חלק ב׳ - מבוא ללמידה (56 נק׳)

✍️חלק א׳ – חלק היבש (28 נק')

kNN – נעים להכיר

בחלק זה תכירו אלגוריתם למידה בשם kNN, או בשמו המלא k-Nearest Neighbors, כאשר ה־k הוא למעשה פרמטר!
יהי סט אימון עם n־דוגמות, $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$, כאשר $\forall i: x_i \in \mathbb{R}^d, y_i \in \mathcal{Y}$.
כלומר הדוגמות הינן וקטורים d־ממדיים והתגיות הינן מדומיין כלשהו, הבעיה היא בעיית קלסיפיקציה (סיווג).
אם לא נאמר אחרת, הקלסיפיקציה תהיה בינארית, כלומר $\mathcal{Y} = \{-, +\}$.
עבור כל דוגמה בסט האימון, ניתן להסתכל על הכניסה ה־i בווקטור כעל הfeature ה־i של הדוגמה, קרי כל דוגמה x_i מיוצגת על ידי d־ערכים: $f_1(x_i), f_2(x_i), \dots, f_d(x_i)$.
תהליך ה"אימון" של האלגוריתם הוא טריוויאלי – פשוט שומרים את סט האימון במלואו.
תהליך הסיווג הוא גם פשוט למדי – כאשר רוצים לסווג דוגמה <u>מסט המבחן</u> מסתכלים על k השכנים הקרובים ביותר שלה במישור ה־d־ממדי <u>מבין הדוגמות בסט האימון</u>, ומסווגים את הדוגמה על פי הסיווג הנפוץ ביותר בקרב k השכנים.
על מנת להימנע משוויון בין הסיווגים, נניח בדרך כלל כי k־אי זוגי, או שנגדיר היטב שובר שוויון.
אם לא נאמר אחרת, במקרה של שוויון בקלסיפיקציה בינארית, נסווג את הדוגמה כחיובית +.

<u>שאלות הבנה</u>

א. (3 נק') כאמור, בתהליך הסיווג אנו בוחרים עבור הדוגמה את הסיווג הנפוץ ביותר של k השכנים הקרובים ביותר, אולם עלינו להגדיר את פונקציית המרחק עבור קביעת סט שכנים זה.
שתי פונקציות מרחק נפוצות הינן מרחק אוקלידי ומרחק מנהטן.

1) עבור איזה ערכים של d, k נקבל שאין תלות בבחירת פונקציית המרחק? (נמקי)
עבור d=1 (לכל k) נשים לב כי מרחק אוקלידי ומרחק מנהטן זהים זה לזה, ועל כן לא יהיה הבדל ביניהם בסיווג.

2) עבור בעיית קלסיפיקציה בינארית תנו דוגמה <u>פשוטה</u> לערכי d, k, סט אימון ודוגמת מבחן בה השימוש בכל אחת מפונקציות המרחק הנ"ל משנה את סיווג דוגמה המבחן.
נבחר $k = 1, d = 2$, בסט האימון הדוגמאות יהיו: (1,4) מסווג לחיובי (+) ו(3,3) תהיה מסווגת לשלילי (-). דוגמאת המבחן תהיה (1,1).
במרחק אוקלידי נקבל:

$$\sqrt{(1-1)^2 + (1-4)^2} = \sqrt{9} = 3$$
$$\sqrt{(1-3)^2 + (1-3)^2} = \sqrt{8} = 2.8$$

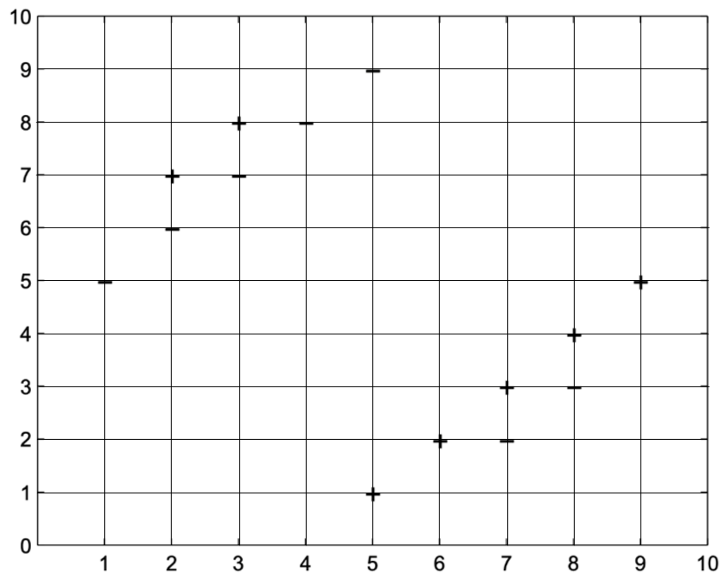
כלומר נקבל שהסיווג יהיה שלילי.
במרחק מנהטן נקבל:

$$|1-1| + |1-4| = 3$$
$$|1-3| + |1-3| = 4$$

כלומר נקבל שהסיווג יהיה חיובי.

מעתה, אלא אם כן צוין אחרת, נשתמש במרחק אוקלידי.

נתונה קבוצת האימון הבאה, כאשר $d = 2$:



3) (1 נק') איזה ערך של k עלינו לבחור על מנת לקבל את הדיוק המרבי **<u>על קבוצת האימון</u>**? מה יהיה ערך זה?

עבור k=1 נקבל את הדיוק המרבי על קבוצת האימון, שכן עבור כל דגימה השייכת לקבוצה זו, האיבר הקרוב ביותר אליה הוא היא בעצמה. נניח כי נקודות הקצה (סהכ 4) הן שגויות, אז הדיוק יהיה $\frac{10}{14}$. בהנחה והקצוות הן לא שגויות, נוכל להגיד כי השגיאה היא 0 ולכן הדיוק שווה ל100 אחוז.

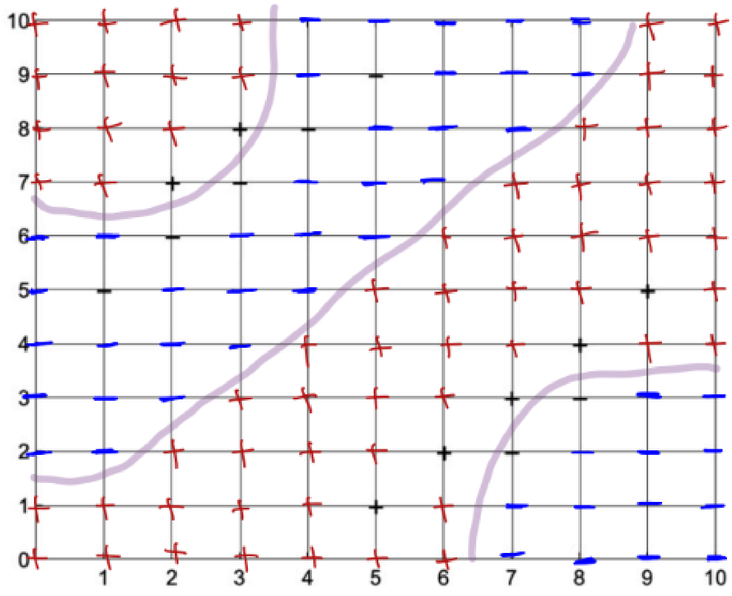
4) (1 נק') עבור איזה ערך של k נקבל מסווג $majority$ של קבוצת האימון? קרי כל דוגמת מבחן תקבל את הסיווג הנפוץ של כלל קבוצת האימון?

עבור k=14 (כלומר kשגודלו כגודל קבוצת האימון). כשנרצה לסווג נקודת מבחן, השכנים אותם נבחן יהיו בעצם כל קבוצת האימון, ולכן מתוך קבוצה זו ילקח הסיווג הנפוץ ביותר עבור דוגמאת המבחן.

5) (2 נק') נמקו מדוע שימוש בערכי k גדולים או קטנים מדי יכול להיות גרוע עבור קבוצת הדגימות הנ״ל.

עבור ערכי K קטנים מדי, אנו נגיע למצב של overfitting, כלומר יהיה משקל גדול מדי למקרי הקצה והם לא יהיו זניחים, וכך ישפיעו על הדיוק, כלומר השפעה גדולה של הרעש.
עבור ערכי K גדולים מדי, אנו נגיע למצב של underfitting, כלומר נתייחס ליותר מדי שכנים, גם כאלה שהם לא באמת קרובים, ועל כן נאבד חלק מהמידע ונקבל תוצאה שגויה.

6) (2 נק') שרטט את גבול ההחלטה של 1-nearest neighbour עבור הגרף.



השוואה בין מודלי למידה:

נתייחס לסעיפים הבאים כפי שראינו בהרצאה (עפי ממוצע features).
שאלת הפיצול של העץ תהיה $f = I\{f(x) \le 0\}$

1) (3 נק') הציגו מסווג מטרה $f(x): R^2 \to \{0,1\}$ וקבוצת אימון בעלת לכל היותר 10 דוגמאות כך שלמידת עץ ID3 תניב מסווג אשר עונה נכון עבור כל דוגמת מבחן אפשרית (כלומר יתקבל מסווג המטרה), אך למידת KNN תניב מסווג שעבורו קיימת לפחות דוגמת מבחן אחת עליה הוא יטעה, לכל ערך K שייבחר.

נגדיר את מסווג המטרה באופן הבא: $f(x) = I\{v_1 > 0, v_2 > 0\}$. כלומר נסווג כחיובי את כל הנקודות שנמצאות ברביע הראשון (בלי הצירים). נגדיר את D באופן הבא:

$$D = \{(x_1, v_1 = 1, v_2 = 1, y = +), (x_2, v_1 = 1, v_2 = -1, y = -),$$
$$(x_3, v_1 = -1, v_2 = 1, y = -), (x_4, v_1 = -1, v_2 = -1, y = -)\}$$

כאשר $x_i, i \in (1,2,3,4)$ הן הנקודות השייכות לקבוצת האימון.
נשים לב כי ב*ID3*, עץ ההחלטה יסתכל על *v1<0* שיתן את ערך ה*IG* המקסימלי, ואז יסתכל על *v2<0* ועל כן הסיווג יהיה מדויק.
נשים לב כי במקרה של *KNN*, עבור דוגמת מבחן $x = (1,0)$ נשים לב כי יטעה ויסווג אותה כחיובית, למרות שע"פ המסווג שהגדרנו, היא צריכה להיות שלילית.

2) (3 נק') הציגו מסווג מטרה $f(x): R^2 \to \{0,1\}$ וקבוצת אימון בעלת לכל היותר 10 דוגמאות כך שלמידת מסווג KNN עבור ערך K מסוים תניב מסווג אשר עונה נכון עבור כל דוגמת מבחן אפשרית (כלומר יתקבל מסווג המטרה), אך למידת עץ ID3 תניב מסווג אשר עבורו קיימת לפחות דוגמת מבחן אפשרית אחת עליה הוא יטעה.

נגדיר את המסווג כך: $f(x) = I\{v1 + v2 \ge 0\}$. כלומר נסווג כחיובי את כל הנקודות שנמצאות מעל הישר y=-x. נגדיר את D כך:

$$D = \{(x1, v1 = 1, v2 = 1, y = +), (x2, v1 = -1, v2 = -1, y = -)\}$$

נשים לב כי במקרה של KNN, עבור k=1, תמיד הסיווג יתבצע על פי השכן היחיד הקרוב ביותר אלינו, ולכן יהיה נכון. עבור הקצוות במקרה זה, שובר השוויון יהיה זה שבקבוצת האימון מסווג כ+ ועל כן גם יסווג נכון.

לעומת זאת, במקרה של ID3, תהיה חלוקה למקרים עבור הערכים של v1,v2 כאשר הם קטנים או שווים 0. כלומר דוגמא שבה נקודת המבחן היא (1,-0.75) תסווג כשלילית למרות שהיא אמורה להיות מסווגת חיובית.

3) (3 נק') הציגו מסווג מטרה $f(x): R^2 \to \{0,1\}$ וקבוצת אימון בעלת לכל היותר 10 דוגמאות כך שלמידת מסווג KNN עבור ערך K מסוים תניב מסווג אשר עבורו קיימת לפחות דוגמת מבחן אפשרית אחת עליה הוא יטעה, וגם למידת עץ ID3 תניב מסווג אשר עבורו קיימת לפחות דוגמת מבחן אחת אפשרית עליה הוא יטעה.

נגדיר את המסווג באופן הבא: $f(x) = I\{v1 \geq 0, v2 \geq 0\}$, כלומר מסווג דומה לסעיף 1 רק שהפעם כללנו את הצירים. נגדיר את D כך:

$$D = \{(x_1, v_1 = 1, v_2 = 1, y = +), (x_2, v_1 = 1, v_2 = -1, y = -),$$
$$(x_3, v_1 = -1, v_2 = 1, y = -), (x_4, v_1 = -1, v_2 = -1, y = -),$$
$$(x_5, v1 = 2, v_2 = 0, y = +)\}$$

עבור ID3 נשים לב כי למשל עבור הנקודה (0.5,-0.5) תסווג כחיובית אך בפועל היא אמורה להיות מסווגת כשלילית ועל כן זו טעות.

עבור KNN נשים לב כי למשל עבור הנקודה (0,0), כאשר k=2, יסווג אותה כשלילית למרות שבפועל אמורה להיות מסווגת כחיובית, על כן זו טעות.

4) (3 נק') הציגו מסווג מטרה $f(x): R^2 \to \{0,1\}$ וקבוצת אימון בעלת לכל היותר 10 דוגמאות כך שלמידת מסווג KNN עבור ערך K מסוים תניב מסווג אשר עונה נכון עבור כל דוגמת מבחן אפשרית (כלומר יתקבל מסווג המטרה), וגם למידת עץ ID3 תניב מסווג עונה נכון עבור כל דוגמת מבחן אפשרית (כלומר יתקבל מסווג המטרה).

נגדיר את מסווג המטרה כך: $f(x) = I\{v1 \geq 0\}$ ונשנה את שאלת הפיצול בעץ שהגדרנו למעלה להיות $f = I\{f(x) < 0\}$, כלומר כעת המגבלה חזקה יותר מהשהייתה בסעיפים הקודמים.

נקבל מצב שבו ID3 יסווג את ציר y כחיובי, ועל כן נקבל כי תמיד יסווג נכון.

עבור KNN, כאשר אינו על ציר y יבחר לפי דגימה בחצי המישור שבו הוא יהיה. כאשר יהיה על הציר, יבחר לסווג כחיובי שכן v1 גדול יותר.

מתפצלים ונהנים

(7 נק') כידוע, בעת סיווג של דוגמת מבחן על ידי עץ החלטה, בכל צומת בעץ אנו מחליטים לאיזה צומת בן להעביר את דוגמת המבחן על ידי ערך סף v שמושווה לfeature של הדוגמה. לפעמים ערך הסף קרוב מאוד לערך הfeature של דוגמת המבחן. היינו רוצים להתחשב בערכים "קרובים" לערך הסף בעת סיווג דוגמת מבחן, ולא לחרוץ את גורלה של הדוגמה לתת־עץ אחד בלבד; לצורך כך נציג את האלגוריתם הבא:

יהיו עץ החלטה T, דוגמת מבחן $x \in \mathbb{R}^d$, ווקטור $\varepsilon \in \mathbb{R}^d$ המקיים $\forall i \in [1,d]: \varepsilon_i > 0$.
כלל אפסילון־החלטה שונה מכלל ההחלטה הרגיל שנלמד בכיתה באופן הבא:
נניח שמגיעים לצומת בעץ המפצל לפי ערכי התכונה i, עם ערך הסף v_i.
אם מתקיים $|x_i - v_i| \leq \varepsilon_i$ אזי ממשיכים **בשני** המסלולים היוצאים מצומת זה, ואחרת ממשיכי לבן המתאים בדומה לכלל ההחלטה הרגיל. לבסוף, מסווגים את הדוגמה x בהתאם לסיווג הנפוץ ביותר של הדוגמאות הנמצאות בכל העלים אליהם הגענו במהלך הסיור על העץ (במקרה של שוויון – הסיווג ייקבע להיות $True$).

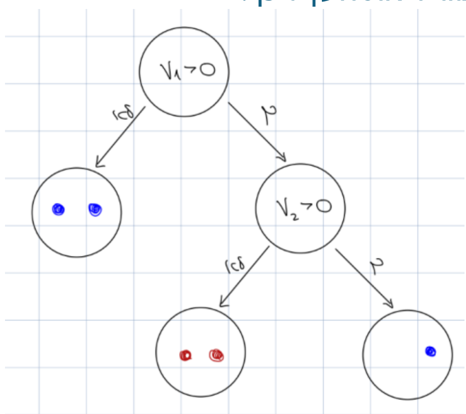
יהא T עץ החלטה לא גזום, ויהא T' העץ המתקבל מ־T באמצעות גיזום מאוחר שבו הוסרה הרמה התחתונה של T (כלומר כל הדוגמות השייכות לזוג עלים אחים הועברו לצומת האב שלהם).

הוכיחו\הפריכו: **בהכרח** <u>קיים</u> וקטור ε כך שהעץ T עם כלל אפסילון־החלטה והעץ T' עם כלל ההחלטה הרגיל יסווגו <u>כל דוגמת מבחן</u> ב-$\mathbb{R}^d$ בצורה זהה.

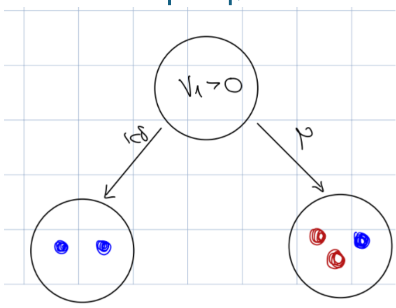
הטענה אינה נכונה, נראה באמצעות דוגמא נגדית:

נגדיר את וקטור אפסילון להיות: $\varepsilon = \binom{\varepsilon_1}{\varepsilon_2}$ כאשר $\varepsilon_1, \varepsilon_2 > 0$.

נגדיר את העץ T כך:



ונגדיר את העץ 'T כך:



נגדיר גם את דוגמת המבחן להיות:

$$x = \binom{0.5\varepsilon_1}{0.5\varepsilon_2}$$

עבור T שמשתמש בכלל אפסילון, נקבל עבור 2 צמתי ההחלטה שהכלל מתקיים, כלומר הסיווג יקבע לפי הרוב בעלים, במקרה של הדוגמא שלנו **כחול**.

עבור 'T, כלל ההחלטה יוביל לכך שבדוגמאת המבחן X נקבל אדום (כי v1 אכן חיובי).

כלומר קיבלנו סיווגים שונים לשני העצים, לכן זו הפרכה.

חלק ב׳ - היכרות עם הקוד

רקע

חלק זה הוא רק עבור היכרות הקוד, עוברו עליו במלואו ווודאו כי הינכם מבינים את הקוד.

בחלק של הלמידה, נעזר ב-$dataset$, הדאטה חולק עבורכם לשתי קבוצות: קבוצת אימון train.csv וקבוצת מבחן test.csv.

ככלל, קבוצת האימון תשמש אותנו לבניית המסווגים, וקבוצת המבחן תשמש להערכת ביצועיהם.

בקובץ utils.py תוכלו למצוא את הפונקציות הבאות לשימושכם:
load_data_set, create_train_validation_split, get_dataset_split
אשר טוענות/מחלקת את הדאטה בקבצי ה־csv למערכי np.array (קראו את תיעוד הפונקציות).

הדאטה של ID3 עבור התרגיל מכיל מדדים שנאספו מצילומים שנועדו להבחין בין גידול שפיר לגידול ממאיר. כל דוגמה מכילה 30 מדדים כאלה, ותווית בינארית **diagnosis** הקובעת את סוג הגידול (0=שפיר, 1=ממאיר). כל התכונות (מדדים) רציפות . העמודה הראשונה מציינת האם האדם חולה (M) או בריא (B). שאר העמודות מציינות כל תכונות רפואיות שונות של אותו אדם (התכונות מורכבות ואינכם צריכים להתייחס למשמעות שלהן כלל).

תיקיית $\underline{ID3 - dataset}$:
- תיקיה זו אלו מכילה את קבצי הנתונים עבור $ID3$.

קובץ $\underline{utils.py}$:
- קובץ זה מכיל פונקציות עזר שימושיות לאורך התרגיל, כמו טעינה של $dataset$ וחישוב הדיוק.
- בחלק הבא יהיה עליכם לממש את הפונקציה $accuracy$. קראו את תיעוד הפונקציות ואת ההערות הנמצאות תחת התיאור **TODO**.

קובץ $\underline{unit\_test.py}$:
- קובץ בדיקה בסיסי שיכול לעזור לכם לבדוק את המימוש.

קובץ $\underline{DecisionTree.py}$:
- קובץ זה מכיל 3 מחלקות שימושית לבניית עץ $ID3$ שלנו.
  - <u>המחלקה $Question$</u>: מחלקה זו מממשת הסתעפות של צומת בעץ. היא שומרת את התכונה ואת הערך שלפיהם מפצלים את הדאטה שלנו.
  - <u>המחלקה $DecisionNode$</u>: מחלקה זו מממשת צומת בעץ ההחלטה.
    הצומת מכיל שאלה $Question$ ואת שני הבנים $true\_branch, false\_branch$ כאשר $true\_branch$ הוא הענף בחלק של הדאטה שעונה $True$ על שאלת הצומת (הפונקציה $match$ של ה$Question$ מחזירה $True$).
    ו־ $false\_branch$ הוא הענף בחלק של הדאטה שעונה $False$ על שאלת הצומת (הפונקציה $match$ של ה$Question$ מחזירה $False$).
  - <u>המחלקה $Leaf$</u>: מחלקה זו מממשת צומת שהוא עלה בעץ ההחלטה. העלה מכיל לכל אחד מהמחלקות בדאטה את מספר הדוגמאות בעלה עבור כל מחלקה (למשל: $\{'B' : 5, 'M' : 6\}$).

קובץ $\underline{ID3.py}$:
- קובץ זה מכיל את המחלקה של $ID3$ שתצטרכו לממש חלקים ממנה, עיינו בהערות ותיעוד המתודות.

קובץ $\underline{ID3\_experiments.py}$:
- קובץ הרצת הניסויים של ID3, הקובץ מכיל את הניסויים הבאים, שיוסברו בהמשך:
  $cross\_validation\_experiment, basic\_experiment$

חלק ג׳ – חלק רטוב ID3 (28 נק')

עבור חלק זה מותר לכם להשתמש בספריות הבאות:

All the built in packages in python, sklearn, pandas ,numpy, random, matplotlib, argparse, abc, typing.

<u>אך כמובן שאין להשתמש באלגוריתמי הלמידה, או בכל אלגוריתם או מבנה נתונים אחר המהווה חלק מאלגוריתם למידה אותו תתבקשו לממש.</u>

1. (3 נק') השלימו את הקובץ utils.py ע"י מימוש הפונקציה $accuracy$.
קראו את תיעוד הפונקציה ואת ההערות הנמצאות תחת התיאור **TODO**.
(הריצו את הטסטים המתאימים בקובץ $unit\_test.py$ לוודא שהמימוש שלכם נכון).
שימו לב! בתיעוד ישנן הגבלות על הקוד עצמו, אי־עמידה בהגבלות אלו תגרור הורדת נקודות.
בנוסף, שנו את ערך הID בתחילת הקובץ מ־123456789 למספר תעודת הזהות של אחד מהמגישים.

2. (10 נק') **אלגוריתם ID3:**

a. השלימו את הקובץ ID3.py ובכך ממשו את אלגוריתם $ID3$ כפי שנלמד בהרצאה.
TODO
שימו לב שכל התכונות רציפות. אתם מתבקשים להשתמש בשיטה של חלוקה דינמית המתוארת בהרצאה. כאשר בוחנים ערך סף לפיצול של תכונה רציפה, דוגמאות עם ערך השווה לערך הסף משתייכות לקבוצה עם הערכים הגדולים מערך הסף. במקרה שיש כמה תכונות אופטימליות בצומת מסוים בחרו את התכונה בעלת <u>האינדקס המקסימלי</u>.
כלל המימוש הנ"ל צריך להופיע בקובץ בשם $ID3.py$, באזורים המוקצים לכך.
(השלימו את הקוד החסר אחרי שעיינתם והפנמתם את הקובץ $DecisionTree.py$ ואת המחלקות שהוא מכיל).

b. ממשו את $basic\_experiment$ שנמצאת ב $ID3\_experiments.py$ **TODO**
והריצו את החלק המתאים ב $main$ ציינו בדו"ח את הדיוק ✍ שקיבלתם.
קיבלנו דיוק של 96.12%

3. **גיזום מוקדם.**
פיצול צומת מתקיים כל עוד יש בו יותר דוגמאות מחסם המינימום m, כלומר בתהליך בניית העץ מבוצע "גיזום מוקדם" כפי שלמדתם בהרצאות. שימו לב כי פירוש הדבר הינו שהעצים הנלמדים אינם בהכרח עקביים עם הדוגמאות. לאחר סיום הלמידה (של עץ יחיד), הסיווג של אובייקט חדש באמצעות העץ שנלמד מתבצע לפי רוב הדוגמאות בעלה המתאים.

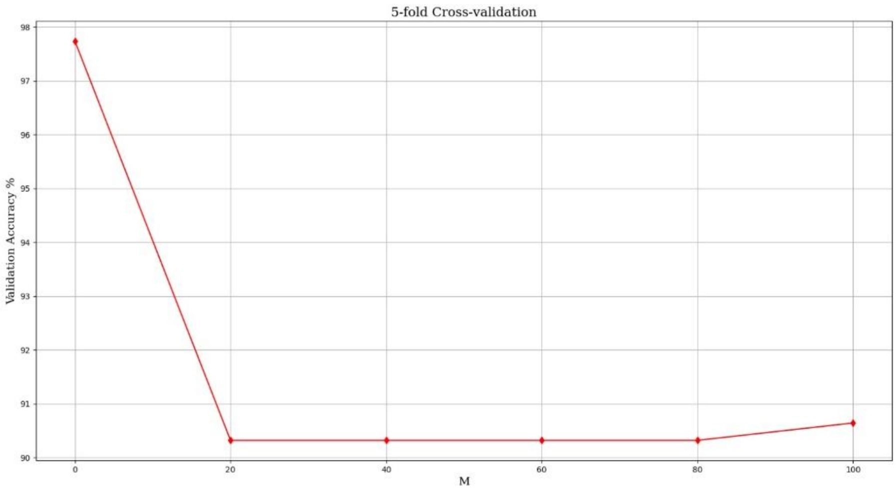
✍ **a.** (2 נק') הסבירו מה החשיבות של הגיזום באופן כללי ואיזה תופעה הוא מנסה למנוע?
גיזום באופן כללי מנסה למנוע תופעת overfitting, כלומר ניסיון להשתמש ביותר מדי דגימות מקבוצת האימון (התאמת יתר) על מנת למצוא את הקלסיפיקציה הנכונה לדגימת הבוחן. המטרה היא ליצור עצים קטנים יותר ולא עקביים, כך שגם אם קיימת טעות בסיווג, היא לא תגרר הלאה, וכן לא תהיה סיטואציה של Overfitting.

b. (3 נק') **<u>עדכנו</u>** את המימוש בקובץ $ID3.py$ כך שיבצע גיזום מוקדם כפי שהוגדר בהרצאה.
הפרמטר $min\_for\_pruning$ מציין את המספר המינימלי בעלה לקבלת החלטה, קרי יבוצע גיזום מוקדם אם ורק אם מספר הדוגמות בצומת קטן שווה לפרמטר הנ"ל. **TODO**

c. (8 נק') **שימו לב, זהו סעיף יבש ואין צורך להגיש את הקוד שכתבתם עבורו.**
בצעו כיוונון לפרמטר M על קבוצת האימון:
1. בחרו לפחות חמישה ערכים שונים לפרמטר M.

2. עבור כל ערך, חשבו את הדיוק של האלגוריתם על ידי K − fold cross validation על קבוצת האימון בלבד.
כדי לבצע את חלוקת קבוצת האימון ל- K קבוצות יש להשתמש בפונקציה sklearn.model_selection.KFold עם הפרמטרים $n\_split = 5$, $shuffle = True$ ו־random_state אשר שווה למספר תעודת הזהות של אחד מהשותפים.

✍ i. השתמשו בתוצאות שקיבלתם כדי ליצור גרף המציג את השפעת הפרמטר M על הדיוק. צרפו את הגרף בדו"ח. (לשימושכם הפונקציה $util\_plot\_graph$ בתוך הקובץ $utils.py$).



2.

✍ i. הסבירו את הגרף שקיבלתם. לאיזה גיזום קיבלתם התוצאה הטובה ביותר ומהי תוצאה זו?
עבור המצב שבו לא נבצע גיזום (כלומר בערך ששווה 0) מתקבלת התוצאה הטובה ביותר, נראה שככל שמתבצע גיזום רחב יותר (כלומר פרמטר הגיזום גדל), כך התוצאה נהיית פחות טובה. מבין הערכים שבחרנו,עבור הערך שבו m=100, אנו מקבלים את התוצאה האופטימלית עם גיזום.

✍ (2 נק׳) השתמשו באלגוריתם ID3 עם הגיזום המוקדם כדי ללמוד מסווג מתוך **כל** קבוצת האימון ולבצע חיזוי על קבוצת המבחן.
השתמשו בערך ה־M האופטימלי שמצאתם בסעיף c. (ממשו $best\_m\_test$ שנמצאת ב $ID3\_experiments.py$ והריצו את החלק המתאים ב $main$). ציינו בדו"ח את הדיוק שקיבלתם. האם הגיזום שיפר את הביצועים ביחס להרצה ללא גיזום?
הרצנו עם הערך m=100, וקילבנו כי התוצאה הינה דיוק של 95.15%. נשים לב כי גיזום זה אינו משפר את הביצועים לעומת הרצה ללא גיזום (כאשר m=0). יתכן כי בחירת ערכי m אותם בדקנו השפיעו על התוצאה, ובחירה שונה של ערכים הייתה יכולה להשפיע אחרת על התוצאה.