תרגיל 1 - IML, הדס אקשטיין 207848615

1.1 - רקע מתמטי:

<u>שאלה 1:</u>

1. Calculate the SVD of the following matrix A. That is, find the matrices $U,\Sigma,V^\top$ where U,V are orthogonal matrices and Σ diagonal.

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

Recall, that to find the SVD of A we can calculate $A^\top A$ to deduce V,Σ and then calculate $AA^\top$ to deduce U. Equivalently, once we deduced V,Σ we can fine U using the equality $AV = U\Sigma$.

- נחשב A^TA:

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

- ערכים עצמיים של A^TA:

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

$$= (2-\lambda)((2-\lambda)(4-\lambda)-4) - 0 + 2(0-2(2-\lambda)) =$$

$$= (2-\lambda)^2(4-\lambda) - 4(2-\lambda) - 4(2-\lambda) =$$

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

$$= (2-\lambda)(\lambda^2-6\lambda) = -\lambda(\lambda-6)(\lambda-2) = 0$$

$\boxed{\lambda=0}$, $\boxed{\lambda=6}$, $\boxed{\lambda=2}$

- נמצא ו"ע של $\lambda=6$:

$$(6I - A^TA)\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{bmatrix} 4 & 0 & -2 \\ 0 & 4 & 2 \\ -2 & 2 & 2 \end{bmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = 0$$

$$\Rightarrow \begin{cases} 4x - 2z = 0 \\ 4y + 2z = 0 \\ -2x + 2y + 2z = 0 \end{cases} \Rightarrow \begin{matrix} 2x - z = 0 \\ 2y + z = 0 \\ -x + y + z = 0 \end{matrix} \Rightarrow \left.\begin{matrix} x = \frac{1}{2}z \\ y = -\frac{1}{2}z \end{matrix}\right\}$$

$x = \frac{1}{2}z$, $y = \frac{1}{2}z$ לכן $\begin{pmatrix}1\\-1\\2\end{pmatrix}$ זה הו"ע של 6

ונקבל: $v_1 = \frac{1}{\left\|\begin{pmatrix}1\\-1\\2\end{pmatrix}\right\|}\begin{pmatrix}1\\-1\\2\end{pmatrix} = \frac{1}{\sqrt6}\begin{pmatrix}1\\-1\\2\end{pmatrix}$

• נמצא ו"ע ל $\lambda=2$: $(2I - A^TA)\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{bmatrix}0&0&-2\\0&0&2\\-2&2&-2\end{bmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = 0$

$$= \begin{cases}-2z=0\\2z=0\\-x+y-z=0\end{cases} \Rightarrow \begin{matrix}z=0\\x=y\end{matrix}$$

נקבל $\begin{pmatrix}1\\1\\0\end{pmatrix}$ ו"ע המתאים ל- $\lambda=2$. ונקבל $v_2 = \frac{1}{\left\|\begin{pmatrix}1\\1\\0\end{pmatrix}\right\|}\begin{pmatrix}1\\1\\0\end{pmatrix} = \frac{1}{\sqrt2}\begin{pmatrix}1\\1\\0\end{pmatrix}$

• נמצא ו"ע ל $\lambda=0$: $(0\lambda - A^TA)\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{bmatrix}2&0&2\\0&2&-2\\2&-2&4\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}$

$$\Rightarrow \begin{cases}x+z=0\\y-z=0\\x-y+2z=0\end{cases} \Rightarrow \begin{matrix}x=-z\\y=z\end{matrix}$$

נקבל $\begin{pmatrix}-1\\1\\1\end{pmatrix}$ ו"ע המתאים ל-0, ונקבל: $v_3 = \frac{1}{\left\|\begin{pmatrix}-1\\1\\1\end{pmatrix}\right\|}\begin{pmatrix}-1\\1\\1\end{pmatrix} = \frac{1}{\sqrt3}\begin{pmatrix}-1\\1\\1\end{pmatrix}$

סה"כ קיבלנו $V^t = \begin{bmatrix}\frac{1}{\sqrt6}&-\frac{1}{\sqrt6}&-\frac{2}{\sqrt6}\\\frac{1}{\sqrt2}&\frac{1}{\sqrt2}&0\\-\frac{1}{\sqrt3}&\frac{1}{\sqrt3}&\frac{1}{\sqrt3}\end{bmatrix}$

• נחשב AA^T: $AA^T = \begin{bmatrix}1&1&0\\1&-1&2\end{bmatrix}\begin{bmatrix}1&1\\1&-1\\0&2\end{bmatrix} = \begin{bmatrix}2&0\\0&6\end{bmatrix}$.

$\Sigma = \begin{bmatrix}\sqrt6&0&0\\0&\sqrt2&0\end{bmatrix} \Leftarrow \sigma_1=\sqrt2,\ \sigma_2=\sqrt6$

שורש הע"ע = ערכים סינגולרים.

• נמצא ו"ע של ע"ע 2,6: אלו הם, $\begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix} \Leftarrow$

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

שאלה 2:

2. Show that the outer product of two vectors $\mathbf{v}\in\mathbb{R}^n, \mathbf{u}\in\mathbb{R}^m$, which is denoted by $\mathbf{v}\otimes\mathbf{u}$ or $\mathbf{v}\cdot\mathbf{u}^\top$ is a matrix $A\in\mathbb{R}^{n\times m}$ with $rank(A)=1$. That is, show that all rows (or columns) in A are linearly dependent.

יהיו $u\in R^m$, $v\in R^n$

אם $v=0$ או $u=0$ אזי $u\otimes v=0$. אחרת-

העמודה ה-i של $u\otimes v$ הינה $\begin{pmatrix} u_i v_1 \\ \vdots \\ u_i v_n \end{pmatrix} = u_i v$

כלומר כל העמודות שייכות ל- $span(v)$ ולכן כ״ל

ומכאן ש- $rank(u\otimes v)=1$.

(*) נוכל לראות זאת גם ע״י פירוק לגורמים

$$A = v\cdot u^T = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}(u_1 \cdots u_m) = \begin{pmatrix} u_1 v_1 & u_2 v_1 & \cdots & u_m v_1 \\ u_1 v_2 & u_2 v_2 & \vdots & \vdots \\ \vdots & \vdots & u_3 v_3 & \vdots \\ u_1 v_n & u_2 v_n & \vdots & u_m v_n \end{pmatrix} =$$

נשים לב כי לכל $1\le i\le n$ [illegible] כל שורה i ב- $v_i\neq 0$

(אם $v_i=0$ $\Leftarrow$ השורה ה-i היא שורת אפסים)

כעת, יש שורה i אם אז שורת אפסים - אזי היא תלויה לינארית.

אחרת, היא שווה ל- u^T , אם יש יותר משורה אחת השווה ל-u^T כ״ל

לינארית עם האחרות (ע״י מספר)

ולכן מתקבל כי הר $1 = rank(A)$

שאלה 3:

3. Show that for any otrthonormal basis $(\mathbf{u}_1, \dots, \mathbf{u}_n)$ and any aribtrary vector $\mathbf{x} \in \mathbb{R}_n$ such that $\mathbf{x} = \sum_{i=1}^{n} a_i \cdot \mathbf{u}_i$, it holds that $a_i = \langle \mathbf{x}, \mathbf{u}_i \rangle$ for any $i \in [1, n]$. That is, show that the i'th coefficient of representing $\mathbf{x}$ in the basis $(\mathbf{u}_1, \dots, \mathbf{u}_n)$, is the inner product between $\mathbf{x}$ and $\mathbf{u}_i$.

יהי $(u_1, \dots, u_n)$ בסיס אורתונורמלי, ויהי $x \in \mathbb{R}_n$ כך ש- $x = \sum_{i=1}^{n} a_i u_i$

צ"ל $a_i = \langle x, u_i \rangle$ לכל $i \in [1, n]$

יהי $i \in [1, n]$.

מכיוון ש- $x = \sum_{j=1}^{n} a_j u_j =$

$$x = \sum_{\substack{j=1 \\ j \neq i}}^{n} a_j u_j + a_i u_i \iff a_i u_i = x - \sum_{\substack{j=1 \\ j \neq i}}^{n} a_j u_j \iff$$

$$\langle u_i, a_i u_i \rangle = \langle u_i, x - \sum_{\substack{j=1 \\ j \neq i}}^{n} a_j u_j \rangle \iff a_i \langle u_i, u_i \rangle = \langle u_i, x \rangle - \sum_{\substack{j=1 \\ j \neq i}}^{n} \langle u_i, u_j \rangle$$

$$\underset{(*)}{\iff} a_i = \langle u_i, x \rangle$$

(*) נשים לב כי מכך ש- $(u_1, \dots, u_n)$ אורתונורמלית אנו מקבלים כי לכל i, j:

$$\langle u_j, u_i \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

שאלה 4:

(a) Let $x = \begin{bmatrix} 3 \\ -4 \\ 1 \\ -2 \end{bmatrix} \in \mathbb{R}^4$. Compute:

- $\|x\|_1$
- $\|x\|_2$
- $\|x\|_\infty$

Which of the norms gives the smallest value? Why?

$\|x\|_1 = 3 + 4 + 1 + 2 = 10$, $\|x\|_2 = \sqrt{3^2 + (-4)^2 + 1^2 + (-2)^2} = \sqrt{30}$, $\|x\|_\infty = 3$

לכן $\|x\|_\infty$ נותן את הערך הנמוך ביותר.

למה? בגלל והוא נותנת ערך נמוך יותר מ- $\|x\|_1$ כיוון שהסכום של ערכים מוחלטים

וחיובי גדול מכל אחד מהם. וגם $\|x\|_2$ מעלה בריבוע כל ערך חיובי ואז מוציא שורש

עד מה גדול יותר.

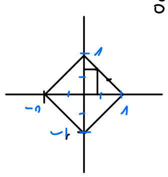
(ב) 2D space, the **unit ball** under a norm is the set of vectors $x \in \mathbb{R}^2$ such that $\|x\| = 1$. The shape of this unit ball depends on which norm is being used.

i. Sketch the unit ball in $\mathbb{R}^2$ for:
   - ℓ_1 norm
   - ℓ_2 norm
   - ℓ_∞ norm

   How do these shapes reflect the way each norm measures "length"? *(Tip: Think of how $\|x\| = 1$ looks geometrically in each case.)*

• ℓ_1 norm :

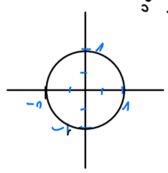
(i)



הנורמה הזאת אורך ה' סכום ערכים מוחלטים, כלומר
מודדת מרחק באותו הצעדים לצורה אלכסונית בין (0,1) (1,0) (0,-1) (-1,0)
6 השבה הזה יותר

• ℓ_2 norm :

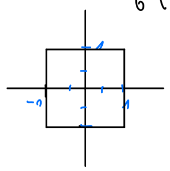
(i)



נראה ש- $\sqrt{x^2+y^2}=1$
הנורמה הזאת אורך באופן אוקלידי. ומודד מעגל הנקודה - מודדת מרחק
באופן אחיד ושווה

• ℓ_3 norm :

(i)



נראה שהמקס' יהיה 1.
הנורמה הזאת אורך ה' לוקחת הערך המוחלט הגדול ביותר ומתעלמת מהשני של הזוג.
כלומר שבתוך ה unit ball 6 כל המרחק הוא עד 1 וגם לכן לא יש נקודה מחוץ

שאלה 1ב

(b) Use the chain rule to calculate the gradient of $h(\sigma) = \frac{1}{2}||f(\sigma) - y||^2$, where $\sigma \in \mathbb{R}^d$ and f is some arbitrary function from $\mathbb{R}^d$ to $\mathbb{R}^d$.

• נגדיר: $g : \mathbb{R}^d \to \mathbb{R}$ ע"י: $g(x) = \frac{1}{2}||x||^2$

כפי שראינו בתרגול, מתקיים: $J_x(g) = \frac{1}{2} \cdot 2 \cdot x^t = x^t$

• כעת, מכלל השרשרת:

$$J_\sigma(h) = J_{f(\sigma)-y}(g) \cdot J_\sigma(f-y) \overset{\text{ה-y קבוע}}{=} (f(\sigma)-y)^t \cdot J_\sigma(f)$$

ולכן:

$$\nabla h(\sigma) = J_\sigma(h)^t = J_\sigma(f)^t (f(\sigma)-y)$$

(c) In recitation 1 we saw the softmax function $S : \mathbb{R}^k \to [0,1]^k$, which is defined as follows:

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

This function takes an input vector $x \in \mathbb{R}^d$ and outputs a probability vector (non-negative entries that sum up to 1), corresponding to the weight of original entries of x.

Question: Calculate the Jacobian of the softmax function S.

$$J_x(S) = \begin{array}{cc} & k \\ & \begin{bmatrix} \frac{\partial S_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_k} \\ \vdots & & \\ \frac{\partial f_k(x)}{\partial x_1} & \cdots & \frac{\partial f_k(x)}{\partial x_k} \end{bmatrix} \\ k & \end{array} =$$

$$\frac{\partial f_m(x)}{\partial x_n} \circ = \underset{\boxed{m \neq n}}{=} \frac{\partial \left(\frac{e^{x_m}}{\sum_{i \in [k]} e^{x_i}} \right)}{\partial x_n} = - \frac{e^{x_m} \circ e^{x_n}}{(\sum_{i \in [k]} e^{x_i})^2} = - \underbrace{\frac{e^{x_m}}{\sum_{i \in [k]} e^{x_i}}} \circ \underbrace{\frac{e^{x_n}}{\sum_{i \in [k]} e^{x_i}}}$$

$$\underset{\boxed{n=m}}{=} \frac{e^{x_m} (\sum_{i \in [k]} e^{x_i}) - e^{x_m} \circ e^{x_m}}{(\sum_{i \in [k]} e^{x_i})^2} = \frac{e^{x_m}}{\sum_{i \in [k]} e^{x_i}} \cdot \frac{(\sum_{i \in [k]} e^{x_i} - e^{x_m})}{\sum_{i \in [k]} e^{x_i}} = \underbrace{\frac{e^{x_m}}{\sum_{i \in [k]} e^{x_i}}} \cdot \underbrace{\left(1 - \frac{e^{x_m}}{\sum_{i \in [k]} e^{x_i}}\right)}$$

$$J_x(S)_{ij} = \begin{cases} S_i(x)(1 - S_i(x)) & i = j \\ -S_i(x) \cdot S_j(x) & i \neq j \end{cases} \quad \Leftarrow$$

רגרסיה לינארית

1.2 Linear Regression

Based on Lecture 1 and Recitation 2

Let $\mathbf{X}$ be the input 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 $\mathbf{X}$.

1.2.1 Solutions of The Normal Equations

5. In (a-d) you will incrementally prove several important properties regarding the solutions of the normal equations.

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

(a) יהי $u \in Ker(X)$. ($u \in \mathbb{R}^d$, $X \in \mathbb{R}_{m \times d}$)

אזי $X \cdot u = 0$ ומתקיים:

$$X^T \cdot X \cdot u = X^T(X \cdot u) = X^T \cdot 0 = 0$$

לכן $u \in Ker(X^TX)$ $\Leftarrow$ $Ker(X) \subseteq Ker(X^TX)$

יהי $u \in Ker(X^TX)$. ($u \in \mathbb{R}^d$, $X^TX \in \mathbb{R}_{d \times d}$)

אזי $X^TXu = 0$ לכן בפרט:

$$u^TX^TX \cdot u = 0 \Rightarrow (Xu)^t \cdot Xu = 0 \Rightarrow$$

$$\|X \cdot u\|^2 = 0 \Rightarrow Xu = 0$$

ולכן $u \in Ker(X)$ $\Leftarrow$ $Ker(X^TX) \subseteq Ker(X)$

הראינו הכלה דו כיוונית $\Leftarrow$ $Ker(X^TX) = Ker(X)$

(b) Prove that for a square matrix A: $Im\left(A^\top\right) = Ker(A)^\perp$

(b) יהי $u \in Im(A^T)$,

אזי קיים v כך ש $A^T \cdot v = u$

אזי לכל $w \in Ker(A)$ מתקיים:

$$\langle u, w \rangle = u^t w = (A^t v)^t w = v^t A w = v^t(Aw) = v^t 0 = 0$$

לכן $u \in ker(A)$ $\Leftarrow$ $Im(A^T) \subseteq ker(A)^\perp$

כעת, נזכר כי $\dim(Ker(A)^{\perp}) = Rank(A) = Rank(A^{t}) = \dim(Im(A^{t}))$

לכן, משיקולי מימד נקבל ש $Im(A^{t}) = Ker(A)^{\perp}$

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

נניח כי למערכת ∞ פתרונות.

אזי מצד שני כי $y \in Im(X)$ (כיון ש-X לא הפיכה אז או שאין פתרון או שיש ∞ פתרונות)

מסעיף קודם - $Im(X) = Ker(X^{t})^{\perp}$

ולכן $y \perp Ker(X^{t})$

נניח כי $y \perp Ker(X^{T})$.

אזי נקבל כי $Im(X) = Ker(X^{t})^{\perp} \ni y$, כלומר קיים w כך ש $Xw = y$.

כיון ש-X אינה הפיכה מתקיים $Ker(X) \neq \{0\}$.

עבור $0 \neq u \in Ker(X)$ מתקיים: $X(w+u) = Xw + Xu = y + 0 = y$

ולכן מקבלים ∞ פתרונות למשוואה.

(d) 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).

אם $X^{T}X$ הפיכה אזי למשוואה יש פתרון יחיד (טריוויאלי)

אם $X^{T}X$ אינה הפיכה אזי מהסעיף הקודם למשוואה יש ∞ פתרונות $\Leftrightarrow X^{T}y \perp Ker(X^{T}X)$

$X^{T}y \perp Ker(X^{T}X) \underset{\text{סעיף a}}{\Leftarrow} X^{T}y \perp Ker(X) \Leftarrow X^{T}y \in Ker(X)^{\perp} \underset{\text{סעיף b}}{\Leftarrow} X^{T}y \in Im(X^{T})$

$X^{T}y \perp Ker((X^{T}X)^{T}) \underset{X^{T}X \text{ סימטרית}}{\Leftarrow}$

ולכן מסעיף (c) נקבל כי למשוואה ∞ פתרונות.

1.2.2 Least Squares

Given a sample $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^m$, the ERM rule for linear regression w.r.t. the squared loss is

$$\hat{\mathbf{w}} \in \underset{\mathbf{w}\in\mathbb{R}^d}{\operatorname{argmin}} \; ||\mathbf{X}\mathbf{w} - \mathbf{y}||^2$$

where $\mathbf{X}$ is the input matrix of the linear regression with rows as samples and y the vector of responses. Let $\mathbf{X} = U\Sigma V^\top$ be the SVD of $\mathbf{X}$, where U is a $m \times m$ orthonormal matrix, Σ is a $m \times d$ diagonal matrix, and V is an $d \times d$ orthonormal matrix. Let $\sigma_i = \Sigma_{i,i}$ and note that only the non-zero σ_i-s are singular values of $\mathbf{X}$. Recall that the pseudoinverse of $\mathbf{X}$ is defined by $\mathbf{X}^\dagger = V\Sigma^\dagger U^\top$ where $\Sigma^\dagger$ is an $d \times m$ diagonal matrix, such that

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

6. Assuming that $\mathbf{X}^\top\mathbf{X}$ is invertible, show that the general solution we derived in recitation 3 $(\mathbf{X}^\dagger\mathbf{y})$ equals to the solution you have seen in lecture 1 $\left([\mathbf{X}^\top\mathbf{X}]^{-1}\mathbf{X}^\top\mathbf{y}\right)$.

נניח כי X^TX הפיכה.

$$X^TX = (U\Sigma V^t)^t(U\Sigma V^t) = V\Sigma^tU^tU\Sigma V^t \underset{\downarrow}{=} V\Sigma^t\Sigma V^t$$

U אורתונורמלית $\Rightarrow U^tU = I$

וכן
$$(X^TX)^{-1} = (V\Sigma^t\Sigma V^t)^{-1} \underset{\downarrow}{=} V(\Sigma^t\Sigma^t)^{-1}V^t$$

V אורתונורמלית

ולכן:
$$[X^TX]^{-1}X^T = V(\Sigma^t\Sigma^t)^{-1}V^t(U\Sigma V^t)^t = V(\Sigma^t\Sigma^t)^{-1}V^tV\Sigma^tU^t \underset{\downarrow}{=}$$

V אורתונורמלית

$$= V(\Sigma^t\Sigma)^{-1}\Sigma^tU^t$$

כעת, מכיוון ש- Σ אלכסונית אזי $\Sigma^t = \Sigma$ (וכן עבור $\sigma_1, \dots, \sigma_d$ הערכים הסינגולריים)

$$(\Sigma^t\Sigma)_{ii} = \begin{cases} \sigma_i^2 & \sigma_i \neq 0 \\ \sigma_i & \sigma_i = 0 \end{cases}, \quad \Sigma^t_{ii} = \Sigma_{ii} = \begin{cases} \sigma_i & \sigma_i \neq 0 \\ 0 & \sigma_i = 0 \end{cases}$$

$$\Leftarrow \quad (\Sigma^t\Sigma)^{-1}_{ii} = \begin{cases} \frac{1}{\sigma_i^2} & \sigma_i \neq 0 \\ 0 & \sigma_i = 0 \end{cases} \Leftarrow \; ((\Sigma^t\Sigma)^{-1}\Sigma^t)_{ii} = \begin{cases} \frac{1}{\sigma_i} & \sigma_i \neq 0 \\ 0 & \sigma_i = 0 \end{cases} \overset{\text{נסק}}{=} \Sigma^+_{ii}$$

ולכן
$$(X^tX)^{-1}X^ty = V(\Sigma^t\Sigma)^{-1}\Sigma^tU^ty = U\Sigma^+U^ty = X^+y$$

החלק הפרקטי:

אימון מודל רגרסיה לינארית

(3) עיבוד מקדים:

- הפיצ'רים שבחרתי להסיר: id, date, lat, long, yr_built, yr_renovated

  שכן הפיצ'רים - id, date לדעתי אינם מספקים מידע שימושי הגורם לתחזית מחיר הבית ולכן אינם רלוונטיים למודל.

  וכן, הפיצ'רים - long, lat הן קואורדינטות גיאוגרפיות ומודל רגרסיה לינארית מניח קשר לינארי בין הפיצ'רים לתגובה. (המיקום הגיאוגרפי משפיע בצורה יותר מורכבת למשל, כדי שיהיו משמעות יותר ניתן היה לפרק אותם לפי- המרחק ממרכז, קירבה לחוף ים/בתי ספר וכו'..)

  בנוסף, הפיצ'רים yr_built, yr_renovated הוסרו כיון שלדעתי לא נותנים מידע שימושי יעיל למודל בצורתם הגולמית. מהם הפקתי פיצ'רים חדשים שמשמעותם תורמת יותר.

  את שאר הפיצ'רים המקוריים שמאפיינים בית וכוללים ערכים תורמים יותר למודל השארתי (כמו condition, grade, bathrooms, sqft_living ועוד) כיון שהם תורמים להסבר המחיר.

- הפיצ'רים שבחרתי להוסיף: decade, renovated_last_decade, bedrooms_to_living_ratio

  משתנה decade - העשור שבו נבנה הבית על בסיס yr_built כדי לשקף סגנון בנייה/איכות חומרים/תכנון שמאפיינים באותה תקופה ולכן השנה עצמה פחות אינדיקטיבית.

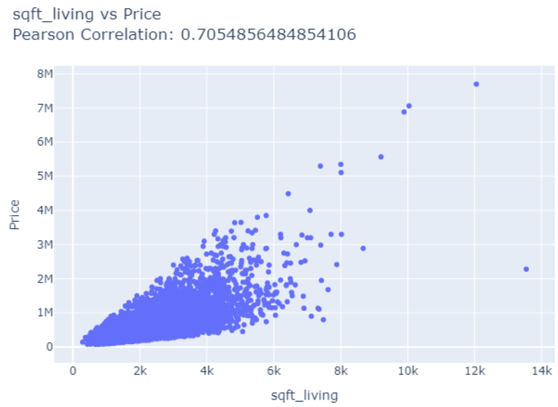
  ו- renovated_last_decade - האם הבית שופץ בעשור האחרון לפי yr_renovated - שיפוץ עדכני יכול להשפיע על מחיר גבוה יותר ועשוי להיות אינדיקטיבי יותר למודל לדעתי.

  וכן bedrooms_to_living_ratio - יחס בין מס' חדרי השינה לשטח המגורים - כאשר פיצ'ר זה יכול לבטא את גודל החדרים הממוצע בבית שלדעתי פיצ'ר זה טוב לאפיון את הבית ושימושי לחיזוי מחיר הבית. (ניתן לראות קורלציה גבוהה)

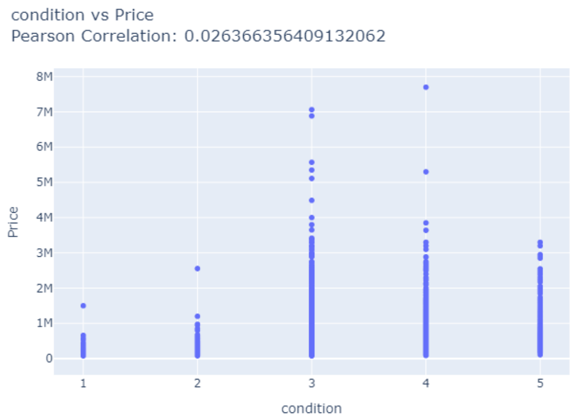
- טיפול בערכים חסרים/לא תקינים - מנקה שורות שבהן יש ערכים חסרים, וכן שורות עם ערכים לא סבירים - למשל ערכים שליליים או אפס במקום גודל או מחיר, ערכים חריגים בטווחים של grade, condition, view וכו'.. (לפי הגבולות של הפיצ'רים המוגדרים על השאלות הקיים ב- kaggle)

  בנוסף, לגבי preprocess_test כיון שלא ניתן למחוק שורות - החלפתי את ערכי ה-NaN לערך הממוצע של הפיצ'ר לפי הדגימות מה-train

(4) פיצ'ר מועיל למודל - sqft_living - פיצ'ר זה מראה קורלציה חיובית חזקה עם המחיר, הגרף מציג מגמה ברורה של עלייה במחיר ככל ששטח המגורים גדל. מקדם פירסון החיובי והגבוה מעיד על קשר חזק בין הפיצ'ר למחיר ולכן הוא מועיל למודל לדעתי.
(מקדם פירסון: 0.705) - קרוב ל-1 ולכן יש קשר חזק



פיצ'ר שפחות מועיל למודל - condition - הגרף מציג פיזור אקראי של נקודות ללא מגמה ברורה. מקדם פירסון הנמוך מעיד על הקשר החלש מאוד בין הפיצ'ר למחיר וכן ביחס לפיצ'רים האחרים ולכן פיצ'ר זה הכי פחות מועיל למודל לדעתי.
(מקדם פירסון: 0.036) - קרוב לאפס ולכן יש קשר חלש



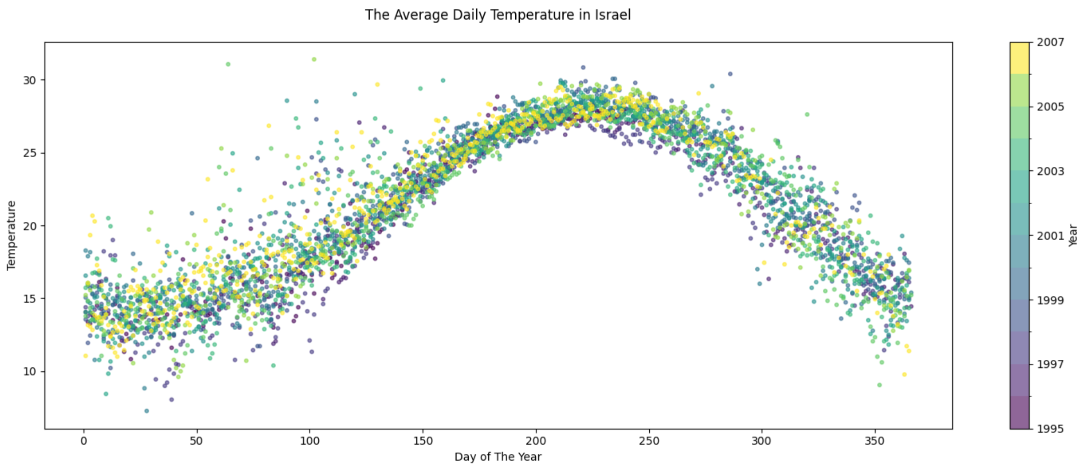
(6) בגרף ניתן לראות שככל שאחוז הדגימות מסט האימון עולה כך ממוצע השגיאה הריבועית הממוצעת יורד ומתייצב. זאת מכיוון שככל שסט האימון גדל יותר, המודל מקבל יותר מידע על הנתונים ולכן מסוגל לבצע תחזיות מדויקות יותר

בנוסף, רצועת הביטחון רחבה יותר עבור אחוזים הקטנים של סט האימון אך מצטמצמת ככל שגודל סט האימון גדל. המשמעות היא שכאשר סט האימון קטן יש יותר שונות בתוצאות בין הדגימות השונות ולכן הביטחון בתחזיות נמוך יותר. ככל שסט האימון גדל, השונות בין הדגימות קטנה והתחזיות הופכות לעקביות יותר.

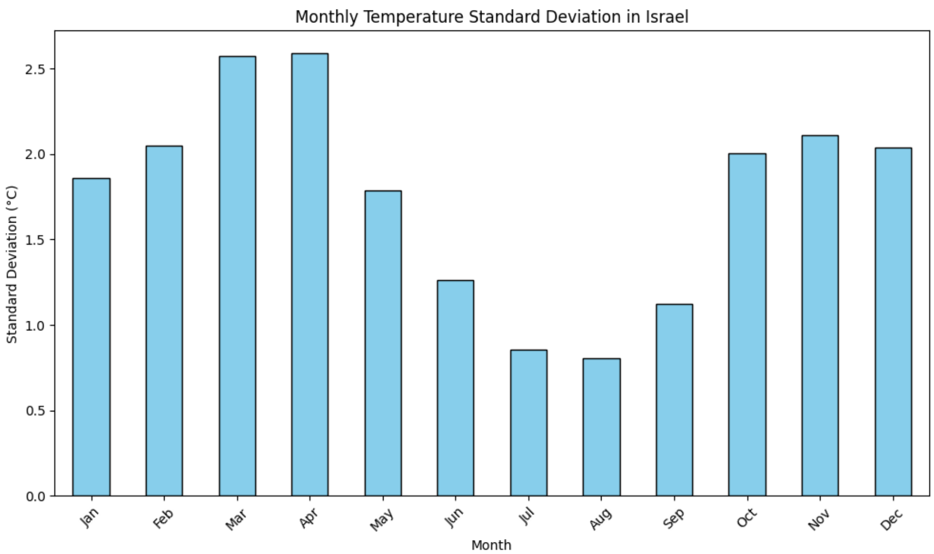
ניתן גם להבחין כי לאחר אחוז מסוים (70%-80%) השיפור בביצועים מתייצב, ולאחר הוספת נתונים נוספים אינה משפרת משמעותית את המודל. ככל שסט האימון גדל יותר המודל הופך ליציב יותר כפי שניתן לראות בצמצום רצועת הביטחון.

אימון מודל פולינומיאלי:

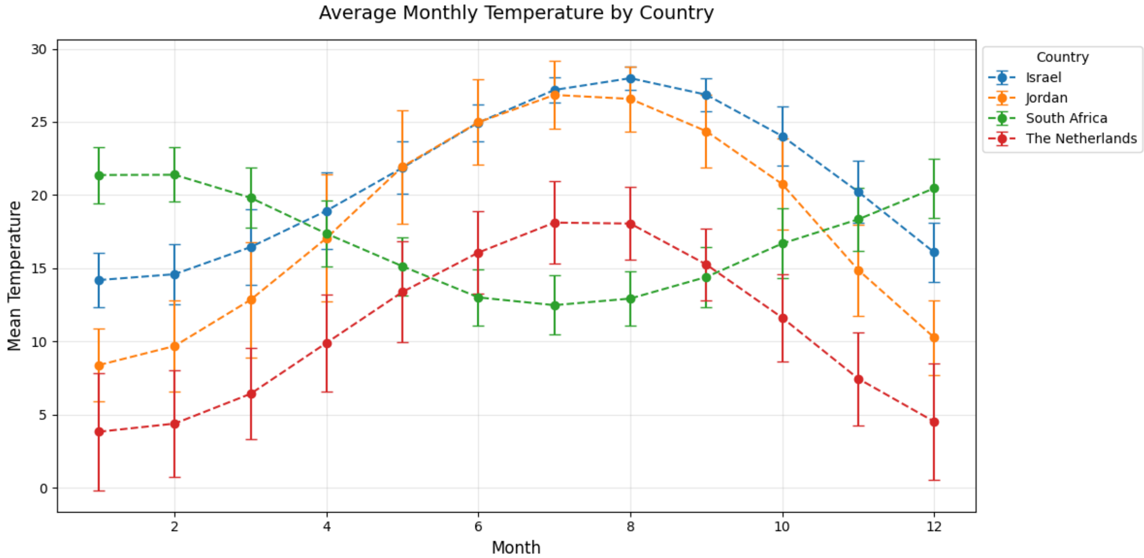
(3) בגרף אנו רואים את הטמפרטורה הממוצעת היומית בישראל כפונקציה של יום בשנה כאשר הצבעים מייצגים את השנים השונות.
דרגת הפולינום גבוהה מ-2 כיוון ש-2/1 לא יוכלו לבטא את העיקול הנראה וכי ניתן לראות כי כל הנראה לגרף יש 2 נק' קיצון ולכן בהכרח מדרגה 3 לפחות.
בנוסף, דרגה גדולה מ-6 כבר עשויה להכניס תנודות מיותרות שלא קיימות באמת. ועל כן, להערכתי דרגת הפולינום תהיה בין 3 ל-6.



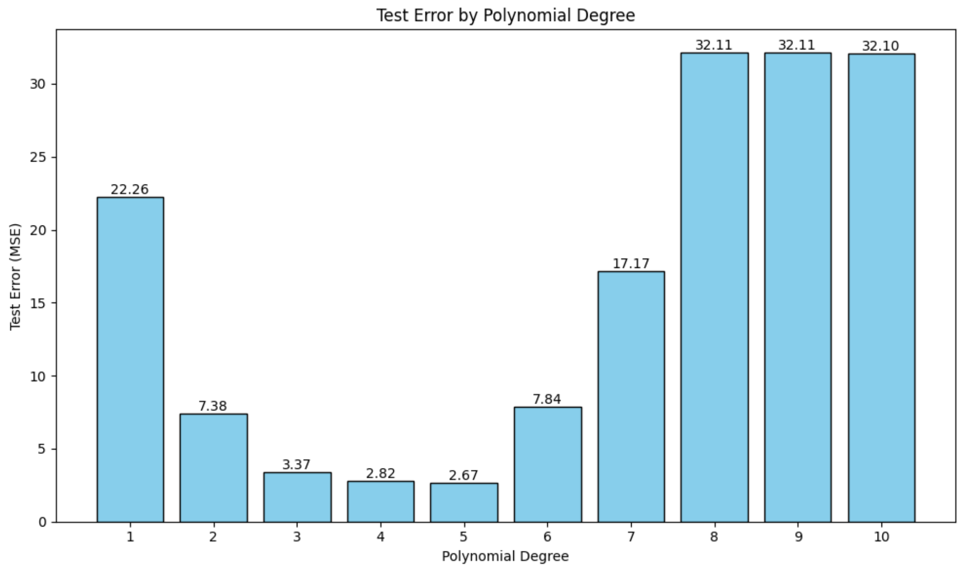
בגרף הזה אנו רואים את סטיית התקן של הטמפ' הממוצעת לפי חודש.
ראינו כי ישנו קשר ישיר בין הסטייה לשונות - סטיית תקן נמוכה - השונות נמוכה ומצביעה על התנהגות יציבה, כלומר קרוב יותר לממוצע
ואילו כאשר סטיית התקן גבוהה - השונות גבוהה וההתנהגות פחות יציבה, לכן קשה יותר לחזות נכון.
ניתן לראות שסטיית התקן גבוהה במיוחד בחודשים מרץ-אפריל וגם באוקטובר-דצמבר, לכן המודל על שונות גבוהה זמן רב יהיה קשה יותר
לחזות את הטמפרטורה במדויק
לעומת שבחודשי הקיץ (יולי-אוגוסט) סטיית התקן נמוכה - השונות נמוכה ולכן המודל צפוי להצליח לחזות בצורה מדויקת יותר את הטמפרטורה בזמן זה.



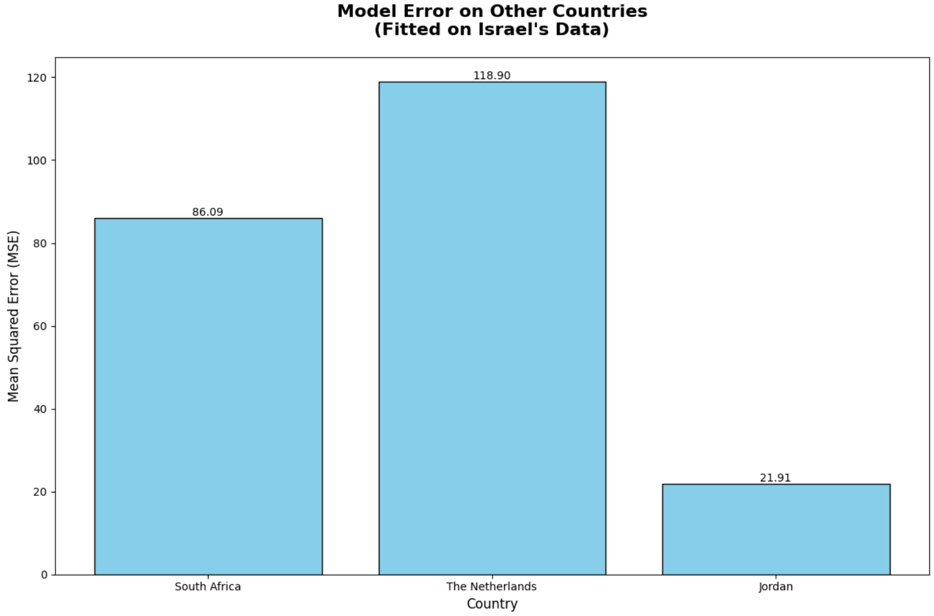
(4) מהגרף ניתן לראות שדפוס השינוי בטמפרטורה בישראל ירדן והולנד מאוד דומה במהלך השנה, כאשר הטמפרטורות גבוהות בחודשי הקיץ יוני - ספטמבר ולעומת זאת בדרום אפריקה בחודשים אלו מגיעות הטמפרטורות הנמוכות ביותר באותה מדינה בנוסף, ניתן להבחין כי אמנם המגמה של השינוי בטמפ' בישראל, ירדן והולנד דומה אך הטמפ' בהולנד הרבה יותר נמוכות במשך כל השנה מבישראל וירדן. אך מן. סביר שהממוצע בישראל יעקוב אחר ירדן (ואולי גם של הולנד אך עם התאמה בשגי')



(5) הגרף הנ"ל מתאר את הקשר בין דרגת הפולינום ל-loss. ניתן לראות כי פולינום ממעלה 5 מספק loss מינימלי, ולכן עבור הנתון הזה בתחום יהיה הכי דומה למודל האמיתי.
נציין כי ניתן להתלבט לבחור גם בפולינומים מדרגה 4/3 שכן הם מספקים loss די דומה ודורשים פחות חישובים לבניית המודל



(ה) הגרף מתאר את ערך ה- mse של המודל שאומן על ישראל לכל אחת משאר המדינות.
ניתן לראות כי בדומה למסקנה בסעיף (ד) המודל מספק תחזית הרבה יותר טובה לירדן בהשוואה לשאר המדינות בשל הקורלציה
הגבוהה בין המדד שלה ושל ישראל בהשוואה לשאר המדינות.
(לעומת זאת, בשאלה (ד) ראינו שהמגמות במדד של ישראל ושל הולנד דומות על אף שהמדד נמוכות בהרבה משל ישראל,
אך נראה שההבדל בין המדדים משקל רב יותר מאשר המגמה מכיוון שבגלל הבדלי את המודל המותאם לישראל על הולנד התקבלו השגיאות הגבוהות ביותר.)



* שימוש בכלי AI:

נעזרתי ב- chatGPT ע"מ לסדר את נראות התרשימים שיצרתי והוספת פרטים שונים אליהם
(כמו הוספת ערך השגיאה מעל התרשימים, הוספת ריבוע עבור המספר מה ט קו בתרשים מסוים)