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 אשר מטיבה עם הדיוק על קבוצת המבחן. אמנם הגדלת יתר של k כפי שאנו שמים לב דווקא גורמת לירידה בדיוק מכיוון שהאלגוריתם כעת מסתמך על שכנים יותר "רחוקים" ולכן הדמיון בין האובייקט הנבדק לשכנים אלו קטן וההנחה כי שכנים קרובים הם דומים כבר לא תקפה.  
   נוסיף כי העלייה בסוף הגרף מערך k=7 לערך k=13 היא מזערית ולא יכולה להעיד כי הדיוק יעלה אם נמשיך להגדיל את k אלא להפך בגלל שאנו מאמנים רק על 500 אובייקטים ובודקים על 500 נצפה לירידה.  
   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 ביחס לאלגוריתם האחרים שנבדק בפרט עבור k=3**, לדעתנו תוצאה זו נובעת מכך שאכן ההנחה כי אובייקטים דומים הם בעלי סיווג זהה חזקה עבור בעיית למידה זו.

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 את כל אחד מתתי המסווגים: Neural Network, KNN, ID3 השתמשנו בחבילת Python: “sklearn” אשר כוללת קוד מוכן של כל מסווגים אלו.

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.

שלבי בניית המסווג:

1. נרמול המידע:

כפי שראינו בהרצאה כדאי לנרמל את המידע זאת מכיוון שנרצה למנוע העדפה של תכונה אחת על גבי אחרת רק בגלל הטווח ערכים שלה. לכן נרמלנו את המידע לפי נרמול Z כלומר לפי תוחלת וסטיית התקן.

1. בחירת פרמטרים עבור כל אחד מתתי המסווגים:

* בחירת פרמטר K למסווג KNN נעשתה בחלק הראשון של התרגיל וכפי שראינו עבור K=3 מתקבל דיוק מקסימלי.
* בחירת פרמטרים עבור ID3 ו Neural Network- נעשתה על ידי hyperparameters optimization. כלומר על ידי כל אחד מהמסווגים עם סטים של פרמטרים והערכתו על ידי k-fold-cross validation כאשר k=4.  
  עבור Neural Network נבדקו 750 סטים של פרמטרים ועבור ID3 נבדקו 100 סטים של פרמטרים.

1. 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 שהמסווג טועה אם ורק אם לפחות שני תתי מסווגים טועים. מכיוון שהם מסווגים את המידע שלנו בצורה שונה אחד מהשני ולכן כמעט בלתי תלויים אחד בשני הסיכוי לכך קטן מאשר ועדה של מסווגים מאותו סוג.