רגרסיה- תרגול 3 :

מטריצת הטלה, המודל הלינארי והתפלגויות משותפות :

מטריצת הטלה אורתוגונלית

מטריצה איידמפוטנטית היא מטריצה $A \in R^{n\times n}$ שדרגתה r המקיימת $A = A^2$.

מטריצה סימטרית ואיידמפוטנטית נקראת **מטריצת הטלה אורתוגונלית.**

טענה- הע"ע של מטריצת הטלה הם 1, בריבוי כדרגת המטריצה, ו-0 בריבוי השווה למימד של גרעין המטריצה.

הוכחה :

מטריצת הטלה למרחב הנפרש על ידי עמודות X:

תהי $X \in R^{n\times p}$ מטריצה מדרגה מלאה ונגדיר $P_X = X(X^TX)^{-1}X^T$. הראו כי P_X היא מטריצת הטלה למרחב הנפרש על ידי העמודות של X. כלומר :

(1) P_X סימטרית.

(2) P_X איידמפוטנטית.

(3) מתקיים שלכל $v \in R^n$: $P_Xv \in IM(X)$.

הוכחה :

תכונות חשובות-אולי הכי חשובות בקורס(!!!)-של מטריצת הטלה:

Proposition 4. *Let $\boldsymbol{X}$ be an $n \times m$ matrix and assume that it has linearly independent columns (i.e., full column rank; remember that this implies $m \leq n$). Then the projection matrix $\boldsymbol{P_X}$ has the following properties.*

1. $\boldsymbol{P}_X$ *is symmetric*
2. $\boldsymbol{P}_X$ *is idempotent,* $\boldsymbol{P}_X^2 = \boldsymbol{P}_X$
3. $\boldsymbol{P}_X \boldsymbol{X} = \boldsymbol{X}$
4. $\boldsymbol{X}^\top (\boldsymbol{I} - \boldsymbol{P_X}) = \boldsymbol{0} \in \mathbb{R}^{m \times n}$
5. $\boldsymbol{P}_X v \in \mathrm{Im}(X)$ *for all* $v \in \mathbb{R}^n$
6. *If* $m = n$ *and* $\boldsymbol{X}$ *is invertible, then* $\boldsymbol{P_X} = \boldsymbol{I}$
7. $(\boldsymbol{I} - \boldsymbol{P_X}) \boldsymbol{v} \in \mathrm{Im}(\boldsymbol{X})^\perp$ *for all* $\boldsymbol{v} \in \mathbb{R}^n$
8. *If* $w \in \mathrm{Im}(\boldsymbol{X})$*, then* $\boldsymbol{P_X w} = \boldsymbol{w}$
9. *If* $\boldsymbol{w} \in \mathrm{Im}(\boldsymbol{X})^\perp$*, then* $\boldsymbol{P}_X \boldsymbol{w} = \boldsymbol{0}$
10. *If* $\boldsymbol{Z}$ *is another* $n \times m$ *matrix s.t.* $\mathrm{Im}(\boldsymbol{Z}) = \mathrm{Im}(\boldsymbol{X})$*, then* $\boldsymbol{P_Z} = \boldsymbol{P_X}$*. This means that* $\boldsymbol{P_X}$ *depends on* $\boldsymbol{X}$ *only through the span of its columns. Hence, for an arbitrary linear space* M*, we can define the projection matrix* $\boldsymbol{P}_M$ *onto* M *(an explicit form for* $\boldsymbol{P}_M$ *can be obtained by taking any basis of* M *and stacking its elements as columns in a matrix* $\boldsymbol{X}$*, then forming* $\boldsymbol{P_X} := \boldsymbol{X} \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top$*)*
11. *If* L *and* M *are two subspaces with* $L \subseteq M$*, then* $\boldsymbol{P}_M \boldsymbol{P}_L = \boldsymbol{P}_L \boldsymbol{P}_M = \boldsymbol{P}_L$*.*

Proposition 6. *We have*

1. $\boldsymbol{I} - \boldsymbol{P_X} = P_{Im(\boldsymbol{X})^\perp}$
2. *if* L *and* M *are two subspaces of* $\mathbb{R}^n$ *with* $L \subseteq M$*, then* $\boldsymbol{P}_M - \boldsymbol{P}_L = \boldsymbol{P}_{M \cap L^\perp}$

Proposition 7. *Let* $\boldsymbol{Q}$ *be an* $n \times n$ *matrix of rank* $m \leq n$ *which is symmetric and idempotent,* $\boldsymbol{Q}^\top = \boldsymbol{Q}, \boldsymbol{Q}^2 = \boldsymbol{Q}$*. Then* $\boldsymbol{Q} = \boldsymbol{P}_M$ *where* $M := \mathrm{Im}(\boldsymbol{Q})$*.*

Proof. Exercise. □

הוכחות ברשימות השיעור.

המרחב המשלים האורתוגונלי: יהי $V \subseteq U$ תת מרחב. נגדיר את תת המרחב המשלים האורתוגונלי של U באופן הבא:

$$U^{\perp} = \{v \in V | u^t v = 0, \forall u \in U\}$$

טענה: $U \oplus U^{\perp} = V$

הוכחה:

__שאלה__

1. הניחו כי A מטריצה ריבועית. הוכיחו כי $IM(A^T) = Ker(A)^{\perp}$.
2. טענה (ללא הוכחה): מטריצה A אינה לכסינה אס״ם קיים לפחות ע״ע אחד של A, λ_i עבורו $(A-\lambda_i)^k = 0, (A-\lambda_i)^m \neq 0, \forall k \geq 2, k \geq m$. השתמשו בטענה זו והראו כי אם A מטריצה סימטרית אז היא ניתנת ללכסון.
3. השתמשו בתוצאות אלו כדי להראות כי $Q = I - P_X$ היא מטריצת הטלה למרחב המשלים האורתוגונלי של $colspace(X)$ וכתבו במפורש את הפירוק הספקטרלי של Q במונחי הו״ע של P_X.
4. הסיקו מכך ש-$\hat{\beta}_{OLS}$ הוא הממזער של $||Y - X\beta||^2$ וכן שככל שדרגת X גדולה יותר, נורמה זו הולכת וקטנה.

פתרון:

1

התפלגויות רב מימדיות:

יהיו $Z_1, \dots, Z_n$ משתנים מקריים המפולגים במשותף (ללא שום הנחות נוספות) בהתפלגות $f_{Z_1,\dots,Z_n}(z_1, \dots, z_n)$. נגדיר $Z \in R^n = (Z_1, \dots Z_n)$ ונגדיר $E(Z) = (E[Z_1], \dots, E[Z_n])$.

באותו האופן, אם A היא **מטריצה מקרית**, כלומר A_{ij} הוא מ"מ: $E[A] = \begin{pmatrix} E[A_{11}] & \dots & E[A_{1m}] \\ \dots & \dots & \dots \\ E[A_{n1}] & \dots & E[A_{nm}] \end{pmatrix}$

תכונות של תוחלות של מטריצות (ובפרט וקטורים) מקריים:

עבור Z, W מטריצות מקריות ו-A, B מטריצות קבועות, לא מקריות (דטרמיניסטיות):

1. $\mathbb{E}[\boldsymbol{Z} + \boldsymbol{W}] = \mathbb{E}[\boldsymbol{Z}] + \mathbb{E}[\boldsymbol{W}]$
2. $\mathbb{E}[\boldsymbol{A}\boldsymbol{Z}\boldsymbol{B}] = \boldsymbol{A}\mathbb{E}[\boldsymbol{Z}]\boldsymbol{B}$
3. $\mathbb{E}[\boldsymbol{A}U + C] = A\mathbb{E}[U] + \boldsymbol{C}$ (from 1+2)

עבור זוג וקטורים מקריים Z, W נגדיר את מטריצת השונות המשותפת בין Z, W:

$$Cov(Z, W) \coloneqq E([Z - E(Z)][W - E(W)]^T)$$

ובאופן דומה את מטריצת השונויות של הוקטור Z:

$$Var(Z) := Cov(Z, Z) \coloneqq E([Z - E(Z)][Z - E(Z)]^T)$$

טענה: $Var(Z)_{ij} = cov(Z_i, Z_j) = cov(Z_j, Z_i) = Var(Z)_{ji}$

"הוכחה":

בתרגיל תוכיחו את התכונות הבאות:

Properties of covariance matrix. $\boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{R}$ random vectors; $\boldsymbol{a}$ fixed vector. Then the following properties hold:

1. $\text{cov}(\boldsymbol{Z}, \boldsymbol{W}) = \text{cov}(\boldsymbol{W}, \boldsymbol{Z})^\top$
2. $\text{cov}(\boldsymbol{Z} + \boldsymbol{R}, \boldsymbol{W}) = \text{cov}(\boldsymbol{Z}, \boldsymbol{W}) + \text{cov}(\boldsymbol{R}, \boldsymbol{W})$
3. $\text{cov}(\boldsymbol{A}\boldsymbol{Z}, \boldsymbol{B}\boldsymbol{W}) = A\text{cov}(\boldsymbol{Z}, \boldsymbol{W})\boldsymbol{B}^\top$
4. $\text{cov}(\boldsymbol{A}\boldsymbol{Z}) = \boldsymbol{A}\text{cov}(\boldsymbol{Z})\boldsymbol{A}^\top$ (from 3)
5. $V\left(\boldsymbol{a}^\top\boldsymbol{Z}\right) = \boldsymbol{a}^\top\text{cov}(\boldsymbol{Z})\boldsymbol{a}$ (from 4)
6. $\text{cov}(\boldsymbol{Z})$ is a nonnegative definite matrix (from 5)

<u>שאלה:</u>

יהיו המשתנים המקריים הבאים: $X \sim Ber(p), Y \sim Ber(q)$ כך שמתקיים: $M := XY \sim Ber(r)$.

א. מצאו את וקטור התוחלות ומטריצת השונויות של הוקטור המקרי $Z = (X, Y, M)^T$

ב. נגדיר את ההעתקה: $A: R^3 \to R$: $A(v) = 2v_1 - 3v_2 + 4v_3 + 7$. האם ההעתקה לינארית? על בסיס הסעיף הקודם חשבו את וקטור התוחלות ומטריצת השונויות של $A(Z)$.

ג. כעת הניחו ש-X, Y ב"ת. חשבו את ההסתברות $P(X = 1, Y = 1, M = 1)$.

פתרון:

יהיו $Z, W \in R^p$ וקטורים מקריים. הראו שהבאים שקולים:

(1) $\forall v \in R^p, Var(v^T Z) \geq Var(v^T W)$

(2) $B := Var(Z) - Var(W)$ היא מטריצה חיובית למחצה.

(3) קיימת המטריצה $B^{\frac{1}{2}}$.

המודל הלינארי:

נתונים: $(\boldsymbol{X}_i, Y_i), \quad i = 1,...,n$

מודל ליניארי:

$$Y_i = \sum_{j=0}^{p} \beta_j X_{ij} + \epsilon_i, \qquad \mathrm{Cov}(\epsilon_i, \epsilon_{i'}) = \begin{cases} \sigma^2, & i = i' \\ 0, & i \neq i' \end{cases}$$

כאשר $\boldsymbol{X_i} = (1, X_{i1}, \ldots, X_{ip})^\top$ הם ממימד $p+1$,

וכאשר $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ ו-σ^2 הם קבועים לא ידועים

אפשר לקבל ייצוג קומפקטי בעזרת כתיב מטריצות. נסמן:

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \qquad \boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{bmatrix}, \qquad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

אז את המודל הליניארי אפשר לכתוב:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}, \qquad \mathbb{E}[\boldsymbol{\epsilon}] = \boldsymbol{0}, \ \ \mathrm{cov}[\boldsymbol{\epsilon}] = \sigma^2 \boldsymbol{I}$$

כמו כן: $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}, \qquad \boldsymbol{e} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$

(הערה: אם לא נציין אחרת, $\hat{\boldsymbol{\beta}}$ זה אומד הריבועים הפחותים)

<u>שאלה</u>

לפניכם מספר טענות. התאימו לכל טענה האם מדובר בהנחה או בתוצאה מתמטית:

1. $\hat{\beta} = argmin_{\beta}||Y - X\beta||^2$
2. $X\beta = E(Y|X)$

3. $0 = E(e_i)$
4. $0 = E(\bar{e})$
5. $X^T e = 0$
6. $Cov(Y) = \sigma^2 I$

פתרון:

1

שאלה:

לפניכם מתוארים מספר מקרים. עבור כל אחד מהם פרטו את ההתפלגויות של $X, Y, Y|X, \epsilon$ וכתבו אילו הנחות של המודל הלינארי כל אחד מהם מקיים:

1. $X_i \in R^p$ **קבועים מראש**, $Y_i = X_i^T \beta + \epsilon_i$ כאשר $\epsilon_i \sim N(0,1)$ ב״ת.
2. $X_i \in R^p$ מ״מ נורמליים סטנרדטיים וב״ת, $Y_i = X_i^T \beta + \epsilon_i$ כאשר $\epsilon_i \sim N(0,1)$ ב״ת.
3. $X_i \in R^1$ מ״מ נורמליים סטנרדטיים וב״ת, $Y_i = X_i^2 \beta + \epsilon_i$ כאשר $\epsilon_i \sim U(-1,1)$ ב״ת.
4. $X_{it} \in R^p$ **קבועים מראש**- כאשר X_{itj} היא הדגימה של האדם ה- i בתקופה ה-t, $Y_{it} = X_{it} \beta + \epsilon_{it}$ כאשר $\epsilon_{it} \sim N(0,1)$ והנדגמים ב״ת.

פתרון:

1

שאלה

1. יהי $0 \neq v \in R^n$. הראו כי $\frac{vv^T}{||v||^2}$ היא מטריצת הטלה אורתוגונלית. מה דרגת המטריצה?
2. יהיו $Y_1, \dots, Y_n \sim (\mu, \sigma^2)$ מ״מ ב״ת ש״ה, כאשר שני הפרמטרים לא ידועים. הראו כי $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$ הוא אומד ח״ה ל-σ^2.

3. הניחו כעת כי $Y_1, \dots, Y_n \sim N(\mu, \sigma^2)$. הוכיחו את התוצאה, שראינו בעבר $(n-1)S_n^2 \sim \sigma^2\chi_{n-1}^2$.

<u>שאלה</u>

2.נתון $Z \in \mathbb{R}^m$ וקטור מקרי. הראו כי מתקיים ש

$$\mathbb{E}\left(||Z||^2\right) = tr(\mathbb{E}[ZZ^T])$$

הסיקו מכך כי אם $\mathbb{E}[Z] = 0$ אזי מתקיים כי

$$\mathbb{E}\left(||Z||^2\right) = tr(cov[Z])$$

פרטו והצדיקו כל שלב בהוכחה.

הוכחה :

1