



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
- נתונה טבלה של צריכת חשמל ביום מסוים (S0) כתלות בצריכת חשמל בחמשת הימים הקודמים. הצריכה ביום הקודם מסומנת ב-"S-1", הצריכה לפני יומיים מסומנת ב-"S-2" וכו'. הסיווג "1"- מייצג צריכה נמוכה בשעה שהסיווג "1" מייצג צריכה גבוהה.

i	S-1	S-2	S-3	S-4	S-5	S0
1	6.26	8.67	9.03	8.80	8.64	-1
2	5.40	6.26	8.67	9.03	8.80	1
3	8.78	5.40	6.26	8.67	9.03	1
4	8.68	8.78	5.40	6.26	8.67	1
5	8.70	8.68	8.78	5.40	6.26	1
6	8.77	8.70	8.68	8.78	5.40	1
7	8.52	8.77	8.70	8.68	8.78	-1
8	8.28	8.52	8.77	8.70	8.68	1
9	8.91	8.28	8.52	8.77	8.70	1
10	8.55	8.91	8.28	8.52	8.77	-1
11	5.92	8.55	8.91	8.28	8.52	-1
12	5.42	5.92	8.55	8.91	8.28	-1
13	6.39	5.42	5.92	8.55	8.91	-1
14	6.56	6.39	5.42	5.92	8.55	-1
15	6.60	6.56	6.39	5.42	5.92	-1
16	6.12	6.60	6.56	6.39	5.42	-1
17	5.65	6.12	6.60	6.56	6.39	1
18	9.05	5.65	6.12	6.60	6.56	1
19	9.11	9.05	5.65	6.12	6.60	1
20	8.81	9.11	9.05	5.65	6.12	1

יש לענות על כל הסעיפים הבאים :

- א. כל אחד מחמשת משתני הקלט חולק לשני אינטרוואלים עפ"י נקודת פיצול מסוימת (Threshold). יש לציין בטבלה הבאה את שיטת החלוקה המתייחסת לנקודת הפיצול של כל משתנה (רוחב שווה / עומק שווה). 5 נקודות

	S-1	S-2	S-3	S-4	S-5
Threshold	7.25	8.28	8.28	7.21	8.52
Discretization method					

ב. יש לחשב את האנטרופיה המותנית, מדד ה-Gini (Gini Split) ומדד ה-twoing של משתנה המטרה S0 ביחס לשני משתני הקלט הראשונים עבור נקודות הפיצול שצוינו בסעיף הקודם. כמו כן, יש לציין את הערך האופטימאלי של כל מדד ולבחור במשתנה הטוב ביותר לפיצול ביחס לכל מדד. **20 נקודות**

אין למלא משבצות בעלות רקע אפור!

	S-1	S-2	Optimal value	The best
Threshold	7.25	8.28		
Prob (\leq Threshold)				
Prob ($-1 / \leq$ Threshold)				
Entropy (\leq Threshold)				
Gini (\leq Threshold)				
Prob ($-1 / >$ Threshold)				
Entropy ($>$ Threshold)				
Gini ($>$ Threshold)				
Conditional entropy				
Gini Split				
Twoing				

ג. יש לחשב את המשקולות של משתני הקלט במודל Linear SVM בעזרת מקדמי α המופיעים בטבלה הנתונה בסעיף הבא. **10 נקודות**

	S-1	S-2	S-3	S-4	S-5
w_j					

ד. יש למלא את העמודות הריקות הבאות בטבלה הבאה (**10 נקודות**):

(1) וקטור תמיכה (כן / לא)

(2) הפונקציה $w^T x_i + b$

(3) הסיווג החזוי (כן / לא). יש להניח שהמקדם $b = -3.950$

4) שגיאה (כן / לא)

i	Actual S0	alpha	Support vector (Yes/No)	$w^T x_i + b$	Predicted S0	Error
1	-1	99.27				
2	1	269.21				
3	1	1.04				
4	1	189.00				
5	1	16.87				
6	1	36.58				
7	-1	183.96				
8	1	127.11				
9	1	67.37				
10	-1	166.71				
11	-1	0.00				
12	-1	76.92				
13	-1	94.99				
14	-1	97.51				
15	-1	142.70				
16	-1	96.47				
17	1	202.54				
18	1	3.94				
19	1	32.55				
20	1	12.32				

ה. יש לחשב את דיוק האימון באחוזים של המודל Linear SVM ואת הרווח בר-סמך שלו ברמת-ביטחון של 95%. **5 נקודות.**

Accuracy (%)	Lower Bound (%)	Upper Bound (%)