**מערכות לומדות – תרגיל מס' 2**

**שם: אליהו צורי**

**ת"ז: 201610672**

1.

|  |  |
| --- | --- |
| תשובת המסווג "תמיד בריא" | |
| Negative | Positive |  | |
| 1 | 0 | Positive  P=1 | סיווג האמת |
| 999 | 0 | Negative  N =999 |  |

2.

מס' השגיאות הוא 1, מאחר שההסתברות האפריורית היא 1 ל-1000 והמסווג אמר "תמיד בריא"

3.

נחשב את הדיוק המסווג בעזרת הנוסחה:

TP = 0

TN = 999

P = 1

N = 999

======== > ACC = (0+999) / (1+999) = 0.999

על כן הדיוק הוא: 99.9 אחוז עפ"י הנוסחה.

4.

|  |  |
| --- | --- |
| תשובת המסווג "תמיד בריא" | |
| Negative | Positive |  | |
| False negative rate = 1 | True positive rate = 0 | Positive  P=1 | סיווג האמת |
| True negative rate = 1 | False positive rate = 0 | Negative  N =999 |  |

5.

|  |  |
| --- | --- |
| תשובת המסווג "תמיד בריא" | |
| Negative | Positive |  | |
| 0 | 1 | Positive  P=1 | סיווג האמת |
| 0 | 999 | Negative  N =999 |  |

6.

מס' השגיאות הוא 999 וזאת מאחר וההסתברות האפריורית היא 1 ל-1000 והמסווג אמר "תמיד חולה" ולכן התשובה היא שמס' השגיאות הוא 999.

7.

הדיוק הוא 0.001 אחוז

8.

|  |  |  |  |
| --- | --- | --- | --- |
| תשובת המסווג "תמיד חולה" | |  | |
| Negative | Positive |
| False negative rate = 0 | True positive rate = 1 | Positive  P= 1 | סיווג האמת |
| True negative rate = 0 | False positive rate = 1 | Negative  N =999 |

9.

|  |  |  |  |
| --- | --- | --- | --- |
| תשובת המסווג "בהסתברות זהה" | |  | |
| Negative | Positive |
| 1 | 1 | Positive  P=2 | סיווג האמת |
| 999 | 999 | Negative  N =1998 |

10.

מאחר וההסתברות האפריורית להיות חולה היא 1 ל-1000 ומאחר והמסווג ענה על מחצית מהפעמים כי הוא חולה, אז מספר השגיאות שם הוא 999 ולכן סה"כ קיבלנו 1000.

11.

נחשב על פי הנוסחה ונקבל

ACC = 1000/2000 = 0.5

כלומר הדיוק הוא 50 אחוז

12.

|  |  |  |  |
| --- | --- | --- | --- |
| תשובת המסווג "בהסתברות זהה" | |  | |
| Negative | Positive |
| False negative rate =0.5 | True positive rate = *0.5* | Positive  P= 2 | סיווג האמת |
| True negative rate = | False positive rate = | Negative  N =1998 |

המסווג עונה על בהסתברות זהה חולה\בריא, לכן ענה על 1000 פעמים חולה וכנ"ל בריא. כמו כן ההסתברות האפריורית להיות חולה הינה 1 לאלף, ז"א שבין הנבדקים 2 חולים. קיבלנו שהמסווג ענה 1000 פעמים "חולה", צדק 1, טעה 999. כמו"כ ענה 1000 "בריא", צדק 999, טעה 1. סה"כ קיבלנו יחסים סימטריים כצפוי.

13.

כאשר נתוני המחלקות זהים, עקומת ה ROC תקביל לאלכסון הראשי.

14.

כאשר אין כלל חפיפה בין נתוני המחלקות, העקומה תהיה על ציר Y ומקבילה ל X בגובה 1.

הסבר, ע"ס ההסבר למעלה: במקרה כזה ניתן להפריד באופן מוחלט בין המחלקות, ע"כ התוצאות של בחירת מסווג כלשהו יכולה לתת אחת מן השלוש:

(0,1) – כאשר נמקם את המסווג בדיוק באמצע, נקבל הפרדה מושלמת בין המחלקות מתאים ל 0 עלות מול 1 תועלת.

((0,0<Y<1 – כאשר נמקם את המסווג ימינה לנקודת ההפרדה ובעקבות כך חלק מהחיוביים יתויגו שלילי, ז"א יעילות לא מושלמת. לעומת זאת כאמור קיימת הפרדה בין המחלקות, ולכן תזוזה ימינה לא תשפיע על תיוגי שליליים כחיובי, ז"א לא תשפיע על ה FPR, ואכן ה FPR קבוע 0.

(0<x<1,1) – באופן סימטרי לתזוזה ימינה רק שהפעם נתון היעילות יישאר קבוע שכן אנו זזים שמאלה. עלות לעומת זאת תעלה מעל 0 ככל שנזוז יותר שמאלה, וכך נתייג יותר שליליים כחיובי.

קיבלנו כי כלל הערכים האפשריים עבור נקודות משלוש הצורות הנ"ל הם אכן הערכים על Y ומקבילים לX בגובה 1.

15.

16.

מדד רגישות או יעילות של המסווג זהו בעצם ה TPR. עבור מחלקה X, ה recall של X יהא: כמות תיוגי ה X הנכונים חלקי כמות ה X בנתונים. במקרה של 8: כמות תיוגי ה 8 הנכונים = 88. כמות ה 8 בנתונים = 88. ולכן = 1.

17.

זהו מדד הדיוק של המסווג. עבור מחלקה X, ה Precision של X יהא: כמות תיוגי ה X הנכונים חלקי כמות תיוגי ה X הכללי. במקרה של 8: כמות תיוגי ה 8 הנכונים = 88. כמות תיוגי ה 8 הכללי = 94.

18.

ע"פ המטריצה. 9 תויג פעמיים שגוי. פעם אחת 3. פעם אחת 5.

19.

ע"פ ההסבר מעלה, הספרה המבוקשת היא הספרה בעלת ה recall הנמוך ביותר. זוהי הספרה 3. היא תויגה 12 פעמים לא נכון. 3 פעמים 5. 4 פעמים 7. 5 פעמים 8.