רגרסיה ומודלים סטטיסטיים- תרגיל 4

שאלה 1:

נתון שהמטריצה $X \in R^{n\times(p+1)}$ מדרגה מלאה. הניחו שהנחות המודל הליניארי מתקיימות.

חשבו את $cov(e), cov(\epsilon, e), cov(\hat{\beta}, \epsilon), Cov(Y_i, \hat{Y}_i)$ וציינו את המימד של כל אחד מהגדלים.
הערה: על מנת למלא את חובת הגשה- הגישו את פתרונכם עבור 2 מתוך 4 השונויות המשותפות הנ"ל.

שאלה 2:

נתון מודל ליניארי, $Y = X\beta + \varepsilon$ כאשר $\varepsilon \sim (0, \sigma^2 I_n)$.

עבור $u, v \in \mathbb{R}^n$ כלשהם נגדיר את נורמת מהלנוביס (Mahalanobis) באופן הבא:

$$||u - v||_\Sigma = \sqrt{(u - v)^T \Sigma^{-1} (u - v)}$$

כאשר $\Sigma \in \mathbb{R}^{n\times n}$ מטריצה סימטרית חיובית.

כמו כן נגדיר את $\hat{\beta}^\Sigma$ באופן הבא:

$$\hat{\beta}^\Sigma = \underset{\beta\in\mathbb{R}^{p+1}}{argmin} ||Y - X\beta||_\Sigma^2$$

א. הראו שמתקיים

$$\hat{\beta}^\Sigma = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$$

והסבירו את האומד שמתקבל במקרה $\Sigma = I_n$.

<u>הדרכה:</u> אפשר (לא חייב ש) להעזר בעובדה הבאה: $A \in \mathbb{R}^{m\times m}$ מטריצה סימטרית חיובית, אזי קיימת מטריצה $B \in \mathbb{R}^{m\times m}$ סימטרית חיובית כך ש: $B^T B = A$. ניתן להפעיל מסקנה זו על Σ כך שיהיה ניתן לייצג את Σ^{-1} באופן הבא: $\Sigma^{-1} = C^T C$ עבור מטריצה C ריבועית כלשהי. מצאו מי היא C, הסבירו מדוע היא מוגדרת היטב, ואז הגדירו משתנים חדשים

$\tilde{X} = \boldsymbol{C}X$, $\tilde{Y} = \boldsymbol{C}Y$, ופתרו את הבעיה (שימו לב שעבור ה **C** הנכון הפתרון מיידי, חישבו מדוע).

ב. הראו כי $\hat{\beta}^\Sigma$ הינו אומד חסר הטיה לβ, וכן מצאו את $cov(\hat{\beta}^\Sigma)$.

ג. מה אומר משפט גאוס מרקוב לגבי המטריצה$cov(\hat{\beta}^\Sigma) - cov(\hat{\beta})$, כאשר $\hat{\beta}$ אומד הריבועים הפחותים?

ד. כעת הניחו מודל ליניארי כללי יותר: $Y = X\beta + \varepsilon$ כאשר $\varepsilon \sim (0, \sigma^2 \Sigma)$ עבור אותה מטריצה סימטרית חיובית $\Sigma \in \mathbb{R}^{n\times n}$ שביחס אליה מוגדר האומד $\hat{\beta}^\Sigma$ (שימו לב ש Σ מטריצה סימטרית חיובית כלשהי, זהו אכן מודל כללי יותר שכן עבור $\Sigma = I_n$ מתקבל המודל "הרגיל").

הראו שבמקרה זה $\hat{\beta}^\Sigma$ הינו האומד הליניארי חסר ההטיה הטוב ביותר מבחינת שגיאה ריבועית מבין האומדים הליניאריים חסרי ההטיה לβ. כלומר, שלכל צירוף ליניארי $\theta = a^T\beta$ האומד $\hat{\theta}^\Sigma = a^T \hat{\beta}^\Sigma$ משיג שונות מינימאלית מבין כל האומדים הליניאריים חסרי ההטיה ל θ.

<u>הדרכה:</u> השתמשו בהדרכה מסעיף א', כתבו את המודל במונחי $\tilde{X}$, $\tilde{Y}$. הראו שמתקבל המודל הליניארי "הרגיל" ומשם ההוכחה שקולה להוכחת משפט גאוס מרקוב "הרגיל".

3 רגולריזציות לרגרסיית ריבועים פחותים (Ridge & Lasso)

Ridge

$$\hat{\beta}_{\text{Ridge}} = \arg\min_{\beta} \|X\beta - y\|_2^2 + \lambda_{\text{Ridge}}\|\beta\|_2^2$$

הניחו מודל לינארי $Y = X\beta + \varepsilon$, עבור $E[\varepsilon] = 0$, $Cov(\varepsilon) = \sigma^2 I$.

וכי הפירוק הספקטרלי של $X^T X = UDU^T$

א. הראו כי

$$\hat{\beta}_{Ridge} = (X^T X + \lambda I)^{-1} X^T Y$$

ב. מצאו ביטוי עבור התוחלת של $\hat{\beta}_{Ridge}$.

ג. בטאו את השונות של $\hat{\beta}_{Ridge}$ באמצעות רכיבי D , עמודות U, λ ו- σ^2.

ד. הוכיחו שלכל וקטור קבוע $a \in R^p$, $a \neq 0$ מתקיים $Var(a'\hat{\beta}_{ridge}) < Var(a'\hat{\beta}_{OLS})$,

והסבירו כיצד התכונה מתיישבת עם משפט גאוס-מרקוב. (כלומר משפט ה"אומד BLUE").

רשות (!!!):

כעת, נבצע מספר סימולציות אשר ימחישו את התוצאות התיאורטיות שראינו. בתיבת ההגשה במודל יש קובץ נתונים *quiz2_df.csv* וקובץ R בשם *Quiz2.R*. קבצים אלו ישמשו אותנו בסעיפים הבאים. בקובץ הנתונים יש משתנים מסבירים X ומשתנה מוסבר Y . בשביל לקבל תוצאות הגיוניות ברגרסית Ridge יש לעשות תקנון (Scale) של X לפי כל עמודה (כלומר, כל עמודה של X , למעט החותך, צריכה להיות עם תוחלת 0 וסטיית תקן 1).

בקובץ *Quiz2.R* נתונים הערכים האמיתיים של β, σ^2 וכן שני וקטורים של ערכי λ שנרצה לבדוק. בנוסף, בקובץ יש את הפונקציה *ridge_aux_functions()* עם הסברים לגבי הקלטים והפלטים שלה. תשלימו את הנדרש בהסבר הפונקציה, כלומר תכתבו את הפונקציה לפי ההדרכה הנמצאת בקובץ R הנתון.

עבור כל אחד מערכי λ בוקטור *lambda_seq* , חשבו את (8) עבור $\hat{\beta}^{OLS}$ ו- $\hat{\beta}^{Ridge}$. את התוצאות שהתקבלו הציגו בגרף שבציר ה-X יש את ערכי λ בסדר עולה (משמאל לימין) ובציר ה-Y את ערכי ה- MSE שהתקבלו עבור כל מודל ועבור כל λ. דונו בתוצאות.

עבור כל אחד מערכי λ בוקטור *lambda_seq_mod* , חשבו את $\hat{\beta}_j^{Ridge}, \ j = 0, \ldots, p$. הציגו את התוצאות בגרף שבו בציר ה-x יש את ערכי λ ובציר ה-Y את ערכי $\hat{\beta}_j^{Ridge}$ השונים. לכל משתנה $j \in \{0, \ldots, p\}$ הציגו את הגרף בצבע שונה. דונו בקצרה בתוצאות.

הערה: הפקודות הנתונות בתרגיל הן ב-*R* אבל אתם יכולים להשתמש בשפת תכנות לבחירתכם (בהתאם להנחיות ועם פונקציות אחרות משלכם בשאלה 3), וכן בכל חומר עזר/מודל שפה ג'נרטיבי. הקפידו לצרף את הקוד והסברים מילוליים משלכם. הגישו קובץ *PDF* אחד הכולל את הפתרון שלכם לכל השאלות וכן את הקוד והתוצאות של השאלה האחרונה.

שאלה 4:

א. יהי $W \in R^n$ וקטור רב נורמלי עם תוחלת $\mu \in R^n$ ומטריצת שונויות $\Sigma \in R^{n\times n} > 0$. זכרו כי במקרה זה קיימת צפיפות ושווה ל:

$$f_{\overrightarrow{W}}(\vec{w}) = (2\pi)^{-\frac{n}{2}} \cdot |\Sigma|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}\overrightarrow{(w-\mu)}^T\Sigma^{-1}\overrightarrow{(w-\mu)}}$$

הוכיחו כי בהתפלגות הרב-נורמלית חוסר מתאם בין רכיבי הוקטור W שקול לאי תלות בין רכיביו (הוכיחו את שני הכיוונים).

הניחו כעת את המודל הלינארי הנורמלי. כלומר: $Y = X\beta + \epsilon,\ \ \epsilon \sim N_n(0, \sigma^2 I)$

ב. השתמשו בסעיף הקודם כדי להראות:

המשתנים $\left|\left|\hat{Y}\right|\right|^2$ ו-$\left|\left|e\right|\right|^2$ ב"ת. **הדרכה:** הראו ש-$\hat{Y}$ ו-e ב"ת, והסיקו על הנורמות תוך שימוש בעובדה הבאה:
יהיו $X \in R^n, Y \in R^k$ מ"מ ב"ת. אז לכל $f: R^n \to R$, $g: R^k \to R$ מתקיים שהמשתנים $f(X), g(Y)$ ב"ת. פרטו מי הן f, g.

ג. מצאו את ההתפלגות של $\hat{\beta}_j$ ובנו רווח סמך ל-β_j כאשר σ^2 ידוע. חזרו על כך כאשר σ^2 איננו ידוע אלא נאמד על ידי אומד ח"ה.