מתרגל: ניר קרת
סמסטר ב' תשפ"א

תרגיל 6

- יש לענות על כל השאלות.
- הגשת התרגילים היא **בזוגות קבועים**.
- שאלות המסומנות ב-** הן שאלות רשות.
- את התרגילים יש להגיש לתיבת ההגשה במודל.
- יש להגיש את כל התשובות והפלטים יחד בקובץ PDF אחד, למשל על ידי שימוש ב Rmarkdown.

1) ב *moodle* ישנו קובץ נתונים בשם *ex6data1*. נניח כי המודל המתאר את הנתונים הללו הינו

$$y_i = 1 + \beta_1 e^{\frac{\beta_1 x_{1_i}}{\beta_2 x_{2_i}}} + \epsilon_i \quad E(\epsilon_i) = 0, V(\epsilon_i) = \sigma^2 \ \ \forall i$$

$$\epsilon_i, \epsilon_j \text{ are independent } \forall i \neq j$$

מצאו אומדים לפרמטרים β_1, β_2 בשיטת אר"פ לא-לינארי בשתי דרכים (על ידי כתיבת פונקציות ב-*R*):

א. בעזרת שיטת גאוס-ניוטון.
ב. על ידי ניוטון-רפסון, אולם את הנגזרות השניות עליכם לחשב בצורה נומרית (כפי שמפורט בשקופית 21 במצגת 4). שימו לב שאת הנגזרות הראשונות עליכם לחשב אנליטית.

2) קרובת משפחה של הרגרסיה הלוגיסטית הינה הרגרסיה הפואסונית. במודל זה המשתנה המוסבר $\mathrm{Y_i}$ סופר את מספר הפעמים שתצפית *i* חוותה אירוע מסויים. בנוסף, לכל תצפית *i* ישנו וקטור של *p* משתנים מסבירים $\underline{X_i}$. אזי, ברגרסיית פואסון אנו מניחים כי לכל *i=1,…,n*

$$Y_i | \underline{X_i} \sim Pois(\lambda_i)$$

ומניחים את המודל הלוג-לינארי הבא:

$$\ln(\lambda_i) = \underline{X_i}^T \underline{\beta}$$

בשאלה זו, נניח כי המודל מכיל רק שני פרמטרים, כלומר:

$$\ln(\lambda_i) = \beta_0 + \beta_1 X_i$$

א. רשמו את פונקציית הנראות כפונקצייה של הפרמטרים β_0, β_1.
ב. הציגו את פונקציית ה*score* והשוו ל-0. האם ניתן לקבל פתרון סגור?
ג. בצעו את הצעדים הנדרשים על מנת למצוא פתרון באמצעות *NR*, וכתבו תוכנית ב-*R* המבצעת את *NR*.

ד. נתחו את קובץ הנתונים הנקרא *count data* שנמצא ב*moodle* בעזרת הפונקצייה שכתבתם. מהם האומדנים עבור הפרמטרים? השוו את התוצאות לפלט המתקבל מהפונקצייה *glm* (לא לשכוח להגדיר בהתאם את ארגומנט ה *family*). מתוך הפלט של *glm* ציינו את סטיות התקן של האומדנים לפרמטרים.

ה. האם במקרה זה יהיה הבדל בין *NR* ל-*FS*? הסבירו.

**3) תהי $g(x)$ פונקצייה גזירה פעמיים בקטע $[a,b]$. ידוע כי קיימת נקודה $c \in [a,b]$ יחידה בה מתקיים $g(c)=0$. כמו כן נתון:

$$g(a) < 0 \quad , g(b) > 0$$

$$\forall x \in [a,b] \; g'(x) > 0 \quad , 0 \le g''(x) \le M$$

הוכיחו כי אלגוריתם NR מתכנס ל-c כאשר נקודת ההתחלה $x_o \in [c,b]$. התחילו בלפתח טור טיילור עם טעות מסדר שני עבור $g(c)$ סביב הנקודה $x_j \in [c,b]$ והראו שהסדרה $\{x_j\}_{j=0}^{\infty}$ חסומה מלרע ע"י c וכן יורדת.

**4)

א. הראו כי כאשר מדובר על פתרון של מערכת משוואות **ליניאריות**, *NR* יסתכם בפעולה בודדת של הפיכת מטריצה, ו *Gauss-Newton* יסתכם בצעד בודד של אומד ריבועים פחותים.

ב. נתונה הפונקצייה: $f(x) = \sqrt{\ln(x^2+1)\ln(x+5)} - 0.5$
ברצוננו למצוא לה שורשים על ידי *NR*. ישנם סה"כ 3 שורשים.

(1) מהו תחום ההגדרה של הפונקצייה?

אם באיטרצייה מסויימת נגיע לנקודה מחוץ לתחום ההגדרה, האלגוריתם ייכשל. זה בעייתי במיוחד אם שורש הפונקצייה נמצא קרוב לשפת תחום ההגדרה.

(2) נסו למצוא שורשים לפונקצייה על ידי שימוש כרגיל ב *NR*. אילו שגיאות אתם מקבלים? מהם השורשים שאתם מצליחים למצוא?

(3) כעת, במידה והאלגוריתם מגיע לניחוש מחוץ לתחום ההגדרה, השתמשו באיטרצייה הבאה בערכו המוחלט של אותו הניחוש במקום זאת. אילו שורשים ושגיאות אתם מקבלים כעת?

(4)פתרון אחר: החל מהנקודה בה הפונקצייה אינה מוגדרת, תחברו לה פונקציית "*rebound*", כלומר פונקצייה שתחזיר את

האלגוריתם להיות בתוך תחום ההגדרה מחד, לא "תזרוק" את הניחוש הבא רחוק מדיי לאינסוף, וגם לא תיצור שורשים חדשים לפונקצייה "יש מאין". נסו לחבר את הפונקצייה הבאה: $a - e^{-0.05x}$. מיצאו את ה- a הדרוש כדי שהפונקצייה הכללית (המקורית + הפונקצייה המשודכת) תהיה רציפה (תיתכן נקודת אי-גזירות).
אילו שורשים אתם מקבלים כעת? האם עדיין מתקבלות שגיאות?

5) א. בצעו את החישובים הדרושים על מנת להשתמש באלגוריתם *NR* למציאת אנ"מ להתפלגות *Gompertz*.
ב. עבור הנתונים *ex6data2*, מצאו אנ"מ זה.

6)
א. ** הוכיחו: ב*gradient descent* , אם לשם מציאת גודל הצעד γ פותרים את בעיית האופטימיזציה – $\gamma_{j+1} = argmin_\gamma f(\overrightarrow{x_j} - \gamma \nabla f(\overrightarrow{x_j}))$ אזי בכל שתי איטרציות עוקבות מתקיים שהכיוון בו נעשה את הצעד יהיה מאונך לכיוון בו התבצע הצעד הקודם, כלומר:
$$\nabla f(\overrightarrow{x_j}) \perp \nabla f(\vec{x}_{j-1})$$
ב. פונקציית *Rosenbrock* הינה הפונקצייה הבאה:
$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2$$

מאחר והפונקצייה מקבלת תמיד ערכים אי-שליליים, ברור שעל ידי הצבת $x = y = 1$ נשיג את המינימום הגלובלי. עם זאת, נסו להגיע אל המינימום הגלובלי בשתי שיטות:
א. Gradient Descent (על ידי אחת השיטות לקביעת גודל הצעד שלמדתם בכיתה)
ב. ניוטון-רפסון

עבור שתי השיטות, בחרו כלל עצירה כרצונכם ונסו נקודות התחלה שונות.

השוו בין מהירות ההתכנסות של שתי השיטות והסבירו את התופעה.