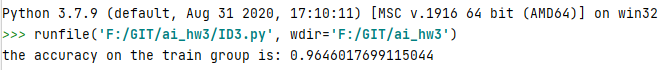
שאלה 1

1. -

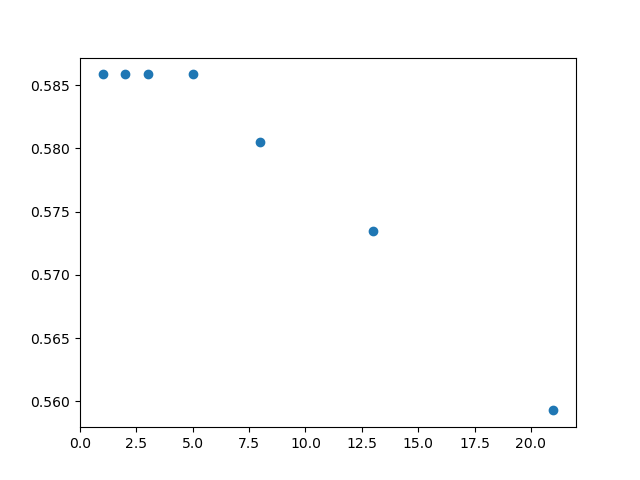


שאלה 2

שאלה 3 – גיזום מוקדם

1. חשיבות גיזום
2. –
3. בחרתי את הערכים: לפרמטר . הדיוק הממוצע עבור כל ערך הוא כדלהלן:

מצורף גרף המציג את השפעת הפרמטר על הדיוק:



ניתן לראות כי עבור , ערכי הדיוק הם זהים והגבוהים ביותר, ואילו עבור ערכים הירידה בדיוק חדה מאוד.

*לפיכך עבור מתקבל הגיזום הטוב ביותר, עם ערך .*

1. ***לימוד הכל***

*שאלה 4*

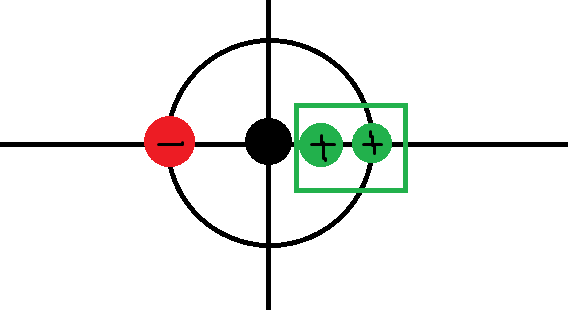
1. *על מנת לשפר את ערך ה- נפעל בצורה הבאה: כאשר מגיעים לכמות דוגמאות נמוכה (למשל 10), לא נבחר בתיוג בעל הכמות הגבוהה ביותר, אלא ניקח מרווח. כלומר, אם הכמות היא כחצי-חצי לטובת ,* ***נבחר ב-*** *מתוך הנחה שייתכן והאבחנה שגויה והובילה לטעויות, ועדיף להתחשב ב-.*

*בפועל, הדבר משפר גם את הדיוק.*

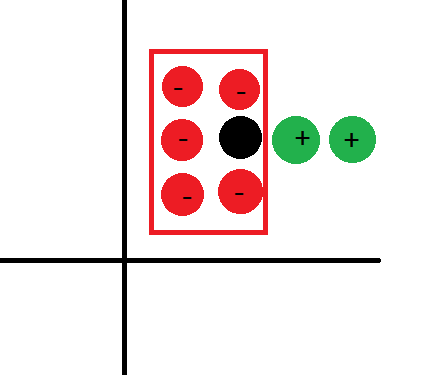
*שאלה 5*

*בכל הדוגמאות הבאות, ירוק ואדום הן דוגמאות האימון, כאשר ירוק זה חיובי ואדום שלילי, ושחור היא נקודת המבחן.* ***כמו כן, כל הגרפים מתארים סיווג שהוא מסווג המטרה.***

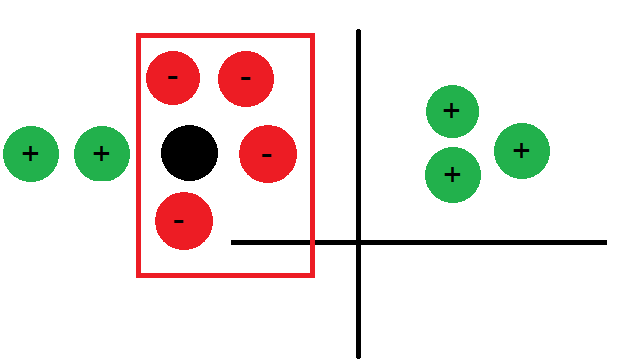
1. *( צודק, טועה) עבור ו- השחור יסווג כירוק, ועבור בגלל שיש שתי נקודות במרחקים זהים, השחור יסווג גם כירוק כי ערך שלו גדול יותר, לכן יטעה עבור כל .*

**

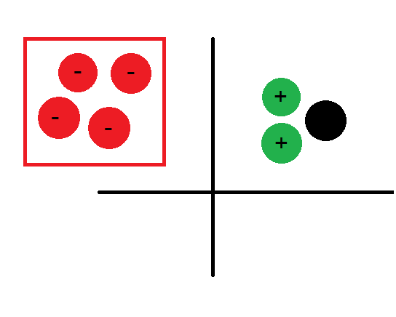
1. *( צודק, טועה) הסיווג של הנקודה השחורה הוא ירוק, אך יסווג אותה כאדומה. עבור ו-, יסווג אותה כירוקה.*

**

1. *(שניהם טועים) הסיווג הנכון של הנקודה השחורה הוא ירוק, אך מסווג אותה כאדומה. עבור לדוגמא, גם יסווג אותה כאדומה.*

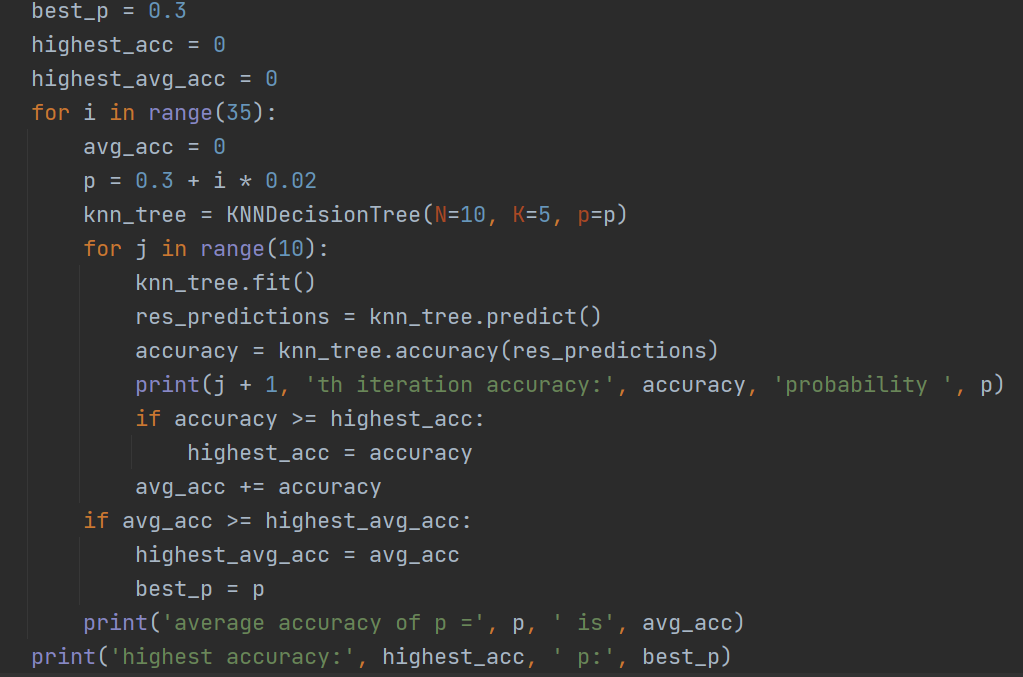
**

1. *(שניהם צודקים) עבור , יסווג את הנקודה השחורה כירוקה, ולכן יתקבל מסווג המטרה כמו , ושני האלגוריתמים צודקים עבור הנקודה השחורה.*

**

שאלה 6

בחרתי באופן שרירותי , ובדקתי לכל בקפיצות של , מתי אני מקבל את הדיוק הממוצע הגבוהה ביותר. לשם הבדיקה הרצתי את הקוד הבא:



טבלה מסכמת לערכי והדיוק הממוצע שהתקבל:

|  |  |
| --- | --- |
|  |  |
| 0.95929203539823 | 0.3 |
| 0.9504424778761061 | 0.32 |
| 0.9548672566371682 | 0.34 |
| 0.9548672566371682 | 0.36 |
| 0.9442477876106196 | 0.38 |
| **0.9610619469026549** | **0.4** |
| 0.9407079646017701 | 0.42 |
| 0.9530973451327434 | 0.44 |
| 0.9486725663716815 | 0.46 |
| 0.9486725663716815 | 0.48 |
| 0.9389380530973451 | 0.5 |
| 0.9557522123893805 | 0.52 |
| 0.9460176991150444 | 0.54 |
| 0.9504424778761061 | 0.56 |
| 0.943362831858407 | 0.58 |
| 0.9460176991150442 | 0.6 |
| 0.954867256637168 | 0.62 |
| 0.9424778761061947 | 0.64 |
| 0.9539823008849557 | 0.66 |
| 0.9486725663716815 | 0.68 |
| 0.9486725663716815 | 0.7 |

הדיוק הממוצע הגבוה ביותר התקבל עבור . אמנם היו ערכי אחרים עבורם התקבל דיוק נקודתי גבוה יותר, אך הממוצע עבור היה גבוה יותר, ולכן נבחר בו.

שאלה 7

להשתמש בRANDOM כדי להגריל אינדקסים של דוגמאות מתוך טווח. לוקחים את הדוגמאות מתוך הכלל, ובונה עץ. לדוגמא ID3. חוזר על התהליך N פעמים, כדי ליצור N עצים. כשבאים לסווג דוגמא, מסווגים לפי K העצים הכי קרובים.

ב-KNN רגיל לא בונים עץ החלטה, אלא מחלקים את הדוגמאות לקבוצות ולוקחים את הקבוצה הקרובה ביותר במרחק אוקלידי.