**תרגיל מעבדה 2**

טען את הקובץ Wisconsin.csv המצורף

\*\*הקובץ נלקח מהאתר: <http://sci2s.ugr.es/keel/dataset.php?cod=73>

היעזרו בקישור המופיע בהערות לשקף הראשון במצגת של מעבדה מספר 2 לטעינת המודל.

This dataset contains 699 cases from a study that was conducted at the University of Wisconsin Hospitals, Madison, about patients who had undergone surgery for breast cancer. The task is to determine if the detected tumor is benign (2) or malignant [4].

attribute ClumpThickness integer [1, 10]

attribute CellSize integer [1, 10] attribute CellShape integer [1, 10] attribute MarginalAdhesion integer [1, 10] attribute EpithelialSize integer [1, 10] attribute BareNuclei integer [1, 10] attribute BlandChromatin integer [1, 10] attribute NormalNucleoli integer [1, 10] attribute Mitoses integer [1, 10] attribute Class {2,4}

1. שמרו את הטבלה במשתנה חדש כך שלא תכיל שורות עם ערכים חסרים, כתבו את הפקודות בהן השתמשתם. יש להציג נתונים סיכומיים של התכונה שעבורה אתם מתכוונים לבצע דיסקרטיזציה ועפ"י נתונים אלו לקבוע ערכי סף ולבצע את הסעיף הנוכחי.
2. כמה שורות נותרו בטבלה?
3. בחרו באחת העמודות בבסיס הנתונים לטובת המשימה הבאה:
   1. הציגו את התפלגות הנתונים בעמודה בה בחרתם.
   2. בחרו באופן מושכל גבולות לביצוע דיסקרטיזציה וציינו מהם הגבולות בהם השתמשתם?
   3. בצעו דיסקרטיזציה ע"י חלוקה לשלושה אינטרוולים לפי בחירתכם.

**הוראות הגשה**

* הגשה בזוגות בלבד.
* עליכם להגיש קובץ word ובו תשובות לשאלות הנ"ל, הכוללות את הפקודות בהם השתמשתם ואת התוצאות שנתקבלו בכל סעיף.
* שם קובץ ה- word אמור לכלול את מספרי תעודות הזהות של שתי המגישים וביניהם קו תחתון.
* רק אחד מבני הזוג יגיש את העבודה באתר הקורס.
* ניתן להגיש את התרגיל עד שבוע לאחר המעבדה בה פורסם לתיבת ההגשה הרלוונטית באתר הקורס.

***בהצלחה!***