

למידת מכונה

מבחן סופי

מועד א' סמסטר ב'

תש"ף

מרצה: פרופסור אילן שמשוני
 מתרגל: דור דביר

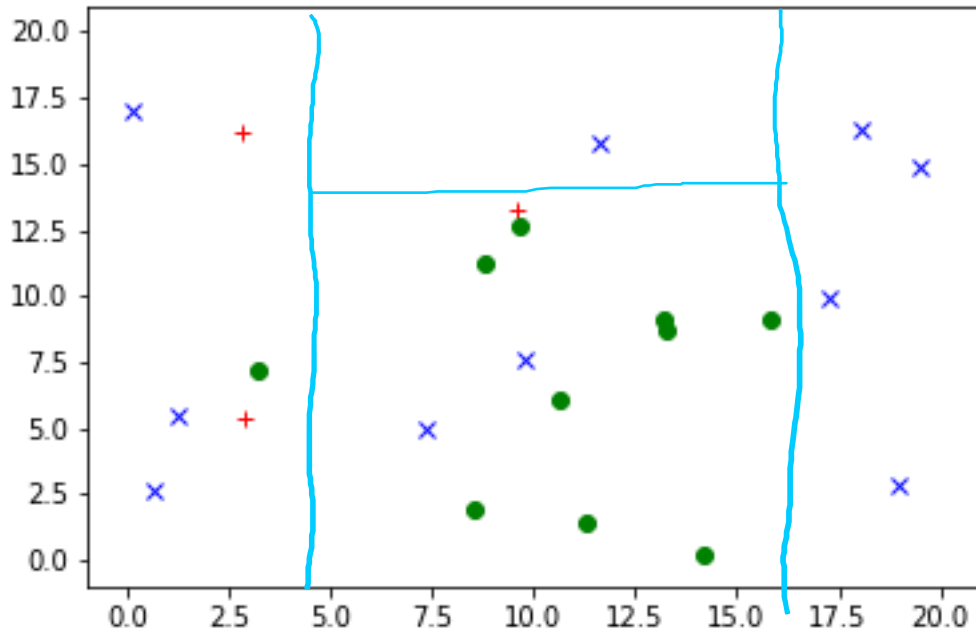
משך הבחינה שבעים דקות.
 ניתן להשתמש בכל חומר עזר.
 יש לענות על כל **שש** השאלות **במחברת המצורפת**.
 שים לב כי בידך 4 דפים כולל דף זה.
 הבחינה מיועדת לגברים ונשים כאחד ומנוסחת בלשון זכר מטעמי נוחות בלבד.
נא לכתוב בצורה מסודרת וברורה בקובץ Word (או על נייר).

בהצלחה!

שאלה מס'	ניקוד מקסימאלי	ניקוד
1. עצי החלטה	25	
2. RKDTree	25	
3. SVM	20	
4. מטריצת בלבול	20	
5. שאלת חשיבה	10	
6. תאוריה	10	

1. עצי החלטה (25 נקודות)

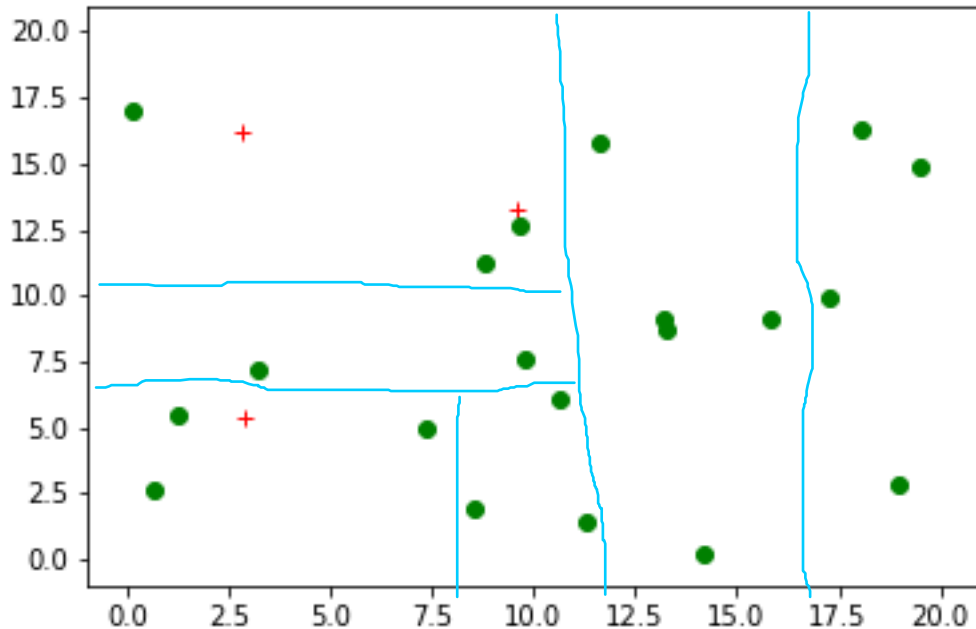
נתונים הנתונים הבאים. חלק שייכים למחלקה X השאר למחלקה O. יש גם שלוש נקודות שמסומנות ב + שהן נקודות לסיווג.



- א) בנה עץ החלטה בעומק 3 (עד 7 צמתים בעץ). נסה שהפתרון יהיה קרוב לעץ שהיה מחושב על ידי אלגוריתם (לא חייבים לנסות את כל האפשרויות).
- ב) לכל צומת חשב את ערכי הג'יני ואת אחוז הנקודות בשתי המחלקות.
- ג) סווג את שלושת הנקודות לסיווג. לכל נקודה חשב את המחלקה ואת ההסתברות להיות באותה מחלקה.

2. Randomized KD trees (25 נקודות)

נתונים הנתונים הבאים. הנקודות המסומנות ב O משמשות לבניית מבנה הנתונים. יש גם שלוש נקודות שמסומנות ב + שהן נקודות לשאלות ל Randomized KD Tree.



- (א) למה משמש RKDTree?
- (ב) בנה את מבנה הנתונים כך שמספר האלמנטים בעלה הוא לכל היותר 5. בנה עץ אחד בלבד.
- (ג) הפעל את האלגוריתם על שלושת הנקודות שסומנו ב + ודווח על התוצאות.
- (ד) האם התוצאות שהתקבלו היו נכונות?
- (ה) האם ב RKDTree ניתן להבטיח שהתוצאות תהיינה תמיד נכונות. אם לא אז מה ניתן לעשות כדי לשפר את ההסתברות שהן תהיינה נכונות ובאיזה מחיר.

3. SVM (20 נקודות)

נתונה הנוסחה הבאה:

$$\frac{1}{2} \langle \mathbf{w}^T, \mathbf{w} \rangle + C \sum \xi_i$$

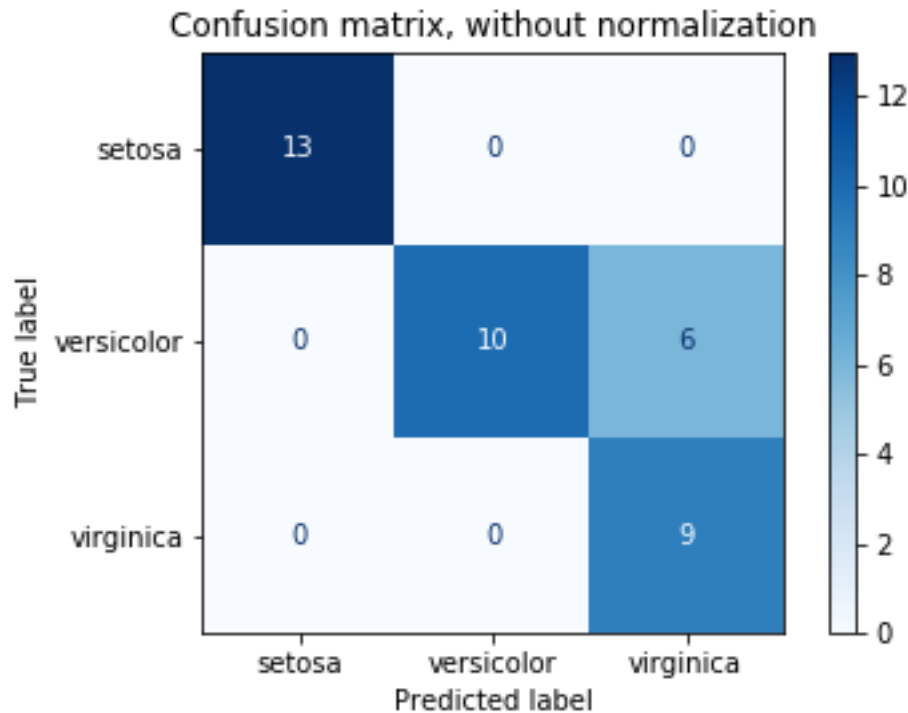
Such that:

$$y_i (\langle \mathbf{w}^T, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0 \text{ for all } i$$

- (א) מה אלגוריתם ה SVM אמור לעשות עם הנוסחה הזאת ומה להחזיר.
- (ב) מה משמעות כל האותיות בנוסחה (משפט לכל אחד).
- (ג) האם ישנם משתנים שלמשתמש יש שליטה על ערכיהם ואם יש איך שינוי הערכים שלהם משפיע על התוצאה.
- (ד) מה משמעות הביטוי $\langle \mathbf{w}^T, \mathbf{x}_i \rangle$ בנוסחה ומדוע יש לזה חשיבות גדולה מאוד ב SVM.

4. מטריצת בלבול (20 נקודות)

הרצתם אלגוריתם סיווג כלשהו וקיבלתם את מטריצת הבלבול הבאה.



- (א) מה משמעות מטריצת בלבול. מה היא מראה.
- (ב) חשב עבור המטריצה הנתונה את הדיוק של המסווג ואת ערכי ה false positive ו false negative.
- (ג) מהסתכלות במטריצה באיזה כיוון היית מנסה לשפר את אלגוריתם הסיווג שלך.

5. שאלת חשיבה (10 נקודות)

תכננת מסווג שבהינתן מידע רפואי מסווג את החולה כחולה סרטן או בריא. המסווג שלך מגיע לדיוק של 96%. למה אתה אמור לא להיות מרוצה מהתוצאות שקיבלת. מה אפשר לעשות בעניין.

6. תיאוריה (10 נקודות)

קיבלת פרויקט לבניית מסווג. ניסית מספר גדול של שיטות וקיבלת מודלים לכל אחת מהן. לפי איזה קריטריונים תבחר את המסווג שבו תשתמש. איך זה תלוי באפליקציה שבו יופעל המסווג.