חלק א : אמת או שקר (12 נקודות)

1. שקר. אם היינו דורשים ש- $\mathcal{H}_n$ ניתנת ללמידת PAC יעילה **ונאותה** (Proper), הטענה הייתה נכונה. מאחר ואיננו דורשים זאת, יתכן שעדיין ניתן לבצע למידת PAC באמצעות מחלקת היפותזות יותר עשירה. דוגמה קלאסית לכך היא נוסחאות 3-term-DNF (ראו ספר הקורס, פרק 8 סעיפים 8.2.4 ו-8.3 להסבר מדוע זוהי דוגמה רלוונטית). התקבלו גם תשובות סבירות על הפער בין למידת CONSISTENT ל-PAC (במקרה ה-non realizable).
2. שקר. במקרה ה-non realizable זהו לא תנאי מספיק על מנת להבטיח למידת PAC, מכיוון שמימד ה- VC של $\mathcal{H}_n$ יכול לגדול באופן super-polynomial (כפונקציה של n).
3. אמת. מאחר והמידע פריד ליניארית, אנו יודעים כי אלגוריתם Adaboost יגיע לשגיאת אימון 0. הרצה של איטרציות נוספות אחר כך תוביל להגדלת $\mathrm{L}_1\ margin$ כפי שהוזכר בכיתה, ולכן אחרי מספיק איטרציות נקבל מסווג זהה.
4. אמת.

חלק ב: זוגות של קטעים (40 נקודות)

1. מימד ה-VC הוא 4.
2. דוגמה לתיוג של 5 נקודות שלא ניתן להשיג: $+ - + - +$
3.
   a. 4 (פולינומים מדרגה 3)
   b. 5 (פולינומים מדרגה 4)
4.
   a. לא, מרחב ההיפותזות של אבי לא מכיל את $\mathcal{H}$
   b. לא, אם לא ניתן לבצע למידת PAC, ברור שגם לא ניתן בפרט לבצע למידת PAC נאותה.
   c. כן, מרחב ההיפותזות של בסאם מסוגל לתייג כל קבוצת אימון באופן מושלם (מרחב ההיפותזות שלו יותר עשיר) ולכן ניתן לבצע למידת PAC.
   d. לא, הסימן של פולינומים ממעלה 4 נכיל היפותזות שלא שייכות ל-$\mathcal{H}$.
   e. לא, אין אפשרות אפילו ללמידת PAC במקרה הלא אגנוסטי, לכן גם במקרה האגנוסטי (הקשה יותר) אין הדבר אפשרי.
   f. כן. כי מצד אחד ה- hinge-loss הוא חסם עליון לשגיאת ה- 0/1, ומצד שני לכל קבוצה סופית של דוגמאות ולכל היפותזה h ב- $\mathcal{H}$ ניתן למצוא פולינום ממעלה 4 ששגיאת ה- hinge-loss שלו ביחס לאוסף הדוגמאות שווה בדיוק לשגיאת ה- 0/1 ביחס ל- h
5. כאשר משתמשים ב-POLYSQUAREFIT, התשובה ל- f ,c משתנה.
6. `Z[i,:]` הינו וקטור של החזקות של x_i, כלומר $(1, x_i, x_i^2, \dots, x_i^r)$. זהו האלמנט העיקרי של שני האלגוריתמים. היתר הוא להגדיר רגרסיה\בעיית מזעור hinge loss כפי שראינו מספר פעמים במהלך הקורס.
7. ???????? התשובה לאף אחד מהסעיפים אינה משתנה.
8. הדבר אפשרי. אלגוריתם טריוויאלי (אך מספיק לקבלת ניקוד מלא) הוא לעבור על רביעיות הנקודות החוקיות ולבדוק את השגיאה לכל רביעייה (יש $O(n^4)$ רביעיות וכל בדיקה היא $O(n)$, סה"כ $O(n^5)$). באמצעות תכנון דינאמי ניתן לפתור בצורה יותר יעילה. גם תשובה שטוענת שמימד ה- VC של $\mathcal{H}$ הוא קבוע (ולכן ה- growth function פולינומיאלי) זוכה לכל הנקודות.

9. שגיאת אימון\סיווג: גרג ובסאם יגיעו תמיד לשגיאה 0, ואבי לא בהכרח מסוגל לכך. לכן: אבי>בסאם=גרג
שגיאת קירוב: שגיאת הקירוב תלויה בגודל מרחב ההיפותזות. גרג משתמש במרחב $\mathcal{H}$ עצמו, לכן שגיאת הקירוב שלו 0. בסאם משתמש במרחב גדול יותר, לכן שגיאת הקירוב שלו גם כן 0. אבי משתמש במרחב שלא מכיל את $\mathcal{H}$ ולכן שגיאת הקירוב שלו יכולה להיות גדולה מ- 0. לכן אבי>בסאם=גרג
שגיאת שיערוך: שגיאת השיערוך (על אותו מדגם) יכולה רק לרדת ככל שמרחב ההיפותזות קטן יותר. לכן בסאם >= גרג >= אבי

חלק ג: חיזוי של סוכרת (30 נקודות)

1. יש לתקן את שני חלקי השגיאה לפי השכיחות היחסית באוכלוסיה:

$$\tilde{L}(h) = L_{S^-}^{01}(h) \cdot \frac{91}{100} + L_{S^+}^{01}(h) \cdot \frac{9}{100}$$
$$= \frac{1}{10000}\sum_{i=1}^{10000} [\![h(x_i) \neq -1]\!] \cdot \frac{91}{100} + \frac{1}{10000}\sum_{i=10001}^{20000} [\![h(x_i) \neq +1]\!] \cdot \frac{9}{100}$$

2.

   a. זוהי שיטה דיסקרימינטיבית, שכן איננו מניחים התפלגות על x בהינתן y.
   b. מימד ה-VC הוא 2^{20}. מכאן שלא נצליח ללמוד היפותזה טובה עם 20,000 דוגמאות בלבד. שימו לב שאין כאן מכשול חישובי בכלל מכיוון שהפונקציה h האופטימלית (במובן של הסעיף) פשוט קובעת את הערך של $h(x)$ לפי הערכים y_i של הדוגמאות x_i ששוות ל- x, וערך שרירותי אם אין דוגמאות כאלה.

3.

   a. זוהי גם כן שיטה דיסקרימינטיבית.
   b. מספר הפרמטרים הוא רק 20.

4. יש לשערך פרמטר לכל תלות אפשרית בין המשתנים- $2^{20}-1$.

5.

   a. השיטה היא גנרטיבית, הפלט ניתן לביטוי כמסווג לינארי וכן הגישה מניחה אי תלות בהינתן סיווג, כלומר שבהינתן הארוע "האדם יסבול מסוכרת בגיל 60", שני הארועים "x[1]=1" ו- "x[2]=0" הינם בלתי-תלויים. כל שאר האפשרויות הן שקר.
   b. יש לשערך את התפלגות כל משתנה בהינתן כל תיוג אפשרי, כלומר $2 \cdot 20 = 40$ משתנים.
   c.

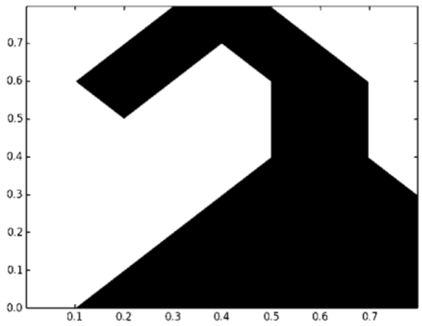
$$\theta_{1,+} = \frac{1}{10000}\sum_{i=10001}^{20000} [\![h(x_i) \neq +1]\!]$$

התקבלו גם תשובות שהגדירו נכונה את בעיית ה-argmax אך לא נתנו פתרון מפורש.

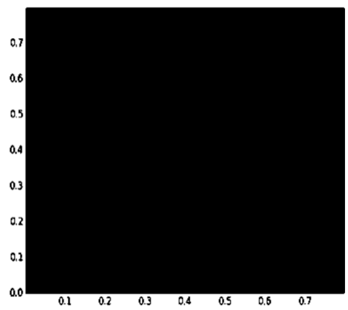
חלק ד: גרעינים (Kernels) (6 נקודות)

1. כן, חיבור של גרעינים הוא גרעין חוקי.
2. כן, כפל של גרעין בקבוע חיובי הוא גרעין, וחיבור גרעינים מהווה גרעין.
3. לא, חיסור אינו בהכרח משמר את תכונת הגרעין.
4. כן, הדבר שקול לאיפוס 20 המאפיינים הראשונים במיפוי $\phi(x)$, אך עדיין נקבל גרעין חוקי עקב שאר המאפיינים.
5. לא, לקיחת הגרעין ב-4 בסימן שלילי לא בהכרח משמרת את תכונת הגרעין.

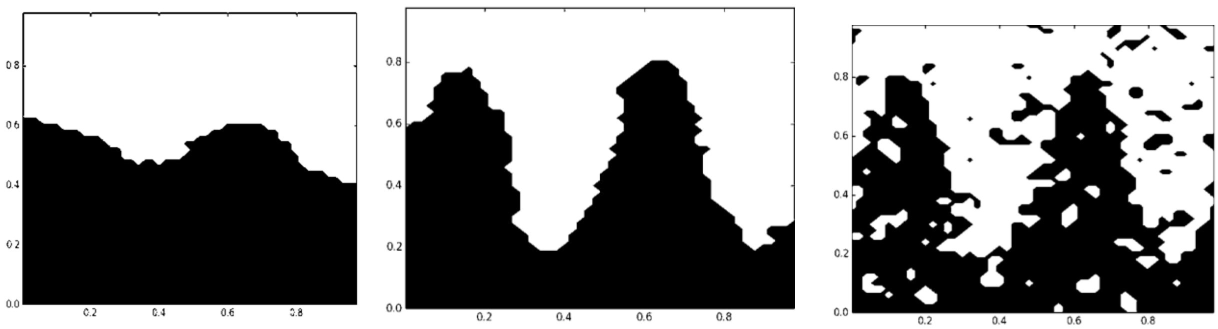
חלק ה: שכנים קרובים (Nearest Neighbors) (12 נקודות)



תמונה זו מתאימה למקרה A. ניתן לשים לב שזוהי דיאגרמת ורנוי (Voronoi diagram) עם קצוות חדים כפי שנצפה עבור $k = 1$ ומספר נקודות קטן.



תמונה זו מתאימה למקרה B. כאשר מספר הנקודות זהה ל-k, הרי שכל הנקודות יקבלו את אותו התיוג- תיוג הרוב.



התמונות הללו מתארות (בהתאמה, מימין לשמאל) את מקרים E ,D ,C.

כאשר k קטן מדי, ישנו רעש רב. הגדלת k מסלקת את הרעשים מגבול ההחלטה ומחליקה את התוצאה. כאשר k גדול מדי (במקרה שלנו $k = 200$), נקבל פונקציה חלקה מדי המאבדת חלק ניכר מהצורה הרצויה.