

336224076 ננוינן ריב

1 place

$$1) A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} \Rightarrow A^T = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{bmatrix}$$

$V, \Sigma, V^T$  גורם קובע,  $A$  ב-SVD  $\Rightarrow$  גורם אלטנקי

$V, \Sigma$  גורם קובע קובע

$$AA^T = \underbrace{\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix}}_{2 \times 3} \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{bmatrix}}_{3 \times 2} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} = \overset{\text{"N" 0}}{\beta}$$

$$A^T A = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{bmatrix}}_{3 \times 2} \cdot \underbrace{\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix}}_{2 \times 3} = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix} = \overset{\text{"N" 0}}{\beta}$$

$\beta$  ב-  $\sigma_1 \sigma_2 \dots \sigma_r$  אלטנקי

$$|\beta - nI| = 0$$

$$\begin{vmatrix} 2-h & 0 & 2 \\ 0 & 2-h & -2 \\ 2 & -2 & 4-h \end{vmatrix} = 0$$

$$(2-h) \begin{vmatrix} 2-h & -2 \\ -2 & 4-h \end{vmatrix} + 2 \begin{vmatrix} 0 & 2-h \\ 2 & -2 \end{vmatrix} = 0$$

$$= (2-h)((2-h)(4-h) - 4) - 4(2-h) = 0$$

$$(2-h) \underbrace{((2-h)(4-h) - 8)}_{\downarrow} = 0$$

$$h=2 \quad h^2 - 6h + 8 - 8 = 0$$

$$h(h-6h) = 0$$

$$h=0, \quad h=6$$

$$\boxed{h=0, 6, 2} \quad \text{10 J''} \quad \text{8 J}$$

"  
δ₁ δ₂ n₁

$n=0$  n₀₀

$$\begin{pmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{pmatrix} \times = 0 \Rightarrow \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \quad \ddot{\delta}$$

$$-x = z, \quad y = z \quad 2z + x = y$$

$n=2$  ת'ור

$$\begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & -2 \\ 2 & -2 & 2 \end{pmatrix} x = 0 \Rightarrow \begin{cases} z = 0 \\ x = y \end{cases} \Rightarrow \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 0 & 2 \\ 0 & -4 & -2 \\ 2 & -2 & -2 \end{pmatrix} x = 0 \Rightarrow \begin{cases} x = z/2 \\ y = -z/2 \end{cases} \Leftrightarrow \begin{pmatrix} 1/2 \\ -1/2 \\ 1 \end{pmatrix} \stackrel{\text{המ}'}{\Leftrightarrow} \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$$

: מינימום ו, נס' המינימום מוגדר בפונקציית

$$V = \begin{bmatrix} 1 & 1/2 & -1 \\ 1 & -1/2 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 6 & 0 \end{bmatrix}$$

, כ' גורם,  $T^T$  פוקטורי  $E_{3N}$

$$AV = V\Sigma$$

$$A v_i = n w_i \Rightarrow \frac{A}{6} v_i = w_i \quad \boxed{n=2G}$$

$$(f_1, f_2, f_3)$$

$$w_1 = \frac{A}{2} \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ 0 \\ 0 \end{bmatrix}$$

$$6 = \sqrt{2}$$

$$U_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{\frac{3}{2}} \end{bmatrix}$$

$$U = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{\frac{3}{2}} \end{bmatrix}$$

אנו מגדירים

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

לצורך

הסברים

הנemme אנו נוגדים בפונקציית ה- $\otimes$

$$V \otimes W = \underbrace{V W^T}_{n \times m}^{m \times n} = B$$

ולפונקציית ה- $\otimes$  נאמר:

$$V \in \mathbb{R}^{n \times r}, W \in \mathbb{R}^{r \times m}$$

המשמעות של ה- $\otimes$  היא:

$$C_{jk} = V_{j1} W_{k1} + V_{j2} W_{k2} + \dots + V_{jr} W_{kr}$$

כלומר רצף וקטור נוצרת מה-

לודג ב' (הדריכת) ב' מילויים

הנורמליזציה כפונקציית נורמליזציה  
 $v_i = \frac{v_i - \bar{v}}{\sigma_v}$  ו-  $\bar{v} = \frac{1}{n} \sum v_i$

$1 \leq i \leq n$  ו-  $v_i$  רצף של נתונים

$\sum_{i=1}^n v_i = w_k$  שווה ל-  $w_k$

• ערך סטטיסטי מסווג לפי סדרה

$\text{Rank}(A) = 1$  ו-  $A^n = A$  ו-  $C = I$

אנו מודדים את ה- $v_i$  ו-  $v_i = 0$  ו-

ולפיה. מבחן סבב נגדי  $\leftarrow a \rightarrow a$  ו-

לעתה גורם לאו

$f_{\mu}$

•  $f_{\mu}$  גן, גן, נספה און ו-  $f_{\mu}$

ונערכו  $w_i, w_j = \int_{\Omega} \phi$

$x = \sum_{i=1}^n \alpha_i w_i$

$\forall i, 1 \leq i \leq n \quad \langle x, w_i \rangle = \langle \sum_{i=1}^n \alpha_i w_i, w_i \rangle$

$= \langle \alpha_1 w_1, w_i \rangle + \langle \alpha_2 w_2, w_i \rangle + \dots + \langle \alpha_n w_n, w_i \rangle$

גפ' כפונקציית

נספה און

$$= \alpha_1 < w_1, w_i > + \alpha_2 < w_2, w_i > + \dots + \alpha_n < w_n, w_i >$$

↑  
כיוון'

$$= 0 + 0 + \alpha_i < w_i, w_i > + 0 = \alpha_i$$

כיוון'

$$\text{לצורך} \quad \alpha_i = \langle x, w_i \rangle \quad \text{הנתק} \quad \text{פוג}$$

4 י.ס.ו.

$$\text{אם } x \in \text{Span } P \text{ אזי } \exists v_1, \dots, v_k \in V \text{ כך ש } x = \sum_{i=1}^k \alpha_i v_i$$

$$P^T = (P)^T = \left( \sum_{i=1}^k v_i \cdot v_i^T \right)^T = \sum_{i=1}^k (v_i \cdot v_i^T)^T = \sum_{i=1}^k (v_i^T)^T \circ v_i^T$$

נזכיר מה  
לעומת

$$\text{משפט} \quad (A \cdot B)^T = B^T A^T$$

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

$v_i^T = v$

$$\text{לכל } P^T = P \quad \text{פוג}$$

2 אם  $n=1, 0 \leq n \leq P$  אזי  $P$  ב- $\mathbb{R}$  נקרא פוג

לפוג מוגדר כ- $\|P\| = \sqrt{\lambda_{\max}(P)}$  ו- $\lambda_{\max}(P) = \max_{1 \leq i \leq n} |P_{ii}|$ .

$$P_{ii} = \left( \sum_j r_i r_j^T \right) v_j = \sum_j v_j \delta_{ij} = v_j$$

הנתק

הנורמליזציה של פונקציית גיבוב היא  $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{\pi}$

$$\text{לפיכך } P(V) = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} V(x) dx = hV$$

לפיכך  $V(x) = h^{-1} P(V)$

$$(h=1 \text{ ו } V=0)$$

במקרה הכללי, ניקח פונקציית גיבוב  $f(x)$  ופונקציית גיבוב  $V(x)$

$$P(V) = \int_{-\infty}^{\infty} f(x) V(x) dx = \int_{-\infty}^{\infty} f(x) \cdot 0 dx = 0 = 0 \cdot V$$

מכיוון ש  $P(V) = 0$ , נקבע  $V(x) = 0$  כפונקציית גיבוב נורמליזованת.

במקרה הכללי, ניקח פונקציית גיבוב  $f(x)$  ופונקציית גיבוב  $V(x)$

$$(f(x) \neq 0, V(x) \neq 0)$$

הנורמליזציה של פונקציית גיבוב היא  $\int_{-\infty}^{\infty} f(x) dx = 1$

$$\text{לפיכך } V(x) = h^{-1} P(V) = h^{-1} \int_{-\infty}^{\infty} f(x) V(x) dx = h^{-1} \int_{-\infty}^{\infty} f(x) \cdot 0 dx = 0$$

$$h=1$$

$v_1, \dots, v_k$  נס.  $\sum_{i=1}^k a_i v_i$  בפ' 3.  $\forall v \in V$  פ' 8.

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

כמ"ג  
אנו מוכיחים  
בנ"ט כפ"ה  
ולא כפ"ג  
כל היותר כפ"ג

משפט. (המשפט הספקטרלי) תהי  $A \in \mathbb{R}^{d \times d}$  ש-סימטרית, או קיימות  $U, D$  ב- $\mathbb{R}^{d \times d}$  ו-עומדים  $A = UDU^T$ .  
 (i)  $U$ : אורתוגונליות שעומודותיה וקטוריהם עצמאיים של  $A$ .  
 (ii)  $D$ : אלכסונית שערכיה מהערבים העצמיים של  $A$ .

$$P^2 = PP = (UDU^T)(UDU^T) = UDU^T \xrightarrow{UU^T=I} UDU^T = P$$

ובנ"ט  $P = D$  כפ"ג  
 $D = I$  כפ"ג  $\Rightarrow P = I$  כפ"ג  
 $\boxed{P^2 = P}$  כפ"ג  $\Sigma$

$$P^2 = P \rightarrow$$

$$(I - P)P = (I - UDU^T)UDU^T \xrightarrow{UDU^T=I} (I - UU^T)UU^T = I$$

$$UU^T = I \quad \text{כפ"ג} \quad \odot$$

## 2.1.2 Multivariate Calculus

Based on Recitation 2

5. 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  $\mathbb{R}^d$ .

6. In recitation 2 we saw the softmax function  $S : \mathbb{R}^k \rightarrow [0, 1]^k$ , 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  $S$ .

12 7.05 2020 11:58 2020

$$\frac{\partial h(\mathbf{f})}{\partial \mathbf{f}_i} = \frac{\partial h}{\partial \mathbf{f}} \circ \frac{\partial \mathbf{f}}{\partial \mathbf{f}_i} = (\mathbf{f}(\mathbf{f}) - \mathbf{y})^\top \cdot \frac{\partial \mathbf{f}}{\partial \mathbf{f}_i} = (\mathbf{f}(\mathbf{f}) - \mathbf{y})^\top \cdot \mathcal{J}_{\mathbf{f}}(\mathbf{f})_i$$

$$\nabla \cdot h(\mathbf{f}) = (\mathbf{f}(\mathbf{f}) - \mathbf{y})^\top \circ \mathcal{J}_{\mathbf{f}}(\mathbf{f})$$

$$2. \quad \mathcal{J}_{\mathbf{x}}(S)_{ij} = \frac{\partial S_i}{\partial x_j}$$

for  $i \neq j$ : sparse off-diagonal terms

$$\frac{\partial S_i}{\partial x_j} = S_i (1 - S_j)$$

$$\boxed{\frac{\partial}{\partial x_i} h = e^{x_i}} \rightarrow \text{softmax} \text{ diff}$$

$j \neq i$  sparse off-diagonal terms

$$\begin{aligned} \frac{\partial S_i}{\partial x_i} &= \frac{\partial}{\partial x_i} \underbrace{\frac{e^{x_i}}{\sum_k e^{x_k}}}_h = - \frac{e^{x_i} \cdot e^{x_i}}{\sum_k e^{x_k}} = - \frac{e^{x_i}}{\sum_k e^{x_k}} \cdot \frac{e^{x_i}}{\sum_k e^{x_k}} \\ &= - \sum_l S_l(x) \cdot S_i(x) \end{aligned}$$

$$\Rightarrow \mathcal{J}(S) = \begin{cases} - \sum_l S_l(x) \cdot S_i(x) & l \neq i \\ S_i (1 - S_i) & l = i \end{cases}$$

### 2.2.1 Solutions of The Normal Equations

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

- Prove that:  $\text{Ker}(\mathbf{X}) = \text{Ker}(\mathbf{X}^\top \mathbf{X})$
- Prove that for a square matrix  $A$ :  $\text{Im}(A^\top) = \text{Ker}(A)^\perp$
- 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  $\infty$  solutions  $\Leftrightarrow \mathbf{y} \perp \text{Ker}(\mathbf{X}^\top)$ .
- 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).

$\mathbf{X} = \mathbf{A}$  גורן שלג פנו כ"י יס. 13 נסן הוכז. ☺

$\mathbf{A}\bar{\mathbf{x}} = \mathbf{0}$  וק  $\text{Ker}(\mathbf{A}) \supset \text{lin } \bar{\mathbf{x}} \in \mathbb{R}^d$  כ"כ כל ר'ז'

וק מ"מ  $\mathbf{A}\bar{\mathbf{x}} = \mathbf{0}$  ו"רנ' וק  $\bar{\mathbf{x}} \in \text{Ker}(\mathbf{A})$  ו"ז  $\Leftarrow$

поб  $\mathbf{A}^\top \mathbf{A}\bar{\mathbf{x}} = \mathbf{0}$  ו"רנ'  $\mathbf{A}^\top \rightarrow$  כ"כ יס. יס'

$\bar{\mathbf{x}} \in \text{Ker}(\mathbf{A})$

$\mathbf{X}\mathbf{X}^\top \mathbf{w} = \mathbf{0}$  ו"רנ'  $\mathbf{U}^\top \mathbf{X}^\top \mathbf{X} \mathbf{w} = (\mathbf{X}\mathbf{w})^2 = 0$  | ס'ל  $\mathbf{X} \in \text{Ker}(\mathbf{X}^\top \mathbf{X}) \Rightarrow$

וק ר'ז ו"כט' מ"מ הנויה ע"נnde כ"מ כ"פ ר'ז'  $\Leftarrow$

$\mathbf{X}\mathbf{w} = \mathbf{0}$  | ס'ל ו"כט' מ"מ כ"מ כ"פ

$\mathbf{X} \in \text{Ker}(\mathbf{X})$  ו"כט'

$\text{Ker}(\mathbf{X}) = \text{Ker}(\mathbf{X}\mathbf{X}^\top)$  ו' ב'ז | ס'ל

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

" סב . b

$A^T x = b$  ו  $\exists v \neq x \in \mathbb{R}^d$  כך  $v \in \text{Im}(A^T)$   $\Leftrightarrow$

$w^T v \neq w^T x$  כי  $w \in \ker(A)$   $\Leftrightarrow$

$$\Rightarrow w^T A^T x = w^T b$$

$$\Rightarrow (Aw)^T x = w^T b \stackrel{Aw=0}{=} 0 \Rightarrow \langle w | b \rangle = 0$$

$$\Rightarrow b \in \ker(A)^+$$

$x^T / A^T x = v$   $\Leftrightarrow x \in \text{Im}(A^T)^\perp$   $\Rightarrow$

$$x^T / A^T x = v$$

$$\Rightarrow x^T A^T x = x^T v$$

$$\Rightarrow (Ax)^T x = 0$$

מכיון שהבנה מאחר  $Ax=0$   $\Leftrightarrow x \neq 0$  ו  $x \in \ker(A)$

$\Leftrightarrow x \in \ker(A)$  ו  $x \in \ker(A)$

$$\text{Im}(A^T)^\perp \subseteq \ker(A)$$

$$\ker(A)^\perp \subseteq \text{Im}(A^T)^\perp = \text{Im}(A^T)^\perp$$

. סב

הנתקה מוקדש בפונקציית האינטגרל כפונקציית גודל נורמי.

$y \in \text{Im}(x)$  אם ורק אם  $y = xw$  עבור  $w \in \text{ker}(x^T)$ .

$xw = 0$  אם ורק אם  $w \in \text{ker}(x^T)$  כי  $x^T w = 0$ .

ההוכחה היא כפונקציית גודל נורמי.

$y \in \text{Im}(x)$  אם ורק אם  $y = xw$  ו- $w \in \text{ker}(x^T)$ .

בנוסף,  $\|y\|_2 = \|xw\|_2 \leq \|x\|_2 \|w\|_2$ .

$y \perp \text{ker}(x^T) \Leftrightarrow y \in \text{Im}(x)^{\perp}$ .

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

ההוכחה מוקדש בפונקציית גודל נורמי.

בנוסף  $\text{ker}(x) = \text{ker}(x^T x)$  כי  $x^T x w = 0 \Leftrightarrow x^T y = 0$ .

בנוסף  $\text{ker}(x^T y) \perp \text{ker}(x)$ .

or infinitely many solutions (otherwise).

## 2.2 Least Squares

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

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

where  $X$  is the input matrix of the linear regression with rows as samples and  $y$  the vector of responses. Let  $X = U\Sigma V^\top$  be the SVD of  $X$ , where  $U$  is a  $m \times m$  orthonormal matrix,  $\Sigma$  is a  $m \times d$  diagonal matrix, and  $V$  is an  $d \times d$  orthonormal matrix. Let  $\sigma_i = \Sigma_{ii}$  and note that only the non-zero  $\sigma_i$ -s are

singular values of  $X$ . Recall that the pseudoinverse of  $X$  is defined by  $X^\dagger = V\Sigma^\dagger U^\top$  where  $\Sigma^\dagger$  is an  $d \times m$  diagonal matrix, such that

$$\Sigma_{ij}^\dagger = \begin{cases} \sigma_i^{-1} & \sigma_i \neq 0 \\ 0 & \sigma_i = 0 \end{cases}$$

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

$$xx^\top = (V\Sigma V^\top)^\top (V\Sigma V^\top) = V\Sigma^\top V^\top V\Sigma V^\top$$

$\underbrace{\Sigma}_{\text{non-zero}} \quad \underbrace{V\Sigma^\top V^\top}_{\text{orthogonal}} \quad \underbrace{V\Sigma V^\top}_{\text{orthogonal}}$

$$= V\Sigma^\top I \Sigma V^\top = V\Sigma^\top \Sigma V^\top = V D V^\top$$

$\uparrow$   
 $D = \Sigma^\top \Sigma$

$$xx^\top (V D^{-1} V^\top) = V D V^\top V D^{-1} V^\top = V D \cdot D^{-1} V^\top = V \cdot V^\top = I$$

$\underbrace{D^{-1} V^\top}_{\text{orthogonal}}$

$$\begin{aligned} [xx^\top]^{-1} x^\top &= V D^{-1} V^\top x^\top = V D^{-1} V^\top (V\Sigma V^\top)^\top \\ &= V D^{-1} V^\top V \Sigma^\top V^\top = V D^{-1} \Sigma^\top V^\top = V (\Sigma^\top \Sigma)^{-1} \Sigma^\top V^\top \\ &= V \Sigma^{-1} V^\top \\ &= V \Sigma^\top V^\top = x^\top \end{aligned}$$

כגלאן

### 3 Practical Part

features ↗ גזען טרינומיאלי ← 3.1.3

לפנינו מתקבלים -

*id, lat, long, date, sqft\_lot, sqft\_living*

כ. ג. ב. כ. ג. ב. נ. כ. ג. ב. כ. ג. ב. כ. ג. ב. כ. ג. ב. כ. ג. ב.

• נציגו את הפונקציה כה פרגו ← פולינום ממעלה ששה משל ביטויים → נציגו כה פרגו

• כ. ג. ב. כ. ג. ב.

(כ. ג. ב. כ. ג. ב.)

## 4

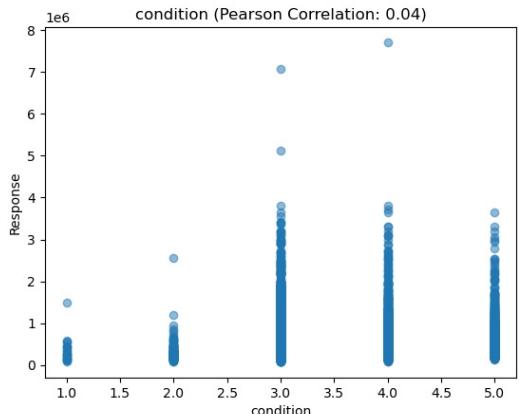
## Basics of feature selection -

(0.04) נינכָה (נינכָה) ↪ נינכָה נינכָה ↪

הנינכָה נינכָה ↪ נינכָה condition condition

הנינכָה ↪ נינכָה ↪ נינכָה ↪

condition ↪ נינכָה ↪ נינכָה ↪



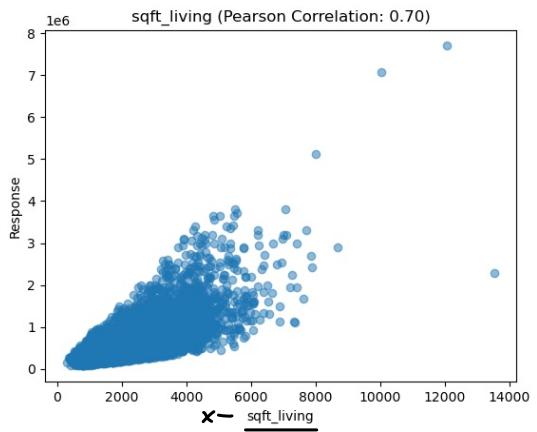
- נינכָה ↪ נינכָה ↪

הנינכָה ↪ נינכָה ↪ condition

condition ↪ sqft\_living ↪ sqft\_living ↪

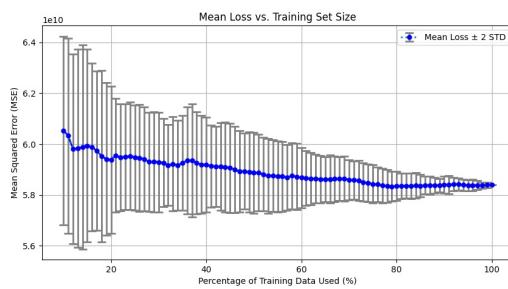
condition ↪ sqft\_living ↪ sqft\_living ↪

condition ↪ sqft\_living ↪ sqft\_living ↪



.sqft\_living ↪ sqft\_living ↪ sqft\_living ↪

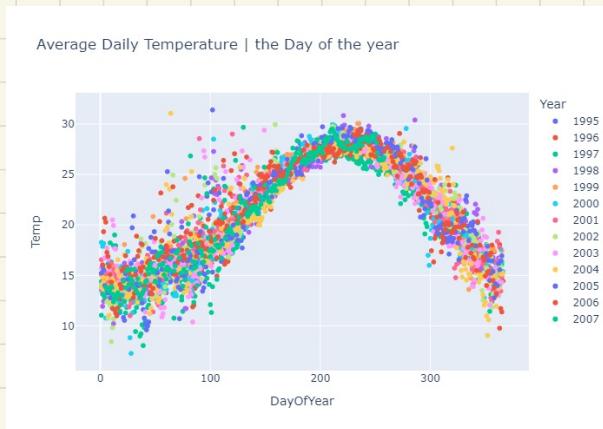
## 6. Fit a linear regression model



השאלה הולכת ונגזרת בפונקציית האפסון  
המינימלית מינימום האפסון מושג על ידי  
ההתקשרות של נסיגת ניוטון-רפלט  
בנוסף לה שיטות מילינריה וריבועית  
הו שיטות מילינריה וריבועית  
הו שיטות מילינריה וריבועית

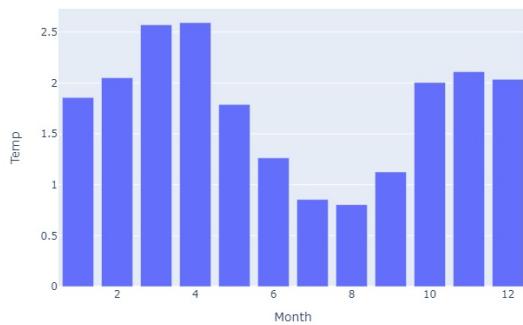
### 3.2 Polynomial Fitting

לפיו, מילויים נאכלים יוצרים פולינום.



המינימום טמפרטורה מתרחש בדצמבר, והמקסימום בטמפרטורה מתרחש ביוני.

Standard Deviation of Daily Temperatures for Each Month



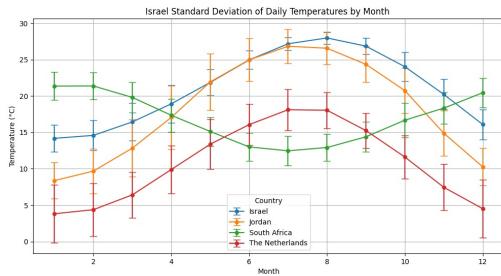
טמפרטורה המינימלית מתרחשת בדצמבר, וטמפרטורה המקסימלית מתרחשת ביוני.

טמפרטורה המינימלית מתרחשת בדצמבר, וטמפרטורה המקסימלית מתרחשת ביוני.

טמפרטורה המינימלית מתרחשת בדצמבר, וטמפרטורה המקסימלית מתרחשת ביוני.

- 6-9 ספטמבר גינס

4



בגדי הרים נסגרו מים, ים ונהר  
הHIGH פארם גראן צהוב נסגרו  
בנורו גראן צהוב נסגרו מים.  
בגדי הרים נסגרו מים, ים ונהר  
הHIGH פארם גראן צהוב נסגרו  
בנורו גראן צהוב נסגרו מים.  
בגדי הרים נסגרו מים, ים ונהר  
הHIGH פארם גראן צהוב נסגרו  
בנורו גראן צהוב נסגרו מים.

לכל  $K=5$

הכי פחות.

$K=3,4$  מוגבר יותר

כ' דוגמ' יד'

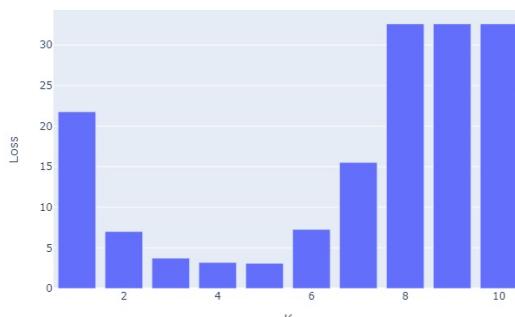
ולא כ' פונק' error

error  $\Rightarrow$  רק מוג'  $K=5$  מוג' כ'.

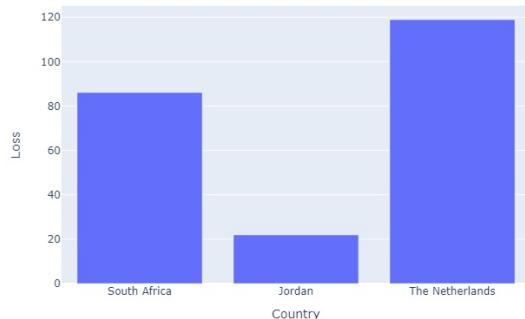
הכי פחות (בזין).

5

Loss of the polynomial model of degree  $k$  over the test set



Loss of the Polynomial Model of Degree 5 Over the Country Set



. 6

טבלה 8.1. LOSS כפונקציית גודל של מודל פולינומי ממעלה 5

הנראה מהלך הנטה של LOSS כפונקציית גודל של מודל פולינומי ממעלה 5. מושגנו שפונקציית LOSS מוגדרת כפונקציית גודל של מודל פולינומי ממעלה 5. פונקציית LOSS מוגדרת כפונקציית גודל של מודל פולינומי ממעלה 5. פונקציית LOSS מוגדרת כפונקציית גודל של מודל פולינומי ממעלה 5.

בכדי למדוד את הנטה של LOSS כפונקציית גודל של מודל פולינומי ממעלה 5, נשתמש בפונקציית LOSS כפונקציית גודל של מודל פולינומי ממעלה 5.

הנטה של LOSS כפונקציית גודל של מודל פולינומי ממעלה 5 היא נטלה של LOSS כפונקציית גודל של מודל פולינומי ממעלה 5.

הנטה של LOSS כפונקציית גודל של מודל פולינומי ממעלה 5 היא נטלה של LOSS כפונקציית גודל של מודל פולינומי ממעלה 5.