פרוייקט

למידה חישובית עם פייתון

# הערות:

1. תאריך הגשה: 23.08.2020
2. הגשה בזוגות או בשלשות.
3. מועד הגנה יקבע בהמשך.

# מבוא

"בשנים האחרונות חלו התפתחויות נוספות בתחום, **רגרסיה לוגיסטית** התגלתה מחדש ועוצבה לתת מענה לבעיות עם טווח רחב יותר. אלגוריתמים כמו "נאיב בייס", רשתות בייסיאניות, ומקסימום אנטרופיה הביאו לתוצאות מצוינות במגוון בעיות. בעקבותיהם פותחו אלגוריתמים משולבים, המשתמשים באנסמבל של מסווגים כגון **אלגוריתם אדה בוסט**, ויערות אקראיים, שגם הם הביאו לשיפור ניכר בתוצאות."

ראה פרק על למידת מכונה בויקפדיה:<https://www.hamichlol.org.il/%D7%9C%D7%9E%D7%99%D7%93%D7%AA_%D7%9E%D7%9B%D7%95%D7%A0%D7%94>

# מטרת הפרויקט

בפרויקט זה נאמת את הכתוב במבוא. בפרט נבנה וריאציה של רגרסיה לוגיסטית ונשלב אותה באלגוריתם אדה בוסט כדי לשפר את התוצאות.

# מטלות הפרויקט

## **חלק 1 - מאמר לקריאה ולשחזור**

קרא את הפרק **"Introduction to AdaBoost**" המצורף לפרוייקט זה. חלקים: 2.1 - 2.3 (כולל). חלקים 2.4 - 2.5 הם רשות, אך מומלצים לקריאה.

תרגם את חלקים 2.1 - 2.3 (כולל) לעברית, תוך שחזור מלא של כל הנוסחאות והמעברים. יש להבין את הפרק היטב.

המאמר המקורי עליו מתבסס פרק זה זיכה את כותביו בפרס גדל לשנת 2003. ראה לינקים הבאים:

<https://en.wikipedia.org/wiki/AdaBoost>

<https://en.wikipedia.org/wiki/G%C3%B6del_Prize>

## **חלק 2 - שלד קוד**

# שחזר את הקוד הנתון בבלוג הבא:

<https://geoffruddock.com/adaboost-from-scratch-in-python/>

בפרט יש לשחזר את כל הגרפים בבלוג. אנחנו נצטרך אותם בהמשך.

הוסף לגרפים בכל איטרציה את ערכי אלפא בנקודת הזמן t ואפסילון בנקודת הזמן t. כלומר את הערכים:

**alpha\_t:** weighted error of the base classifier

**epsilon\_t**: coefficient of the base classifier

(ראה שורות 4 - 5 בקוד הנתון בעמוד 12 חלק 2.2)

## **חלק 3 - רגרסיה לוגיסטית עם משקולות**

הקלט לאלגוריתם DecisionTreeClassifer בו השתמש כותב המאמר כקופסה שחורה, מקבל סט של וקטורים עם משקולות (כלומר, וקטורים ממושקלים). ראה המערך: **curr\_sample\_weights**

הרחב את האלגוריתם לרגרסיה לוגיסטית שכתבנו בכיתה כדי שיתמוך בוקטורים ממושקלים. על פונקציית העלות של הרגרסיה הלוגיסטית החדשה לקחת בחשבון משקלים אלו.

שימו לב: אם משקלי כל הוקטורים הם 1, על הפונקציה החדשה לקרוס לפונקציה המקורית.

## **חלק 4 - אדה בוסט עם רגרסיה לוגיסטית עם משקולות**

החלף את הקופסה השחורה DecisionTreeClassifer שבו השתמש כותב הבלוג עם הגרסה החדשה של רגרסיה לוגיסטית עם משקולות.

שחזר את הגרפים מחדש עבור הגרסה החדשה.