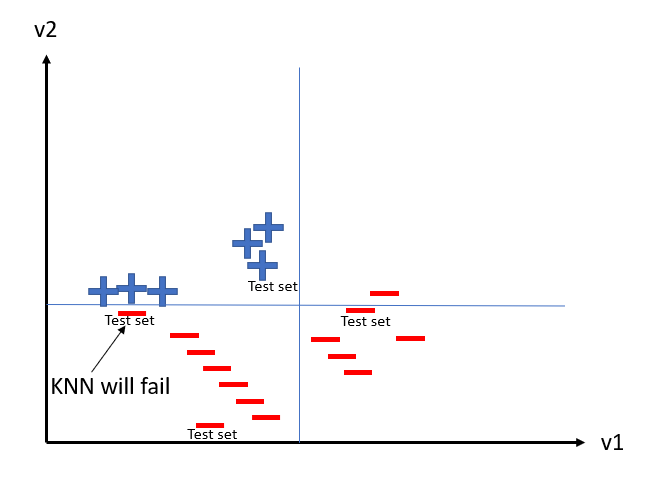
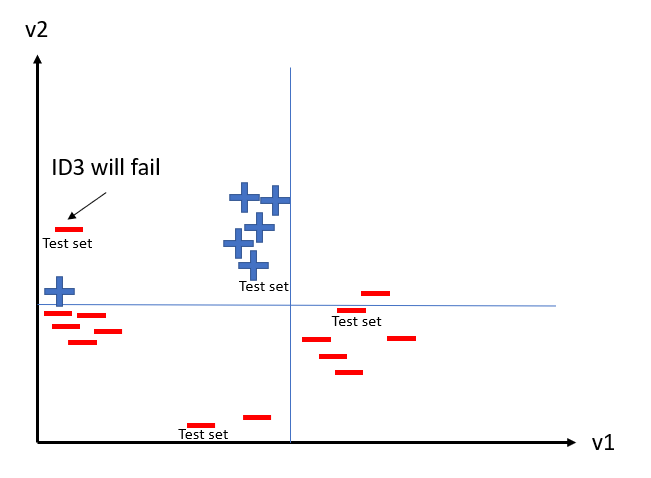
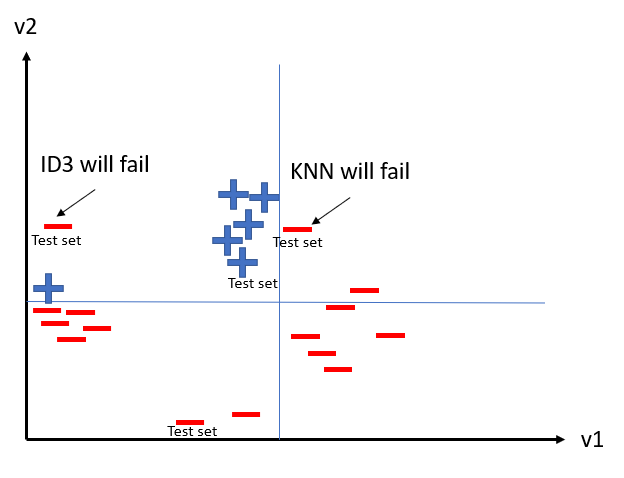
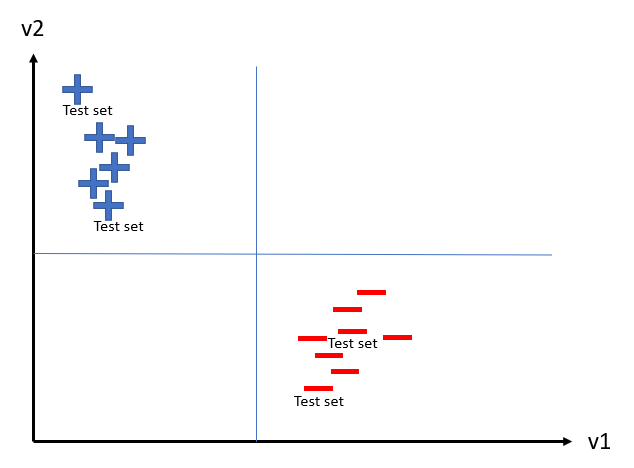
**חלק ב'**

1. **הוכחה:**
   1. בהינתן דאטה כלשהו עם תכונות רציפות ותיוגים בינאריים, הפעלה של פונקציית נירמול MINMAX, על הדאטה אינה משפיעה על דיוק של מסווג ID3, הנלמד על קבוצת האימון והנבחן על קבוצת המבחן וזאת מכיוון שאלגוריתם זה מחלק בכל צומת את כל האיברים לפי פונקציה סקלרית ולכן אם נפעיל את פונקציית הנירמול minmax גם על ערכי הפונקציה הנ"ל נקבל כי החלוקה למחלקות תישאר זהה ומכיוון ש – ID3 פועל עם הפונקציה הנ"ל נקבל עץ זהה לאחר הנירמול.
2. KNN מול ID3:
   1. ID3 מוגדרת ע"י הקווים הכחולים (קטן מv1 מסויים וגדול מv2 אחר, עבור נקבל כי KNN ייכשל בחץ המסומן - דגימה מקבוצה הבחינה תסווג כחיובית כחולה למרות שהיא שלילית ואדומה, אך מכיוון שערך v2 שלה קטן מההפרדה בid3 העץ יצליח לחזות אותה בתור דגימה אדומה.
   2. עבור K=3 במקרה הנ"ל חזאי KNN יחזה את כל הנקודות כפי שהן מכיוון שליד כל דגימת ניסוי יש לפחות 3 דגימות מאותו סוג שהכי קרובות אליה. נשים לב כי עבור סט האימון בבניית עץ ע"י ID3 נקבל אותו סיווג כמו בסעיף הקודם (הנק' האדומה השמאלית עליונה לא בסט אימון ולכן לא תשפיע על בניית העץ) אך עץ זה לא ינבא נכון בהפעלת החיזוי על הדגימה הנ"ל – ינבא כי הדגימה חיובית כחולה ולא שלילית אדומה.
   3. בקבוצה זו, כפי שהוסבר בסעיף 1.3 ID3 ייכשל בדגימה האדומה- שמאלית עליונה. ניתן לראות כי הוספנו עוד דגימת ניסוי אדומה שלילית בה KNN ייכשל עבור k=3 וזאת מכיוון שדגימה זו נמצאת בקרבת דבוקה של 3 דגימות חיוביות ורחוקה מדגימות אדומות שליליות ולכן תסווג ע"י KNN כדגימה כחולה חיובית למרות שהיא דגימה אדומה שלילית. לכן שני המסווגים ייטעו עבור קבוצת אימון וניסוי זאת.
   4. ניתן לראות בקבוצה זו כי עבור נקבל כי KNN הינו מסווג מטרה ויסווג נכון את כל דגימות הניסוי (מעל 8 נקבל כי כל דגימת ניסוי תקבל תוצאה אדומה מכיוון שזהו הרוב של הדגימות) וניתן לראות בקלות את סיווג העץ ע"י ID3 המאפשר הפרדה מלאה בין קבוצות הדגימות ולכן גם כאן תהיה דגימה נכונה לכל דגימות הניסוי.
3. שאלה 3:
   1. מכיוון שבקבוצת האימון יש 5 דגימות עם תווית '0' ו5 דגימות עם תווית '1' נקבל כי ערך הדיוק של מסווג הרוב הינו 50% מכיוון שאין רוב לאף תווית המסווג ייבחר רנדומלית בפילוג אחיד ומכיוון שיש 50% דגימות מאותו סיווג הוא יסווג ב50% סיווג נכון וב50% סיווג שגיאה.
   2. ע"י הרצת 2-fold validation כאשר נחלק את שני הקבוצות ל – 2 ע"פ החלוקה הנתונה יתקבל שקבוצה 1 : תחזה תמיד 1 (Majority classifier)  
      קבצוה 2: תחזה תמיד 0 (Majority classifier)  
      מכיוון שבקבוצה 1 80% מהדגימות בעלות תווית "1" אז כאשר נבדוק אותה לפי המסווג של קבוצה 2 נקבל דיוק של 20%  
      כנ"ל על קבוצה 2 (20% דיוק)  
      ולכן כאשר נעשה ממוצע על שניהם נקבל שערך הדיוק הוא 20%

**חלק ג'**

1. שאלה 4 – רטוב
2. שאלה 5 – רטוב
   1. הדיוק המתקבל הינו: 97.35% !!
3. שאלה 6 –
   1. גיזום עץ נועד למנוע תופעת Over fitting הנובעת מכך שבנינו את העץ ע"י קבוצת דגימות אימון ספציפית שאינה מכילה את כל הדגימות בעולם. אם לא נגזום את העץ נקבל מסווג מושלם עבור קבוצת האימון אך כנראה שהסתברות השגיאה עבור דגימה שאינה מקבוצת האימון תגדל. באופן כללי ע"פ עקרון Occam's Razor נרצה לבחור בפתרון הפשוט ביותר הנותן תוצאה טובה על קבוצת האימון אך לאו דווקא התוצאה המיטבית.
   2. רטוב
   3. רטוב + יבש – להשלים
   4. רטוב - להשלים

**חלק ד'**

1. אלגוריתם KNN בוחן את התווית הנפוצה ביותר מבין k הדגימות השכנות אליו במרחב של דגימות האימון ובוחר בתווית זו. **יתרונות האלגוריתם** הינם 1) זמן אימון מהיר ו2) אלגוריתם פשוט למימוש ופשוט להסבר - ניתן להסביר בקלות למה האלגוריתם בחר נק' מסויימת. **חסרונות האלגוריתם** הינם 1) זמן סיווג ארוך מאוד על חשבון זמן אימון קצר ו2) רגישות לתכונות לא רלוונטיות - תכונות שאינן משפיעות על סיווג של דוגמא מסויימת ישפיעו בכל מקרה על מיקומו במרחב ולכן על אופציה לשכנים אחרים מהסיווג ללא התכונה הנ"ל.
2. שאלה 8:
   1. קבוצת תת הקבוצות הינן קבוצת החזקה - קבוצה בגודל - ולכן זהו מס' תתי הקבוצות X.
   2. ישנן תתי קבוצות וזאת מכיוון שנסכום את כל האופציות לתתי קבוצות בגודל משתנה בין 1 ל D וכל תת קבוצה יכולה להכיל b מספרים מתוך הקבוצה הראשית S.