יובל דלוס 305211880

מבוא למערכות לומדות – תרגיל 2

Solutions of the Normal Equations

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

1. נראה הכלה דו כיוונית של שתי הקבוצות:

- $Ker(X^T) \subseteq Ker(XX^T)$ –

יהי $v \in Ker(X^T)$, אזי:

$$X^T \cdot v = \overline{0}$$

ולכן:

$$XX^T \cdot v = X \cdot \overline{0} = \overline{0} \Longrightarrow v \in Ker(XX^T)$$

- $Ker(XX^T) \subseteq Ker(X^T)$–

יהי $v \in Ker(XX^T)$, אזי:

$$XX^T \cdot v = \overline{0} = v^T \cdot (XX^T \cdot v) = v^T XX^T v = \langle X^T v, X^T v\rangle = \|X^T v\|^2 \Longrightarrow$$
$$\Longrightarrow \|X^T v\|^2 = 0 \Leftrightarrow X^T v = 0 \Leftrightarrow v \in Ker(X^T)$$

לכן סה״כ הראינו הכלה דו כיוונית והקבוצות שוות.

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

2. אראה הכלה דו כיוונית:

* $Im(A^T) \subseteq Ker(A)^{\perp}$ –

יהי $v \in Im(A^T)$, אזי, קיים u כך ש-

$$\exists u\ s.t\ \ A^T \cdot u = v$$

בנוסף, קיים $x \in \mathrm{Ker}(A)$, אזי:

$$\langle v, x\rangle = \langle A^T \cdot u, x\rangle = u^T A x = \langle u, A \cdot x\rangle = \left\langle u, \overline{0}\right\rangle = \overline{0} \Longrightarrow \langle v, x\rangle = \overline{0}$$

אזי, עבור כל וקטור $v \in Im(A^T)$ קיים וקטור $x \in \mathrm{Ker}(A)$ כך ש- $\langle v, x\rangle = \overline{0}$, לכן לפי הגדרה $v \in Ker(A)^{\perp}$, ומכאן קיבלנו כי

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

$* \quad Ker(A)^{\perp} \subseteq Im(A^T)$ –

יהי $v \notin Im(A^T)$ נרצה להוכיח כי $v \notin Ker(A)^{\perp}$.

$$v \notin Im(A^T) \overset{(*)}{\Longrightarrow} v \in Im(A^T)^{\perp} \Longrightarrow \forall b \in Im(A^T) \mid \langle b, v \rangle = 0$$

$$b \in Im(A^T) \Longrightarrow \exists u \in R^n \ s.t \ A^T \cdot u = b \Longrightarrow$$

$$\Longrightarrow 0 = \langle v, b \rangle = \langle v, A^T u \rangle = v^T A^T u = \langle Av, u \rangle \Longrightarrow Av = 0 \Longrightarrow$$

$$\Longrightarrow v \in Ker(A) \Longrightarrow v \notin Ker(A)^{\perp}$$

(*) $if\ Im(A^T) = U,\ Im(A^T) = U^{\perp} \Longrightarrow U \oplus U^{\perp} = R^n$

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

3. נתון לנו כי $y \perp Ker(X)$, כלומר, $y \in Ker(X)^{\perp}$, ולכן מהסעיף הקודם:

$$y \perp Ker(X) \Longrightarrow y \in Ker(X)^{\perp} \Longrightarrow y \in Im(X^T)$$

נתון לנו כי המטריצה אינה הפיכה, לכן $\dim\left(Ker(X^T)\right) \neq 0$, לכן ממשפט שלמדנו בכיתה יש לה אפס פתרונות או אינסוף פתרונות.

הראינו כי $y \in Im(X^T)$ אזי קיים $w \in R^n$ כך ש-

$$y = X^T w$$

ומכיוון שיש פתרון יחיד ישנם אינסוף פתרונות.

יובל דלוס 305211880

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

.4 נניח תחילה כי XX^T הינה הפיכה, אזי למערכת יש פיתרון יחיד והוא:

$$XX^Tw = Xy \Leftrightarrow w = (XX^T)^{-1}Xy$$

כעת נניח כי XX^T איננה הפיכה, ניעזר בסעיף 3 בו ראינו כי למע' המשוואות $y = X^Tw$ יש אינסוף פתרונות אמ"מ $y \perp Ker(X)$, אזי:

$$y = X^Tw \Leftrightarrow Xy = XX^Tw \Longrightarrow \infty\ solutions \ \Leftrightarrow Xy \perp Ker(XX^T)$$

ומסעיף אחד קיבלנו כי $Ker(XX^T) = Ker(X^T)$ אזי:

$$y = X^Tw \Leftrightarrow Xy = XX^Tw \Longrightarrow \infty\ solutions \ \Leftrightarrow Xy \perp Ker(X^T)$$

ולכן, עבור $w \in Ker(X^T)$:

$$w \in Ker(X^T) \Longrightarrow \langle w, Xy\rangle = \langle X^Tw, y\rangle = \langle \bar{0}, y\rangle = \bar{0}$$

ולכן למע' משוואות אינסוף פתרונות.

יובל דלוס 305211880

Projection Matrices

5. 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 $v_1, \ldots, v_k$ be an orthonormal basis of V. Define the orthogonal projection matrix $P = \sum_{i=1}^{k} v_i v_i^\top$ (Notice this is an outer product). Show that:

   (a) P is symmetric

.5

(a) בכדי להראות כי P סימטרית נרצה להראות כי $P^T = P$, אזי לפי הגדרה:

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

   (b) The eigenvalues of P are 0 or 1 and that $v_1, \ldots, v_k$ are the eigenvectors corresponding the eigenvalue 1

(b) בכדי להראות כי כל וקטור v_i הוא ו״ע עם ע״ע 1 נרצה שיקיים $Pv_i = 1 \cdot v_i = v_i$ עבור כל $1 \le i \le k$.
נבחר v_j כלשהו $j \in [1, k]$ ונראה:

$$Pv_j = \sum_{i=1}^{k} v_i v_i^T v_j = \sum_{i=1}^{k} v_i \langle v_i, v_j \rangle = v_1 \langle v_1, v_j \rangle + v_2 \langle v_2, v_j \rangle + \cdots + v_k \langle v_k, v_j \rangle = \sum_{i=1}^{k} v_i \delta_{ij} = v_j$$

כאשר הדלתא של קרונקר נובעת מהגדרתה של ההטלה האורתוגונלית

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

   (c) $\forall v \in V \;\; Pv = v$

(c) עבור כל וקטור $v \in V$ הוא באחת משתי צורות, או שהוא וקטור ״טהור״ מאחד מווקטורי הבסיס, או שניתן לבטא אותו כצירוף לינארי של וקטורי הבסיס ואז $v = \sum_{i=1}^{k} a_i v_i$ עבור $a_i \in \mathbb{R}$. עבור המקרה הראשון הוכחנו בסעיף הקודם, עבור המקרה השני:

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

(d) $P^2 = P$

(d) יהי $P = UDU^T$ פירוק ה- EVD של מטריצה P. אזי

$$P^2 = P \cdot P = UDU^T \cdot UDU^T = UD^2U^T$$

אבל מכיוון ש- D מורכבת מע״ע של P בלבד, והוכחנו בסעיף קודם כי ע״ע אלה הם 0 ו-1, ו-D אלכסונית, נקבל כי

$$D^2 = D$$

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

כנדרש.

(e) $(I - P)\,P = 0$

(e) מהסעיף הקודם למדנו כי $P^2 = P$ אזי,

$$(I - P)P = P - P^2 = P - P = 0$$

כנדרש.

6. We will first show that if $XX^\top$ is invertible, the general solution we derived in recitation is equal to the solution you have seen in class. For this part, assume that $XX^\top$ is invertible.

- Show that $\left(XX^\top\right)^{-1} = UD^{-1}U^\top$, where $D = \Sigma\Sigma^\top$.

.6

$$XX^T = U\Sigma \overbrace{V^TV}^{I} \Sigma^T U^T = U\Sigma\Sigma^T U^T = UDU^T$$

כעת נרצה להראות כי $(XX^T)^{-1} = UD^{-1}U^T$, נשתמש בתכונה של מטריצות עבור מטריצה $M \in \mathbb{R}_{m\times n}$ כלשהי, $M \cdot M^{-1} = I$
אזי:

$$XX^T(UD^{-1}U^T) = (UD\overbrace{U^T)(U}^{I}D^{-1}U^T) = U\overbrace{DD^{-1}}^{I}U^T = \overbrace{UU^T}^{I} = I = XX^T(XX^T)^{-1}$$

ולכן

$$(XX^T)^{-1} = (UD^{-1}U^T)$$

- Use this to show that $\left(XX^\top\right)^{-1}X = X^{\top\dagger}$.

$$(XX^T)^{-1}X = (UD^{-1}U^T)(U\Sigma V^T) = UD^{-1}\Sigma V^T = U(\Sigma\Sigma^T)^{-1}\Sigma V^T = U\overbrace{\begin{pmatrix} \sigma_1^{-2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_m^{-2} \end{pmatrix}}^{(\Sigma\Sigma^T)}\overbrace{\begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_m \end{pmatrix}}^{\Sigma}V^T =$$

$$= U\Sigma^{\mathrm{T}^{-1}}V^T = U\Sigma^{T^\dagger}V^T \overset{(*)}{=} X^{T^\dagger}$$

$$* \quad X^{T^\dagger} = \left(X^\dagger\right)^T = \left(V\Sigma^\dagger U^T\right)^T = U\Sigma^{T^\dagger}V^T = U\Sigma^{\dagger^T}V^T$$

יובל דלוס 305211880

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

7. ראשית נציג את XX^T ע״פ הגדרה:

$$XX^T = \begin{pmatrix} \vdots & \vdots & \vdots \\ x_1 & \cdots & x_m \\ \vdots & \vdots & \vdots \end{pmatrix}\begin{pmatrix} \cdots & x_1^T & \cdots \\ \cdots & \vdots & \cdots \\ \cdots & x_m^T & \cdots \end{pmatrix} = \begin{pmatrix} \Sigma x_i^1 x_i^1 & \cdots & \Sigma x_i^1 x_i^m \\ \vdots & \cdots & \vdots \\ \Sigma x_i^m x_i^1 & \cdots & \Sigma x_i^m x_i^m \vdots \end{pmatrix}$$

אם XX^T הפיכה אז ממשפט שלמדנו בלינארית גם $(XX^T)^T = X^TX$ הפיכה, וזה קורה אס״ם:

$$\dim\big(Ker(X)\big) = \dim\,(Im(XX^T) = d \Leftrightarrow \dim\big(Im(X)\big) = \dim\,(Im(X^TX)$$

נשים לב כי XX^T הינה מגודל $d \times d$ לכן נוכל לפרקה בפירוק SVD ונקבל:

$$\Sigma = \begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_d \end{pmatrix}$$

כאשר Σ מגודל $d \times d$ ומההפיכות נקבל שדרגתה מלאה ויש לה d ע״ע ולכן גודל התמונה שלה הוא d.
וזה קורה אס״ם גודל המימד של X הינו מימד העמודות הנדרשות, כלומר $Span\{x_1, \ldots, x_m\} = \mathbb{R}^d$ כנדרש.

8. Recall that if $XX^\top$ is not invertible then there are many solutions. Show that $\hat{\mathbf{w}} = X^{\top\dagger}\mathbf{y}$ is the solution whose L_2 norm is minimal. That is, show that for any other solution $\overline{\mathbf{w}}$, $\|\hat{\mathbf{w}}\|_2 \leqslant \|\overline{\mathbf{w}}\|_2$ (Why is it ok to do so?)

8. ראשית נשים לב כי

$$XX^Tw = Xy \implies UDV^Tw = U\Sigma V^Ty \implies DV^Tw = \Sigma V^Ty$$

כלומר, בכתיב מטריציוני:

$$\begin{pmatrix} \sigma_1^2 & & & \\ & \ddots & & \\ & & \sigma_r^2 & \\ & & & 0 \\ & & & & \ddots \end{pmatrix} V^Tw = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \\ & & & & \ddots \end{pmatrix} V^Ty$$

לכן, נגדיר עבור כל $r + 1 \le i$ כי $w_i = 0$ ונכפול את שני האגפים ב- D^T ונקבל

$$\underbrace{\begin{pmatrix} 1 & & \\ & 1 & \\ & & 0 \end{pmatrix}}_{D^T D} V^T w = \underbrace{\begin{pmatrix} \sigma_1^{-1} & & \\ & \sigma_r^{-1} & \\ & & 0 \end{pmatrix}}_{D^T \Sigma} = \Sigma^\dagger \quad V^T y$$

לכן,

$$V^T w = \Sigma^\dagger V^T y \implies w = V\Sigma^\dagger V^T y = X^{T^\dagger}$$

נשים לב כי הגדרנו $w_i = 0$ עבור $i \geq r+1$ על מנת לקבל תוצאה יחידה, לכן כל $\overline{w}$ זהה ל- $\hat{w}$ בכל קורדינאטה עבורה $r \leq i$ מכיוון שהע״ע מסודרים מהגדול לקטן, וכן σ_i, d_i שונים מאפס. ובנוסף, עבור $i \geq r+1$ נקבל שהמכפלה שווה ל-0 ומכאן לכל פיתרון $\overline{w}$ מתקיים $\overline{w}_i = \hat{w}_i$ עבור $r \geq i$.

נתבונן על נורמה 2 שניתנה לנו :

$$\|\hat{w}\|_2 = \langle \hat{w}, \hat{w} \rangle = (\hat{w})^T \cdot \hat{w} = \sum_{i=1}^{d} \hat{w}_i \hat{w}_i = \sum_{i=1}^{r} \hat{w}_i \hat{w}_i$$

ובאופן דומה עבור $\overline{w}$:

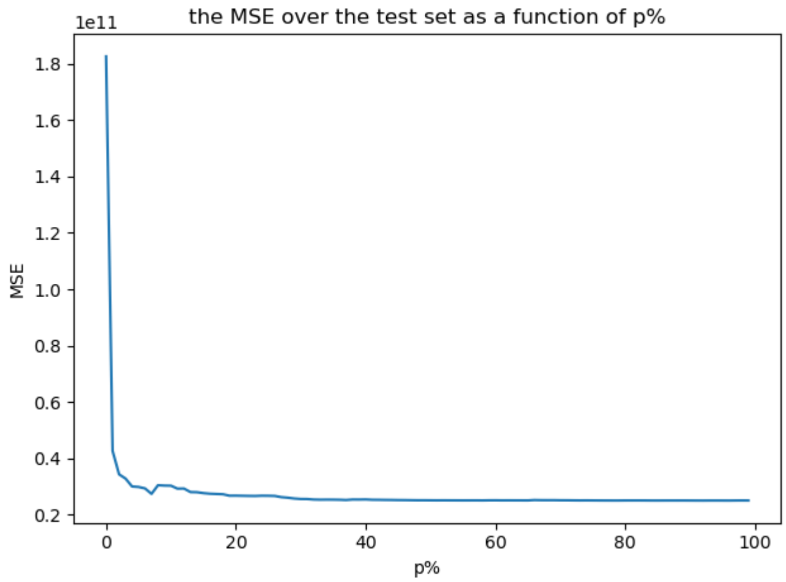
$$\|\overline{w}\|_2 = \langle \overline{w}, \overline{w} \rangle = \sum_{i=1}^{d} \overline{w}_i \overline{w}_i = \sum_{i=1}^{r} \overline{w}_i \overline{w}_i + \overbrace{\sum_{i=r+1}^{d} \overline{w}_i \overline{w}_i}^{\geq 0}$$

ולכן קיבלנו $\|\overline{w}\|_2 \leq \|\hat{w}\|_2$ וכן $\hat{w}$ בעל נורמה מינימלית כנדרש.

13. בחרתי את הנתון לגבי הזיפ-קוד כערך קטגורי, המספר שלו כשם עצמו אינו מייצג נתון אמיתי (לדוגמא אין סיבה ש-100 יהיה טוב יותר מ-20) אלה בעיקר נתון שרירותי ולכן לא רציתי שהדגם יתייחס אליו כמספר. לעומת זאת, בנתון של מתי הבית שופץ אמנם יכולתי להתייחס גם אליו כקטגורי האם הבית שופץ או לא, אך ראיתי שניתן להגיע לתוצאות מעניינות אם מתחשבים בשנה עצמה של שיפוץ הבית לעומת שנים אחרות.

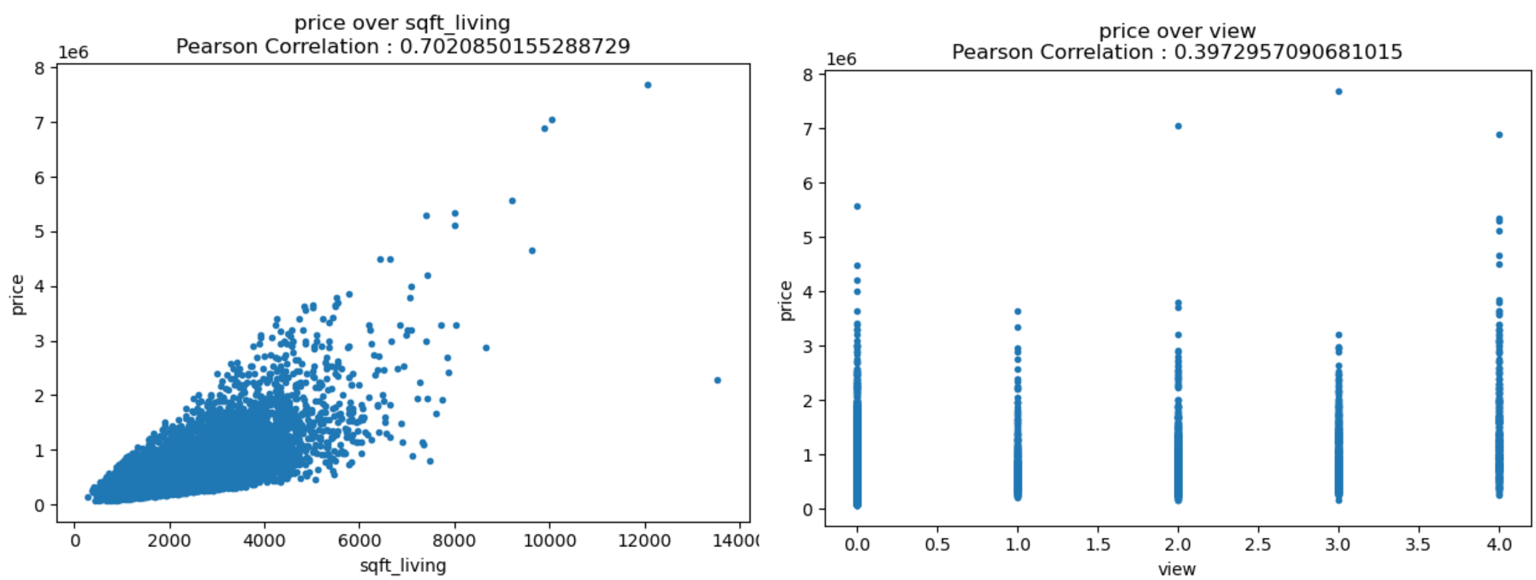
15. ניתן ללמוד מה המגמות וכמה כל גורם (מספר חדרי שינה, מקלחות וכו׳) משפיע על המחיר.

.16



ניתן לראות מהגרף שככל שהמערכת שלנו מקבלת יותר נתונים להתאמן עליהם ככה התוצאה שלה יותר מדויקת ומתכנסת אל התוצאה האמיתית.

.17



ניתן לראות מגמה ברורה כאשר ככל שהבית יותר גדול, כך המחיר שלו יותר גבוהה, לכן זה פיצ׳ר משמעותי בחיזוי ערכו של בית. לעומת זאת ניתן לראות שהגרף של המראה מפוזר פחות או יותר אחיד, ובתים בכל הדרגות נמכרים באותו טווח מחירים, לכן זה לא מוסיף לנו מידה חדש ווקטור התוצאות שלנו לא היה שונה במידה רבה אם לא היה לנו את הנתון הזה.

305211880 יובל דלוס

.21