**תרגיל בית רטוב 2 – מבוא למערכות לומדות (02360766)**

*בן הייטנר – 213930175*

*לילך ביטון – 205764517*

חלק 1: בחירת מודל בסיסי עבור kNN

**(Q2)**  
ה-k הטוב ביותר הוא: 13 (הוא האחד שעבורו מתקבל הדיוק הגבוה ביותר על ה-test data).  
עבור k קטן מאוד (לדוג' 1) נקבל overfitting שכן אזורי ההחלטה תלויים במעט נקודות ועל כן רגישים מאוד לסטיות בנתונים - כל סטיה מקבלת יחס של מקרה רגיל ונפוץ בלי קשר במגמה של הנתונים.  
עבור k גדול מאוד נקבל underfitting מהסיבה ההפוכה – כל מקרה שלא רווח מאוד מקבל יחס של סטייה.

Attach plots + write average training and test accuracy

**(Q3)**

attach plots test accuracy +

**(Q4)**

ב-Q1 השתמשנו במודל עם k=1 וכפי שציינו ב-Q2 זה גורר ל-overfitting קיצוני כתוצאה מרגישות יתר של המודל לסטיות. לעומת זאת, ב-Q3 ע"י tuning של ה-k קיבלנו אזורי החלטה שמייצגים את המגמה של הנתונים ע"י איזון הרגישות לסטיות נקודתיות. עבור k-ים גדולים מ-13 נתחיל לקבל underfitting וכפי שראינו עבור ב-Q1 עבור k-ים קטנים מ-13 נקבל overfitting.

חלק 2: עצי החלטה

**(Q5)**

attach plot + training accuracy

**(Q6)**

Attach plots  
קומבינציה טובה: 12,8

Underfit עבור 19,5 – קשה מאוד ליצור פיצול בגלל min leaf ויש מעט פיצולים בגלל depth.

Overfitting עבור 1,24 - מפצלים על כל שטות בגלל min leaf ויש הרבה פיצולים בגלל depth.

**(Q7)**

מספר הקומבינציות השונות שווה למכפלת גדלי הטווחים של ההיפר-פרמטרים. במקרה שלנו עבור שתי היפר-פרמטרים עם טווחים 20 ו-19 קיבלנו 380 קומבינציות. בהינתן היפרפרמטר נוסף, מספר הקומבינציות היה מוכפל בגודל הטווח של ההיפר-פרמטר החדש.

**(Q8)**

Test accuracy of optimal tree

חלק 3: SVM לינארי והגרעין הפולינומי

**Q10))**

קיבלנו underfit מפחיד עם דיוק מזעזע ששומר את ה-loss גבוהה כי בחרנו lr קטן מדי.

**Q11))**

קצב הלמידה שהיינו בוחרים הוא כי הוא מקנה התכנסות יציבה ל-loss נמוך ודיוק גבוה ביחס לאחרים. עבור ערכים קטנים יותר לא מקבלים תוצאות טובות באותה המידה ועבור רערכים גדולים יותר נקבל רעש אקססיבי שגם פוגע בתוצאות.