**AI-HW3**

שאלה 1

1. בכדי לקבל מדד יחיד נבצע ממוצע על כל ההרצות, כלומר נסכום את הדיוק של כל ההרצות ונחלק ב-4. הדיוק שקיבלנו הינו: **0.709263714303** .
2. בדומה לסעיף א' גם כן נבצע ממוצע, כלומר נסכום למטרציה אחת את כל המטריצות המתקבלות מההרצות ונחלק כל תא ב-4. המטרציה שהתקבלה:

|  |  |  |
| --- | --- | --- |
|  | + | - |
| + | 59.75 | 10.75 |
| - | 20 | 15.25 |

שאלה 2

1. נגדיר את הבעיה כך:

True Positive – מעבד תקין שהאלגוריתם אומר שהוא תקין.

True Negative – מעבד מקולקל שהאלגוריתם אומר שהוא מקולקל.

False Positive – מעבד תקין שהאלגוריתם אומר שהוא מקולקל.

False Negative – מעבד מקולקל שהאלגוריתם אומר שהוא תקין.

1. נבדוק עבור כל מסווג את גודל ההפסד (סכום הכסף אותו החברה מפסידה) כתוצאה משימוש במסווג זה.

את ההפסד נחשב כסכום של מספר המעבדים המקולקלים שהמסווג סיווג כתקינים כפול מחיר הוולדציה (כלומר מחיר הוולדציות שהרצנו על מעבדים שאינם תקינים כפול המחיר של וולדציה) ומספר המעבדים התקינים שהמסווג סיווג כלא תקינים אבל היו תקינים כפול המחיר של מעבד פחות מחיר הוולדציה.

הסבר לחישוב:

על כל מעבד מקולקל שהחברה מעבירה וולדציה היא מפסידה את מחיר הוולדציה (זה סכום כסף שהיא לא תרוויח בחזרה ממכירת המעבד בשונה ממעבד תקין). בנוסף, על כל מעבד תקין שנזרק הפסידה החברה את מחיר המעבד שאינו נמכר אבל היא חסכה את הוולדציה שלא ביצעה עליו.

נשים לב שחישוב זה שקול לחישוב הרווח המקסימלי שניתן

לפי החישוב הזה נקבל שעל כל אלפיים יחידות מסווג A יגרום להפסד של מיליון ו400 אלף שקלים.

מסווג B יגרום להפסד של 920 אלף שקלים ומסווג C יגרום להפסד של 990 אלף שקלים.

**לכן, המסווג העדיף הינו מסווג B.**

שאלה 3

1. בכדי לגרום להתאמת יתר ביצענו מיון של המידע כך שהמידע שהמסווג קיבל לצורך אימון היה יחסית דומה לעצמו בעוד המידע עליו בוצע המבחן היה שונה ברובו.
2. בכדי לגרום לתת התאמה גרמנו למסווג להתייחס רק לתכונה הראשונה ולכן המסווג אינו מצליח ללמוד את מושג המטרה.
3. התוצאה שקיבלנו במקרה של התאמת יתר:
   1. עבור סט האימון: 0.911949685535
   2. עבור סט המבחן: 0.409523809524

הסבר: המידע עבר מיון ואז חולק ל75% הראשונים ו-25% הנותרים. חלוקה זו גרמה לכך שהשונות של הערכים בסט האימון נמוכה ובסט המבחן היו תכונות בעלות ערכים שהמסווג לא ראה באימון (למשל, עבור התכונה f1\_0 כל הערכים בסט האימון היו 0 אך בסט המבחן היו שורות עם הערך 1 בתכונה זו). לכן, המסווג הצליח ברמה גבוהה לסווג את סט האימון אך נכשל לסווג את סט המבחן משום שהיו ערכים שהוא לא ידע לסווג ולא נתקל בהם לפני כן.

התוצאה שקיבלנו במקרה של תת התאמה:

1. עבור סט האימון: 0.685534591195
2. עבור סט המבחן: 0.609523809524

הסבר: במקרה זה למסווג חסרים נתונים, על פי תכונה אחת בלבד הוא לא מסוגל ללמוד את הסיווג, נשים לב שהיחס בין סיווג True ל-False הוא 1:2 ולכן נקבל המסווג נותן תוצאות קרובות לכך (כ66% דיוק).

שאלה 4

שאלה 5

יש לבדוק את הביצועים של כל מסווג (המסווגים שונים זה מזה בתת הקבוצה של התכונות הנבחרות במצב הנוכחי) על קבוצת ולידציה ולא על קבוצת המבחן מכיוון שתמיד נרצה לבחון את המסווג ה"סופי"\"נבחר" על קבוצת אובייקטים שמעולם לא למד\ נבחן עליהם כדי שנוכל לקבל תמונת מצב קרובה ככל האפשר על ההתנהגות של המסווג על אובייקטים מהעולם החיצוני.

שאלה 6

מומש

שאלה 7

1. 0.7075471698113207 KNN=5 , no feature choosing , precision =
2. [0, 4, 28, 30, 1, 2, 8, 26, 29]features\_subset =

knn = 5 , b = 8 precision is 0.8018867924528302