**מגישים:**

מיכאל תורג'מן – 307910984

בני לודמן - 307270215

**חלק 1**

1. א. עשינו ממוצע

ב. גם עשינו ממוצע

1. א. True Negative – מעבד מקולקל שסווג כמקולקל

True Positive – מעבד תקין שסווג כתקין

False Negative – מעבד תקין שסווג כמקולקל

False Positive – מעבד מקולקל שסווג כתקין

ב. נתסכל על ההפסד הממוצע\* של כל מסווג:

מסווג A:

מסווג B:

מסווג C:

לכן נבחר במסווג B.

\*הפסד ממוצע – כמה כסף החברה הוציאה על בדיקות של מעבדים מקולקלים שהמסווג אמר שהם תקינים (False Positive) ועוד הכסף שהחברה הוציאה על ייצור של מעבד תקין שלא יצא לשוק מכיוון שהוא הוגדר כמקולקל ע"י המסווג (False Negative).

1. א. לא שינינו כלום. האלגוריתם הרגיל מספיק כדי לקבל overfitting, כי מקבלים את העץ הכי גדול שאנחנו יכולים לקבל – לא מתבצע גיזום, והעץ מפותח לגודל מקסימלי.

ב. שינינו את הפרמטר max\_depth להיות 1, כי כפי שראינו בהרצאה ככל שהעץ קטן יותר כך מקבלים underfitting. בנוסף, שינינו את האלגוריתם כך שיתייחס רק למאפיין אחד, כדי להוריד כמה שיותר את ההתאמה לקבוצת האימון.

ג. עבור התאמת יתר קיבלנו:

סט האימון – 0.91

סט המבחן – 0.71

עבור התאמת חסר קיבלנו:

סט האימון – 0.66

סט המבחן – 0.72

התוצאות להתאמת יתר מתאימות למצופה משום שעבור סט האימון אנו מקבלים אחוזים גבוהים (עדיין לא 100% משום שה-dataset שקיבלנו אינו עקבי), ולעומת זאת עבור סט המבחן אנו מקבלים אחוזים נמוכים יותר – כלומר יכולת ההכללה שלו נמוכה.

התוצאות להתאמת חסר לא כל כך מתאימות לציפיות שלנו – היינו מצפים לקבל בערך 50% דיוק גם בסט האימון וגם בסט המבחן. אנו מערכים שהתוצאה שקיבלנו נובעת מכך שאין איזון בין דוגמאות חיוביות ושליליות ב-dataset. בכל מקרה אנו מבחינים כי עבור סט המבחן אנו מקבלים דיוק טוב יותר מאשר מסט האימון, כצפוי למקרה של התאמת חסר.

חלק 2

1. 1. מספרי תתי הקבוצות של הקבוצה S הוא הגודל של קבוצת החזקה שלה - .

2. מספר תתי הקבוצות של הקבוצה S בגודל b הוא: (מספר האפשרויות לבחור b מאפיינים מתוך קבוצה של N, ללא חשיבות לסדר ביניהם).

1. את הביצועים נמדוד על קבוצת **הוולידציה** – במקרה שלנו אנחנו מבצעים כוונון פרמטרים, שנחשב כחלק מתהליך הלמידה, לכן אנחנו משתמשים בקבוצת האימון לצורך בחירת הפרמטרים (ולצורך המדידה משתמשים בקבוצת הוולידציה שהיא מתוך קבוצת האימון).
2. מימוש בקוד.
3. 1. הדיוק שהתקבל הוא 0.698.

2. הדיוק שהתקבל הוא 0.65.