Let $\mathbf{X}$ be the design matrix of a linear regression problem with m rows (samples) and d columns (variables/features). Let $\mathbf{y} \in \mathbb{R}^m$ be the response vector corresponding the samples in $\mathbf{X}$. Recall that for some vector space $V \subseteq \mathbb{R}^d$ the orthogonal complement of V is: $V^\perp := \{\mathbf{x} \in \mathbb{R}^d | \langle \mathbf{x}, \mathbf{v} \rangle = 0 \quad \forall \mathbf{v} \in V\}$

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

נוכיח בהכלה דו-כיוונית:

$\subseteq$: (כיוון קל)

יהי $v \in Ker(X)$ צ"ל: $v \in Ker(X^TX)$

מכיוון ש $v \in Ker(X)$ אזי $X(v)=0$ לכן ברור [illegible] ש $X^TX(v)=0$

$\supseteq$: יהי $v \in Ker(X^TX)$ צ"ל $v \in Ker(X)$

מכיוון ש $v \in Ker(X^TX)$ נקבל ש $X^TXv=0$ אם נכפיל את המשוואה (משני הכיוונים) ב- v^T נקבל $v^TX^TXv = 0 \Leftarrow (Xv)^T \cdot Xv = 0$ לפי ההגדרה של נורמה קיבלנו ש $\|Xv\|^2 = (Xv)^T(Xv) \Leftarrow = 0$

$Xv = 0 \iff v \in Ker(X)$.

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

גם כאן נוכיח דרך הכלה דו כיוונית

$\subseteq$: כיוון ראשון

יהי $y \in Im(A^T)$ לכן קיים x וקטור כך ש: $y = A^Tx$, ויהי $v \in Ker(A)$ עלינו להראות ש $\langle y, v \rangle = 0$ מכיוון ש $v \in Ker(A)$ אנו יודעים ש $Av=0$ ולכן:

$\langle y, v \rangle = \langle A^Tx, v \rangle = x^T(Av) = 0 \Leftarrow \langle y, v \rangle = 0$ לכל $v \in Ker A$ מהכיוון ש y שייך להשלמה האורתוגונלית ולכן $Im(A^T) \subseteq Ker(A)^\perp$

כיוון שני: יהי $v \in Ker(A)^\perp$ צ"ל: $v \in Im(A^T)$ נניח בשלילה ש. $v \notin Im(A^T) \Leftarrow v \notin Ker(A)^\perp$

במילים אחרות נניח ש $v \notin Im(A^T)$ אזי עלינו להראות קיים $c \in Ker(A)^\perp$ כך ש: $\langle v, c \rangle \neq 0$ מכיוון שאנו יודעים ש $v \in Ker(A)^\perp$ כי על מנת ש v יהיה ב- $Ker(A)^\perp$ עליו לקיים $\langle v, u \rangle = 0$ לכל $u \in Ker(A)$, בוא נמצא את וקטור c המקיים

את הדרוש:

נניח ש $v \notin Im(A^T)$ זה אומר שקיים $c \in Im(A^T)^{\perp}$ כך ש עבור c אנו יודעים ש $\langle v,c\rangle \neq 0$ כי $v \notin Im(A^T)$, c וקטור אורתוגונלי לכל וקטור כלשהו ב- $Im(A^T)$ ולכן:

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

עבור וקטור $x = Ac$ אנו יודעים ש $A^T x \in Im(A^T)$ ובגלל ש c אורתוגונלי עבור כל וקטור ב $Im(A^T)$ הוא אורתוגונלי ל- $A^T x = A^T A c$ ולכן המכפלה הינה 0

$\Rightarrow Ac = 0 \Rightarrow c \in Ker(A)$

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 $\Leftrightarrow y \perp Ker\left(\mathbf{X}^{\top}\right)$.

אם $y \perp Ker(X^T)$ אמ"ם $y \in Ker(X^T)^{\perp}$ אמ"ם $y \in Im((X^T)^T)$ $\Leftarrow$ $y \in Im(X)$

$\Leftarrow$ למערכת יש אין סוף פתרונות.

הסבר:

אם יודעים X איננו הפיך זה אומר שאין פתרון יחיד ויש שני מצבים

אין פתרון בכלל, יש אינסוף פתרונות, בנוסף לכך נתון ש X איננה מערכת

הומוגנית כלומר אומרת לפחות קיים פתרון אחד למערכת ולכן שאם $y \in Im(X)$ אז

יש לפחות פתרון אחד למערכת

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).

מסעיף הקודם אנו יודעים ש $X^Ty \perp Ker(X^TX) \underset{\text{סעיף 1}}{=} Ker(X)$ אמ"ם

למערכת יש אינסוף פתרונות לכן אנו נראה ש $X^Ty \perp KerX$

אם $v \in KerX$ $\Leftarrow$ $Xv = 0$ $\Leftarrow$ $\langle Xv, y\rangle = 0$

ואנו יודעים מנתוני השאלה ש $0 = \langle Xv, y\rangle = \langle v, X^Ty\rangle$

ולכן הוכחנו ש אם X^TX הפיך אז קיים פתרון יחיד (unique)

ואם איננו הפיך אז ישנם אינסוף פתרונות למערכת.

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 $\mathbf{v}_1, \dots, \mathbf{v}_k$ 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).

(1) Show that P is symmetric.

צריך להראות ש $P^T = P$ לפי הגדרה

$$P^T = \left(\sum_{i=1}^{k} V_i V_i^T\right)^T = (V_1V_1^T + V_2V_2^T + \dots + V_kV_k^T)^T \underset{(AB)^T = B^TA^T}{=} \sum_{i=1}^{k} (V_i^T)^T V_i^T$$

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

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

אנו יודעים ש P מטריצה סימטרית ולכן קיימת מטריצה אורתוגונלית
U ומטריצה אלכסונית D כך ש $P = UDU^T$ ולכן

$$P^2 = P \cdot P = UDU^T UDU^T \underset{U^TU = I}{=} UDDU^T \underset{\substack{D^2 = D \\ \text{מטריצה אלכסונית}}}{=} UDU^T = P$$

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

נשתמש בסעיף הקודם ש $P^2 = P$ ונקבל:

$$(I - P)P = IP - P \cdot P = P - P^2 \underset{P = P^2}{=} P - P = 0$$

(4) (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.

יהי λ ע"ע ו- X ו"ע לכן $PX = \lambda X$ נפעיל על שני הצדדים ב- P ונקבל

$$P^2X = P(\lambda X) = \lambda PX \underset{\substack{P^2 = P \\ \text{הוכחנו לפני}}}{=} PX$$

כעת אם X אינו וקטור אפס נקבל $\Leftarrow$ $\lambda X = X$ זה אומר $\lambda = 1$

כלומר נקבל ש $X=0$ ולכן כל ע"ע הינו או 0 או 1.

כעת נראה ש $v_1,\dots,v_k$ הינם ו"ע ל-ע"ע 1, לכל $1 \le i \le k$ נקבל

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

ולכן v_i הינו ו"ע ל-P עבור ע"ע 1 כך ש $v_1,\dots,v_k$ בסיס אורתונורמלי

ל-V כל וקטור ב-V ניתן לכתוב אותו כצירוף לינארי של $v_1,\dots,v_k$ אורתונורמלי

כבר הראינו ש $v_1,\dots,v_k$ הינם וקטורים עצמיים ולכן כל צירוף לינארי שלהם הינו ו"ע של

P עבור $\lambda=1$ ולכן $v_1,\dots,v_k$ הינם ו"ע עצמים כנ"ל.

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

יהי $v \in V$ מכיוון ש $v_1,\dots,v_k$ בסיס אורתונורמלי עבור V אזי נכתוב

$v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_k v_k$, $\alpha_1,\dots,\alpha_k$ סקלרים

$$\Rightarrow Pv = P(\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha v_k) \underset{\text{לינאריות}}{=} P\alpha_1 v_1 + P\alpha_2 v_2 + \dots + P\alpha_k v_k$$

$$= \alpha_1 P v_1 + \dots + \alpha_k P v_k$$

$$\underset{\text{סעיף קודם}}{=} \alpha_1 v_1 + \dots + \alpha_k v_k$$

$$= v.$$

מ.ש.ל

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.

יהי $X = U\Sigma V^T$ ה- svd של X אנו יודעים מתוך הגדרה ש X^TX הפיך

- $\mathbf{X}$ can be written as
$$\mathbf{X} = \mathbf{U}\cdot\Sigma\cdot\mathbf{V}^\top$$
  - $\mathbf{U}\in\mathbb{R}^{m\times m}$ is orthonormal $\Rightarrow \mathbf{U}^\top = \mathbf{U}^{-1}$
  - $\Sigma\in\mathbb{R}^{m\times(d+1)}$ is diagonal with **singular** values $\sigma_1\geq\ldots,\geq\sigma_{d+1}\geq 0$
  - $\mathbf{V}\in\mathbb{R}^{(d+1)\times(d+1)}$ is orthonormal

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

$(AB)^T = B^TA^T$ $\rightarrow$ $$= \left(\left((V\Sigma^T U^T)\cdot(U\Sigma V^T)\right)^{-1}\cdot V\Sigma^T U^T\right)\cdot y$$

$U^T = U^{-1}$ $\rightarrow$ $$= \left((V\Sigma^T \underbrace{U^T U}_{I}\Sigma V^T)^{-1}\cdot V\Sigma^T U^T\right)\cdot y$$

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

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

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

נבין כמה דברים $\Sigma^T\Sigma$ הינה מטריצה אלכסונית, כי Σ אלכסונית וגם $\Sigma\in R^{m\times d}$

מימדים
$$\underset{d\times m}{\Sigma^T}\cdot\underset{m\times d}{\Sigma} = C_{d\times d}$$

אנו יודעים ש $\sigma_i = \Sigma_{i,i}$ ולכן $\Sigma^T\Sigma_{i,i} = \sigma_i^2$ כך ש $\sigma_1 \cdots \sigma_d > 0$

אנו יודעים שהערכים של X הינם חיובית ולכן אנו יודעים ש:

$$(\Sigma^T\Sigma)^{-1}\Sigma_{i,i} = (\sigma_i^2)^{-1}\cdot\sigma_i = \sigma_i^{-1} \underset{\uparrow}{=} \Sigma^\dagger_{i,i}$$

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

לכן אנו נקבל $(X^TX)^{-1}\cdot X^Ty = V\Sigma^\dagger U^Ty$

$$\Rightarrow \quad (X^TX)^{-1}X^T y = X^\dagger y$$

$\mathbf{X}^\dagger = V\Sigma^\dagger U^\top$ where $\Sigma^\dagger$ is an $d\times m$ diagonal matrix,

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

נשים לב ש X^TX מטריצה סימטרית מימד $d\times d$ קיימת לה: מטריצה הפיכה אמ"מ הע"ע שלה שונים מאפס כך ש $\lambda_i = \sigma_i^2$

$\Leftarrow$: נניח $\operatorname{span}\{x_1,\ldots,x_m\}=\mathbb{R}^d$ זה אומר שכל וקטור ב-$\mathbb{R}^d$ יכול להיות צירוף לינארי של $x_1,\ldots,x_m$ במילים אחרות $\operatorname{rank}(X)=d$ לכן לפי משפט המימדים

$$\dim(\operatorname{null}(X)) = m-d \le 0 \Rightarrow \operatorname{null}(X)=0 \Rightarrow Xw=0 \Rightarrow w=0$$

כלומר לכל $0\neq w\in\mathbb{R}^d$ לכן $Xw\neq 0$ וכן וקטורים שהעמודות של X פורשים את $\mathbb{R}^d$ ולכן

$w^TX^TXw = \|Xw\|^2 > 0 \quad \Leftarrow \quad X^TX$ חיובי וסופי ולכן כל הע"ע הינם חיובים כי אומר X^TX הפיך.

$\Rightarrow$: נניח X^TX הפיך ולכן הע"ע שלו הינם שונים מאפס.

יהי $w\in\mathbb{R}^d$ אנו רוצים להראות שניתן לרשום את w כצירוף לינארי של $x_1,\ldots,x_m$

קיים $v\in\mathbb{R}^d$ כך ש $Xv=w$ לא מהכרח v unique וע"י נקבל

$$X^TXv = X^Tw \;/\cdot(X^TX)^{-1} \Rightarrow v=(X^TX)^{-1}X^Tw$$

לכן אנו יכולים לכתוב את v כצירוף לינארי של העמודות של X ומכיוון ש $w=Xv$

אנו הצלחנו לכתוב את w כצירוף לינארי של $x_1,\ldots,x_m$ ולכן העמודות של X

הם span של $\mathbb{R}^d$.

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 $\overline{\mathbf{w}}$, $||\hat{\mathbf{w}}|| \le ||\overline{\mathbf{w}}||$

Hints:

- Recall that the rank of $\mathbf{X}$ and the rank of $\mathbf{X}^\top\mathbf{X}$ are determined by the number of singular values of $\mathbf{X}$. If you are not sure why this is true, go over recitation 1.
- Which coordinates must satisfy $\widehat{w}_i = \overline{w}_i$? What is the value of $\widehat{w}_i$ for the other coordinates? If you are not sure, go back to the derivation of $\widehat{\mathbf{w}}$ (see recitation 4).

אנו יודעים שלכל $i\in\{1,\dots r\}$ $\bar{w}_i = \hat{w}_i$ ועבור $i\in\{r+1,\dots d\}$

$\hat{w}_i = 0$ וכמו כן $\bar{w}_i$ יכול להיות כל ערך אפשרי בפרט

שווה לאפסס כך אולם ה L_2 הנורמה מינימלית :

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

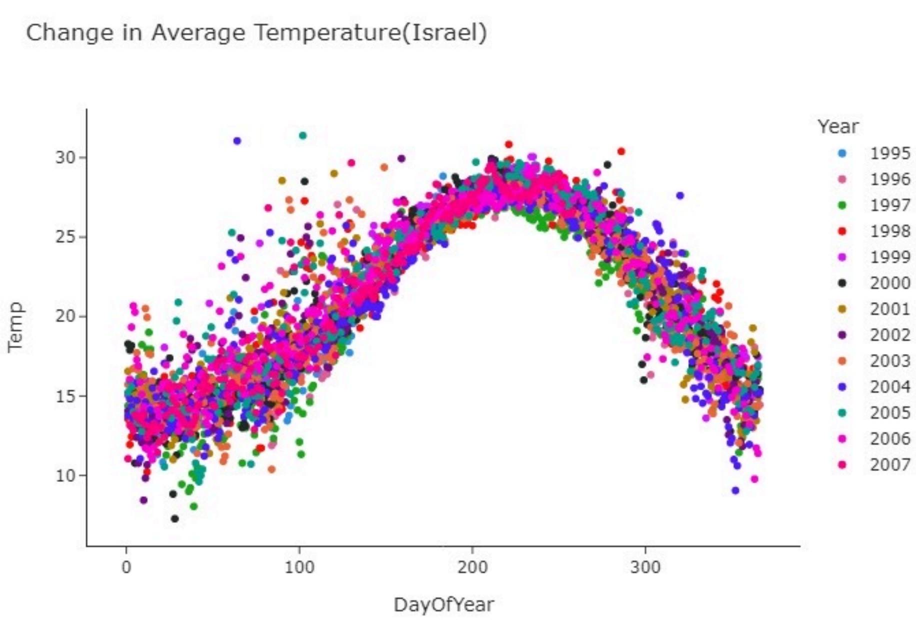
$$= \|\bar{w}\|^2.$$

Polynomial fitting

2. Subset the dataset to caintain samples only from the country of Israel. Investigate how the average daily temperature (‘Temp‘ column) change as a function of the ‘DayOfYear‘.
   - Plot a scatter plot showing this relation, and color code the dots by the different years (make sure color scale is discrete and not continuous). What polynomial degree might be suitable for this data?
   - Group the samples by ‘Month‘ (have a look at the pandas ‘groupby‘ and ‘agg‘ functions) and plot a bar plot showing for each month the standard deviation of the daily temperatures. Suppose you fit a polynomial model (with the correct degree) over data sampled uniformly at random from this dataset, and then use it to predict temperatures from random days across the year. Based on this graph, do you expect a model to succeed equally over all months or are there times of the year where it will perform better than on others? Explain your answer.

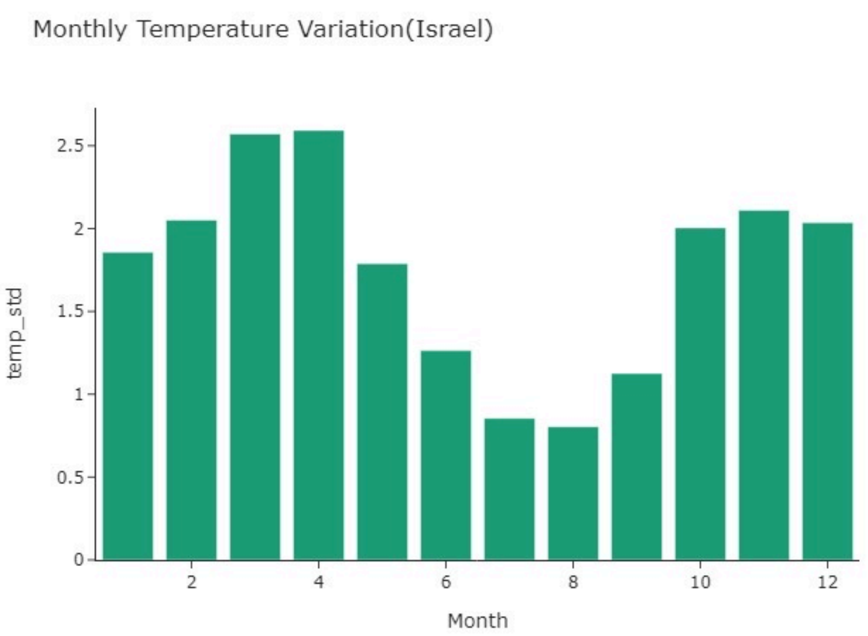
   Add both plots and answers to the `Answers.pdf` file.

אם נסתכל על גרף שלמעלה נוכל לראות הטמפרטורה הגבוהה ביותר הינה בימים באינטרוול של 200 אבל עם זאת ברור שהטמפרטורה לא ממשיכה לרדת משני צידי הפסגה, אלא מתייצבת לכן התאמת פולינומים מדרגה גבוהה מ- 2 נמוכה ביותר ולכן פולינום $\geq 3,4$ יהיו בחירה ומתאימים ביותר על מנת למדל ולהציג את הנתונים בצורה הטובה ביותר.



חשוב לציין שהביצועים של המודל לא יכול להיות עקבי
לכל השנה וסביר להניח שהחודשים עם השונות הנמוכה
מיוני עד ספטמבר יהיו צפויים להיות מדוייקים יותר
מכיוון שהמודל יכול להתאים קרוב יותר לערכי
הנתונים בתקופה זו מכיוון שההנחה ש-Cos בקירוב
נכונה באותה.

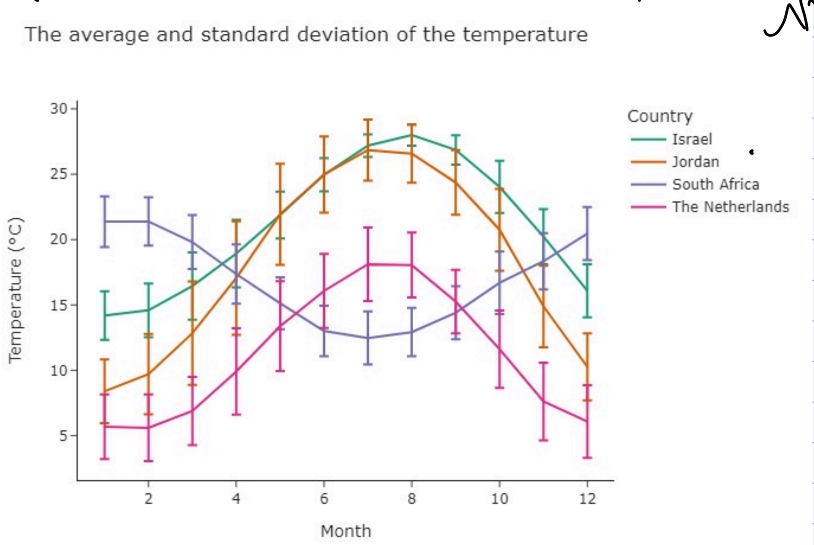
סביר להניח הדיוק של המודל בגלל חודשים אלה ישופר
במודל למשל אפשר לראות שיש שונות גבוהה ולכן בתקופה
זכך סביר לצפות לביצועים נמוכים יותר מתקופה אחרת.



3. Returning to the full dataset, group the samples according to `'Country'` and `'Month'` and calculate the average and standard deviation of the temperature. Plot a line plot of the average monthly temperature, with error bars (using the standard deviation) color coded by the country. If using `plotly.express.line` have a look at the `error_y` argument.

Based on this graph, do all countries share a similar pattern? For which other countries is the model fitted for Israel likely to work well and for which not? Explain your answers.

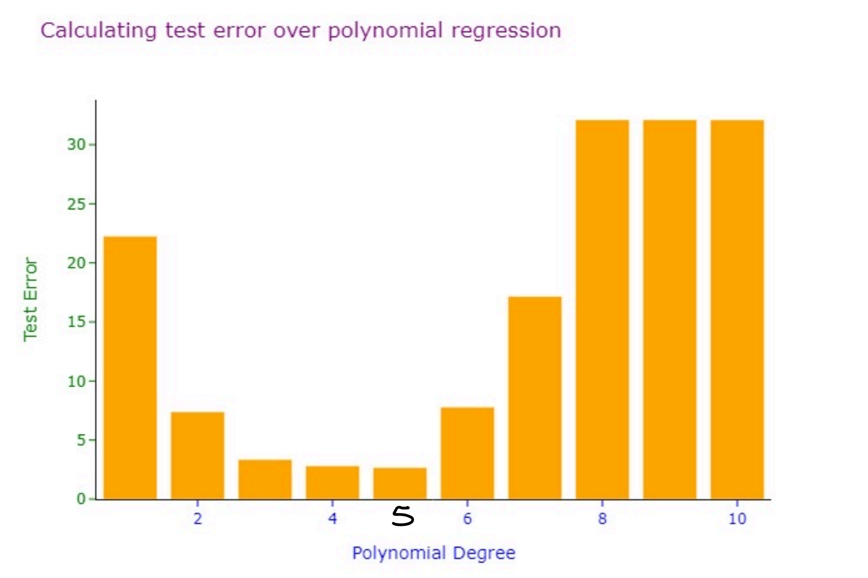
מבדיקת הגרף, נשים לב כי לכל מדינה יש התפלגות ייחודית של ממוצע הטמפרטורות היומי לאורך השנה.
כאשר אנו מקבלים את נתוני הטמפרטורה של ישראל אנו יכולים לצפות לתוצאות מדויקות יותר בעת עבודה עם שאר המדינות.
התפלגות הטמפרטורה בדרום אפריקה למשל הינה שונה בהרבה מישראל עם טמפרטורות פחות קיצוניות להפך, לכן יהיה ראוי להחיל את אותו מודל על שתי המדינות אבל לבצע שינויים מהותיים.
לעומת זאת הולנד למשל יש לה התפלגות טמפרטורה דומה לזו של ישראל בתבנית שלה, ניתן להשתמש במודל הישראלי אבל צריך להתאים משמעותית את הפרש הטמפרטורות



לעומת זאת ירדן היא המדינה הדומה ביותר לישראל ולכן המודל יפעל בצורה טובה יותר על הנתונים הירדניים מאשר על הנתונים של הולנד ודרום אפריקה.

4. Over the subset containing observations only from Israel perform the following:
   - Randomly split the dataset into a training set (75%) and test set (25%).
   - For every value $k \in [1, 10]$, fit a polynomial model of degree k using the training set.
   - Record the loss of the model over the test set, rounded to 2 decimal places.

   Print the test error recorded for each value of k. In addition plot a bar plot showing the test error recorded for each value of k. Based on these which value of k best fits the data? In the case of multiple values of k achieving the same loss select the simplest model of them. Are there any other values that could be considered?

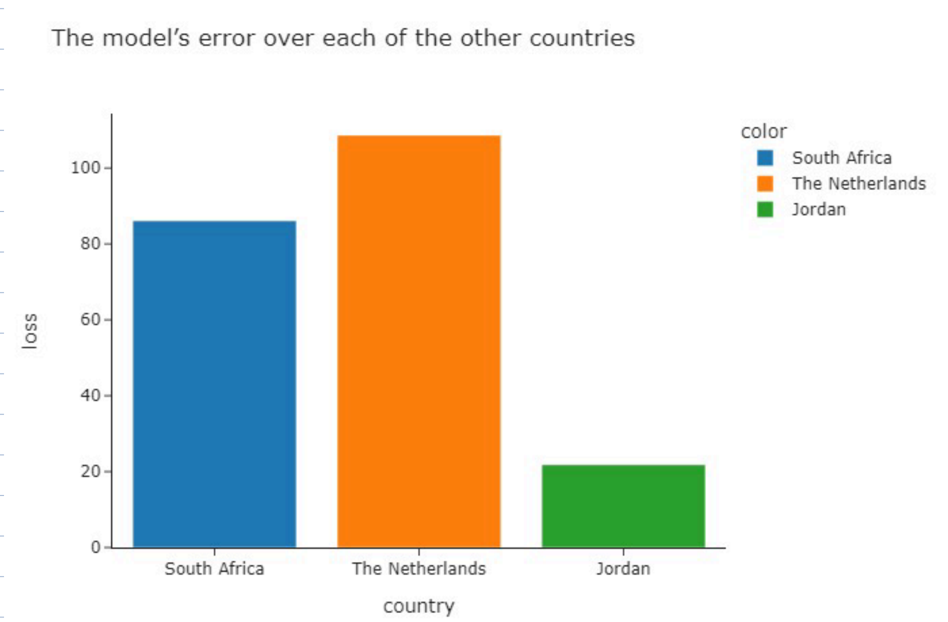


תוצאות הבדיקה מראות ומעידות על כך ש אכן K=5
המודל משום ששגיאת הבדיקה הנמוכה ביותר לעומת שאר המודל
עם K=4 יותר שקול ב· 0.2 ל- K=5 וגם K=3 יש כל
שגיאות בדיקה כמעט 4 ולכן עם מידע הזה אני נכנס לומר
K=5 כבחירה סופית שלנו

```
C:\Users\97254\anaconda3\envs\iml.en
degree=1,error=22.26
degree=2,error=7.38
degree=3,error=3.37
degree=4,error=2.82
degree=5,error=2.67
degree=6,error=7.85
degree=7,error=17.17
degree=8,error=32.11
degree=9,error=32.11
degree=10,error=32.10

Process finished with exit code 0
```

5. Fit a model over the entire subset of records from Israel using the *k* chosen above. Plot a bar plot showing the model's error over each of the other countries. Explain your results based on this plot and the results seen in question 3.



אנו ראינו שבשאלה 3 התפלגות הטמפרטורה של ירדן דומה
מאוד לזו של ישראל, התפלגות הטמפרטורה בדרום אפריקה
ובהולנד חורגת משמעותית מזו של ישראל ומעידה מכך הביצועים
של המודל המותאם על נתוני המדינות הללו היו גרועים
מעניין לציין כי למרות שהתפלגות הטמפרטורה של הולנד דומה
למדי לזו של ישראל, המודל הציג ביצועים טובים יותר על
הנתונים בדרום אפריקה, ייתכן שהדבר נובע מכך שהמדידות ה
הישראליות הדרום אפריקאיות דומות זו לזו בממוצע, יותר מאשר

לתצפיות מהפולינג, מאחר שהדיוק עדיין היה פחות מסטנדרטי מאשר
המקרה לא מושפע מהפולינג.

Fitting A Linear Regression Model

Question 2

- What sort of values are valid for different types of features? Can house prices be negative? Can a living room size be too small?

הערכים של תכונות, שנה, תאריך, Sqft_living, sqft_lot, Sqft_basement, yr_built, yr_renovated --- כולם צריכים להיות ערכים אי שליליים

waterfront : ערך בינארי 0,1

View : מס' שלם בין 0 ל-4

Condition : מצב הנכס בין 1 ל-5

Grade : מס' שלם בין 1-13 המעיד על הניקוד של איכות הבית

Price : ערכים עשרוניים המציגים את מחיר המכירה של הבית שאינם יכולים להיות שליליים.

יש לנו תכונות כמו Sqft_above, Sqft_lot, Sqft_living, Sqft_basement המציגות את שטח המרחב הכולל של יצור המגורים, תכונות אלו צריכות להיות בעלות מינימום ערך שכל שורה שבה ערכים אלו שווים לאפס או קטנים ממנו.

- Some of the features are categorical with no apparent logical order to their values (for example zip-code). Correctly address these features such that it will make sense to fit a linear regression model using them. For assistance you may refer to the following StackOverflow question.

עבור תכונות קטגוריות, יש לנו מיקוד שאין לו סדר לוגי.
אנו יכולים להמיר תכונות קטגוריות אלה למשתנה דמה:
באמצעות קידוד חם אחד, כלומר: לכן תיצור תכונה בינארית
עבור כל ערך מיקוד ייחודי, ותכונות בינאריות האלה יכולות לשמש
כמנבאים למודל רגרסיה לינארית.

- Are there any additional features that might be beneficial for predicting the house price and that can be derived from existing features?

כן אנו יכולים להוסיף כמה תכונות נוספות שעשויות להיות מועילות
בחיזוי מחיר בית: ניתן לחשב את שנים נבנה והיא ההפרש בין השנה
הנוכחית לשנה בה נבנה
מספר חדרי האמבטיה: אנו יכולים לספור את מספרי חדרי האמבטיה
המלאים וחצי חדר אמבטיה כדי ליצור תכונה חדשה המייצגת את:
מספר הכולל של חדרי האמבטיה...

- Which features to keep and which not?

אנו השמטנו את התכונות "id", "sqft_living15", "sqft_lot15",
"long", "date", "lat" מכיוון שהם עשויים לא לתרום הרבה
למשתנה היעד תכונה שלא עשויה לספק לנו רלוונטית או להכיל
מידע כפול שכבר יש אותו בתכונות אחרות.

- Which features are categorical how how did you treat them?

התכונות הקטגוריות היחידה בנתונים היא "Zipcode" רק שהמרנו אותה באמצעות קידוד חם זו תכונה נומינלית המשמשת למפות נתונים הקטגוריים לספרות שניתן להשתמש בו במודלים של רגרסיה.

- What other features did you design and what is the logic behind creating them?

לא הוספנו עוד תכונות

- How did you treat invalid/missing values?

עבור ערכים חסרים, אנו ממלאים אותם בערך (main() ערך ממוצע של העמודה המתאימה, עבור ערכים לא חוקיים כגון ערכים שליליים או ערכים גבוהים במיוחד, אנו זיהינו את האינדקסים של ערכים אלה והחלפנו אותם בערך main ממוצע של העמודה המתאימה.

- Explain any additional processing performed on the data.

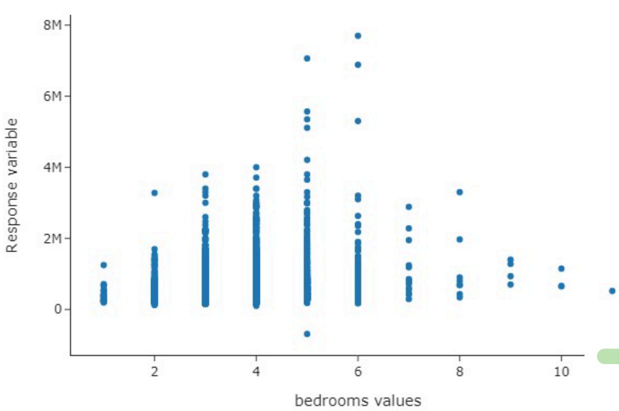
לא בוצע עיבוד נוסף של נתונים אבל אחרי העיבוד עשינו reindex() בין הtrain ו-test על מנת שיהיו עם אותו סדר. כמו כן בעת העיבוד אם מחקנו שורה מה Dataframe עלינו גם למחוק מוקטור y את השורות "המתאימות"

Question 3

Choose two features, one that seems to be beneficial for the model and one that does not. In your Answers.pdf add the graphs of these two chosen features and explain how do you conclude if they are beneficial or not.

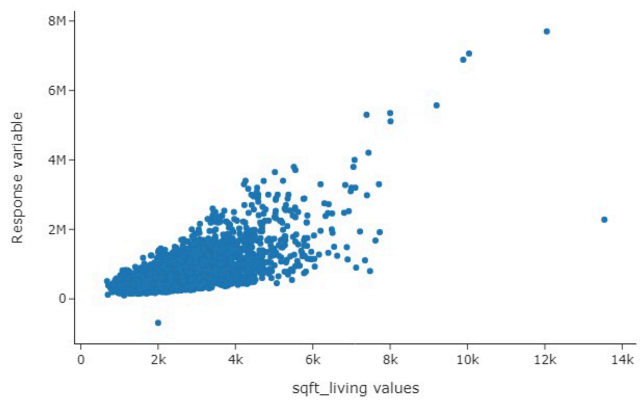
לאחר בחינת חלוק הפיזור ניתן להסיק ש sqft-living הוא מאפיין מיטיבי עם המודל משום שהוא מראה מתאם חיובי ברור עם מחירי קנייה, ככל sqft-living עולה כך עולה מחיר הבית, ומצד שני, מספר חדרי שינה לא נראה תכונה מועילה מאוד כפי שהוא מראה מתאם חלש עם מחירי הבתים. יש שונות רבה במחירי הבתים עבור בתים עם אותו מספר חדרי השינה

Correlation between bedrooms and variable 0.248916444266802



אנו מניחים שמצב אמת פלוני בגלל יותר מתאים לבית יקר יותר, ובגלל רק המתאם בין שני אלה גבוה (≈ 0.71)

Correlation between sqft_living and variable 0.714834305066782



4. Fit a linear regression model over increasing percentages of the *training set* and measure the loss over the *test set*:
   - Iterate for every percentage $p = 10\%, 11\%, \ldots, 100\%$ of the training set.
   - Sample $p\%$ of the train set. You can use the `pandas.DataFrame.sample` function.
   - Repeat sampling, fitting and evaluating 10 times for each value of p.

   Plot the mean loss as a function of $p\%$, as well as a confidence interval of $mean(loss) \pm 2 * std(loss)$. If implementing using the `Plotly` library, see how to create the confidence interval in Chapter 2 - Linear Regression code examples.

   Add the plot to the `Answers.pdf` file and explain what is seen. Address both trends in loss and in confidence interval as function of training size. What can we learn about the estimator $\hat{y}_i$ in terms of estimator properties?

הדגמה הראשית היא עבור ערכת נתונים ב קטנה יותר השגיאה גבוהה יותר וזה נכון עד כדי ספק אך

נראה השגיאה מגיעה לערך קבוע כלשהו זה יכול להיות בגלל הסיבה כי יותר מידע יותר דגימות לא מצביע על התמונה קוגה יותר.

כאשר שגיאת הבדיקה פוחתת משמעות הדבר שהמודל משתפר בחיזוי התוצאות של נקודות נתונים חדשות שהוא לא ראה בעבר זה דבר טוב שכן הוא מצביע על כך שהמודל הופך למדויק יותר. עם זאת אם השונות בתחזית גם פוחתת זה מעיד על כך שהמודל הופך עקבי יותר ורגיש פחות אבל צריך לאזן בין הדיוק לרגישות.

בציור אנו שככל הבדיקה פוחתת גם כן השונות פוחתת