חלק ב' – פיתוח KNN

מימוש KNN- מצורפי קבצי הקוד להגשה.

הרצת ניסויים:

**תרגיל 5:**

1. מצורף *experiments*6.*csv* להגשה.
2. מבין המדידות שעשינו עבור k=3 התקבל הערך המקסימלי של דיוק 0.944 כלומר 94.4%.
3. ערך המינימום הינו: 0.935 ומתקבל כאשר k=13, ערך המקסימום הינו 0.944 ומתקבל כאשר k=3. כפי שניתן לראות מהגרף ישנה מגמת עלייה עד k=3 וישנה מגמת ירידה החל מk=3. מגמת העלייה בדיוק הראשונית נובעת מכיוון שהגדלת הערך של k מונעת תופעות של overfitting . אמנם הגדלת יתר של k כפי שאנו שמים לב דווקא גורמת לירידה בדיוק מכיוון שהאלגוריתם מסתמך על שכנים "רחוקים" יותר ולכן הדמיון בין האובייקט הנבדק לשכנים אלו קטן.  
   נוסיף כי העלייה בסוף הגרף מערך k=7 לערך k=13 היא מזערית ולא יכולה להעיד כי הדיוק יעלה אם נמשיך להגדיל את k.  
   כפי שניתן לראות הצלחנו למצוא k אופטימלי בעזרת k – cross - fold validation , מבין הערכים שנבדקו.

**תרגיל 7:**

1. צורף בקוד.
2. צורף בקוד.
3. צורף כקובץ להגשה.
4. התוצאות הכי טובות התקבלו עבור אלגוריתם KNN **לכל k שנבדק בפרט עבור k=3.**

חלק ג' – תחרות ובונוס

בחלק זה נתאר את המסווג אשר בנינו לתחרות ואת בחירת הפרמטרים עבורו.

אנו השתמשנו במסווג אשר כולל ועדה של 3 מסווגים: Neural Network, KNN, ID3 .  
על מנת לממש את המסווג Neural Networkהשתמשנו בחבילה “sklearn” אשר כוללת קוד מוכן של מסווג זה.

למרות שמסווג מסוג Neural Network לא נלמד בקורס זה, הרקע לבחירתו הוא על ידי חיפוש ברשת ורקע קודם מקורס מערכות לומדות (של הפקולטה לחשמל). קוד זה למעשה בונה מסווג מסוג Neural Network על פי הפרמטרים שהוא מקבל ומאמן את המסווג על פי קבוצת אימון נתונה.

* נרמול המידע:   
  כפי שראינו בהרצאה כדאי לנרמל את המידע זאת מכיוון שנרצה למנוע תיעדוף של תכונה אחת על גבי אחרת רק בגלל הטווח ערכים שלה נרמלנו את המידע לפי נרמול Z לפי תוחלת וסטיית התקן.
* בחירת פרמטרים:  
  בחירת פרמטר K למסווג KNN נעשתה בחלק הראשון של התרגיל ובחרנו בK=3.  
  בחירת פרמטרים עבור ID3, Neural Network נעשתה על ידי hyperparameters optimization על ידי בניית המסווג עם סטים של פרמטרים והערכתו על ידי k-fold-cross validation כאשר k=4.
* קבלת החלטה:

לבסוף עטפנו את שלושת המסווגים במסווג יחיד אשר מסווג את האובייקט על סמן 3 המסווגים שקבענו ומחליט את הסיווג הסופי על פי החלטת הרוב. המוטיבציה לממש מסווג כזה היא שעל מנת שהמסווג יטעה לפחות שני מסווגים שעטף צריכים לטעות ומכיוון שהם מסווגים בצורה שונה אחד מהשני הסיכוי לכך קטן.