Exercise 2 IML

Eyal Perets 209541903

1. Prove that: $Ker(\mathbf{X}) = Ker(\mathbf{X}^\top\mathbf{X})$

1. נזכיר v שייך לגרעין של X אומר $X\cdot r=0$,

יהי V וקטור שונה מאפס השייך לגרעין X, כלומר:

$$X\cdot v=0 \implies X^T X\cdot v=0 \implies v\in ker(X^TX)$$

כלומר לכל $v\in ker(X)$ מתקיים $v\in ker(X^TX)$

כעת נוכיח את הכיוון השני: יהי $u\in ker(X^TX)$, אזי:

$$X^TXu=0 \implies u\cdot X^TX\cdot u=0 \implies u^T\|X\|^2\cdot u=0$$

$$\implies \|Xu\|^2=0 \implies X_u=0$$

כלומר $u\in ker(X)$ לכל $u\in ker(X^TX)$,

ולכן $ker(X)=ker(X^TX)$

2. Prove that for a square matrix A: $Im(A^\top) = Ker(A)^\perp$

2. יהי $V\in\mathbb{R}^m$, וקיים $b\in Im(A^T)$ כך ש-

$$A^Tv=b$$

נוכיח הכלה בשני כיוונים,

$Im(A^T)\subset ker(A)^\perp$:

יהי $u\in ker(A) \implies A\cdot u=0 \implies \langle Au,v\rangle=0$

$$\langle u,A^Tv\rangle=0 \implies \langle u,b\rangle=0$$

כלומר לכל $b\in Im(A^T)$ מתקיים $b\in ker(A)^\perp$

:$\mathrm{Im}(A^T) \supset \ker(A)^{\perp}$

נניח ש $b \notin \mathrm{Im}(A^T)$ אזי $b \in \mathrm{Im}(A^T)^{\perp}$

יהי $x \in \mathrm{Im}(A^T)^{\perp}$ כך ש $\langle b, x \rangle \neq 0$, ואז ש x אורתוגונלי.

לכן נקבל b כולו ב $\mathrm{Im}(A^T)$ אזי | שייך

$$\|A \cdot c\|^2 = \langle Ac, Ac \rangle = \langle c, A^T A c \rangle = 0$$

לפי הנתון מקיים, כלומר $\|Ac\|^2 = 0$ ולכן $Ac = 0$, $c \in \ker(A)$

וקיבלנו שאם $b \notin \mathrm{Im}(A^T)$ אזי $b \notin \ker(A)^{\perp}$, כנדרש ולכן

$$\ker(A)^{\perp} \subset \mathrm{Im}(A^T)$$

ולסיכום $\mathrm{Im}(A^T) = \ker(A)^{\perp}$

3. Let $\mathbf{y} = \mathbf{X}\mathbf{w}$ be a non-homogeneous system of linear equations. Assume that $\mathbf{X}$ is square and not invertible. Show that the system has ∞ solutions $\Leftrightarrow y \perp Ker(\mathbf{X}^{\top})$.

3. X לא הפיכה, כלומר $\det(X) = 0$ ולמערכת משוואות

יש אינסוף פתרונות, או אין פתרון כלל.

למערכת משוואות הזו יש אינסוף פתרונות אמ"מ

$y \in \mathrm{Im}(X)$, ולפי סעיף 2 ב3 קודם אזי $y \in \ker(X^T)^{\perp}$

כלומר $y \perp \ker(X^T)$.

4. Consider the (normal) linear system $\mathbf{X}^\top\mathbf{X}\mathbf{w} = \mathbf{X}^\top\mathbf{y}$. Using what you have proved above prove that the normal equations can only have a unique solution (if $\mathbf{X}^\top\mathbf{X}$ is invertible) or infinitely many solutions (otherwise).

4.

לכן: המסקנה היחידה הינה שאם X^TX לא הפיך:

אזי: למערכת אינסוף משוואות אמ"מ $X^+y \perp ker(X^TX)$

לפי סעיף 1 $ker(X^TX) = ker(X)$ ולכן $\Updownarrow$ $X^Ty \perp ker(X)$.

כי: $v \in ker(X)$ אזי $\langle v, X^+y\rangle = \langle Xv, y\rangle = 0$ ($y \in Im(X)$)

ולכן $X^Ty \perp ker(X)$ ולכן יש לה אינסוף משוואות אם X^TX

לא הפיך, ופתרון יחיד אם X^TX הפיך.

5. Based on Recitation 1 In this question you will prove some properties of orthogonal projection matrices seen in recitation 1. Let $V \subseteq \mathbb{R}^d$, $dim(V) = k$ and let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be an orthonormal basis of V. Define the orthogonal projection matrix $P = \sum_{i=1}^k \mathbf{v}_i\mathbf{v}_i^\top$ (notice this is an outer product).

Prove the following properties in any order you wish:

(a) Show that P is symmetric.

(b) Prove that the eigenvalues of P are 0 or 1 and that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are the eigenvectors corresponding the eigenvalue 1.

(c) Show that $\forall \mathbf{v} \in V\ P\mathbf{v} = \mathbf{v}$.

(d) Prove that $P^2 = P$.

(a) Show that P is symmetric.

5. a.

$V \subseteq R^d$

בסיס אורתונורמלי של V

$B = (v_1, \ldots, v_k)$

$P = \sum_{i=1}^k v_i v_i^T$

והרי המכפלה החיצונית הינה קומוטטיבית,

כלומר $v \otimes u = u \otimes v$ כאשר v ו- u בפרט אורתונורמליים,

ולכן P הינה סימטרית

(b) Prove that the eigenvalues of P are 0 or 1 and that $\mathbf{v}_1, \dots, \mathbf{v}_k$ are the eigenvectors corresponding the eigenvalue 1.

ב.

יהי: v_j ו״ע שמתאים לע״ע λ_j אזי

$$\lambda_j v_j = P\cdot v_j = \sum_{i=1}^{k} v_i v_i^T \cdot v_j \overset{\circledast}{=} \sum_{i=1}^{k} v_i \cdot \underset{\text{אורתונורמלי כמו } \circledast}{a_{ij}} = v_j$$

$$v_i^T \cdot v_j = a_{ij} \quad \circledast$$

רק הרכיב ה $i=j$ שווה

ולכן $\lambda = 1$ או $\lambda = 0$

כמו כן ו״ע המתאימים הם הפורשים $v_1, \dots, v_k$

(c) Show that $\forall \mathbf{v} \in V \ P\mathbf{v} = \mathbf{v}$.

ג.

אנו יודעים כי נובע ישירות מהסעיף הקודם, נקבל ש $v_1, \dots, v_k$

וע נורמ אינם 0 כי הם המתאימים לע״ע 1,

ולכן $Pv = \lambda v = 1 \cdot v = v$.

(d) Prove that $P^2 = P$.

ד.

P מטריצה סימטרית ולכן ניתן לכתוב כך:

$P = UDU^T$ כאשר $UU^T = I$

$$P^2 = UDU^TUDU^T = UD^2U^T = UDU^T = P$$

והרי D היא מטריצה אלכסונית שבאלכסונה הע״ע של P -

כלומר 0 או 1, ולכן $D^2 = D$

(e) Prove that $(I-P)P=0$.

ה.

$$(I-P)\cdot P = P - P^2 \underset{\text{מסעיף הקודם}}{=} 0$$

6. Show that if $\mathbf{X}^\top\mathbf{X}$ is invertible, the general solution we derived in recitation equals to the solution you have seen in class. For this part, assume that $\mathbf{X}^\top\mathbf{X}$ is invertible.

.6

מתרגול ראינו שהפתרון כאשר X^TX הפיך הינו $\hat{w}=X^+y$

בכיתה ראינו שהפתרון הוא $w=[X^TX]^{-1}X^Ty$

אנו צריכים להראות כי $X^+y=[X^TX]^{-1}X^Ty$

ה $X=U\Sigma V^T$ ה SVD של X כמשמעו כעת:

אזי

$$[X^TX]^{-1}X^Ty = [(U\Sigma V^T)^T\cdot U\Sigma V^T]^{-1}(U\Sigma V^T)^T\cdot y$$

$$=[V\Sigma^T \underbrace{U^TU}_{I}\Sigma V^T]^{-1}\cdot V\Sigma^T U^T y = [V\Sigma^T\Sigma V^T]^{-1}V\Sigma^T U^T y$$

מכיוון ש V אורתוגונלית:

$$V^{-1}=V^T \overset{\Downarrow}{=} V(\Sigma^T\Sigma)^{-1}\underbrace{V^TV}_{I}\Sigma^T U^T y = V(\Sigma^T\Sigma)^{-1}\Sigma^T U^T y = V(\Sigma^2)^{-1}\Sigma^T U^T y$$

כאשר Σ^2 הינה מטריצה שבאלכסון הערכים הסינגולריים

בריבוע, אזי $(\Sigma^2)^{-1}$ הינה מטריצה אלכסונית שבאלכסונה:

$\frac{1}{\sigma^2}$, ו Σ $(\Sigma^2)^{-1}$ הינה מטריצה אלכסונית, נקרא לה

Σ^+ שבאלכסונה: $\frac{1}{\sigma^2}\cdot\sigma = \frac{1}{\sigma}$

ולכן

$$[X^TX]^{-1}X^T = V\Sigma^+U^T = X^+$$

ובפרט $[X^TX]^{-1}X^Ty = X^+y$ והפתרונות הינן שווים.

7. Show that $\mathbf{X}^\top\mathbf{X}$ is invertible if and only if $\text{span}\{\mathbf{x}_1,\dots,\mathbf{x}_m\} = \mathbb{R}^d$.

.7

X^TX מטריצה בגודל $d\times d$ והפיכה אמ"מ

דרגתה היא d, כלומר בעלת פתרון יחיד.

דרגת X היא כדרגת המרחב הנפרש ע"י $\{x_1,\dots,x_m\}$

לפי שאלה 1: דרגת X ו-X^TX זהה, ולכן

X^TX הפיכה אמ"מ דרגתה היא d, כלומר נפרש

ע"י $\text{span}\{x_1,\dots,x_m\}=\mathbb{R}^d$

8. Recall that if $\mathbf{X}^\top\mathbf{X}$ is not invertible then there are many solutions. Show that $\hat{\mathbf{w}} = \mathbf{X}^\dagger\mathbf{y}$ is the solution whose L_2 norm is minimal. That is, show that for any other solution $\overline{\mathbf{w}}$, $\|\hat{\mathbf{w}}\| \le \|\overline{\mathbf{w}}\|$.

.8

$\hat{w} = X^+y$ נסמן SVD של X הינו $X=U\Sigma V^T$

$\hat{w} = V\Sigma^+U^Ty$

אנו בכך קיבלנו שהפתרון לפי נוסחה זו הינו מינימלי.

אנו מנסים למזער את המינימום של $\|Xw-y\|_2$

$$\|Xw-y\| = \|U\Sigma V^Tw-y\| \overset{*}{=} \|\Sigma V^Tw-U^Ty\|$$

$*$ כפל ב-U^T, והרי $UU^T=I$ נכון כי U אוניטרית

נגדיר $z=V^Tw$, והרי $\|w\|=\|V^Tw\|=\|z\|$ כי V אוניטרי.

ולכן אנו צריכים למזער את המינימום של הנורמה הבאה:

$\|\Sigma z-U^Ty\|$

מכן ש-Σ תלויה לשתי ואלכסונית, המינימום בהכרח

מתקיים עבור (הנורמה תלויה רק בשווה לאפס)! $z=\Sigma^+U^Ty$

כמו שהפתרון המינימלי יתקבל עבור $w^+ = V \cdot z^+ = V \Sigma^+ U^T y$

וזה בדיוק הפתרון w^+ הרצוי.

חלק הפרקטי

2.

1. אילו פיצ'רים להשאיר ואילו למחוק?

מחקתי פיצ'רים שרירותיים שאין בהם כל קשר למחיר, כמו המיקוד, קווי אורך ורוחב, והשארתי פיצ'רים שנראה כי יכול להיות בינם לבין המחיר קשר.

2. אילו פיצ'רים הם קטגוריאליים?

הפיצ'רים הקטגוריאליים הם ה Lat ו Long אשר חשבתי תחילה לחלק אותם לאזורים הנקודתיים יותר, ולאחר לאנקד אותם במספרים, כאשר 1 הוא האיזור הזול יותר ו 5 הכי יקר. אך הסתפקתי לוותר בסוף (גרפית) כי אין כל קשר בין האזורים ב dataset הנבחר.

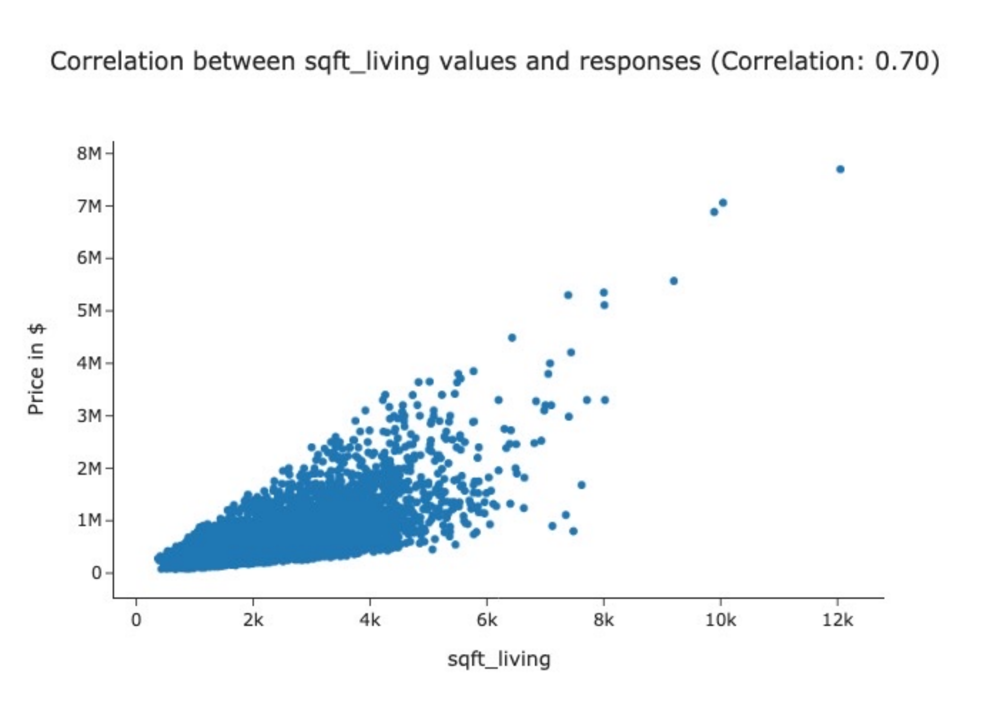
כמו כן שנת השיפוץ היא גם פיצ'ר קטגוריאלי, ועבורי הוא מסמן רק אם שעברו שיפוץ ונכנסו להם ערך 1, ולבתים שלא עברו שיפוץ נתתי להם ערך 0.

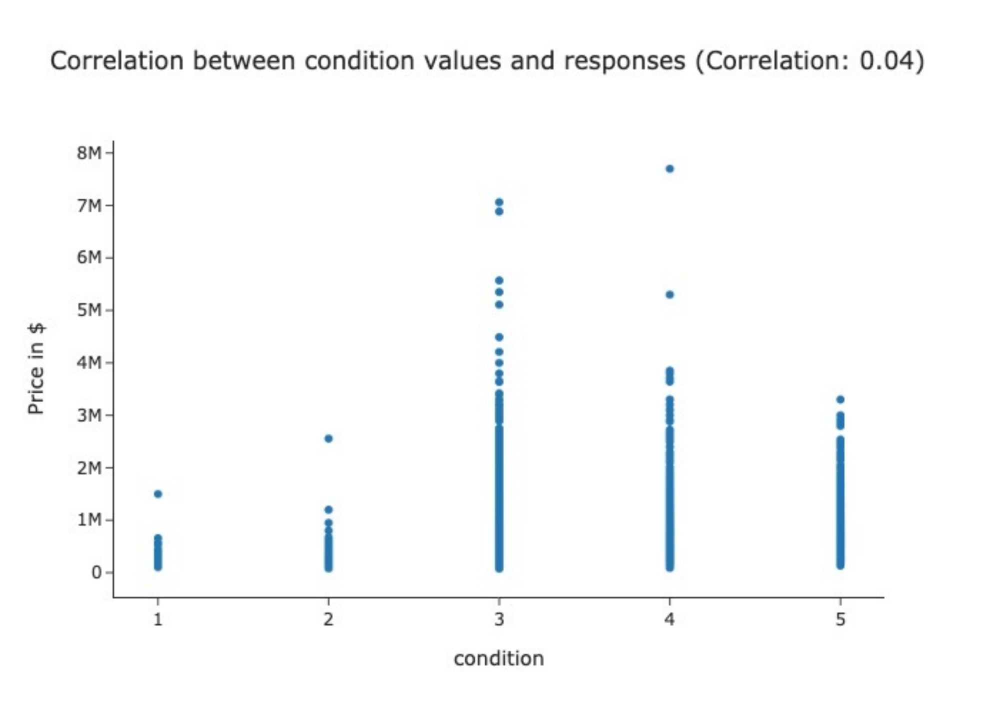
3. יצרתי פיצ'ר חדש בשם decade ושלפתי מ- שנת הבנייה והמבחן אותו על מנת שיהיה יותר רגיש לכל ערך וכך יוכל לבנות Linear regression טובה יותר.

4. מחקתי אינדקסים שונים חריגים - מחיר שלילי, חדרים וכו'.

5. ראינו שב dataset אין מספיק נתונים עבור מחיר גבוה יותר
מ- 13 מיליון ולכן הורדתי אותם מה dataset.
כמו כן עשיתי פעולה דומה עבור sqft_lot שגדול מ- 120,000

3.

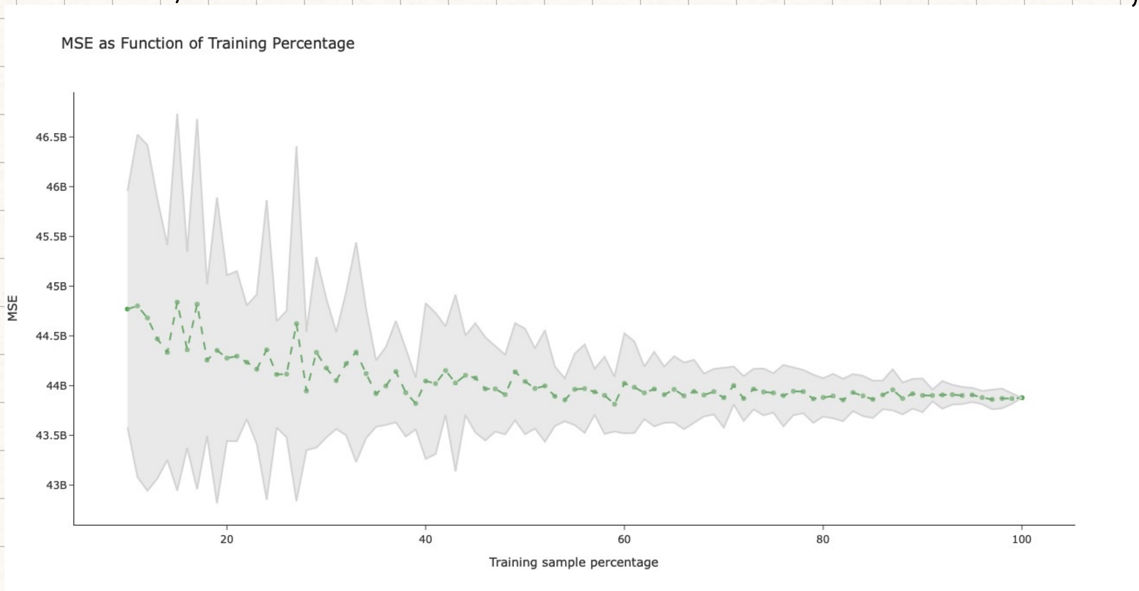




אך מסיקים מהגרף הפיצ'רים הנ"ל קשורים וגם לא לפי הקורלציה למחיר,
ניתן לראות שה sqft_living ישנה קורלציה גבוהה ולכן ניתן להסיק כי הוא

משפיע עיקרי על המחיר ואולי עבור condition הקורלציה מאוד נמוכה.

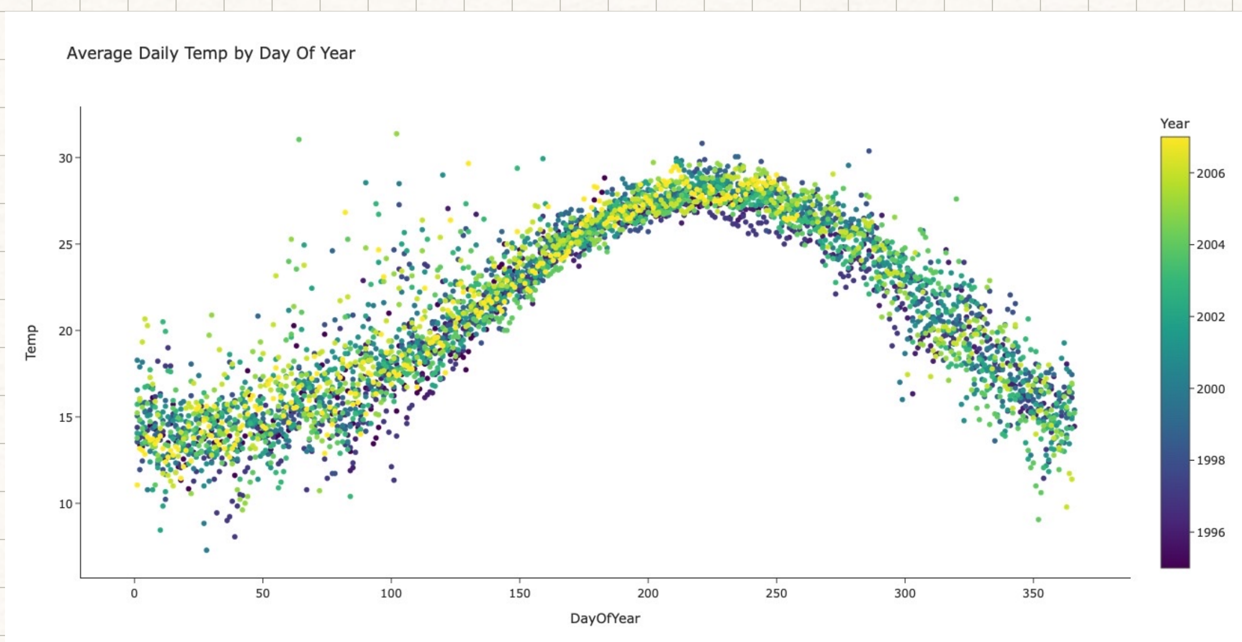
4.



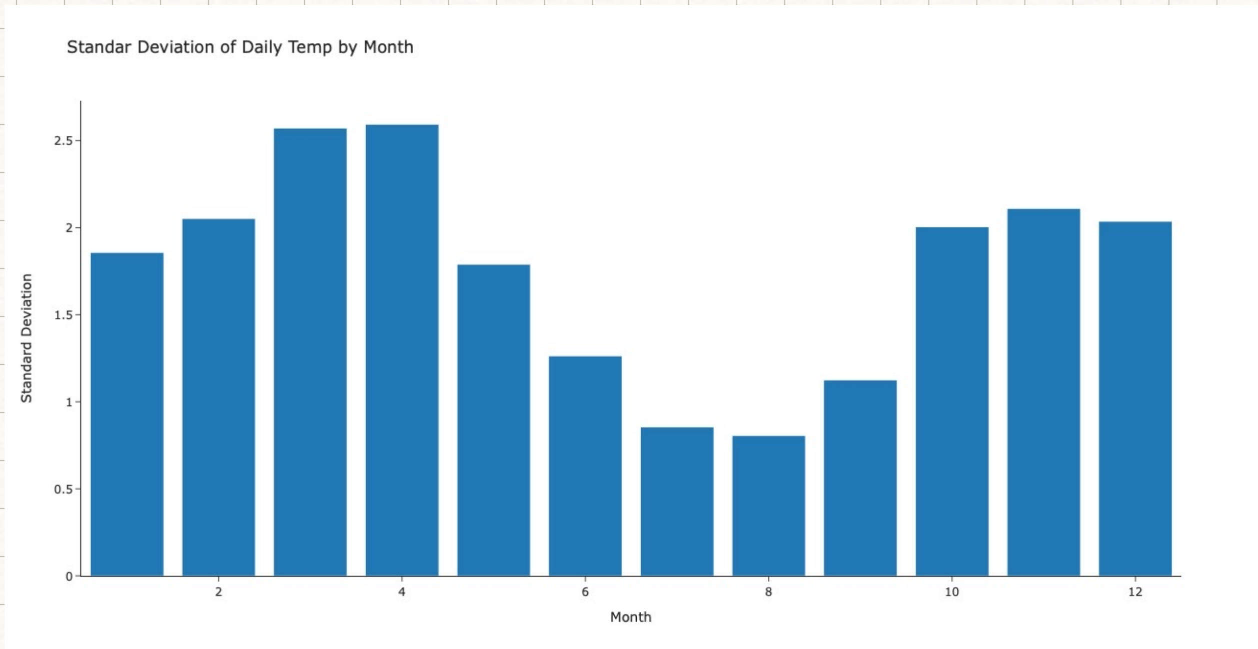
ניתן לראות מן הגרף שככל שאחוז הדגימות גדל יותר, כך ה-confidence interval יורד, כלומר לשערוך יש שונות קטנה.

Polynomial Fitting

2.

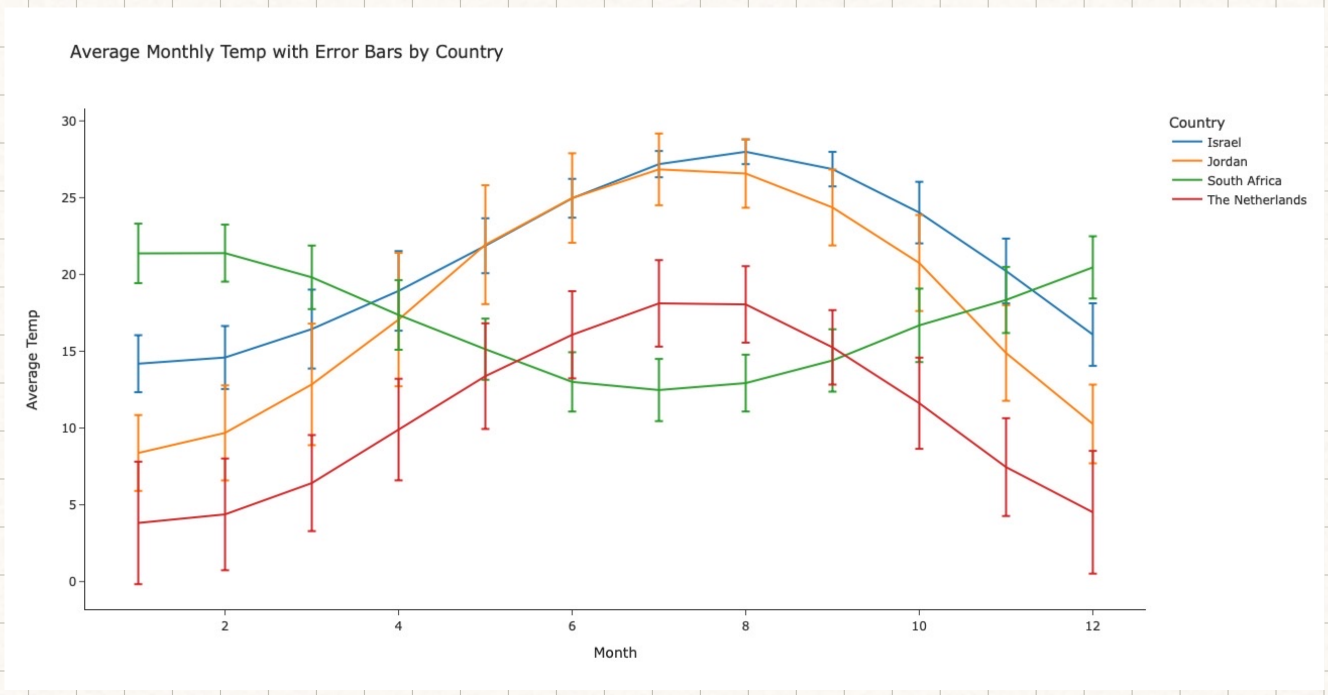


ניתן לראות כי הגרף אינו דומה לפולינום ממעלה 2 שכן בימים 0-50 הגרף נראה אופקי, ולכן הייתי אומר שהמעלה הקטנה ביותר המתאימה היא 3.



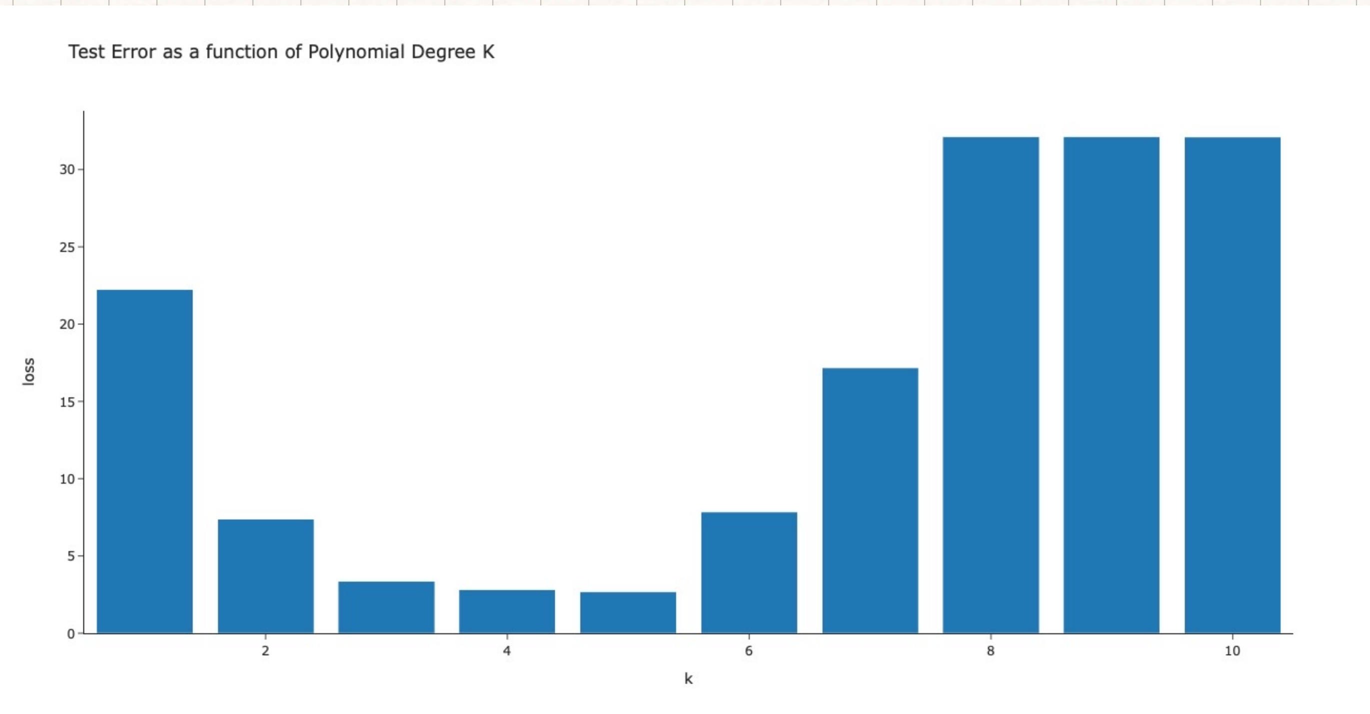
לפי הגרף אני מצפה שהמודל יצליח יותר בחודשים יוני עד ספטמבר,
ובשאר החודשים פחות, שכן השונות גדולה יותר.

3.



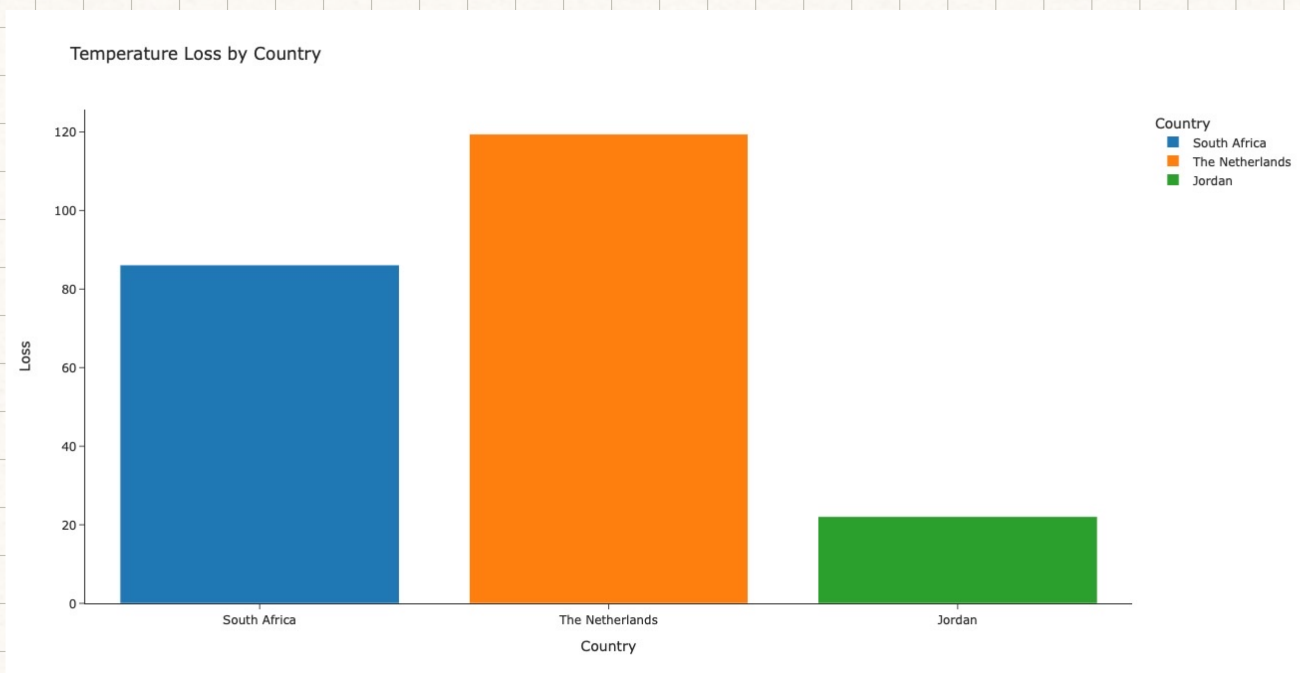
על סמך הגרף ניתן לראות כי לירדן המודל ישב טוב, גרוע להולנד
ולדרום אפריקה פחות, ולמה? מפני שירדן וישראל בעלי טמפ' קרובות.
ואילו דרום אפריקה נמצאת בחצי הדרומי של כדור הארץ ולכן
העונות שנה בה הפוכות.

4.



ערך ה-k שעבורו המודל ישיג את הדיוק הכי טוב הוא k=5,
ועדיין k=3,4 מקבלים loss נמוך כמעט באותה המידה ולכן ניתן לבחור
גם אותם.

5.



התוצאות הן כמו שציפיתי בשאלה 3, שעבורה ירדן קיבלה מודל כמעט
זהה לישראל ולכן ה-loss היה נמוך, ואילו עבור דרום אפריקה והולנד
ה-loss היה גבוה בהרבה (ציפיתי שעבור דרום אפריקה יהיה loss
גבוה יותר משל הולנד)