חלק 1

1

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} = U \Sigma V^T$$

נמצא ע"ע של A^TA

$$A^TA = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix}$$

$$\det(A^TA - \lambda I) = \begin{vmatrix} 2-\lambda & 0 & 2 \\ 0 & 2-\lambda & -2 \\ 2 & -2 & 4-\lambda \end{vmatrix} = (2-\lambda)\begin{vmatrix} 2-\lambda & -2 \\ -2 & 4-\lambda \end{vmatrix} + 2\begin{vmatrix} 0 & 2-\lambda \\ 2 & -2 \end{vmatrix}$$

$$= (2-\lambda)(8-6\lambda+\lambda^2-4) + 8 - 4\lambda = -\lambda^3 + 8\lambda^2 - 16\lambda + 8 - 8 + 4\lambda$$

$$-\lambda(\lambda^2 - 8\lambda + 12) = -\lambda(\lambda-6)(\lambda-2) = 0 \rightarrow \lambda = 0, 2, 6$$

$$\Sigma = \begin{bmatrix} 0 & \sqrt{2} & 0 \\ 0 & 2 & \sqrt{6} \end{bmatrix}$$

נמצא ו"ע

$$A^TA \vec{V_0} = 0$$

$\lambda = 0$:

$$\begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0 \rightarrow y = z, \; x = -z$$

$$\vec{V_0} = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$$

$$:\lambda=2$$

$$\begin{bmatrix}0&0&2\\0&0&-2\\2&-2&2\end{bmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=0 \to z=0\,,\ x=y$$

$$V_2=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\\0\end{pmatrix}$$

$$:\lambda=6$$

$$\begin{bmatrix}-4&0&2\\0&-4&-2\\2&-2&-2\end{bmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=0 \to z=2x\,,\ y=-x$$

$$\vec{V}_6=\frac{1}{\sqrt{6}}\begin{pmatrix}1\\-1\\2\end{pmatrix}$$

$$V^T=\begin{pmatrix}-1/\sqrt{3}&1/\sqrt{3}&1/\sqrt{3}\\1/\sqrt{2}&1/\sqrt{2}&0\\1/\sqrt{6}&-1/\sqrt{6}&2/\sqrt{6}\end{pmatrix}$$

לכן נמצא את מטריצה $V=\begin{bmatrix}u_1&u_2\\u_3&u_4\end{bmatrix}$ כמובן ועוד $AV=U\Sigma$

$$AV=\begin{bmatrix}1&1&0\\1&-1&2\end{bmatrix}\begin{bmatrix}-1/\sqrt{3}&1/\sqrt{2}&1/\sqrt{6}\\1/\sqrt{3}&1/\sqrt{2}&-1/\sqrt{6}\\1/\sqrt{3}&0&2/\sqrt{6}\end{bmatrix}=\begin{bmatrix}0&\sqrt{2}&0\\2/\sqrt{3}&0&4/\sqrt{6}\end{bmatrix}$$

$$=\begin{bmatrix}0&\sqrt{2}&0\\0&0&\sqrt{6}\end{bmatrix}=\begin{bmatrix}u_1&u_2\\u_3&u_4\end{bmatrix}\begin{bmatrix}0&\sqrt{2}&0\\0&0&\sqrt{6}\end{bmatrix}=\begin{bmatrix}0&\sqrt{2}u_1&\sqrt{6}u_2\\0&\sqrt{2}u_3&\sqrt{6}u_4\end{bmatrix}$$

$$U=\begin{bmatrix}1&0\\0&1\end{bmatrix}$$

$$A=\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}0&\sqrt{2}&0\\0&0&\sqrt{6}\end{bmatrix}\begin{bmatrix}-1/\sqrt{3}&1/\sqrt{3}&1/\sqrt{3}\\1/\sqrt{2}&1/\sqrt{2}&0\\1/\sqrt{6}&-1/\sqrt{6}&2/\sqrt{6}\end{bmatrix}$$

ב) המכפלה הטנזורית בין שני וקטורים שרירותיים $\vec{u}$ / $\vec{v}$ הינה:

$$V \otimes U = \begin{bmatrix} V_1 u_1 & \cdots & V_1 u_n \\ V_2 u_1 & \cdots & V_2 u_n \\ \vdots & \ddots & \vdots \\ V_m u_1 & \cdots & V_m u_n \end{bmatrix} = \begin{bmatrix} | & | & & | \\ u_1 \vec{V}, & u_2 \vec{V}, & \ldots, & u_n \vec{V} \\ | & | & & | \end{bmatrix}$$

מכאן ניתן לראות שכל העמודות נקבעות ע"י מצו רק $\vec{v}$ כפול בסקלר, כלומר כולן תלויות לינאריות אחת בשנייה.

ג) $X = \sum_{i=1}^{n} a_i u_i$

$$\langle X, u_i \rangle = \langle \sum a_i u_i, u_i \rangle = \langle a_i u_i, u_i \rangle + \sum_{\substack{j=1 \\ j \neq i}}^{n} \langle a_j u_j, u_i \rangle$$

מאורתונורמליות של $\vec{u}$ מתקיים $\langle u_i, u_j \rangle = \delta_{i,j}$ ולכן:

$$\langle X, u_i \rangle = a_i$$

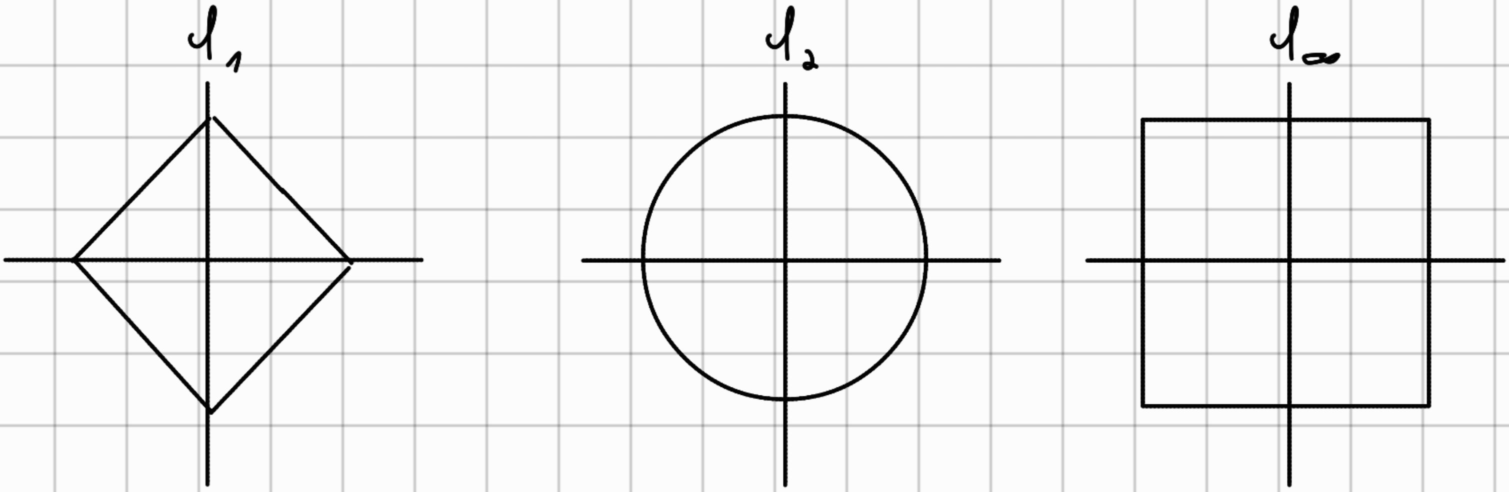
ד) א) $\vec{X} = \begin{pmatrix} 3 \\ -4 \\ 1 \\ -2 \end{pmatrix}$

$$\|X\|_1 = 3 + 4 + 1 + 2 = 10$$

$$\|X\|_2 = \sqrt{9 + 16 + 1 + 4} = \sqrt{30}$$

$$\|X\|_\infty = \max\{x_i\} = 4$$

נשים לב ששלוש הפונקציות הנורמות מזדהות עבור וקטור במימד יחיד.
כאשר מוסיפים עוד איברים נגלה ש φ_1, φ_2 גדלו בעוד ש φ_∞
נקבע רק ע״י איבר יחיד, ולכן φ_∞ הוא הקטן ביותר.
$\|x\| = 1$ עבור כל אחת מן הפונקציות ב- R^2:



φ_1 מתייחס למרחק המשוכלל מוגדרים

φ_2 מתייחס למרחק מהראשית

φ_∞ מתייחס רק לקואורדינטה אחת שבה המרחק מהראשית
מירבי.

ג)

$$h(\sigma) = \frac{1}{2}\|f(\sigma)-y\|^2 = \frac{1}{2}\left(\|f(\sigma)\|^2+\|y\|^2-2f^T(\sigma)y\right)$$

$$= \frac{1}{2}f^2(\sigma)+\frac{1}{2}y^2-y^Tf(\sigma)$$

נחשב את הנגזרות של h בכיוון לפי אחד מן הרכיבים של σ

$$\frac{\partial h}{\partial \sigma_i} = \frac{\partial f}{\partial \sigma_i}\frac{\partial}{\partial f}\left(\frac{1}{2}f^2(\sigma)+\frac{1}{2}y^2-y^Tf(\sigma)\right) = \frac{\partial f}{\partial \sigma_i}\left(f^T(\sigma)-y^T\right)$$

ד)

$$S(X)_j = \frac{e^{x_j}}{\sum_{l=1}^{k} e^{x_l}}$$

נחשב את הנגזרות לפי עבור $x_i \neq x_j$:

$$\frac{\partial S}{\partial x_i} = -\frac{e^{x_i+x_j}}{\left(\sum_{l=1}^{k} e^{x_l}\right)^2}$$

עבור $x_i = x_j$:

$$\frac{\partial S}{\partial x_i} = \frac{e^{x_i}\sum e^{x_l}-e^{2x_i}}{\left(\sum_{l=1}^{k} e^{x_l}\right)^2} = \frac{\sum e^{x_l}-e^{x_i}}{\left(\sum_{l=1}^{k} e^{x_l}\right)^2}e^{x_i}$$

נסמן את הסכום $a = \sum_{l=1}^{k} e^{x_l}$

$$J_x(S) = \frac{1}{a^2}\begin{bmatrix} ae^{x_1}-e^{2x_1} & e^{x_1+x_2} & \cdots & e^{x_1+x_k} \\ e^{x_2+x_1} & ae^{x_2}-e^{2x_2} & \cdots & e^{x_2+x_k} \\ \vdots & & \ddots & \vdots \\ e^{x_k+x_1} & & & ae^{x_k}-e^{2x_k} \end{bmatrix}$$

$$= \frac{1}{a^2}\underbrace{\begin{bmatrix} ae^{x_1}-e^{2x_1} & 0 & & \\ & ae^{x_1}-e^{2x_1} & & 0 \\ & & \ddots & \\ 0 & & 0 & ae^{x_k}-e^{2x_k} \end{bmatrix}}_{\text{חיובית}} + \begin{bmatrix} 0 & e^{x_1+x_2} & e^{x_1+x_3} & \cdots & e^{x_1+x_k} \\ e^{x_2+x_1} & 0 & e^{x_2+x_3} & & \vdots \\ e^{x_3+x_1} & e^{x_3+x_2} & 0 & & \vdots \\ \vdots & & & 0 & \vdots \\ & & & \ddots & \\ e^{x_k+x_1} & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

3

אם נניח הכלל קו-ליניארי, כלומר שקיים $X\vec{v}=0$ יש גם $X^TX=0$.

כלומר, אם $X^TX\vec{v}=0$ אז:

$$v^TX^TXv=0$$

$$(Xv)^TXv=0$$

$$|Xv|^2=0$$

$$Xv=0$$

ג) נראה תחילה ש- $Im(A^T) \subseteq ker(A)^\perp$

יהיו x, y כך ש- $Ax = 0$, $A^T y = w$ (כלומר $w \in Im(A^T)$
$x \in ker(A)$), אז $y^T x = 0$.

$$Ax = 0 \to y^T A x = (A^T y)^T x = wx = 0 \quad \checkmark$$

כיוון שני $ker(A)^\perp \to Im(A^T)$, נסמן את מימדי המטריצה A $n \times n$.

$$dim(ker(A)^\perp) + dim(ker(A)) = n$$

$$dim(ker(A)^\perp) = n - dim(ker(A)) = dim(Im(A))$$

בנוסף $dim(Im(A)) = dim\, Im(A^T))$ ולכן $dim(ker(A)^\perp) = dim(Im(A))$.

בנוסף הוכחנו ש $Im(A^T) \subseteq ker(A)^\perp$ ולכן $Im(A^T) = ker(A)^\perp$

ד) X אינו הפיך ולכן $dim(ker(X)) > 0$

לפי סעיף הקודם $y \perp ker(X^\perp) \longleftrightarrow y \in Im(X)$.

וקטור y בתמונה של X אומר קיים w כך ש $Xw = y$

בנוסף מכיוון ש $dim(ker) > 0$ קיימות קבוצה אינסופית של וקטורים
k כך ש $Xk = 0$ ולכן לכל וקטור כזה $X(w + k) = y$

כלומר יש אינסוף פתרונות.

ק) אם X^TX הפיכה אז לפי סעיף א גם X הפיכה ולכן למערכת המשוואות $Xw=y$ יש פתרון יחיד ואז גם ל- $X^TXw=X^Ty$ פתרון יחיד.

אם לא אז לפי סעיף ג' למערכת המשוואות $Xw=y$ יש אינסוף פתרונות ולכן גם ל $X^TXw=X^Ty$.

6) נחשב את הביטוי $(X^TX)^{-1}X^T$:

מאורתוגונליות $U^{-1}=U^T$

$$(X^TX)^{-1}X^T=(V\Sigma U^TU\Sigma V^T)^{-1}V\Sigma U^T=(V\Sigma^2V^T)^{-1}V\Sigma U^T$$

$$=(V^T)^{-1}(\Sigma^2)^{-1}V^{-1}V\Sigma U^T=V\begin{pmatrix}\sigma_1^2 & & 0\\ & \sigma_2^2 & \\ 0 & & \ddots\end{pmatrix}^{-1}\Sigma U^T$$

$$=V\begin{pmatrix}1/\sigma_1^2 & & 0\\ & 1/\sigma_2^2 & \\ 0 & & \ddots\end{pmatrix}\Sigma U^T=V\Sigma^+U^T=X^+$$

לכן $X^+y=(X^TX)^{-1}X^Ty$

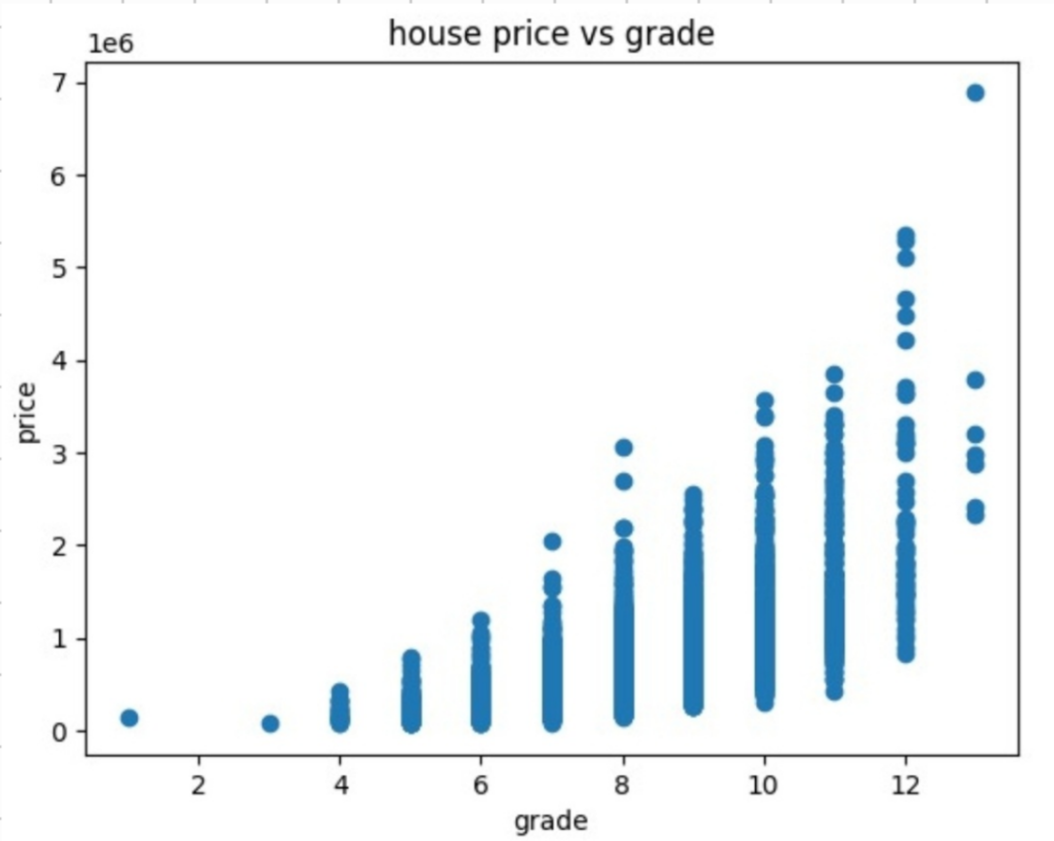
חלק שני - רגרסיה ליניארית

3) - כאשר מבצעים התאמה צריך להתחשב בכל התכונות של הדוגמאות, לכן הסרתי דוגמאות שחסר להן אחד השדות.

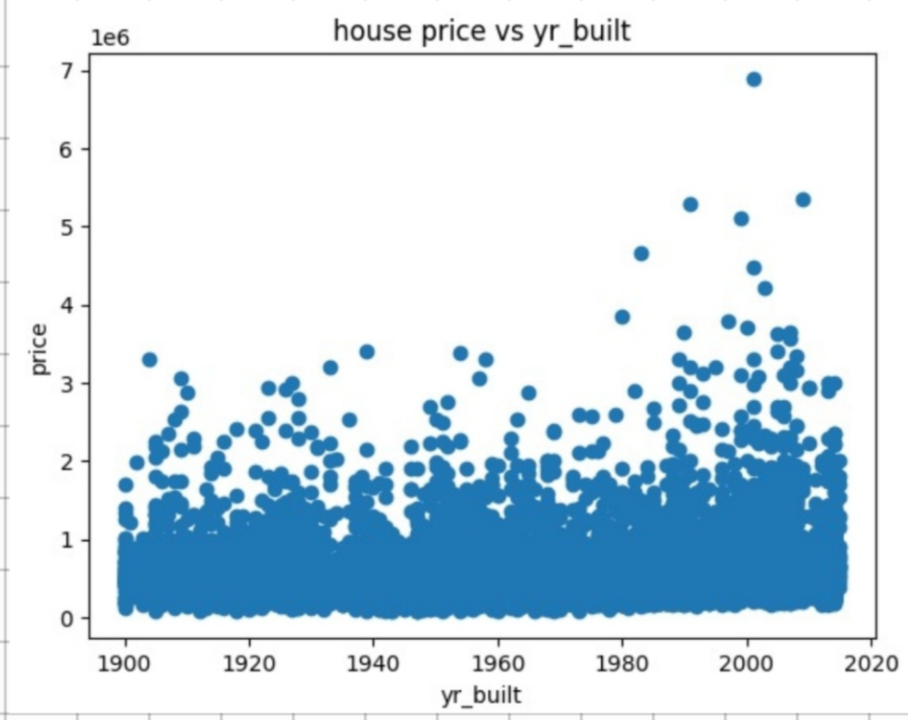
- התאריך הוא משתנה שקשה לעבוד איתו על מנת להסיק קשר, לכן הוספתי את תכונת תאריך הקנייה כמשתנה שסופר את מספר הימים מה-1 במאי 2014.

4) ערך פירסון הינו אומדן לקשר ליניארי בין משתנים, ככל שהערך רחוק מה-0 הקשר יותר קרוב להיות ליניארי כאשר 1, -1 הינם קשר מושלם. ערך קרוב ל-0 מצביע על חוסר קשר (אי תלות) בין שני המשתנים.

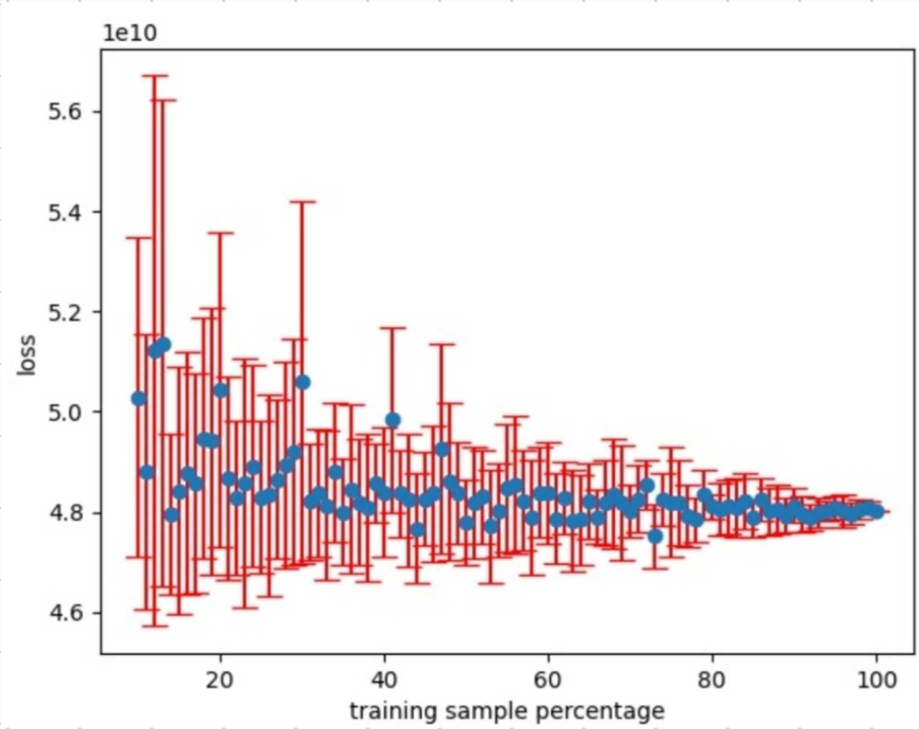
לכן תכונה שיש לה מקדם גבוה תתרום למודל.
לדוגמה התכונה grade בעלת מקדם פירסון של 0.67:



לעומת זאת לשנת הבנייה יש מקדם פירסון של 0.06



6)



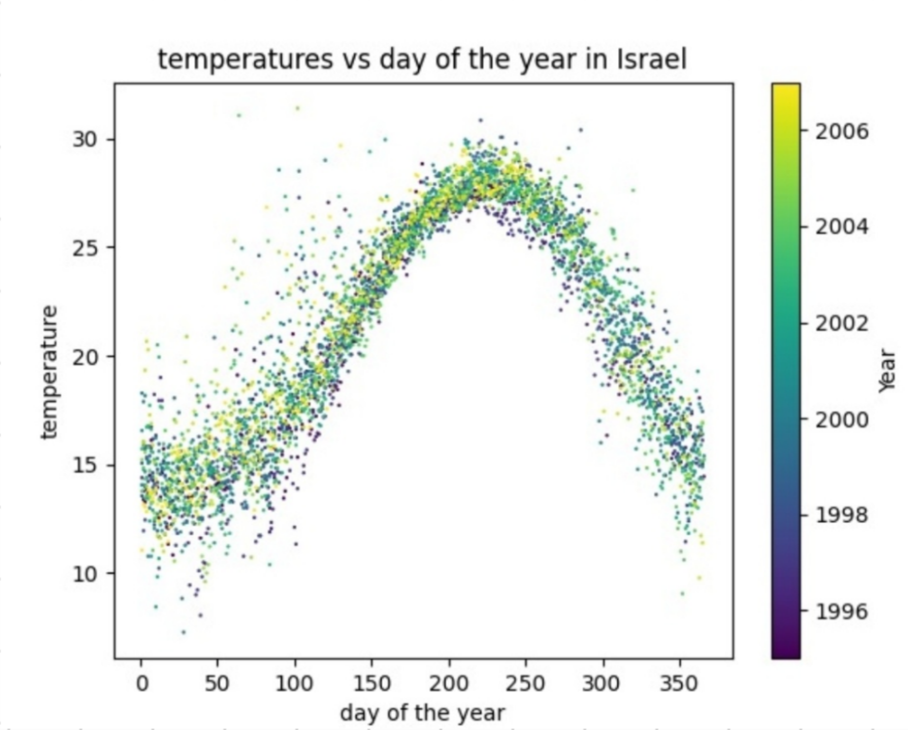
הקו האדום מסמל את סטיית התקן של כל אימון.

ניתן לראות שככל שמגדילים את קבוצת המדגם כך סטיית

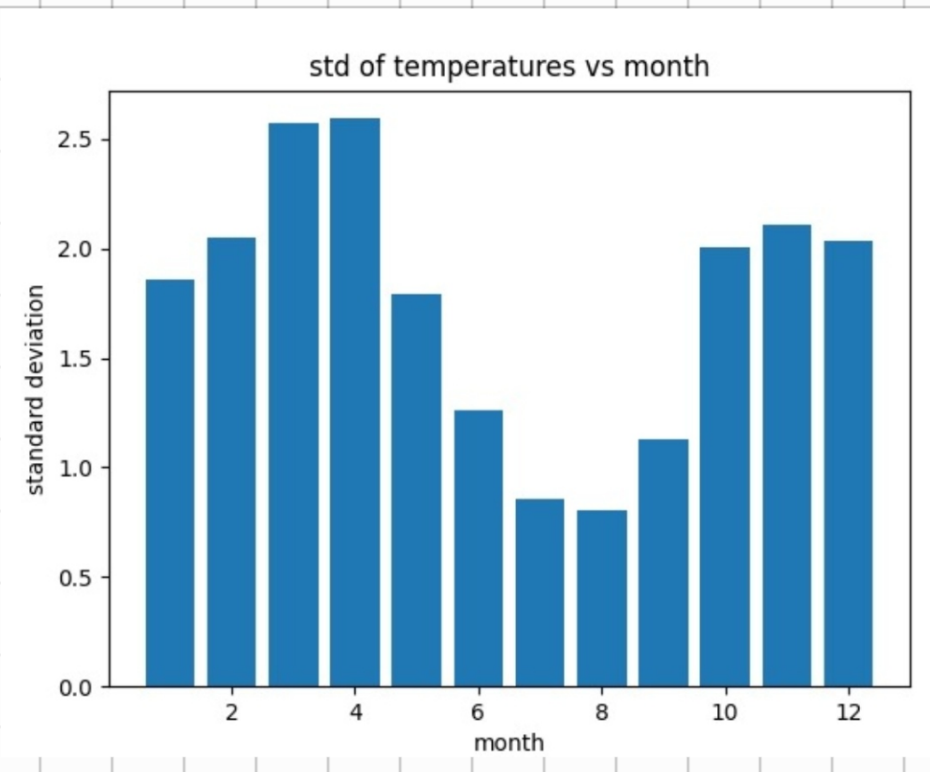
התקן קטנה, ובנוסף גם ערך ה- loss קטן (אם כי לא באופן משמעותי)

<u>חלק שני - התאמה פולינומיאלית.</u>

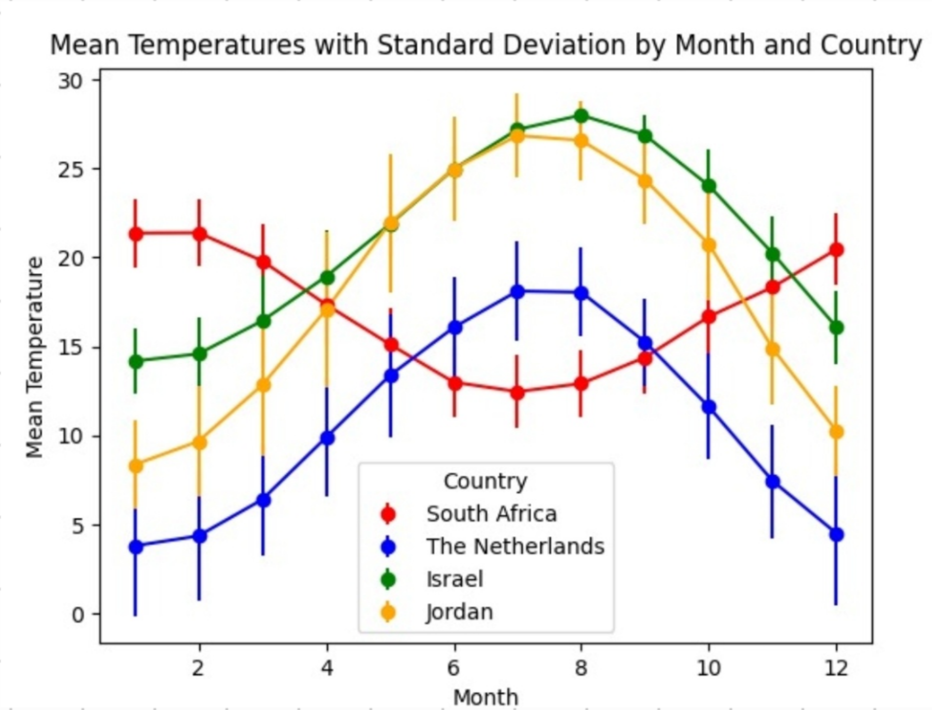
ג. נראה שאכן הנה מתאימים פולינום מדרגה זוגית



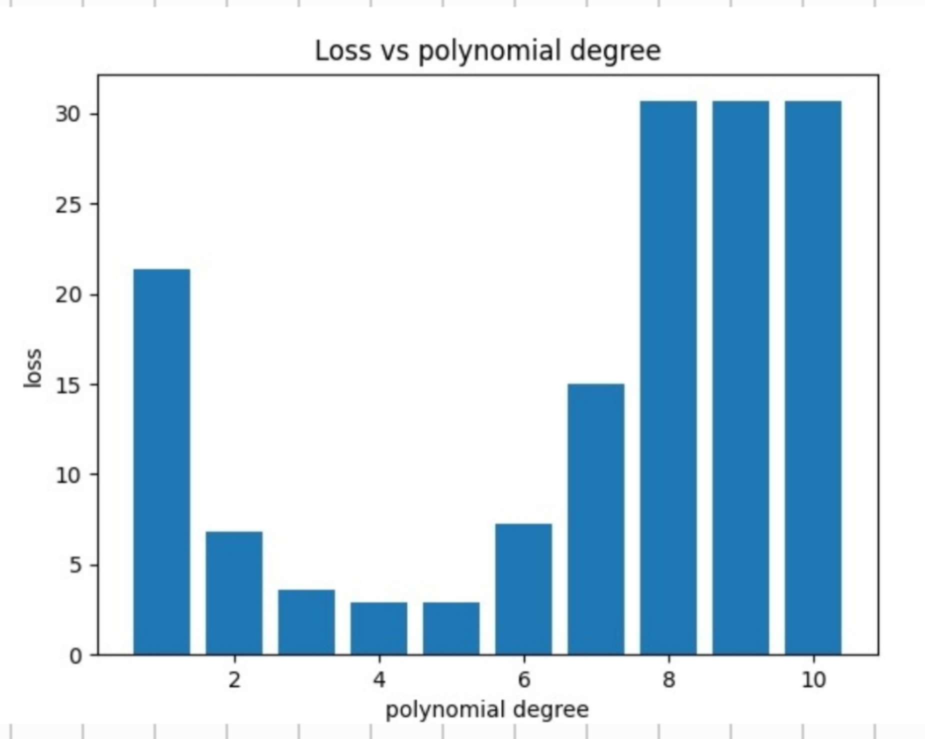
על ההתאמה נבדוק בעזרת ה std, ניתן לראות בבירור שעונות לא
משתנות בכל חודש. באוגוסט נקבל התאמה מיטבית ובאפריל נקבל
את ההתאמה הגרועה ביותר.



4) נראה כי לא לכל המדינות יש את אותה המגמה. הטמפרטורות בירדן יחסית קרובות לישראל ולכן המודל שמתבסס על הנתונים של ישראל ככל הנראה יתאים גם לירדן



5) עבור $k = 4, 5$ מקבלים ערכים מאוד קרובים (2.91, 2.93 בהתאמה). לכן בחרתי $k = 4$, כיוון שעדיף לפשט את המודל ולא להשתמש ב $k=3$ עם ערך של 3.6.



6) כיון ניתן לראות שבירדן יש את ה loss הנמוך ביותר
בדיוק כפי שראינו בסעיף 4.