דוח הפרויקט

שאלה 1

קבוצת התנאים שהשתמשנו בהם בגרסה השנייה של המימוש כוללים את התנאים שהיו לנו בגרסה הראשונה, ובנוסף לכך הוספנו 100 תנאים נוספים, כאשר כל אחד הוא מהצורה הזאת:
האם התמונה הנתונה היא במקום ה- i מבחינת מרחק מהתמונה הממוצעת של הספרה j?
כאשר $0 \le j \le 9$ $1 \le i \le 10$

- נגדיר מרחק בין תמונה לתמונה כמרחק בין שתי נקודות במרחב האוקלידי כאשר כל תמונה מיוצגת כנקודה במרחב $(x1,\ldots,xn)$ ובמקרה שלנו :
n=784, וכל קורדינטה מייצגת את גודל הפיקסל בתמונה.

- נגדיר תמונה ממוצעת של ספרה k כתמונה שבה בכל פיקסל i,j הנע בין :
- $0 \le i,j \le 27$, ניקח את גודל הפיקסל הממוצע מכל התמונות של הספרה k במדגם האימון.

- לדוגמא, עבור תמונה נתונה אלו המרחקים מכל תמונה ממוצעת:

מרחק מהתמונה הנתונה לתמונה הממוצעת	450	10	50	70	90	11	100	170	300	400
הספרה של התמונה הממוצעת	0	1	2	3	4	5	6	7	8	9

עבור השאלה:
האם התמונה הנתונה היא במקום ה2 מבחינת מרחק מהתמונה הממוצעת של הספרה 5?
נקבל שהתשובה היא כן.
האם התמונה הנתונה היא במקום ה1 מבחינת מרחק מהתמונה הממוצעת של הספרה 5?
נקבל שהתשובה היא לא.

- בחרנו תנאים אלו מכיוון שהבנו שכל תמונה ניתן לייצג כנקודה במרחב, ומהבנה שתמונות בעלות אותה
הספרה יהיו נקודות קרובות יותר במרחב מאשר תמונות בעלות ספרה שונה.
לכן אם ניקח את הנקודה הממוצעת של כל ספרה אז המרחק מכל נקודה ממוצעת כזאת לתמונה הנתונה יהווה פקטור משמעותי בהחלטה איזו ספרה מתאימה לתמונה הנתונה.

- תנאים אלו מומשו בעזרת כך שיצרנו תמונות ממוצעות עבר כל ספרה בעזרת התמונות ממדגם האימון, וכך בהינתן תמונה ניתן יהיה לחשב מרחק לתמונה הממוצעת של כל ספרה.

- בהתחלה התייחסנו רק לתנאים שבודקים האם התמונה הנתונה הכי קרובה לתמונה הממוצעת, של הספרה ,i ואז הבנו שיש גם משמעות ל-עד כמה היא רחוקה מהתמונה הממוצעת והבנו שיש להתייחס גם למיקום מבחינת קירוב לכל אחת מהתמונות הממוצעות, ולכן הוספנו עוד תנאים.
למשל אם תמונה מסוימת הכי רחוקה מתמונה ממוצעת מסוימת אז יש לכך משמעות, וזה יכול להועיל בהחלטה איזו ספרה להתאים לתמונה.

שאלה 2

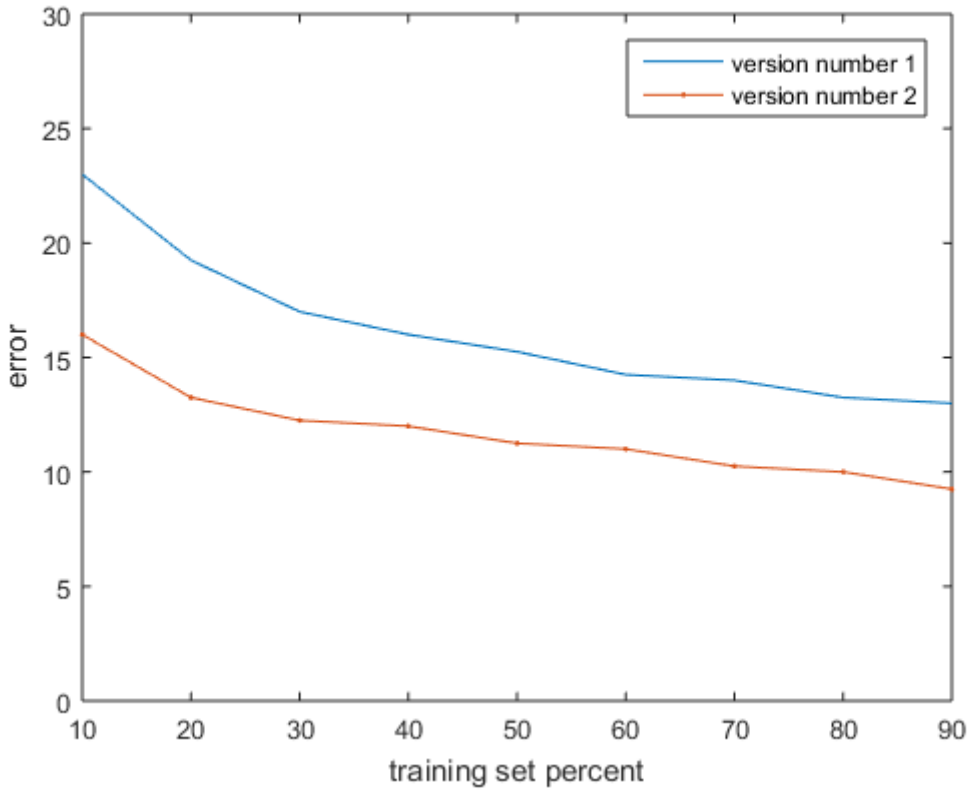
מבני הנתונים שהשתמשנו בהם:

- ראשית מימשנו עץ בינארי אשר ייצג את עץ ההחלטה.
- מימשנו node אשר יהווה קדקוד בעץ ההחלטה.
- על מנת לייצג תמונה ייצרנו את האובייקט trainingimage אשר ישמור את כל הפיקסלים של התמונה ואת הלייבל המיוחס לו.
- בגרסה השנייה על מנת לממש את התנאים היא צורך בלחשב את התמונה הממוצעת של כל ספרה ולכן יצרנו את האובייקט trainingimagegroup המייצג את כל התמונות של ספרה מסוימת, כלומר הוא מחזיק את כל ה trainingimages של ספרה מסוימת.
- על מנת להחליט בכל שלב איזה עלה לפתח מימשנו אובייקט מסוג leaflist המחזיק את כל העלים בעץ הנוכחי.

האתגרים שנתקלנו בהם היו בעיקר בבחירת התנאים שהניבו לנו את השגיאה הנמוכה ביותר בגרסה השנייה, והתמודדנו איתם באמצעות הבנה של הבעיה וגם בעזרת ניסוי וטעיה.

שאלה 3

א.



גודל מדגם	10	20	30	40	50	60	70	80	90
טווח	0	1	0	0	1	1	0	1	0

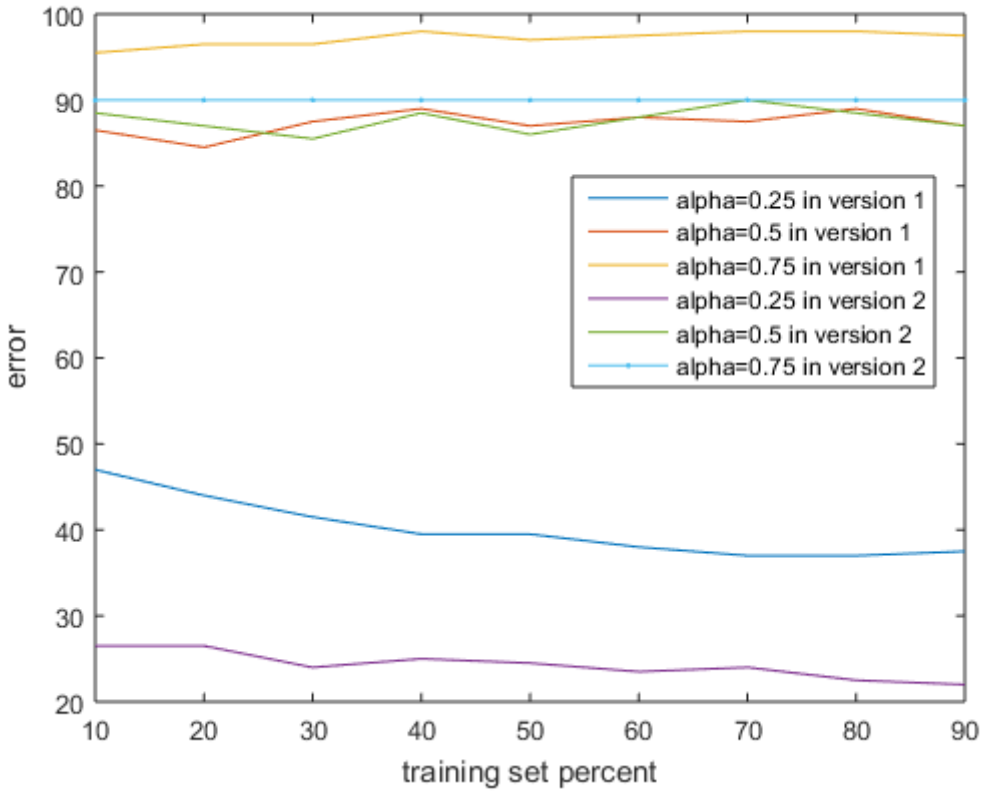
הרצנו כל גודל מדגם 3 פעמים.

ב. ניתן לראות שבגרסה השנייה השגיאה לכל גודל מדגם האימון היא קטנה יותר מהשגיאה בגרסה הראשונה.
מבחינת תלות השגיאה בגודל מדגם האימון ניתן לראות שככל שגודל מדגם האימון יותר גדול אז השגיאה יותר קטנה.

ג. קיבלנו שבגרסה השנייה השגיאה לכל גודל מדגם האימון היא קטנה יותר מהשגיאה בגרסה הראשונה מכיוון שבגרסה השנייה ישנם יותר תנאים מאשר בגרסה הראשונה.
קיבלנו שככל שגודל מדגם האימון יותר גדול אז השגיאה יותר קטנה זאת מכיוון שאלגוריתם הלמידה מקבל יותר מידע ולכן יכול לייצר עץ החלטה יותר מדויק מבחינת התאמה של תמונה נתונה לתווית המתאימה לו.

שאלה 4

א.



Version number 1:

Alpha=0.25

גודל מדגם	10	20	30	40	50	60	70	80	90
טווח	2	2	1	1	1	0	0	2	1

Alpha=0.5

גודל מדגם	10	20	30	40	50	60	70	80	90
טווח	7	5	3	0	2	0	3	0	4

Alpha=0.75

גודל מדגם	10	20	30	40	50	60	70	80	90
טווח	1	1	1	0	0	1	0	0	1

Version number 2:

Alpha=0.25

גודל מדגם	10	20	30	40	50	60	70	80	90
טווח	1	1	0	2	1	1	0	1	0

Alpha=0.5

גודל מדגם	10	20	30	40	50	60	70	80	90
טווח	1	2	1	3	2	2	0	3	4

Alpha=0.75

גודל מדגם	10	20	30	40	50	60	70	80	90
טווח	0	0	0	0	0	0	0	0	0

- הרצנו כל גודל מדגם 3 פעמים

ב.

- מבחינת תלות בגודל הרעש ראשית ניתן לראות שככל הגודל הרעש גדל כך גודל השגיאה גדל משמעותית בשני הגרסאות.
- מבחינת ההבדלים בין שני הגרסאות ניתן לראות שכל גודל רעש גרם לשגיאה יותר גדולה בגרסה הראשונה גם מבחינת גודל השגיאה וגם מבחינת הטווח מהשגיאה שהתקבלה בלי הפרעה(בשאלה 3).

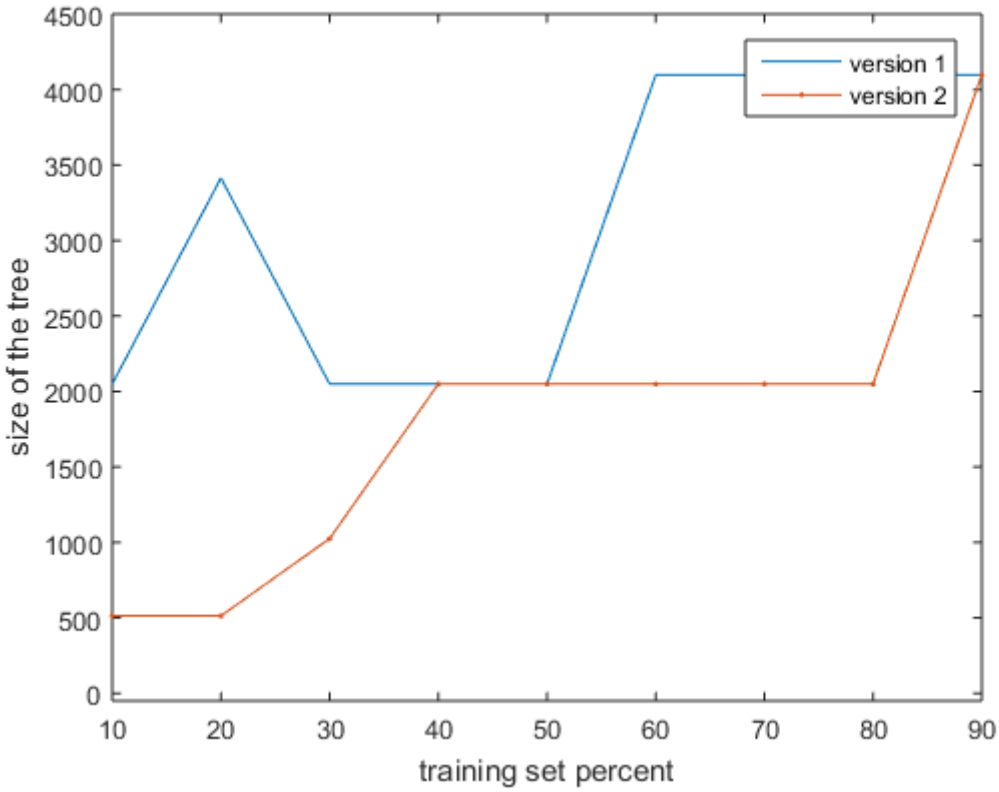
- מבחינת תלות בגודל המדגם ניתן לראות שכאשר גודל הרעש הוא קטן יותר אז ככל שגודל המדגם גדול יותר מקבלים שגיאה יותר קטנה כלומר מקבלים <u>מגמה יורדת.</u>
וכאשר גודל הרעש הוא גדול יותר אז ככל שגודל המדגם גדול יותר מקבלים שגיאה יותר גדולה כלומר מקבלים <u>מגמה עולה</u>.
בעצם מה שמקבלים זה שככל שגודל הרעש יותר גדול אז כאשר אלגוריתם הלמידה מקבל יותר מידע(גודל מדגם גדול יותר) אז הוא מייצר עץ החלטה פחות טוב כלומר בעל שגיאה גדולה יותר.

ג.

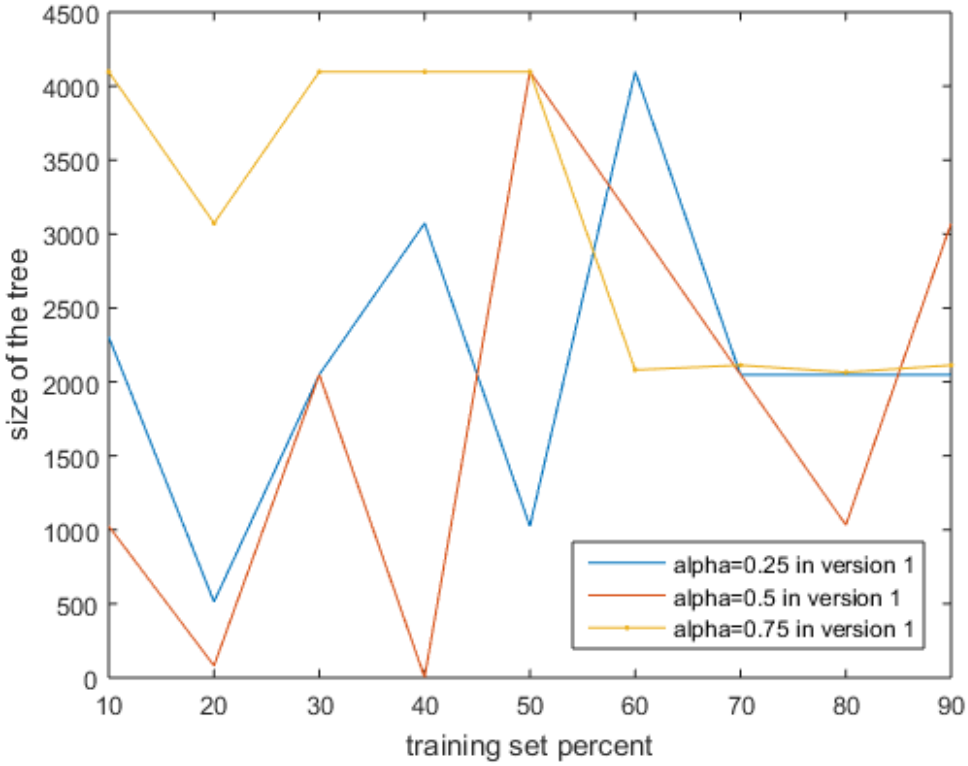
- ראינו שככל הגודל הרעש גדל כך גודל השגיאה גדל משמעותית בשני הגרסאות דבר שהוא מאוד הגיוני ואנחנו נצפה לו מכיוון שככל המעוותים יותר את המידע אז כמובן עץ ההחלטה שנייצר באמצעות מידע מעוות זה יחזה בצורה שגויה יותר את התווית של תמונה נתונה שהוא יקבל.
- קיבלנו שבהבדלים בין שני הגרסאות ניתן לראות שכל גודל רעש גרם לשגיאה יותר גדולה בגרסה הראשונה גם מבחינת גודל השגיאה וגם מבחינת הטווח מהשגיאה שהתקבלה בלי הפרעה(בשאלה 3)
דבר שנצפה לו מכיוון שבגרסה יש לנו יותר תנאים ואלגוריתם הלמידה הוא יותר טוב ,הוא כמובן עדיין יושפע מאוד מכל גודל רעש שהוא יקבל אך הוא פחות רגיש לרעש.
- ראינו שמה שמקבלים זה שככל שגודל הרעש יותר גדול אז כאשר אלגוריתם הלמידה מקבל יותר מידע(גודל מדגם גדול יותר) אז הוא מייצר עץ החלטה פחות טוב כלומר בעל שגיאה גדולה יותר דבר שנצפה לו מכיוון שכאשר אלגוריתם הלמידה מקבל מידע שהוא שגוי יותר אז ככל שהוא יקבל יותר מידע שגוי כזה הוא ייצר עץ החלטה התואם יותר למידע השגוי וכך יטעה יותר כאשר יקבל תמונה וירצה לחזות את התווית שלה.

שאלה 5

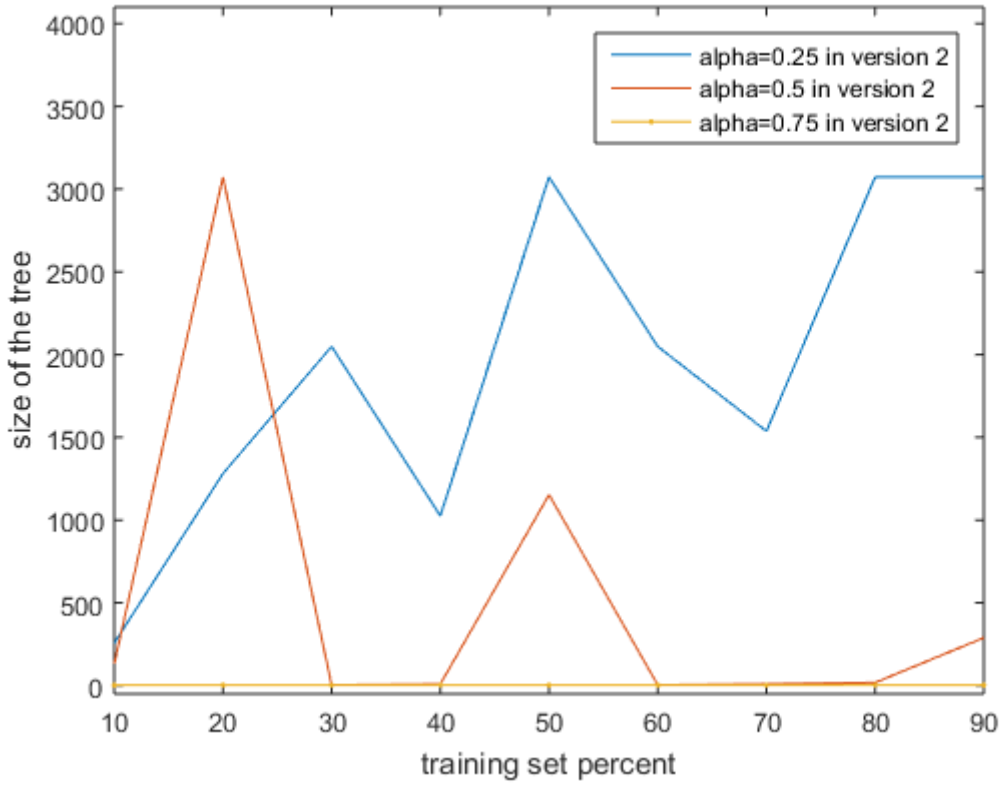
א. גרף עבור שאלה 3(ללא רעש)



גרף עבור שאלה *4(עם רעש) גרסה* 1:



גרף עבור שאלה 4(עם רעש) גרסה 2



ב.

- מבחינת ההבדלים בין שני הגרסאות ניתן לראות שבגרסה הראשונה אלגוריתם הלמידה מייצר עץ החלטה גדול יותר גם כאשר יש הפרעה.
- כאשר יש הפרעה ניתן לראות בגרסה הראשונה שכאשר גודל הרעש גדול יותר אז האלגוריתם משתמש בגדלי עצים יותר גדולים.
- כאשר יש הפרעה ניתן לראות בגרסה השנייה שכאשר גודל הרעש גדול יותר אז האלגוריתם משתמש בגדלי עצים יותר קטנים.
- כאשר יש הפרעה ניתן לראות שמבחינת תלות בגודל המדגם כאשר גודל הרעש הוא קטן יותר אז ככל שגודל המדגם גדול יותר מקבלים עץ החלטה יותר גדול כלומר מקבלים <u>מגמה עולה.</u>
  וכאשר גודל הרעש הוא גדול יותר אז ככל שגודל המדגם גדול יותר מקבלים עץ החלטה יותר קטן כלומר מקבלים <u>מגמה יורדת</u>.
  בעצם מה שמקבלים זה שככל שגודל הרעש יותר גדול אז כאשר אלגוריתם הלמידה מקבל יותר מידע(גודל מדגם גדול יותר) אז הוא מייצר עץ החלטה קטן יותר.

ג.

- ראינו מבחינת ההבדלים בין שני הגרסאות ניתן לראות שבגרסה הראשונה אלגוריתם הלמידה מייצר עץ החלטה גדול יותר גם כאשר יש הפרעה דבר שנצפה לו מכיוון שבגרסה השנייה יש לנו יותר תנאים וגם תנאים שהם יותר טובים כלומר עוזרים יותר בחיזוי התווית ולכן בגרסה השנייה גודל העץ האופטימלי מתקבל בגדלי עץ יותר קטנים כלומר נצטרך פחות שאלות על מנת לחזות בצורה טובה את התווית.
- ראינו שכאשר יש הפרעה ניתן לראות בגרסה הראשונה שכאשר גודל הרעש גדול יותר אז האלגוריתם משתמש בגדלי עצים יותר גדולים ,ניתן להסביר זאת בכך שכאשר גודל הרעש הוא גדול אז בכל שלב באלגוריתם מידת אי הוודאות בכל עלה כלומר האנטרופיה היא גדולה יותר ולכן נעדיף לפתח עוד ועוד עלים כדי להגיע למידת וודאות טובה יותר ובסופו של דבר נעדיף גדלי עצים יותר גדולים.
- ראינו שכאשר יש הפרעה ניתן לראות בגרסה השנייה שכאשר גודל הרעש גדול יותר אז האלגוריתם משתמש בגדלי עצים יותר קטנים, ניתן להסביר זאת בכך שבגרסה השנייה אלגוריתם הלמידה הוא יותר טוב בגלל כמות התנאים שיש לו ולכן כאשר גודל הרעש גדול יותר אז בשלב יותר מוקדם באלגוריתם הלמידה הוא "מבין" שכאשר מפתחים את העץ יותר ויותר אז זה לא יעזור בשיפור השגיאה על מדגם הולידציה ולכן הוא פשוט יעדיף להשתמש בגדלי עצים קטנים קטנים יותר. ניתן לראות זאת במיוחד כאשר גודל הרעש הוא 0.75 ושם הוא פשוט תמיד משתמש בגודל עץ הכי קטן כי השגיאה המתקבלת עבור כל גודל עץ היא אותו דבר.
- מה שקיבלנו זה שככל שגודל הרעש יותר גדול אז כאשר אלגוריתם הלמידה מקבל יותר מידע(גודל מדגם גדול יותר) אז הוא מייצר עץ החלטה קטן יותר ,זה קורה מכיוון שכאשר מקבלים יותר מידע שגוי אלגוריתם הלמידה מפתח עצים גדולים יותר אשר תלויים יותר ויותר במידע שהוא שגוי וכך גודל השגיאה על מדגם הולידציה המתקבל עבור עצים גדולים יותר אלו הוא גדול יותר ולכן בסופו של דבר האלגוריתם יעדיף להשתמש בעץ החלטה קטן יותר.