

פרויקט סיום מבוא ללמידה חישובית- רגרסיה לינארית ושיטות למציאת מינימום

- 1) נתון קובץ עם נתונים שקשורים לחולי סרטן כאשר העמודה האחרונה היא ערך ה- y שאותו רוצים לחזות (תוחלת החיים לחולה). קרא את הנתונים והכנס למטריצה X ווקטור y .
 - א) נרמל את הנתונים (בצע בדיקה שאחרי הנרמול אכן הממוצע 0 וסטית התקן 1).
 - ב) הוסף עמודה של אחדים למטריצה X
 - ג) רשום פונקציה שמקבלת את θ ווקטור x ומחזירה את $h_{\theta}(x)$ (במקרה של רגרסיה לינארית)
 - ד) רשום פונקציה שמקבלת וקטור θ ואת המטריצות X ו- y ומחזירה את הערך של $J(\theta)$
 - ה) רשום פונקציה שמקבלת וקטור θ ומטריצות X ו- y ומחזירה את הערך של $\nabla J(\theta)$
 - ו) הרץ את אלגוריתם Gradient Descent עם כמה ערכים של α (למשל 1, 0.1, 0.01, 0.001) וצייר את הגרף שמראה את ירידת הערך של $J(\theta)$ כפונקציה של צעדי זמן.
 - ז) הרץ את אותו הקוד עם mini-batch השווה לסעיף הקודם. מה מסקנותיך מהריצה?
 - ח) הרץ SVD על הדאטה X והשאר עם 3 פיצ'רים בלבד. (הורדת מימד ל-3). הרץ שוב את האלגוריתם על X אחרי הורדת מימד. מה מסקנותיך על הריצה מבחינת תוצאות וזמני ריצה?

יש להגיש את הפרויקט בזוגות, להוסיף לעבודה (כקובץ נפרד או בתוך מחברת הפייתון) מסקנות מההרצות שקשורות להשוואה בין ריצות, התכנסות וכו