

למידת מכונה

מבחן סופי

מועד ב' סמסטר ב

תש"ף

מרצה: פרופסור אילן שמשוני
 מתרגל: דור דביר

משך הבחינה שבעים דקות.
 ניתן להשתמש בכל חומר עזר.
 יש לענות על כל שש השאלות במחברת המצורפת.
 שימו לב כי בידך 3 דפים כולל דף זה.
נא לכתוב בצורה מסודרת וברורה בקובץ Word (או על נייר).

בהצלחה!

שאלה מס'	ניקוד מקסימאלי	ניקוד
1. אשכול	25	
2. LSH	25	
3. Loss Function	20	
4. Bagging+	20	
5. תיאוריה	10	
6. תאוריה	10	

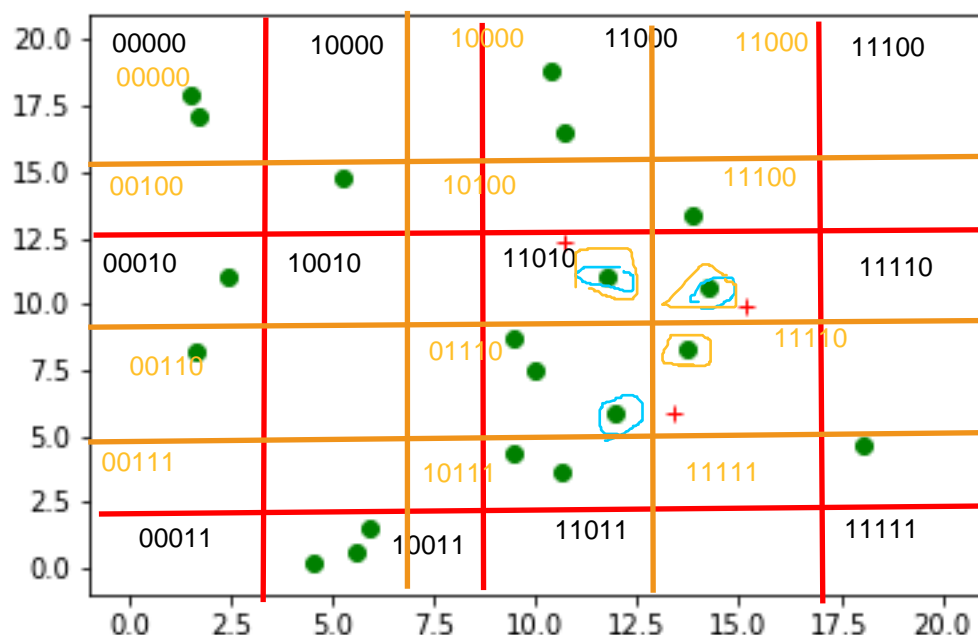
1. אשכול היררכי Agglomerative or Hierarchical Clustering (25 נקודות)
 לחמשת הנקודות הבאות נתונה טבלת המרחקים הבאה.

	1	2	3	4	5
1	0				
2	2	0			
3	8	9	0		
4	10	7	5	0	
5	11	3	6	9	0

- (א) הסבירו את המושגים Single link ו Complete link
 (ב) בנו את האשכול לפי Single Link ו Complete Link
 (ג) בשני המקרים תנו את התוצאות כשיש שני clusters ושלושה clusters.
 (ד) אם המרחק המקסימלי המותר בין clusters הוא 5 מה תצא תוצאת האשכול בשני המקרים.

2. Locality Sensitive Hashing (LSH) (25 נקודות)

נתונים הנתונים הבאים. הנקודות המסומנות ב O משמשות לבניית מבנה הנתונים. יש גם שלוש נקודות שמסומנות ב + שהן נקודות לשאילתות ל LSH.



- (א) בנו את מבנה הנתונים לפי $L=2$ ו $K=5$.

- (ב) הפעילו את האלגוריתם על שלושת הנקודות שסומנו ב + ודווח על התוצאות.
- (ג) האם התוצאות שהתקבלו היו נכונות? אם לא אמרו מהי התשובה הנכונה.
- (ד) מצאו דוגמא נוספת (X, Y) שלה התשובה לא נכונה. הסבירו איך בחרתם אותה.
- (ה) איך הערכים של K ו L משפיעים על הדיוק וזמן הריצה של האלגוריתם.

3. Loss Function (20 נקודות)

הרבה מהאלגוריתמים שלמדנו מנסים למזער פונקצית מחיר loss function.

לכל אחד מהאלגוריתמים הבאים ענו על השאלות הבאות:

- (א) מהי פונקצית המחיר שהיא מנסה למזער. ניתן לתת נוסחא או הסבר.
- (ב) האם האלגוריתם מוצא מינימום לוקאלי, מינימום גלובלי, או משהו אחר. הסבירו.

האלגוריתמים הם:

- (1) SVM
- (2) K means
- (3) t-SNE
- (4) עץ החלטה.
- (5) SVD

4. Bagging, AdaBoost, Random Forest (20 נקודות)

- (א) מה המשותף ל Bagging, AdaBoost ו Random Forest. לאיזה קבוצת אלגוריתמים הם שייכים?
- (ב) מצאו איך אחד האלגוריתמים יוצא דופן יחסית לשניים האחרים. הסבירו.
- (ג) מצאו איך אחד האלגוריתמים יוצא דופן יחסית לשניים האחרים. הסבירו. (שוב).

5. תיאוריה (10 נקודות)

הסבירו את משמעות המושגים False Positive ו False Negative. לכל מושג תנו דוגמא שתדגימו בה את הסכנה המעשית שקורית כשהמקרים האלה קורים.

6. תיאוריה (10 נקודות)

בניתם Random Forest עם 1000 עצים. שגיאת הסיווג על ה training data יצא 0.0. לעומת זאת שגיאת הסיווג על ה validation data יצא 34.22. מה עשיתם לא נכון. איך ניתן לשפר את המודל שבניתם (אין הכוונה לעבור למסווג מסוג אחר).