\Sigma)\,V^T\right)^T y = \left(\left(V\,(U\Sigma)^T\right)\left(U\Sigma V^T\right)\right)^{-1} \left(V\,(U\Sigma)^T\right) y =$$

$$= \left(\left(V\Sigma^T U^T\right)\left(U\Sigma V^T\right)\right)^{-1} \left(V\Sigma^T U^T\right) y = \left(V\Sigma^T U^T U\Sigma V^T\right)^{-1} \left(V\Sigma^T U^T\right) y \overset{(1)}{=}$$

$$\overset{(1)}{=} \left(V\Sigma^T\Sigma V^T\right)^{-1} \left(V\Sigma^T U^T\right) y = \left(V\Sigma^T\Sigma V^T\right)^{-1} V\Sigma^T U^T y =$$

$$= V\left(\Sigma^T\Sigma\right)^{-1} V^T V\Sigma^T U^T y = V\left(\Sigma^T\Sigma\right)^{-1} \Sigma^T U^T y$$

$\left(\Sigma^T\Sigma\right)$ זו מטריצה ריבועית כך שלכל $i \in [d+1]$ מתקיים $\left(\left(\Sigma^T\Sigma\right)\right)_{i,j} = \mathbb{1}_{i=j}\cdot\sigma_i^2$ והערכים על האלכסון מסודרים בסדר יורד. בנוסף, העמודות בלתי תלויות לינארית. והמטריצה הפיכה, לכן $\left(\left(\Sigma^T\Sigma\right)\right)_{i,j}^{-1} = \mathbb{1}_{i=j}\cdot\frac{1}{\sigma_i^2}$.
כמו כן,

$$\left(\left(\Sigma^T\Sigma\right)^{-1}\Sigma^T\right)_{i,j} = \mathbb{1}_{i=j}\cdot\frac{1}{\sigma_i^2}\cdot\sigma_i = \mathbb{1}_{i=j}\cdot\frac{1}{\sigma_i} = \Sigma_{i,j}^\dagger$$

לכן $\left(\Sigma^T\Sigma\right)^{-1}\Sigma^T = \Sigma^\dagger$ ולכן $\left(X^T X\right)^{-1} X^T y = V\Sigma^\dagger U^T y = X^\dagger y$.

(7) פיתרון:

נראה ש-X^TX הפיכה אם"ם $Span\left\{x_1,...,x_m\right\}=\mathbb{R}^d$.

נניח ש-X^TX הפיכה. אם ורק אם $dim\left(Ker\left(X^TX\right)\right)=0$. כמו כן, הוכחנו ש-$Ker\left(X^TX\right)=Ker\left(X\right)$. אם ורק אם $dim\left(Ker\left(X\right)\right)=0$.
לכן למטריצה X יש דרגה מלאה. מכיוון שאנו מניחים ש-$m\geq d+1$ אז ע"פ הגדרת דרגה, מספר השורות הבלתי-תלויות המקסימלי הוא $max\left(m,d\right)=d$.
כלומר $Row\left(X\right)=Span\left\{x_1,...,x_m\right\}=\mathbb{R}^d$.

(8) פיתרון:

אם X^TX הפיכה, אז יש למשוואה הנורמלית פיתרון יחיד והוא המינימלי.
אחרת, X^TX לא הפיכה.
יהא $\hat{w}=X^\dagger y$ פיתרון שממזער את $RSS\left(w\right)$.
יהא $\overline{w}$ פיתרון אחר לבעיית הרגרסיה. נראה ש-$\parallel\hat{w}\parallel\leq\parallel\overline{w}\parallel$.
יהא $X=U\Sigma V^T$ פירוק SVD של X. נסמן $rank\left(X\right)$.
נכתוב:

$$V=\left[\begin{array}{cc}V_1 & V_2\end{array}\right],U=\left[\begin{array}{cc}U_1 & U_2\end{array}\right],\Sigma=\left[\begin{array}{cc}\Sigma_1 & 0\\ 0 & 0\end{array}\right]$$

V_1 ו-U_1 הן r העמודות הראשונות של המטריצות V ו-U בהתאמה. V_2 ו-U_2 הן שאר העמודות של V ו-U בהתאמה.
Σ_1 היא מטריצה אלכסונית שאיברי האלכסון הראשי שלה שונים מ-0.
בהינתן w נגדיר $b=U^Tw,b_1=U_1^Tw,b_2=U_2^Tw$.
לכן $b=\left[\begin{array}{c}b_1\\ b_2\end{array}\right]$.
U אורתוגונלית לכן איזומטרית לכן $\parallel b\parallel=\parallel w\parallel$.
לכן מספיק להראות שהנורמה של b מינימלית.
V גם איזומטרית, לכן:

$$\parallel y-X^Tw\parallel^2=\parallel y-V\Sigma U^Tw\parallel^2=\parallel V\left(V^Ty-\Sigma b\right)\parallel^2=\parallel V^Ty-\Sigma b\parallel^2=$$

$$=\parallel\left[\begin{array}{cc}V_1 & V_2\end{array}\right]^Ty-\left[\begin{array}{cc}\Sigma_1 & 0\\ 0 & 0\end{array}\right]\left[\begin{array}{c}b_1\\ b_2\end{array}\right]\parallel^2=\parallel V_1^Ty-\Sigma_1b_1\parallel+\parallel V_2^Ty\parallel^2$$

כדי לקבל ממזער על $\parallel y-Xw\parallel$ נצטרך לקבל:

$$\Sigma_1b_1-V_1^T=0\iff\Sigma_1b_1=V_1^Ty\iff b_1=\Sigma_1^{-1}V_1^Ty$$

וגם ש-$b_2=0$.
נשים לב שעבור $\hat{w}=X^{T\dagger}y$ מתקיים:

$$U_1^T\hat{w}=U_1^TX^{T\dagger}y=U_1^TU_1\Sigma_1^{-1}V_1^Ty=\Sigma_1^{-1}V_1^Ty$$

$$U_2^T\hat{w}=U_2^TX^{T\dagger}y=U_1^TU_1\Sigma_1^{-1}V_1^Ty=0$$

לכל פיתרון אחר $\overline{w}$, אנו מקבלים שהתנאי הראשון מתקיים אבל לא בהכרח השני. לכן $\parallel\hat{w}\parallel\leq\parallel\overline{w}\parallel$.

$$\parallel\hat{w}\parallel=\parallel X^\dagger y\parallel=$$

חלק פרקטי

(2) פיתרון:

(1) ה-features שנשמרו היו מה שלא "id", "date", "lat", "long", "sqft_living15, "sqft_lot15". כי ה-features האלה יש השפעה מכרעת על מחירי הדיור.
כל שאר ה-features לא נשמרו משום שההשפעה שלהם על מחיר הדיור עבור לקוח יחיד הייתה זניחה או שלא הייתה קיימת כלל.

(2) ה-date וה-zipcode היו features קטגוריים. את ה-date הורדנו מהטבלה, ואל ה-zipcode הוספנו עמודות שממפות את הערכים שלו לאוסף של features בינאריים שמייצגים את ערכי ה-zipcode.

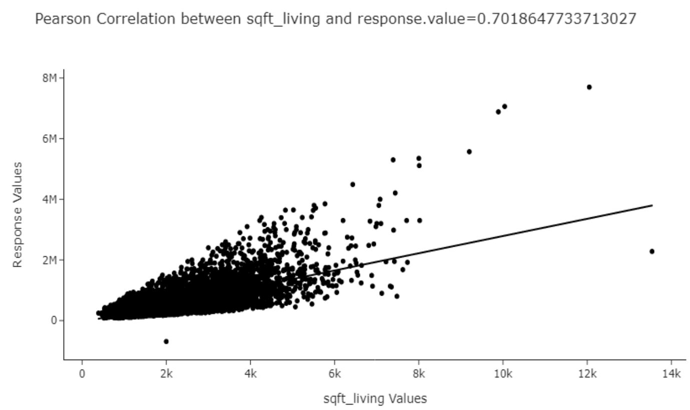
(3) לא הוספתי features נוספים.

(4) מחקתי דגימות בהן הופיעו ערכים לא תקינים או לא קיימים.

(5) בדקתי אם הקלט של features היה תקין. לדוגמא, בדקתי אם מספר חדרי השינה ומספר חדרי השירותים חיובי. בנוסף, וידאתי שהציון של דירה הוא מספר שלם בין 1 ל-13 כולל. כמו כן, בדקתי ששאר ה-features מקבלים ערכים אי-שליליים.

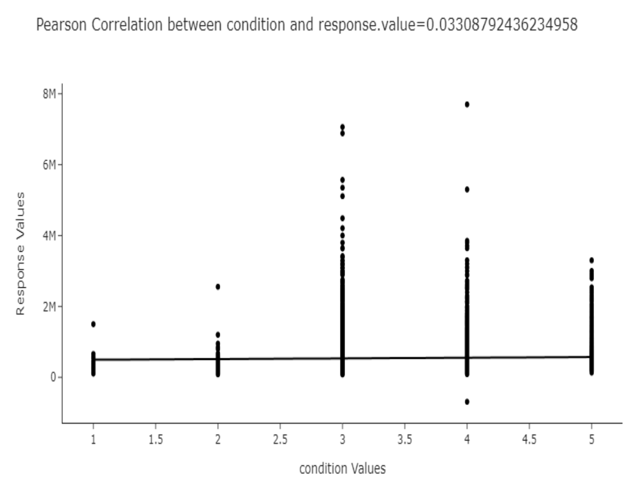
(3) __פיתרון:__

feature אחד אותו בחרנו להציג הוא sqft_living כי יש התאמה גבוהה בין הערכים שלו לערכי ה-set response.
ניתן לראות זאת בתמונה :

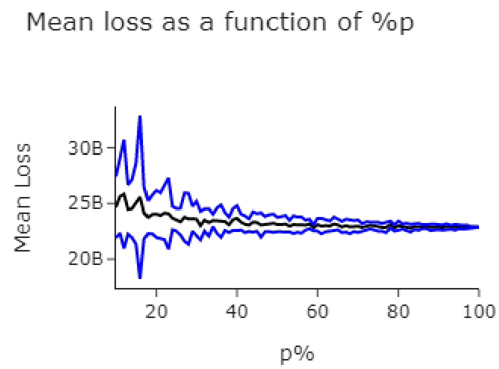


מדד הקורלציה של פירסון הוא 0.7 בערך. ככל שמדד כזה עבור feature כלשהו קרוב ל-1, אז הוא מידת ההשפעה שלו על התוצאה יותר גדולה. מדד הקורלציה של פירסון עבור sqft_living הוא 0.7 וקרוב ל-1 ביחס למדדי פירסון של שאר ה-featureים. לכן ה-feature אותו המדד מייצג משפיע במידה רבה יחסית על הערכים של ה-set response, כלומר על מחירי הבתים.

feature אחר אותו בחרתי להציג הוא condition, משום שישנה השפעה נמוכה של ה-featureים על תוצאות של ה-set response. מדד הקורלציה של פירסון הוא 0.02 וככל שמדד הקורלציה קרוב יותר ל-0 אז השפעת ערך ה-feature על תוצאות ה-set response קטנה. להלן הגרף שמייצג את הקורלציה :



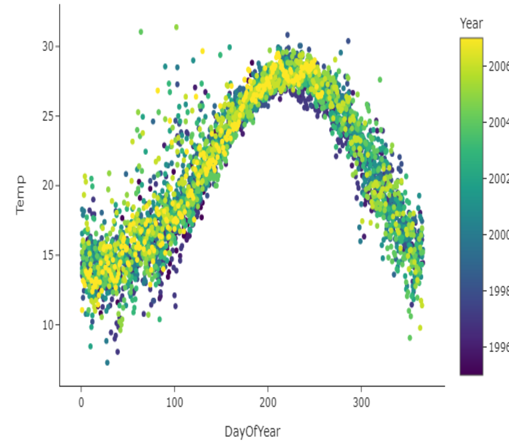
(4) <u>פיתרון:</u>



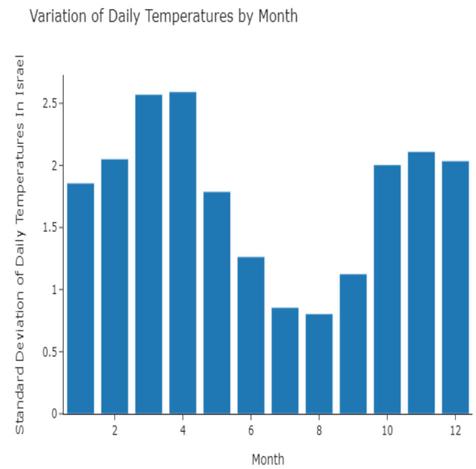
כשמגדילים את ה-training set אז השגיאה והשונות בתחזית קטנים.

חלק פרקטי 2

(2) <u>פיתרון:</u>



פולינום שמתאים לנתונים הללו הוא פולינום מדרגה 3.



בחלק מהחודשים המודל יצליח יותר ובחלק אחור יצליח פחות.
עבור ערכי שונות גבוהים יותר (שניתן לראות בעיקר בחודשים 3-4, 10-12) ההערכה תהיה פחות מדויקת.
לעומת זאת, עבור ערכי שונות נמוכים יותר (חודשים 6-9) המודל יהיה יותר קרוב לנקודות.

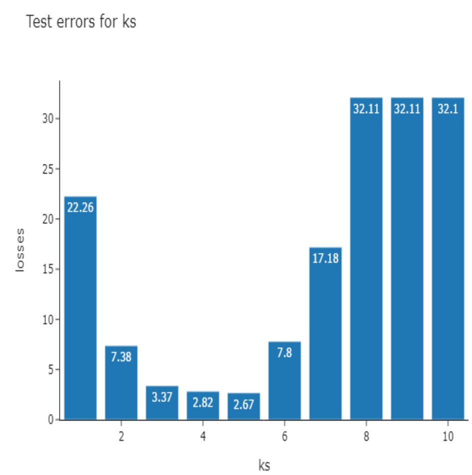
(3) <u>פיתרון:</u>

ניתן לראות שיש מגמה דומה עבור ישראל, ירדן והולנד. ישנה עלייה קמורה מינואר ועד לאפריל. לאחר מכן היא הופכת לקעורה ומגיעה לשיא בסביבות חודשים יולי ואוגוסט. ואז יש ירידה עד דצמבר.

לעומת זאת, בדרום אפריקה ישנה מגמה הפוכה לחלוטין - היא מתחילה בירידה קעורה עד לאפריל, שהופכת לקמורה ומגיעה למינימום סביב חודש אוגוסט. לאחר מכן, ישנה עלייה עד דצמבר.

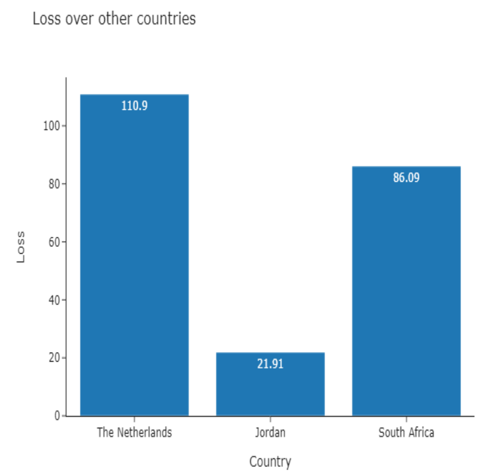
המודל יתאים עבור ירדן והולנד, אבל לא עבור דרום אפריקה בגלל הצורה של הגרפים.

(4) <u>פיתרון:</u>



ניתן לראות מהגרף שה-k שמשיג את הערך הכי נמוך הוא 5.
היה אפשר לחשוב על ערך אחר אם הפיצול הרנדומלי היה אחרת.

(5) פיתרון:



התוצאות של המודל עבור ישראל עם set training של ישראל היה יותר טוב מאשר עבור מדינות אחרות.
השגיאה הכי נמוכה מבין שאר המדינות היא אצל ירדן, עם שגיאה של 21.91. אחריה מגיעות דרום אפריקה והולנד עם שגיאות של 86.09 ו-110.9 בהתאמה. להן המודל פחות מתאים מאשר ירדן.