שאלה 1:

|  |  |  |  |
| --- | --- | --- | --- |
| תשובת המסווג: תמיד בריא | |  | |
| Negative | Positive |
| FN: 1 | TP: 0 | Positive:  P=1 | סיווג האמת |
| TN: 999 | FP: 0 | Negative: N=999 |

שאלה 2:

האלגוריתם ביצע רק שגיאה אחת שכן הוא סיווג רק אדם אחד חולה כבריא מתוך 1000 הנבדקים.

שאלה 3:

שאלה 4:

|  |  |  |  |
| --- | --- | --- | --- |
| תשובת המסווג: תמיד בריא | |  | |
| Negative | Positive |
| FNR (FN/P): | TPR (TP/P): | Positive:  P=1 | סיווג האמת |
| TNR (TN/N): | FPR (FP/N): | Negative: N=999 |

שאלה 5:

|  |  |  |  |
| --- | --- | --- | --- |
| תשובת המסווג: תמיד חולה | |  | |
| Negative | Positive |
| FN: 0 | TP: 1 | Positive:  P=1 | סיווג האמת |
| TN: 0 | FP: 999 | Negative: N=999 |

שאלה 6:

האלגוריתם ביצע כמעט 1000 שגיאות (999 שגיאות זה כמעט 1000😊), והוא סיווג 999 אנשים בריאים כחולים.

שאלה 7:

מידת הדיוק של המסווג היא:

שאלה 8:

|  |  |  |  |
| --- | --- | --- | --- |
| תשובת המסווג: תמיד חולה | |  | |
| Negative | Positive |
| FNR (FN/P): | TPR (TP/P): | Positive:  P=1 | סיווג האמת |
| TNR (TN/N): | FPR (FP/N): | Negative: N=999 |

שאלה 9:

|  |  |  |  |
| --- | --- | --- | --- |
| תשובת המסווג: הטלת מטבע הוגן | |  | |
| Negative | Positive |
| FN: 1 | TP: 1 | Positive:  P=2 | סיווג האמת |
| TN: 999 | FP: 999 | Negative: N=1998 |

שאלה 10:

האלגוריתם ביצע 1000 שגיאות בתוחלת, שכן לכל אדם (מתוך ה 2000) קיבל תוצאה שגויה בהסתברות של 50%. לפיכך, בתוחלת, על פני כל מרחב המדגם אנו מצפים שהמסווג יטעה כ 50% מהפעמים.

שאלה 11:

מידת הדיוק של המסווג היא:

שאלה 12:

|  |  |  |  |
| --- | --- | --- | --- |
| תשובת המסווג: הטלת מטבע הוגן | |  | |
| Negative | Positive |
| FNR (FN/P): | TPR (TP/P): | Positive:  P=2 | סיווג האמת |
| TNR (TN/N): | FPR (FP/N): | Negative: N=1998 |

שאלה 13:

עקומת ה ROC תהיה קו ישר לינארי, שכן לשני המסווגים הסתברות שווה לתת תשובה נכונה או לא נכונה, ולכן ה TPR וה FPR שלהם יהיה זהה עבור כל נקודה שנחליט לסווג ממנה.

שאלה 14:

מכיוון שיש הבחנה ברורה מבחינת המאפיינים בין הנתונים, ואין חפיפה כלל בין העקומות שלהם, אז העקומה תהיה מעין קו אנכי העולה ל -1 שהופך לפונק' קבועה (קו אופקי) במקביל לציר ה X (גרף שנראה כמו שיקוף של ר). זאת משום מיקום קו הבחירה בין הגרפים נותן לנו מסווג שנותן רק TPR ואף פעם לא FPR, ומיקום קו בחירה זה הוא מאוד ברור (בין העקומות). כל הזזה של קו הבחירה כך שהוא יחתוך את העקומות רק יוסיף לנו סיווג שהוא FPR, ולכן הגרף הופך לקבוע – רק גורעים מדיוק התוצאה ע"י הוספת FPR לתוצאות.

שאלה 15:

עקומות שנמצאות מתחת לאלכסון מעידים שהמאפיינים מצליחים לזהות בבירור תוצאות שהן FPR, ולכן כל שיש לעשות הוא להפוך את הבחירה (את מה שהמאפיין מזהה חיובי כשלילי – נקבע את ההפך).

שאלה 16:

ה recall של הספרה 8 הוא 1.00 (כלומר 100%). המשמעות היא שמספר הפעמים שספרות 8 תויגו לא נכון היא 0. במילים אחרות, למאפיין היה TP של 100 אחוז בזיהוי הספרות 8 ו FN של 0 אחוז – לא ניתן סיווג שונה מ 8 לספרה שהיא 8. ניתן לראות זאת בשורה ה-9 (ששייכת לספרה 8) שבה נראה כי אף ספרה 8 לא תויגה כספרה אחרת.

שאלה 17:

ה - precision של הספרה 0 היא 1.00 (כלומר 100%). המשמעות היא שכמות הספרות שסווגו כ 0 במקום הספרה הנכונה היא 0%, במילים אחרות למאפיין היה 0 מקרים של FP ולכן קיבל ציון של 100%. ניתן לראות זאת בעמודה הראשונה – שבה אף ספרה אחרת לא סווגה כ 0 בטעות.

שאלה 18:

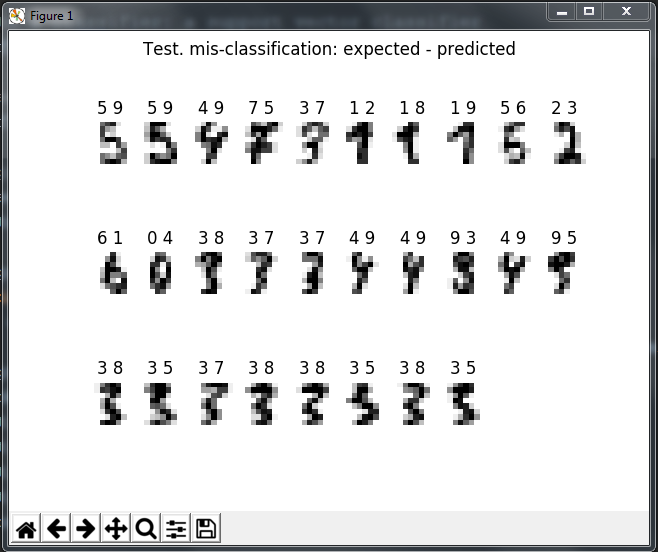
7 פעמים הספרה 9 סווגה בטעות כספרות אחרות, המייצגות 2% ירידה בערך ה recall. הספרות הן 1,5, ו-4.

שאלה 19:

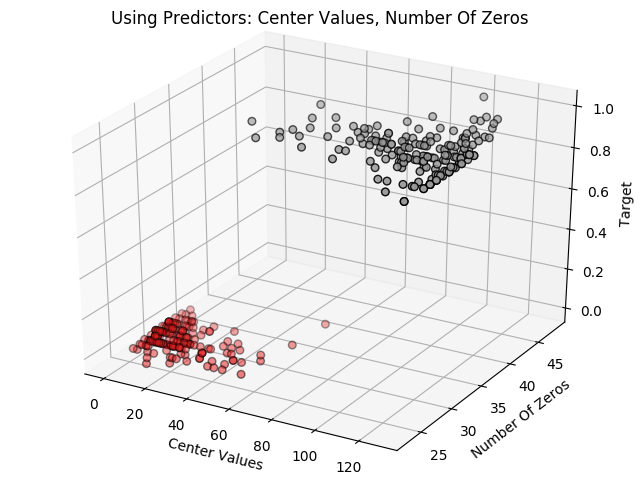
הספרה בעלת ה recall הנמוך ביותר – 87% - היא עבור הספרה 3, כאשר 12 פעמים היא סווגה לא נכון. 5 פעמים סווגה כ – 8, 4 פעמים סווגה כ – 7 ו – 3 פעמים סווגה כ 5.

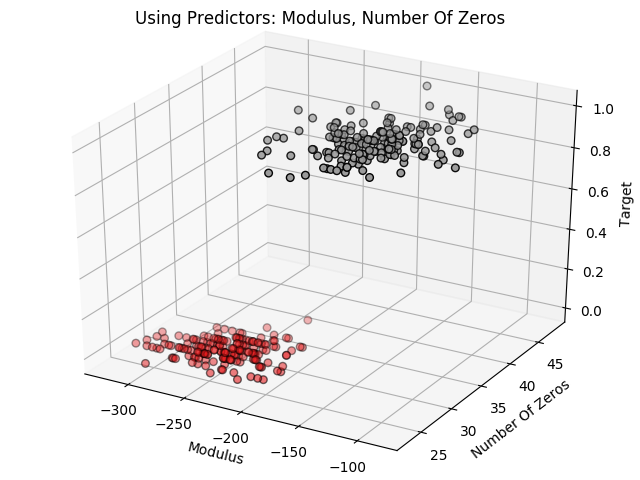
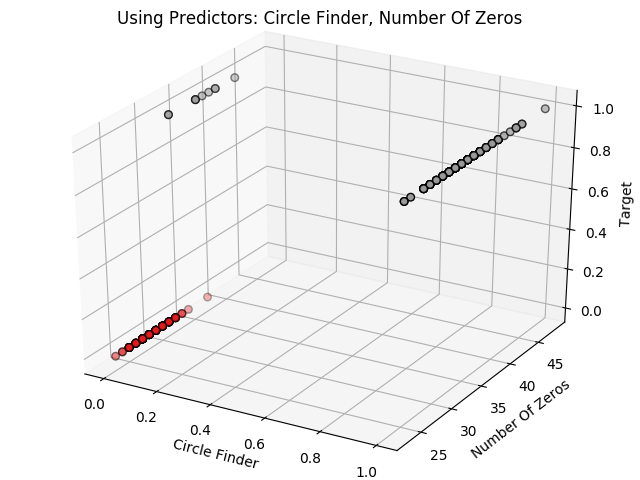
שאלה 20:

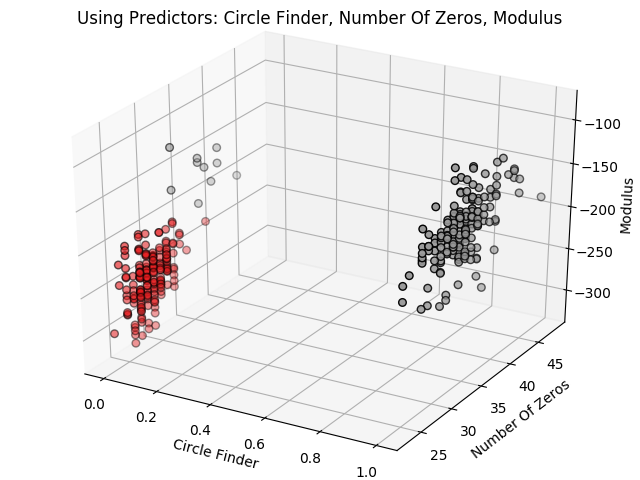
מצורף קוד בקובץ Pyhton. זו התוצאה:

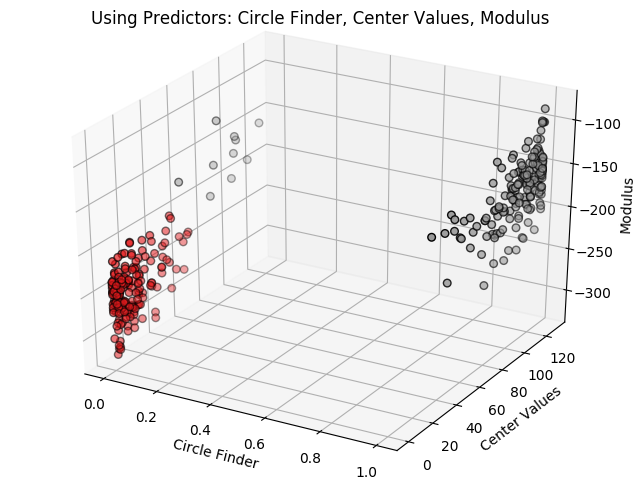


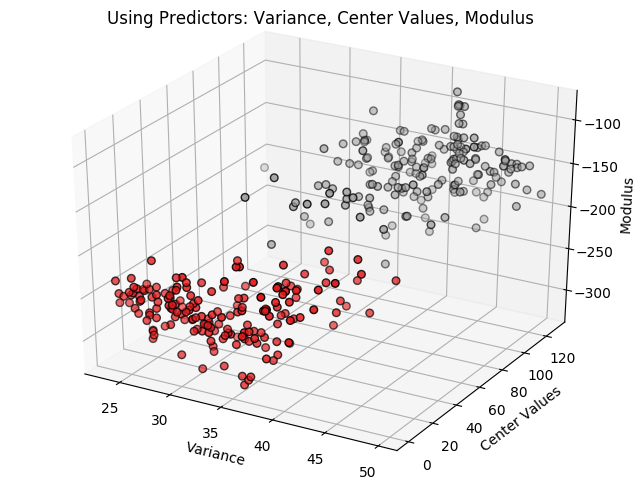
שאלה 21:

ה. שלושה תרשימים עם שני מאפיינים ושלושה תרשימים עם שלושה מאפיינים. אנחנו הרצנו על הרבה יותר וריאציות אך כמה מייצגים.

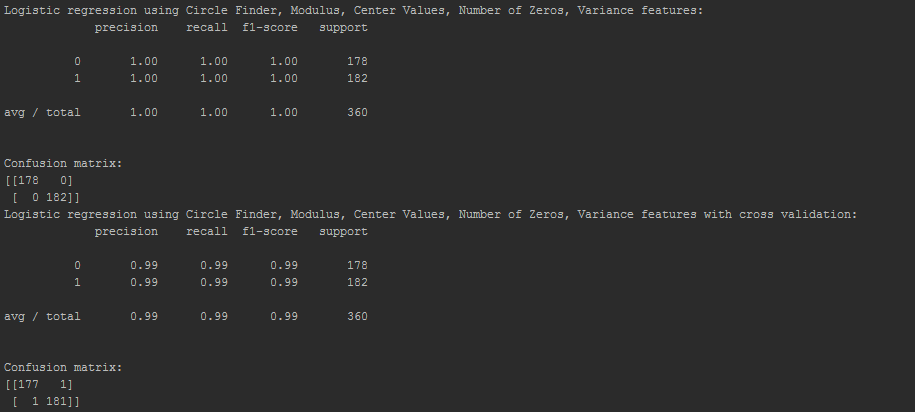
 







ו.



ז. בסופו של דבר, בהסתמך על התרשימים מסעיף ה וסעיף ו, החלטנו על שימוש בכל המאפיינים:

