**תרגיל בית מס' 2 (הגשה עד 18.12.2019)**

יש לבחור באחת מהאופציות הבאות:

אופציה 1: שפת Python

לשנות את הקוד של האלגוריתם RandomForestClassifier בחבילת scikit learn כך שכל עלה בעצים המתקבלים יהיה מורכב ממודל סיווג מסוג Naïve Bayes (מכונה GaussianNB בחבילת Scikit). בניית העץ תעשה כפי שהיא היום, אך במקום להחזיק את התפלגות ה-Classes בכל עלה, העץ החדש יחזיק מודל סיווג כאמור. בזמן סיווג ישולבו תוצאות הסיווג של מודלי ה-Naïve Bayes על בסיס ההסתברות שמודל ה-Naïve Bayes מחזיר לכל Class.

אופציה 2: שפת R

לשנות את הקוד של האלגוריתם RandomForest בחבילת party כך שכל עלה בעצים המתקבלים יהיה מורכב ממודל רגרסיה מסוג naïve Bayes (naiveBayes בחבילת e1071), כלומר במקום להחזיק את הממוצע בכל עלה, העץ החדש יחזיק מודל סיווג כאמור. ניתן להיעזר במנגנון ההרחבה הקיים בפונקציית mob.

ללא קשר לאופציה שנבחרה:

1. השינוי המוצע צריך לתמוך גם בבעיות סיווג בינאריות וגם בבעיות סיווג מרובות Classes (MultiClass)
2. איזה שינוי יש לדעתכם לעשות בערכי ברירת המחדל של הייפר-פרמטרים של האלגוריתם RandomForest כדי להבטיח לרוב תוצאות טובות יותר.
3. יש להשוות את ביצועי החיזוי של המודל עם ובלי התיקון המוצע על גבי 10 בסיסי נתונים שיבחרו מתוך האתר <https://archive.ics.uci.edu/ml/datasets.html>
4. הציגו את מסקנותיכם

הגשה בזוגות