רגרסיה ומודלים סטטיסטיים- תרגיל 1

שאלה 1:

א. תהי $A \in R^{n\times n}$ מטריצה סימטרית ויהיו $(\lambda_1,u_1),(\lambda_2,u_2)$ זוגות של וקטורים עצמיים וע"ע של A כך שמתקיים שהערכים העצמיים $\lambda_1 \neq \lambda_2$. הראו כי $u_1^t u_2 = 0$.

ב. נסמן $V = I + \theta A, \theta \in R$. הוכיחו ש-$u_1$ הוא ו"ע של V ומצאו את הע"ע המתאים לו.

ג. נניח כעת כי A הפיכה ונכתוב $A = U\Lambda U^T$ כאשר המטריצות מוגדרות באותו האופן בו הגדרנו בכיתה. בטאו את A^{-1} במונחי $\lambda_1,\dots,\lambda_n;\ u_1,\dots,u_n$.

א. מהגדרת הע"ע והו"ע:

$$Au_1 = \lambda_1 u_1 \Leftrightarrow u_1^t A^t = u_1^t A = \lambda_1 u_1^t \Leftrightarrow u_1^t A u_2 = u_1^t \lambda_2 u_2 = \lambda_1 u_1^t u_2 \Leftrightarrow (\lambda_2 - \lambda_1) u_1^t u_2 = 0$$

כאשר השוויון השני נובע מהיותה של A סימטרית, והשיוויון הרביעי מכך ש- λ_2 ע"ע של A המתאים לו"ע u_2. כעת כיוון ש-$\lambda_1 \neq \lambda_2$ בהכרח מתחייב כי $u_1^t u_2 = 0$.

ב. $Vu_1 = (I+\theta A)u_1 = u_1 + \theta A u_1 = u_1 + \theta\lambda_1 u_1 = (1+\theta\lambda_1)u_1$

ולכן הזוג $(1+\theta\lambda_1, u_1)$ הוא זוג עצמי (eigenpair) של V.

ג. מהגדרת הע"ע והו"ע וכיוון ש-A הפיכה:

$$Au_i = \lambda_i u_i \Leftrightarrow A^{-1}Au_i = \lambda_i A^{-1}u_i \Leftrightarrow \frac{1}{\lambda_i}u_i = A^{-1}u_i$$

כלומר $(\frac{1}{\lambda_i}, u_i)$ הם זוגות עצמיים של A^{-1}.

רק כדי שתראו שזה עובד גם לפי משפט הפירוק הספקטרלי:

$$A^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i}\ U_i U_i^T$$

$$AA^{-1} = \sum_{j=1}^{n} \lambda_j\ U_j U_j^T \sum_{i=1}^{n} \frac{1}{\lambda_i}\ U_i U_i^T = \sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_j \left(\frac{1}{\lambda_i}\right) U_j \boldsymbol{U_j^T U_i} U_i^T$$

כאשר $i \neq j$, כיוון שהמטריצה סימטרית והמטריצה U אורתוגונלית, הביטוי המודגש מתאפס, ואחרת, הוא שווה ל-1 ולכן:

$$= \sum_{i=1}^{n} \lambda_i \left(\frac{1}{\lambda_i}\right) U_i U_i^T = \sum_{i=1}^{n} U_i U_i^T = I$$

שאלה 2:

א. תהי $X \in R^{n\times p}$, כך ש- $n > p$ ונסמן $A = X^T X$. הוכיחו כי הבאים שקולים:

(1) A הפיכה.

(2) עמודות X בת"ל.

(3) A חיובית מוגדרת.

(4) $\lambda_1,\dots,\lambda_p$ ע"ע של A, אז $\lambda_i > 0, \forall p \geq i \geq 1$.

ב. הסיקו מהסעיף הקודם כי המטריצה $A + \theta I$ הפיכה $\forall\theta > 0$.

א. הערה: יש דרכים רבות להוכיח זאת, נציג שתיים מהן:

(1) ← (2): נניח $A = X^T X \in R^{p\times p}$ הפיכה, אז דרגתה מלאה ושווה ל-p. מתכונות של דרגה:

$rank(CB) \leq \min\{rank(C), rank(B)\}$

$rank(B) = rank(B^T)$

לכן:

$$p = rank(I_p) = rank(X^T X (X^T X)^{-1}) \leq rank(X^T X) \leq rank(X^T, X) \leq rank(X)$$
$$\leq \min(n,p) = p$$

ולכן בהכרח אי השוויונות מתקיימים בשוויון ודרגת X היא p כיוון שמספר השורות הבת"ל במטריצה

שווה למספר העמודות הבת"ל במטריצה, זה אומר של-X יש p עמודות בת"ל. כלומר כל העמודות שלה בת"ל.

דרך אחרת (קלה יותר): נניח בשלילה שעמודות X תלויות לינארית אז קיים פתרון לא טריוויאלי למערכת המשוואות $Xv = 0$. אבל אז קיים פתרון לא טריוויאלי גם למערכת המשוואות $Av = X^T Xv = X^T \cdot 0 = 0$ בסתירה לכך ש-A הפיכה.

(2) ← (3): לכל $v \in R^p$ מתקיים:

$$v^T Av = v^T X^T Xv = ||Xv||^2 \geq 0$$

מתכונות של נורמה, שיוויון מתקיים אם"ם $Xv = 0$. אם עמודות X בת"ל, קיים פתרון טריוויאלי בלבד למערכת המשוואות ההומוגנית ולכן בהכרח מתקיים שלכל $v \neq 0$: $v^T Av > 0$ כלומר A חיובית.

(3) ← (4): A סימטרית ולכן קיים לה פירוק ספקטרלי . כיוון ש-$A > 0$:

$$v^T Av = v^T U \Lambda U^T v = \tilde{v}^T \Lambda \tilde{v} = \sum_{i=1}^{p} \lambda_i \tilde{v_i}^2 = (\lambda_1, \dots, \lambda_p)(\tilde{v_1}^2, \dots, \tilde{v_p}^2)^T > 0$$

זה נכון לכל v, ובפרט עבור $v = e_j$ שיתן $\sum_{i=1}^{p} \lambda_i \tilde{v_i}^2 = \lambda_j$ ולכן $\forall j, \lambda_j > 0$.

(4) ← (1): מטריצה היא הפיכה אם"ם הדטרמיננטה שלה שונה מ-0. כמו כן הדטרמיננטה של מטריצה לכסינה היא מכפלת הע"ע ולכן אם כולם חיוביים ממש, אז גם המכפלה שלהם.

ב. באופן דומה לסעיף ב' בשאלה 1 נקבל כי הע"ע ה-i של $A + \theta I$ הוא $\lambda_i + \theta$, עבור λ_i ע"ע של A. באותו האופן שבו הוכחנו את (2) ← (3) נוכל לקבל ש-A אי שלילית. כמו כן באותו האופן שהוכחנו את (3) ← (4) נקבל שהע"ע של A הם אי שליליים. לכן: $\lambda_i \geq 0 \Rightarrow \lambda_i + \theta > 0, \forall \theta > 0$.

שאלה 3:

א. הניחו את הבסיס S ל-R^3 ואת המטריצה $A \in R^{3\times 3}$:

$$A = \begin{bmatrix} 1 & -2 & 1 \\ 3 & -1 & 0 \\ 1 & 4 & -2 \end{bmatrix} \quad \text{and} \quad S = \{u_1, u_2, u_3\} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\}$$

תזכורת: A היא מטריצה המייצגת את ההעתקה A ביחס לבסיס הסטנדרטי.

מצאו את המטריצה $[A]_S$ המייצגת את ההעתקה A ביחס לבסיס S (בתחום ובטווח).

רמז: לכל העתקה לינארית G ובסיסים C, B מתקיים:

$$[G]_C = [I]_C^B [G]_B [I]_B^C$$

ב. - נניח $T: V \to W$ וכן $F: W \to U$ העתקות לינאריות, ונניח B, C, D בסיסים ל-V, W, U בהתאמה.

הראו כי:

$$[F \circ T(v)]_D = [F(T(v))]_D = [F]_D^C [T]_C^B [v]_B$$

כלומר, שניתן לייצג הרכבת העתקות לינאריות על ידי כפל במטריצות המייצגות את ההעתקות.

- הניחו כעת כי $V = W$ וכן $B = C$. הראו ש T הפיכה אם"ם $[T]_B$ הפיכה; וכן שמתקיים במקרה כזה $[T]_B^{-1} = [T^{-1}]_B$.

א. ראינו בתרגול כי לכל העתקה לינארית G ובסיסים C,B מתקיים:

$$[G]_C = [I]_C^B [G]_B [I]_B^C$$

בפרט עבור: $B = E, C = S, G = A$. כמו כן ראינו כי $[I]_E^S = [u_1|u_2|u_3]$ ולכן:

$$[A]_S = [I]_S^E [T]_E [I]_E^S = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 1 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 1 & -2 & 1 \\ 3 & -1 & 0 \\ 1 & 4 & -2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & -4 & -2 \\ 1 & -3 & -1 \\ 1 & 3 & 2 \end{pmatrix}$$

ב. - $[F(T(v))]_D = [F]_D^C [T(v)]_C = [F]_D^C [T]_C^B [v]_B$

- נניח כי T הפיכה. מכאן שלכל v: $T^{-1}\circ T(v) = v$. ולכן ביחס לכל בסיס, ובפרט עבור B:

$$[T^{-1}\circ T(v)]_B = [T^{-1}]_B[T]_B[v]_B = [v]_B = I[v]_B$$

כיוון שזה קורה לכל $v \in V$ בהכרח מתקיים: $[T^{-1}]_B[T]_B = I$.

מהכיוון השני, נניח כי $[T]_B$ הפיכה. לכן קיימת מטריצה A עבורה לכל v:

$$A[T]_B[v]_B = A[T(v)]_B = [v]_B$$

כיוון שכפי שציינו, קיים קשר חח"ע ועל בין העתקה לינארית ובין המטריצה המייצגת אותה ביחס לבסיס B, נקבל כי: $v = A\big(T(v)\big)$ ולכן T הפיכה ו-$A = [T]_B^{-1}$.

הערה: שימו לב שניתן להכליל זאת לכל בסיס.

שאלה 4:

תהי $X \in R^{n\times p}$ מטריצה מדרגה מלאה, $Y \in R^n$ וקטור כלשהו ו-$\beta \in R^p$ וקטור מקדמים. בתרגול הוכחנו את הנגזרות הבאות:

נגזרת של מכפלה סקלרית של וקטורים:

$$\frac{\partial}{\partial x}(b^T x) = \frac{\partial}{\partial x}(x^T b) = b$$

$$\frac{\partial}{\partial x}(x^T x) = 2x$$

וכן:

עבור מטריצה סימטרית A נקבל

$$\frac{\partial}{\partial x}(x^T A x) = 2Ax$$

א. השתמשו בתכונות אלו על מנת להראות:

$$\beta^* = argmin_{\beta\in R^p} \left|\left|Y - X\beta\right|\right|^2 = (X^TX)^{-1}X^TY$$

וכן:

$$X\beta^* = P_X Y$$
$$P_X = X(X^TX)^{-1}X^T$$

ב. מצאו ביטוי מפורש עבור $\beta^* = (\beta_0^*, \beta_1^*)$ למקרה בו $X \in R^{n\times 2}$ מהצורה:

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & x_5 \\ 1 & x_6 \\ 1 & x_7 \end{bmatrix}$$

הסבירו מדוע למעשה כבר פתרתם את הבעיה הזו בקורסים קודמים.

א.

$$\left|\left|Y - X\beta\right|\right|^2 = Y^TY - 2\beta^TX^TY + \beta^TX^TX\beta$$

נגזור לפי β. הביטוי השמאלי לא תלוי ב-β ולכן מתאפס. לפי הטענה הראשונה, הנגזרת של הביטוי האמצעי היא $-2X^TY$, ולפי הטענה השלישית, כיוון ש-X^TX סימטרית, נקבל שהנגזרת של הביטוי הימני היא $2X^TX\beta$. כמו כן נגזרת הסכום היא סכום הנגזרות. לאחר שנשווה ל-0 נקבל:

$$X^TX\beta = X^TY$$

X מדרגה מלאה לכן X^TX הפיכה, ונקבל:

$$\beta^* = (X^TX)^{-1}X^TY$$

ב. נחשב את X^TX:

$$\begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix} = n\begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{pmatrix}$$

וההופכית:

$$\frac{1}{n(\overline{x^2} - (\bar{x})^2)}\begin{pmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} = \frac{1}{nVa\hat{r}(x)}\begin{pmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

ואילו

$$X^T Y = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

ובסה"כ נקבל:

$$\beta^* = \frac{1}{Va\hat{r}(x)}\begin{pmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}\begin{pmatrix} \bar{y} \\ \overline{xy} \end{pmatrix} = \frac{1}{Va\hat{r}(x)}\begin{pmatrix} \overline{x^2} \cdot \bar{y} - \bar{x} \cdot \overline{xy} \\ \overline{xy} - \bar{x}\bar{y} \end{pmatrix}$$

שימו לב כי $\beta_1^* = \frac{\widehat{cov(x,y)}}{\widehat{var(x)}}$ שזהו בדיוק הביטוי לשיפוע שקיבלנו ברגרסיה על משתנה מסביר אחד בקורס בעקרונות.

כעת:

$$\frac{\overline{x^2} \cdot \bar{y} - \bar{x} \cdot \overline{xy}}{\widehat{var(x)}} = \frac{\overline{x^2} \cdot \bar{y} - \bar{x} \cdot \overline{xy} + (\bar{x})^2\bar{y} - (\bar{x})^2\bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{\bar{y}\left(\overline{x^2} - (\bar{x})^2\right) - \bar{x}(\overline{xy} - \bar{x}\bar{y})}{\overline{x^2} - (\bar{x})^2}$$
$$= \bar{y} - \bar{x} \cdot \beta_1^*$$

כלומר הביטוי המוכר לחותך.

כמובן, זה לא מקרי:

$$argmin_{\beta \in R^2} \left|\left|Y - X\beta\right|\right|^2 = argmin_{a,b \in R^2} \sum_{i=1}^{n} \left(Y_i - X_i^T \begin{pmatrix} b \\ a \end{pmatrix}\right)^2$$
$$= argmin_{a,b \in R^2} \sum_{i=1}^{n} (Y_i - 1 \cdot b - ax_i)^2$$

וזו בדיוק הבעיה אותה פתרנו בקורס בעקרונות.