

תרגיל בית 3

הנחיות להגשת התרגיל

- לצורך פתרון שאלה 4 ניתן להשתמש בספריות לקריאת והצגת נתונים, numpy וספריות סטטיסטיות עבור חישוב השיברונים, אך לא בספריות שמבצעות ישירות רגרסיה לינארית או מוציאות פלט סטנדרטי עבורה.
- יש להגיש שני קבצים נפרדים, אחד עבור הקוד ואחד לפתרון החלק היבש. בתוך קובץ פתרון החלק היבש הוסיפו גם את הפלט של הקוד.

שאלה 1

יהי X וקטור מקרי בעל וקטור תוחלת $\mathbb{E}[X] = \mu_X \in \mathbb{R}^n$. מטריצת השונות של X נתונה על ידי $\Sigma = \text{Var}(X) = \mathbb{E}[(X - \mu_X)(X - \mu_X)^T]$. הוכיחו כי $\text{Var}(AX + b) = A\Sigma A^T$ עבור $A \in \mathbb{R}^{m \times n}$ ($b \in \mathbb{R}^m$).

שאלה 2

- נאמר ש- A היא מטריצה אידמפוטנטית אם היא מקיימת $A^2 = A$.
- ראינו בהרצאה ש- $\hat{y} = (X(X^T X)^{-1} X^T)y$. הראו שמטריצת הכובע, $P = X(X^T X)^{-1} X^T$, היא מטריצה אידמפוטנטית סימטרית. מצאו את הממד והדרגה של P .
 - הראו שהערכים העצמיים של מטריצות אידמפוטנטיות הם 0 או 1.
 - במודל רגרסיה לינארית אנו מניחים $n \gg p$. הסבירו מדוע לא ניתן למצא אומד ריבועים פחותים כפי שלמדנו בקורס כאשר $n > p$.

שאלה 3

עבור מודל רגרסיה לינארית פשוט $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, הראו כי כאשר המודל נכון נצפה לסטטיסטי מבחן F גדול.

הדרכה: הראו כיצד סכומי הריבועים הרלוונטיים תלויים ב- $\hat{\beta}_1$.

שאלה 4

חוזק ההלחמה של החוט הוא מאפיין חשוב במעגלים חשמליים. בקובץ ex3.csv מצורפת טבלה המתארת את חוזק ההלחמה (Y), גובה הרכיב (X_1), גובה המקור (X_2), גובה הלולאה (X_3), גובה החוט (X_4), רוחב החיבור (X_5), רוחב החיבור במקור (X_6). נרצה להתאים מודל רגרסיה לינארית לחוזק ההלחמה.

העזרו בפייתון וענו על הסעיפים הבאים.

- מצאו אומד ריבועים פחותים עבור המשתנים המסבירים X_2, X_3, X_4, X_5 . רשמו במפורש את המודל.

- מצאו את האומד חסר ההטיה לשונות הרעש σ_ε^2 .

- מלאו את טבלת ניתוח שונות עבור המודל

Source	Sum of Squares	df	MS	F
Regression				
Residuals				
Total				

תחת העמודה F יש להציג את סטטיסטי מבחן F ואת מובהקות התוצאה (p-value).

- ד. חשבו את התחזית לחוזק ההלחמה עבור $X_2 = 20, X_3 = 30, X_4 = 90, X_5 = 2$.
- ה. חשבו רווח סמך ברמת סמך 0.95 עבור תוחלת תחזית חוזק ההלחמה המתקבלת מהנתונים בסעיף ד'.
- ו. חשבו רווח סמך ברמת סמך 0.95 עבור תחזית חוזק ההלחמה המתקבלת מהנתונים בסעיף ד'.
- ז. האם הנחות המודל הלינארי מתקיימות? הסבירו והדגימו באמצעים המתאימים (אין צורך לבדוק הנחות שלא למדנו איך לבדוק אותן).