

חלק 2 [50 נקודות]

- יש להציג את כל התוצאות עם שלוש ספרות אחרי נקודה עשרונית אלא אם צוין אחרת!
- יש לרשום את כל התשובות על-גבי שאלון הבחינה בלבד
- טיוטות החישוב ייגרסו ללא בדיקה
- להלן נתונים של צריכת חשמל שעתית במשך יממה אחת:

Hour	Tcons-1	Tcons-2	Tcons-1-2	Tcons-24	Class
00:00	46	61	-15	36	Low
01:00	40	46	-6	31	Low
02:00	33	40	-7	62	Low
03:00	34	33	1	23	Low
04:00	29	34	-5	33	Low
05:00	38	29	9	38	Low
06:00	39	38	1	51	High
07:00	115	39	76	59	High
08:00	59	115	-56	52	High
09:00	94	59	35	58	High
10:00	58	94	-36	46	Low
11:00	38	58	-20	49	High
12:00	101	38	63	48	Low
13:00	48	101	-53	50	High
14:00	105	48	57	52	Low
15:00	52	105	-53	45	High
16:00	58	52	6	64	High
17:00	68	58	10	80	High
18:00	78	68	10	155	High
19:00	73	78	-5	65	High
20:00	67	73	-6	68	High
21:00	76	67	9	65	High
22:00	73	76	-3	61	High
23:00	59	73	-14	46	Low

- העמודה Tcons-1 מייצגת את הצריכה השעתית בשעה הקודמת
- העמודה Tcons-2 מייצגת את הצריכה השעתית בשעה שלפני השעה הקודמת
- העמודה Tcons-1-2 מייצגת את ההפרש בין הצריכות השעתיות בשעתיים הקודמות
- העמודה Tcons-24 מייצגת את הצריכה השעתית באותה שעה ביום הקודם
- העמודה Class מייצגת את הצריכה השעתית החזויה (נמוכה / גבוהה)
- כל השאלות הבאות מתייחסות לטבלה הנ"ל

נמצא
!!

א. יש לבצע דיסקרטיזציה של המשתנה **Tcons-1-2** לשני אינטרוולים בעלי עומק שווה

(equal depth). 5 נק'

Min	Max	Interval Code
		0
		1

Hour	Tcons-1-2	Code
00:00	-15	
01:00	-6	
02:00	-7	
03:00	1	
04:00	-5	
05:00	9	
06:00	1	
07:00	76	
08:00	-56	
09:00	35	
10:00	-36	
11:00	-20	

Hour	Tcons-1-2	Code
12:00	63	
13:00	-53	
14:00	57	
15:00	-53	
16:00	6	
17:00	10	
18:00	10	
19:00	-5	
20:00	-6	
21:00	9	
22:00	-3	
23:00	-14	

ב. יש לבצע דיסקרטיזציה של המשתנה **Tcons-24** לשלושה אינטרוולים בעלי רוחב שווה

(equal width). 5 נק'

Min	Max	Interval Code
		0
		1
		2

Hour	Tcons-24	Code
00:00	36	
01:00	31	
02:00	62	
03:00	23	
04:00	33	
05:00	38	
06:00	51	
07:00	59	
08:00	52	
09:00	58	
10:00	46	
11:00	49	

Hour	Tcons-24	Code
12:00	48	
13:00	50	
14:00	52	
15:00	45	
16:00	64	
17:00	80	
18:00	155	
19:00	65	
20:00	68	
21:00	65	
22:00	61	
23:00	46	

טבעה !!

ג. יש לחשב את האנטרופיה המותנית, מדד ה-Gini (Gini Split) ומדד ה-Twoing של משתנה המטרה Class ביחס למשתני הקלט Tcons-1 ו-Tcons-2 שעברו דיסקרטיזציה וקיבלו שמות Code-1 ו-Code-2 בהתאמה. כמו כן, יש לציין את הערך האופטימאלי של שני המדדים הראשונים (אנטרופיה מותנית ו-Gini) ולבחור במשתנה הטוב ביותר לפיצול ביחס לכל מדד. **20 נקודות**

נא להציג את החישובים בטבלה הבאה (אין למלא משבצות בצבע אפור):

Hour	Class	Code-1	Code-2
00:00	Low	0	1
01:00	Low	0	0
02:00	Low	0	0
03:00	Low	0	0
04:00	Low	0	0
05:00	Low	0	0
06:00	High	0	0
07:00	High	1	0
08:00	High	0	2
09:00	High	1	1
10:00	Low	0	2
11:00	High	0	1
12:00	Low	1	0
13:00	High	0	2
14:00	Low	1	0
15:00	High	0	2
16:00	High	0	1
17:00	High	1	1
18:00	High	1	1
19:00	High	1	2
20:00	High	1	2
21:00	High	1	1
22:00	High	1	2
23:00	Low	0	2

	Code-1	Code-2
Count (0)		
Count (1)		
Count (2)		
Prob (0)		
Prob (1)		
Prob (2)		
Count (Low; 0)		
Prob (Low/0)		
Count (High; 0)		
Prob (High/0)		
Ent (0)		
Gini (0)		
Count (Low; 1)		
Prob (Low/1)		
Count (High; 1)		
Prob (High/1)		
Ent (1)		
Gini (1)		
Count (Low; 2)		
Prob (Low/2)		
Count (High; 2)		
Prob (High/2)		
Ent (2)		
Gini (2)		
Cond. Entropy		
Gini Split		
Abs (0)		
Abs (1)		
Twoing		

Best Conditional Entropy: _____

Selected attribute: _____

Best Gini Split: _____

Selected attribute: _____



ד. בהנחה שמשנתה המטרה (Class) מייצג השתייכות של כל תצפית לאשכול (cluster) יש לחשב את מרכז

הכובד (centroid) של כל אחד משני

האשכולות. 5 נקודות.

	Tcons-1	Tcons-2	Tcons-24
Centroid Low			
Centroid High			

ה. בהינתן שני מרכזים התחלתיים של שני האשכולות Low ו-High (שאינם קשורים למרכזים שחושבו בסעיף הקודם), יש לבצע איטרציה אחת של האלגוריתם K-MEANS. יש להשתמש במרחק האוקלידי. 15 נקודות.

	Tcons-1	Tcons-2	Tcons-24
Centroid Low	50	45	35
Centroid High	70	75	65

Hour	Tcons-1	Tcons-2	Tcons-24	Class	Distance to Low	Distance to High	New class	Change (Yes / No)
00:00	46	61	36	Low				
01:00	40	46	31	Low				
02:00	33	40	62	Low				
03:00	34	33	23	Low				
04:00	29	34	33	Low				
05:00	38	29	38	Low				
06:00	39	38	51	High				
07:00	115	39	59	High				
08:00	59	115	52	High				
09:00	94	59	58	High				
10:00	58	94	46	Low				
11:00	38	58	49	High				
12:00	101	38	48	Low				
13:00	48	101	50	High				
14:00	105	48	52	Low				
15:00	52	105	45	High				
16:00	58	52	64	High				
17:00	68	58	80	High				
18:00	78	68	155	High				
19:00	73	78	65	High				
20:00	67	73	68	High				
21:00	76	67	65	High				
22:00	73	76	61	High				
23:00	59	73	46	Low				