

## מדור בחינות ומערכת שעות

המחלקה להנדסת תוכנה

25/6/2019

09:00-12:00

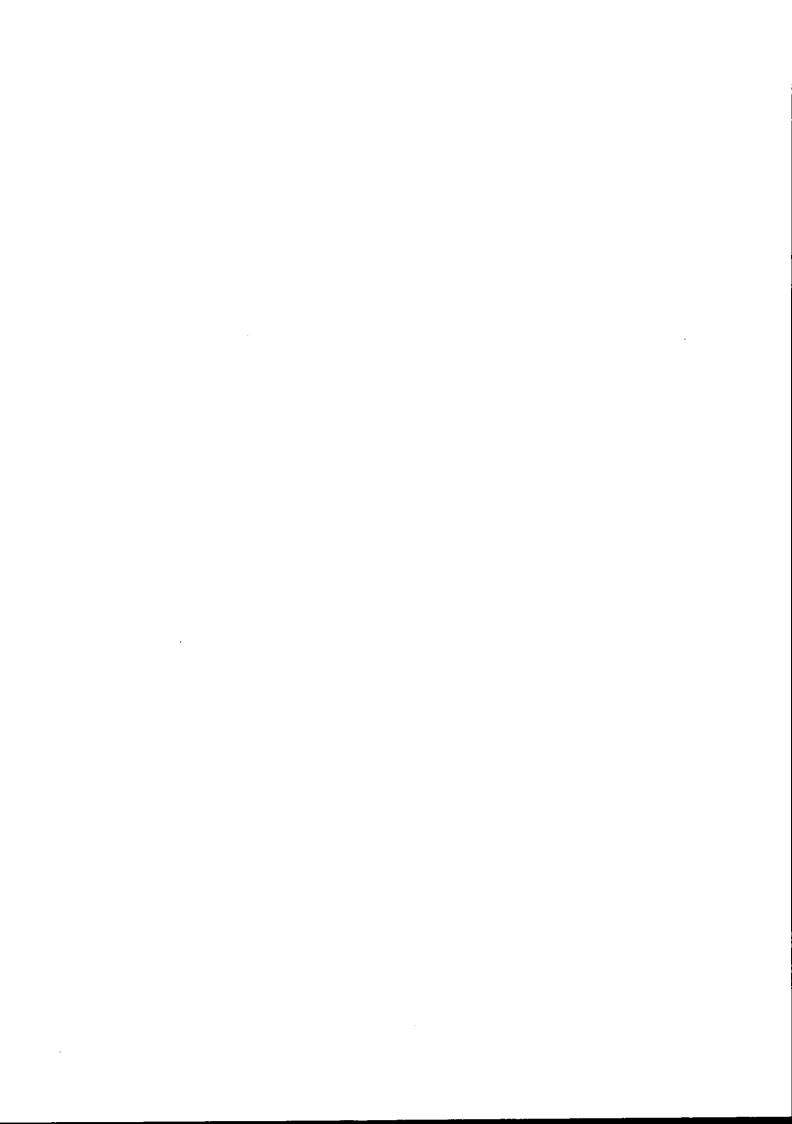
## מבוא לכריית נתונים 'מועד א יניב הדר

'תשע"ט סמסטר ב

אות	מחא	ון כ	דו –	עזר	חומר
2111	<b>1110</b>	14		יענו	11/4/11/4

השאלון מכיל 8 עמודים.
בהצלחה ! ====================================
<u>חומר עזר : נא סמן במשבצת המתאימה את המתאים</u> * X ניתן להשתמש בכל מחשבון (ללא יכולות תכנות) * לא ניתן להשתמש במחשבון Casio FX-991EX * <b>לא</b> ניתן להשתמש במחשבון
* לא ניתן להשתמש בחומר עזר X * מותר שימוש בדף נוסחאות, כמפורט: דף אחד משני הצדדים (סטודנטים מביאים עימם) * הבחינה בחומר פתוח – מותר להשתמש בכל חומר עזר מודפס או כתוב
<u>הערות</u> X יש לענות על <u>כל</u> השאלות במקומות המיועדים ע"ג טופס השאלון בלבד X יש להחזיר את השאלון ביחד עם הכריכה/מחברת.
אחר: 1.
.2 .3
השאלון מכיל 8 עמודים (כולל עמוד זה).
בהצלחה !

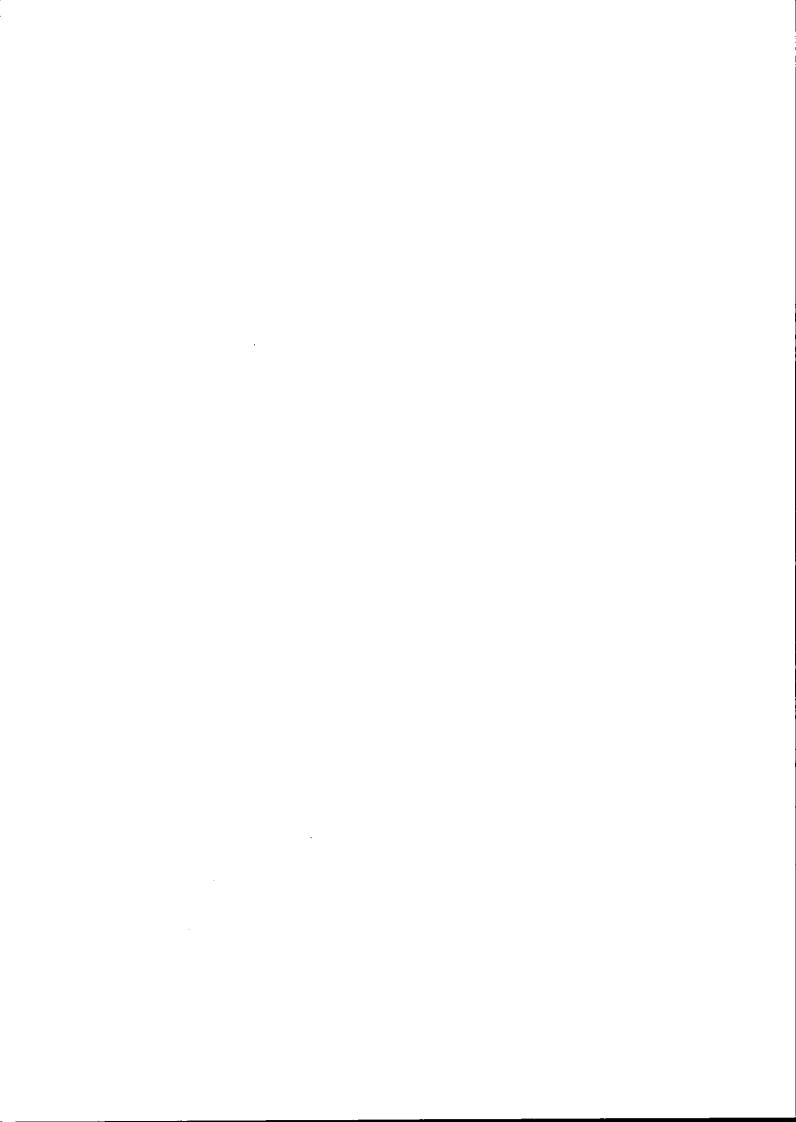
המכללה האקדמית להגדסה סמי שמעון



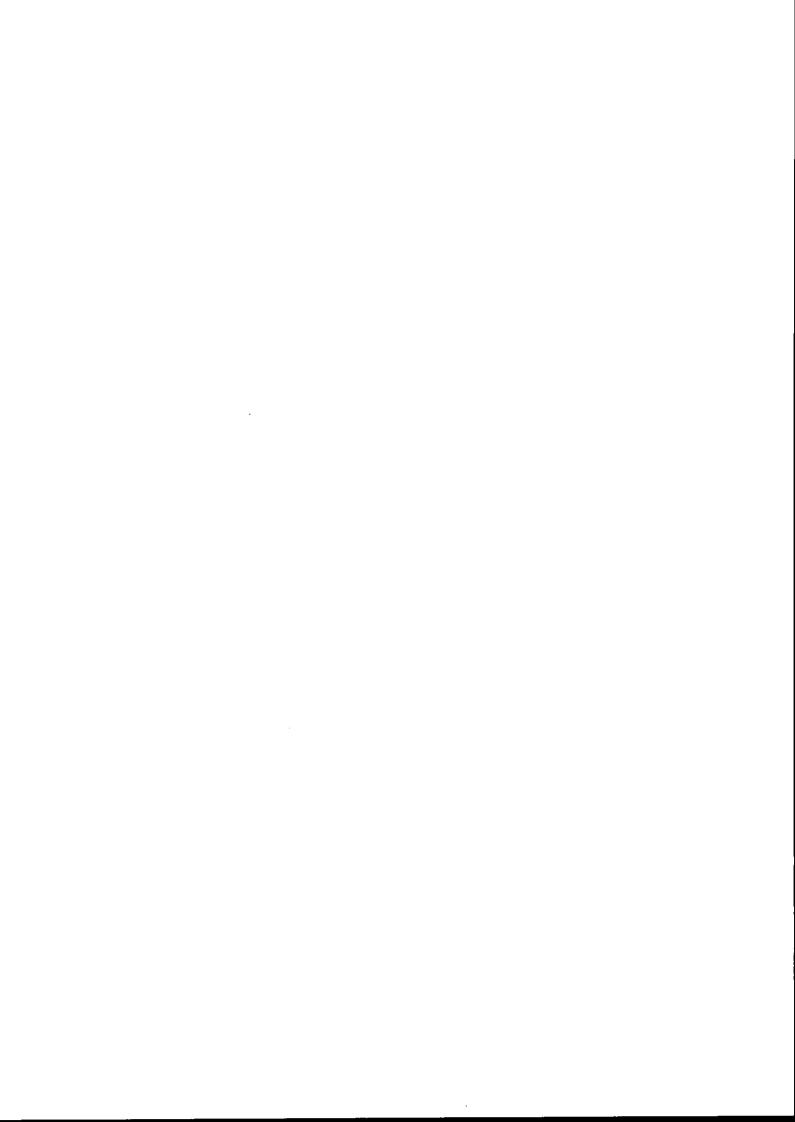
## חלק א' (40 גקודות, 4 גקודות לכל שאלה)

## יש לבחור תשובה אחת בלבר לכל שאלה ולסמן על גבי השאלון בלבר – תשובות במחברת לא ייבוקי.

- 1. אדם ביצע בדיקה לנוכחות מחלה מסוימת, הבדיקה יצאה ודובית והוא אכן ווולה. זוהי דוגמה ל:
  - Faise negative .x
  - False positive .a
  - True positive ...
  - True negative .7
  - 2. המרוזק בין שני אשכולות בשיטת Centroid מדשב עפ"י
    - א. המרוזק בין מרכזי האשכולות
  - ב. המרחק הארוך ביותר בין שני אובייקסים בשני האשכולות
    - ג. המרחק הממוצע בין שני האשכולות
  - ד. המרווק הקצר ביותר בין שני אובייקטים בשני האשסלות
    - 3. חלוקת ערכים לאינטרוולים לפי עומק שווה
    - א. מקסינה את ססיית התקן של הנתנים
    - ב. יוצרת אינסרוולים בעלי שכיוצת שוה
      - ב. יוצרת אינטרוולים בעלי טווה שווה
        - ד. תשובות א' ו-ב' נכתות
  - .4 עפ"י תורת האינפורמציה, אי וודאות האירוע שווה למינימום אם
    - א. כל העוצאת הן עוך דטרמיניסטי (קבוע) יווד
      - ב. התרטאות מתפלנות התפלנות מעריכית
      - ב התרצאות מהפלנות התפלנות נורמאלית
        - ד. התרצאות מתפלבות התפלבות אחדה
          - ? למה אלגוריתם Naïve Bayes מאדני .5
  - א. התושה שקיימת תלות וושה בין כל משתנה למשתנה היא נאיבית.
  - ב. הזבונה שלא קיימת תלות בין אף אחד מהמשתבים היא נאיבית.
    - ב. ההנחה שניתו לסכום את המשתנים היא סדבית.
  - ד. התנותו שקבוצת על של קבוצת פריטים שכיחים חייבת להיות שכיוה היא מייבית
    - 6. אלגוריתם מסוג eager לעומה אלגוריתם מסוג lazy:
    - א. eager שומר את כל נתוני האימון ו-lazy שומר חלק מנתוני האימון
    - ב. eager שומר הלק מנתוני האימון ו-lazy שומר את כל נתוני האימון
      - ג eager שומר את נתוני האימון ו-lazy שומר מחל
      - ד. eager שומר מחל ו-lazy שומר את נתוני האימון
- בית אריוה משניין להפריד בעורת צילום תמננות, בין תפחדים לבננות וללא סיווג ראשוני. מדובר במשימה של:
  - א. ניתוח אשכולות (Clustering)
    - (Classification) mo \_\_
      - (Prediction) ארוי ב
        - ד. למדה מתוית.



- 8. כדי לדעת מה הסיכויים שקבוצת פריטים אכן תידכש ביחד , המדד שצריך לחשב הוא:
  - confidence .x
  - mutual information .3
  - conditional probability .3
    - support .7
  - 9. בבעיית סיווג בינארי, דיוק האימון של מודל בעל אנטרופיה השווה ל-1 הוא:
    - 50% .א
    - 0% .⊐
    - د %001
    - ד. לא ניתן לדעת
    - 10. נחונה הפונקציה Y=2X+3, האנטרופיה של Y בהינחן X שווה ל:
      - 1 .x
      - د. 0.5
      - د 0
      - ד. לאניתן לושו



חלק ב' (60 נקידות)

בכל הסעיפים בחלק זה יש להראות את כל הדישובים הרלוונטיים במתברת, יש לרשום תשובה סופית על גבי השאלון בלבד, הניקוד הוא על תשובה סופית מדויקת בלבד, אולם תשובה ללא חישובים רלוונטיים במוברת תיפסל

וואולוגים מעוניינים לחזות האם נחש הוא ארסי. לשם כך, נאסף מידע מ 10 נחשים המכיל 5 משתנים מעמדים ומשתנה מטרה אחד – סיווב. הנתנים בטבלה:

רשומה	כנין	אמרך	צבע קשקשים עיקרי	בבע אחיד	ניבים	סיווג
1	male	24	yellow	TRUE	TRUE	yes
2	male	32	black	FALSE	FALSE	no
3	male	15	black	FALSE	TRUE	no
4	male	18	brown	TRUE	TRUE	yes
5	female	27	yellow	FALSE	FALSE	yes
6	male	9	yellow	TRUE	FALSE	yes
7	female	3	black	FALSE	TRUE	no
8	female	18	black	FALSE	FALSE	yes
9	male	22	black	TRUE	FALSE	no
10	male	21	brown	FALSE	FALSE	yes

א. (4 נק') יש לבצע נרמל בשיטת decimal scaling לאורך הווואים



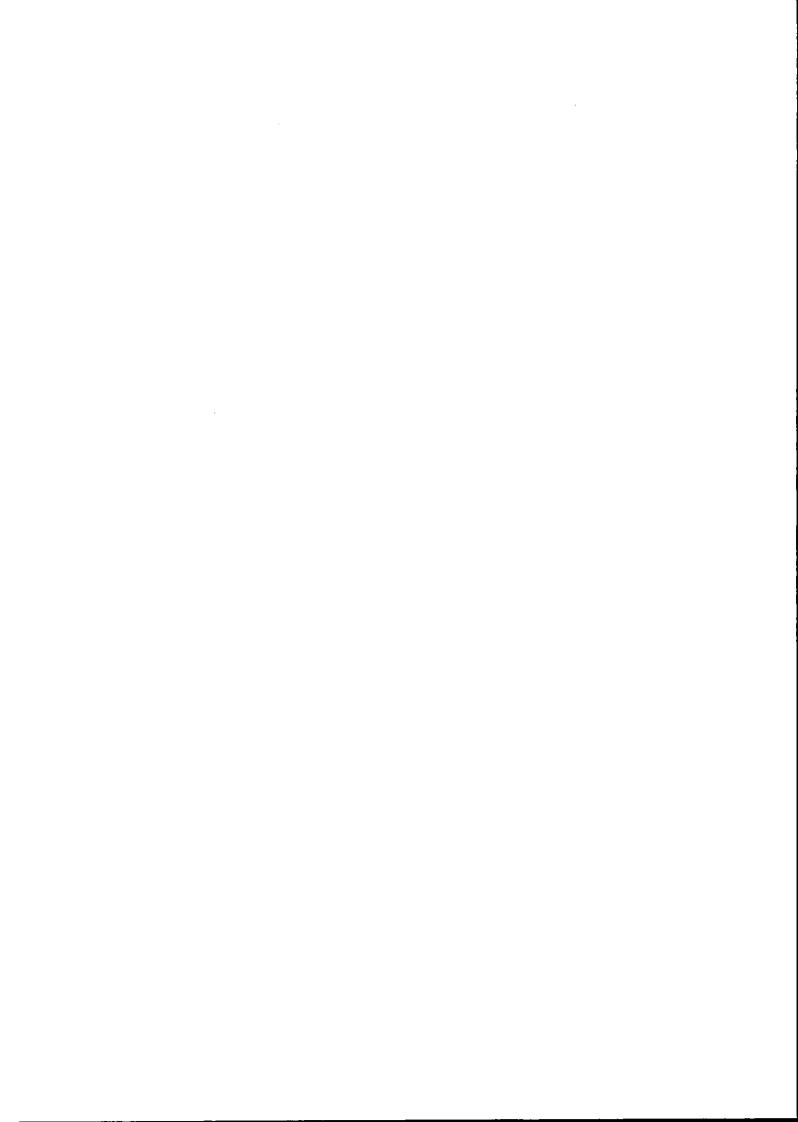
(4 נק') יש לבצע דיסקרטיוציה לשני טווזדם (BINS) בשיטת שומק שווה לערכי אודך הווזולים המגורמל.	٦.
(4 נק') מהו דיוק חוק הרוב בנחונים שלהלן?	
(6 נק') מהי האנטרופיה של משתנה המטרה סיווג?	.7
(6 נק') האם מין בערך male ו- צבע אחיד כערך FALSE הם חבנית שכיזה: (Minimum support = 30%)?	71
- Stantantin Sulpant - 50/6)	



עבור כל אות מותכונות	ו. ( 12 נקי) יש לחשב את מדד ה-Gint index) (Gint הבטרה – "סיוזג" "ציבים" ו – "צבע אחיד" ולהואיט לפי איזו תכונה עדיף לפצל את קדקוד השורש.
	•



ל אחת מהתכונות "צבע ש.	ו. (12 נק) יש לחשב את מדד ה— gain ratio של משתנה המטרה — <b>"סיווג"</b> עבור כ קשקשים" ו — "צבע אחיד" ולהחליט לפי איזו תכונה עדיף לפצל את קדקוד השורי	t



תבאה התצפית לסווג את (K nearest neighbors) אוא התצפית הבאה את נק") יש להשתמש באלגוריתם (K nearest neighbors) (K=1):

רשומה	מין	THE	צבע קשקשים עיקרי	צבע אוזיד	ניבים	סיווג
1	female	50	yellow	FALSE	TRUE	???

על מנת לחשב מרחק בין תצפיות יש להשתמש במרחק מנהגון כאשר המרחקים מוגדרים באופן הבא:

- עבור המשתנים מין, ניבים וצבע אחיד המרחק בין ערכים שונים הוא 1.
- עבור המשחנה אורך יש להשחמש במשחנה לאזור הדיסקרטיוציה (מסעיף ב') המרחק בין
  אינטרוולים שונים הוא 1.
  - עבור המשתנה "צבע קשקשים עיקרי" יש להשתמש בטכלת המרחקים הבאת:

צבע קשקשים עיקרי	brown	yellow	black
brown	0		
yellow	1	0	
black	2	1	0