

## מדעי הנתונים ובינה עסקית

## תרגיל מס' 3

תאריך ההגשה (דרך המודל): 2.6.2020.

רופאים מ-Cleveland Clinic Foundation מעוניינים לחזות נוכחות של מחלת לב בחולים. לשם כך, נאסף מידע על 100 חולים. קובץ ה-CSV המצורף (dataset) מכיל 4 משתנים מועמדים ומשתנה מטרה אחד (prediction).

הגדרות המאפיינים:

Attribute	Domain	Type	Use In Network number
gender	male, female	Nominal	Candidate Input
fbs	false, true	Nominal	Candidate Input
slope	upsloping, flat, downsloping	Nominal	Candidate Input
age	37-71 ( $T_1 = 57$ , $T_2 = 65$ )	Continuous	Candidate Input
prediction	absence, presence	Nominal	Target

•  $T_1$  ו-  $T_2$  הם שני ערכים אפשריים (Thresholds) עבור חלוקת המשתנה age לשני טווחים.

1. יש לבנות רשת IFN, תוך שימוש בכל הנתונים כ-training set.

א. Minimum significance level = 5%

ב. Maximal number of layers = 3

ג. יש להציג את כל החישובים בכל איטרציה של האלגוריתם ( $MI$ ,  $G^2$ ,  $\chi^2$ ), מסקנה

האם דוחים את השערת האפס וטבלה המסכמת את ה-  $MI$  של כל המאפיינים).

ד. יש לצייר את הרשת שהתקבלה בסיום הריצה כולל המשקולות.