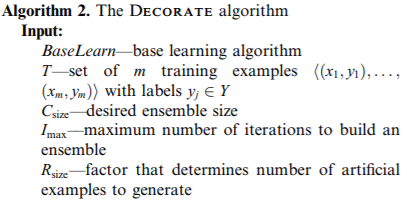
תרגיל מס' 3

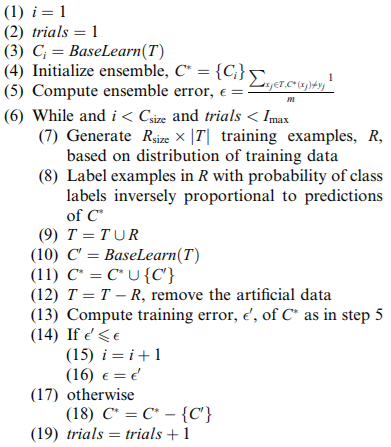
הגשה: בזוגות מועד אחרון להגשה: 13.1.2019

הצגנו בכתה את האלגוריתם DECORATE (Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples) אשר פורסם במסגרת המאמר:

Prem Melville, Raymond J. Mooney, Creating diversity in ensembles using artificial data, Information Fusion, Volume 6, Issue 1, 2005, Pages 99-111 (<https://doi.org/10.1016/j.inffus.2004.04.001>)

להלן האלגוריתם אליו מתייחס התרגיל:





מטרת תרגיל זה היא לממש את הרעיון המופיע באלגוריתם עם שתי גרסאות ליצירת תצפיות מלאכותיות (צעד 7 באלגוריתם הנ"ל):

1. על פי השיטה המתוארת במאמר המקורי ליצירה אקראית של תצפיות - תיאור מדויק מופיע בסעיף 4.1. Construction of artificial data במאמר הנ"ל.
2. על ידי יצירת תצפיות מלאכותיות באמצעות אלגוריתם GAN (generative adversarial network)

**דגשים לתרגיל:**

1. הבעיה מתמקדת בבעיות סיווג בלבד עבור נתונים טבלאיים (בדומה לתרגיל התחרות)
2. מודל הבסיס הוא עץ החלטה - אין צורך לממש מחדש את האלגוריתם הלמידה של עץ ההחלטה אלא להשתמש בחבילה מוכנה ללימוד עצים למטרות סיווג.
3. ניתן להיעזר במימושים קיימים של GAN לצורך יצירה של תצפיות מלאכותיות (אפשר להשתמש במימוש כפי שהוא ללא צורך לממש מחדש). קישורים מומלצים לקריאה / קוד

<https://arxiv.org/pdf/1811.11264.pdf>

וכן הקישור למימוש: <https://github.com/DAI-Lab/TGAN>

<http://www.rricard.me/machine/learning/generative/adversarial/networks/keras/tensorflow/2017/04/05/gans-part2.html>

<http://www.vldb.org/pvldb/vol11/p1071-park.pdf>

**הנחיות הגשה**

ההגשה צריכה לכלול את הפרטים הבאים + דגשים חשובים.

1. קוד מקור + תיעוד
2. מסמך המתאר את אופן המימוש, כיצד להריץ את הקוד וה-Evaluation שעשיתם
3. ה-Evaluation צריכה לכלול השוואה לביצועי החיזוי של מסווג מסוג עץ החלטה (על בסיס חבילה קיימת) על גבי 10 Dataset (רצוי לבחור ב- Datasets גדולים דיו כדי ששיטת GAN תהיה אפקטיבית). שימו לב לדגשים הבאים:
   1. ה-Evaluation צריך לכלול השוואה של האלגוריתם DECORATE+GAN לעומת DECORATE רגיל (ללא שימוש ב-GAN)
   2. ה-Evaluation צריך לכלול לפחות שני מדדי דיוק ובנוסף ביצועי זמן ריצה (זמן סיווג וזמן החיזוי).
   3. ה-Evaluation יתבסס על בדיקה של 10 Fold Cross Validation.
4. הציגו את תוצאות ה-Evaluation ומסקנותיכם