

1. [10] השתמשו בספליינים של לוח דק בדרגות גמישות שונות כדי לבצע למידה מקבוצת אימון בדקו את שגיאת ה MSE על קבוצת האימון ואת שגיאת ההכללה (בעזרת Cross Validation) בדרגות גמישות שונות ויצרו כך את הגרף שבתחתית השאלה.

a. האם לדעתך מדובר בקלסיפיקציה או רגרסיה?

**רגרסיה**

b. מיהו לדעתך הגרף המתאר את שגיאת האימון ומיהו הגרף המתאר את שגיאת הוולידציה?

**האדום(העליון) שגיאת ההכללה**

c. באילו דרגת גמישות נאמר שהמודל מבצע התאמת יתר? ובאילו נאמר שיש התאמת חסר?

**מעל 5 התאמת יתר**

**מתחת ל 5 התאמת חסר**

d. צייר על הגרף כיצד לדעתך יראה הגרף המתאר את שגיאת ריבוע הביאס ושגיאת הוואריאנס (בהנחה שהשגיאה האינהרנטית – שאינה ניתנת להפחתה – היא  $\epsilon=0$ ).

**ריבוע הביאס ירד מונוטונית וישאף ל 0 ככל שנגדיל את הגמישות**

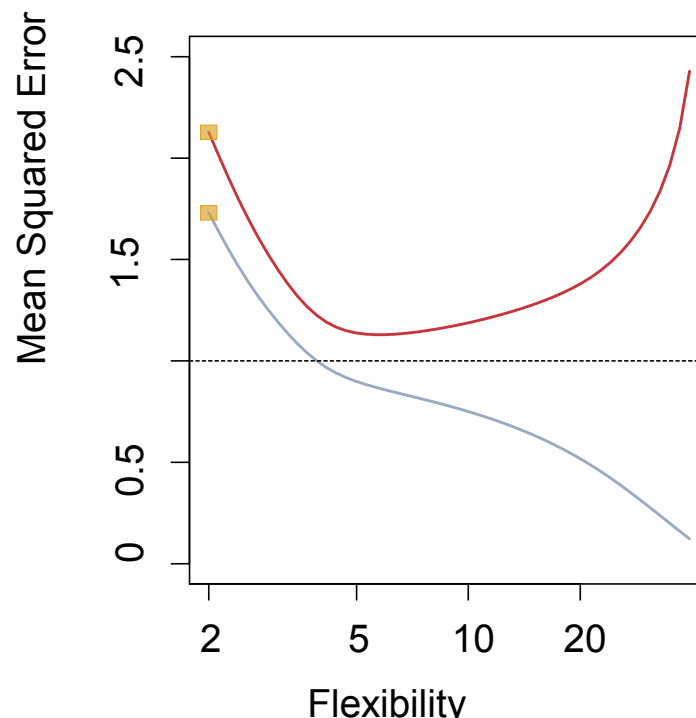
**הוואריאנס יעלה מונוטונית. 2 הגרפים יהיו חסומים ע"י הגרף של שגיאת ההכללה**

e. ע"פ הגרף ישנן כמה נקודות שבהן שגיאת ההכללה היא 1.5. בהנחה שהשגיאה הבלתי ניתנת להפחתה הוא  $\epsilon=0$  ובהנחה ששגיאת הביאס בנקודה הראשונה היא 1, מה יהיה שיערוך שגיאת הוואריאנס בנקודה הראשונה (השמאלית)?

**0.5**

f. בהמשך לסעיף הקודם, האם שגיאת הביאס בנקודה השנייה תהיה גבוהה או נמוכה מ 1?

**נמוכה**



2. [10] רוצים ללמוד פונקציות בוליאניות שצורתם 1CNF בעלות  $n$  משתנים. ששמותיהם בסדר לקסיקוגרפי (A,B,C,...). רק ליטרלים חיוביים אפשריים. (זיכרו כי ההיפוטזה הריקה- גם היא נכללת)

לדוגמא: עבור  $n=3$ , הליטרלים הם A,B,C

הליטרלים ב 1CNF נרשמים בסדר לקסיקוגרפי (כך ש A and B שקול ל A and B וירשם כמכפלה AB)  
a. רשום את כל ההיפוטזות במרחב עבור  $n=3$  (אל תכלול היפוטזות שקולות לוגית)

**","A,B,C,AB,AC,BC,ABC,**

b. מה גודלו של מרחב ההיפוטזות ממימד  $n$ ? (אל תכלול היפוטזות שקולות לוגית)

**$2^n$**

c. עבור  $n=3$  קיבלו  $D=\{(111,1),(011,1),(001,0)$

i. מהו ה version Space? (קבוצת ההיפוטזות הקונסיסטנטיות עם D)

**{B,BC}**

ii. איזו היפוטזה נבחר במידה ונשתמש בעיקרון התער של אוקם כדי לבחור היפוטזה?

**נבחר את ההיפוטזה הקונסיסטנטית הכי קצרה B**

3. [6] סמן איזה משפט נכון ואיזה אינו נכון:

a. בהינתן  $m$  דוגמאות, במידה ונבצע שיערוך שגיאת הכללה באמצעות 2-fold Cross Validation, ההפרש בין שגיאת האימון ושיערוך ההכללה יקטן ככל ש  $m$  גדל. **נכון** / לא נכון

b. הסיכוי ל overfitting קטן, ככל שגודל מרחב ההיפוטזות קטן. נכון / **לא נכון**

c. בהינתן קבוצת אימון  $D$ , קיים אלגוריתם למציאת עץ החלטה אופטימלי (העץ הכי קטן מבחינת מספר הצמתים אשר נותן accuracy מקסימלי על  $D$ ). **נכון** / לא נכון. הסבר בקצרה את תשובתך

4. [5] ברצוננו לסווג תמונה ל 3 קטגוריות A,B,C בעזרת רגרסיה לוגיסטית One Vs Rest. cross entropy I.

איזה וקטור targets,  $t$  יהיה שימושי עבור התמונות המסווגות?

a.  $t$  הוא תו המייצג את הקטגוריה ומכיל את התווים A או B או C

- b. t הוא וקטור בוליאני shape=[3] בקידוד One-Hot
- c. t הוא וקטור בוליאני Shape=[3] המורכב מאפסים ואחדים.
- d. t הוא מספר שלם המייצג את הקטגוריה ומכיל את המספרים 1 או 2 או 3

5. [5] מה לא נכון לגבי היפר פארמטר C (Capacity) ב SVM?
- d. ככל ש C יגדל יהיו פחות נקודות support והוואריאנס יקטן
- e. ככל ש C יגדל יהיו יותר נקודות support והוואריאנס יקטן
- f. ככל ש C יגדל יהיו יותר נקודות support והוואריאנס יגדל
- g. ככל ש C יגדל יהיו פחות נקודות support והוואריאנס יקטן

6. [6] נתון הקוד הבא בפיתון.
- a. מה הפלט שמתקבל?

```
from sklearn.metrics import confusion_matrix
target_labels = [1,1,2,3,1,1,3,2,1,3]
predicted_labels = [1,2,2,3,1,1,2,1,1,3]
matrix = confusion_matrix(target_labels,predicted_labels)
print(matrix)
```

```
[[4 1 0]
 [1 1 0]
 [0 1 2]]
```

- b. איזה תכונה היינו רוצים שמטריצת הפלט תקיים?
- מטריצה אלכסונית

7. [6] כידוע ניתן לעשות רגרסיה פולינומיאלית (או רגרסיה לוגיסטית פולינומיאלית) על מרחב features מותמר אשר מכיל (גם) את כל המכפלות של ה features מקוריים (כולל חזקות)
- a. כאשר מימד הקלט המקורי הוא  $n=3$   $x=(x_1, x_2, x_3)$  ומעוניינים להתאים פולינום מדרגה 2. כמה מאפיינים חדשים נוסף? רשום אותם

6

- b. כאשר מימד הקלט המקורי  $x=(x_1, x_2)$  הוא 2 ומעוניינים למצוא גבול החלטה מעגלי, מי הם המאפיינים (מספר מינימלי של מאפיינים) אשר חייבים להיות במרחב המותמר?
- $x_1^2, x_2^2$

8. [6] מבצעים Forward Feature Selection על מרחב קלט תלת מימדי  $(x_1, x_2, x_3)$  על מנת לצמצם את מימד הקלט למימד 2. באיטרציה הראשונה בודקים את ה Cross Entropy (CE) של 3 מודלים. כאשר בודקים את  $x_1$  מתקבל  $CE=5$ , כאשר בודקים את  $x_2$  מתקבל  $CE=6$  וכאשר בודקים את  $x_3$  מתקבל  $CE=7$ . באיטרציה השנייה בודקים הוספת ה features שלא התווספו באיטרציה הקודמת, מתקבלות תוצאות CE הקטנות ב 1 מהתוצאות שהתקבלו לאותם features באיטרציה הקודמת. איזו קבוצת features לדעתך תיבחר?
- $\{x_1, x_2\}$

9. [8] בהינתן מסווג של naive Bayes בעל 3 קלטים בוליאניים  $X_1, X_2, X_3$  ופלט בוליאני  $Y$ .  
**a.** מהו מספר הפרמטרים (המינימלי) שהמסווג צריך לשערך לצורך למידה? רשום אותם.

$p(+), p(x_1=1|+), p(x_2=1,+), p(x_3=1|+)$   
 $p(x_1=1|-), p(x_2=1|-), p(x_3=1|-)$   
 היתר ניתנים לחישוב למשל  $p(x_i=0|+) = 1 - p(x_i=1|+)$ ,  $p(-) = 1 - p(+)$ ,  
 סה"כ 7

**b.** מה מספר הפרמטרים המינימלי שהמסווג צריך לשערך מתוך  $D$ , אם לא נשתמש בהנחה הנאיבית?

$p(x|+) \mid p(x|-)$  לכל אחד מ 7 קלטים שיכולים להינתן (את השמיני ניתן לחשב). סה"כ  
 $14 = 7 + 7$ . את ה Priors ניתן לחשב.

10. [6] מבצעים leave One Out Cross validation. ומשתמשים במודלים שנוצרו עבור אנסמבל בשיטת bagging.  
 בהנחה שהיו  $n$  דוגמאות מתוגות.

**a.** כמה דוגמאות יש בכל קבוצת אימון?  
 **$n-1$**

**b.** בכמה מודלים משתמש האנסמבל?  
 **$n$**

**c.** בהנחה שמשתמשים ברגרסיה וב MSE. כיצד משתמשים בכל המודלים שניבנו כדי לקבל שיערוך לשגיאת הטסט הכוללת. בחר את התשובה הנכונה

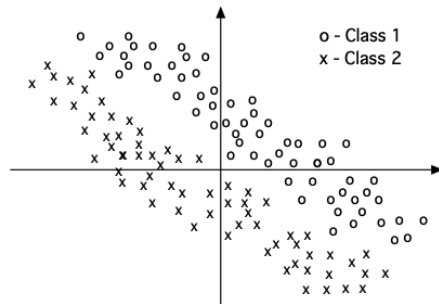
- כל מודל חוזה את ערך הרגרסיה על דוגמאות הוולידציה. מסכמים את ריבועי ההפרשים המצוי והרצוי (label) של כל המודלים.
- כל מודל חוזה את ערך הרגרסיה על דוגמאות הוולידציה, מבצעים הצבעה (voting) בין המודלים.
- כל מודל חוזה את ערך הרגרסיה על דוגמת הוולידציה ali. מחשבים ממוצע של ערכי החיזוי על כל המודלים.
- כל מודל חוזה את ערך הרגרסיה על דוגמת הוולידציה שלו. מחשבים את ריבוע ההפרש בין החזוי לרצוי (label) וממצעים על כל המודלים.

11. [8] בהינתן מישור הנפרש ע"י 2 ה principle components הראשונים של קבוצת נקודות  $D$  במרחב 4 מימדי.

**a.** רשום איזו תכונה מיוחדת מקימים המרחקים האוקלידיים של הנקודות ב  $D$  מהמישור הנ"ל (בשונה מכל המישורים האפשריים האחרים)?

- סכום ריבועי המרחקים (של הנקודות מהמישור) הינו מינימלי
- סכום ריבועי המרחקים (של הנקודות מהמישור) הינו מקסימלי
- הוואריאנס של ההטלות (של הנקודות על המישור) הוא מקסימלי
- הוואריאנס של המרחקים (של הנקודות מהמישור) הוא מינימלי

b. בהינתן סט הדוגמאות הבא ללא התייגים המופיעים בגרף במידה ונרצה להפריד את הקבוצות הבאות לשתי המחלקות הנ"ל ע"י PCA המוריד למימד אחד בלבד על איזה Principal component



נצטרך להטיל את הנתונים לקבל ההפרדה?.

**על ה PC השני: ווקטור עצמי בעל זווית של 45 מעלות.**

12. [9] עצי החלטה: בהינתן סט דוגמאות בעל 5 קלטים בוליאנים. לומדים פונקציה בוליאנית  $F: X \rightarrow Y$ . כאשר  $x$  וקטור ממימד 5.

a. בהינתן  $y = x_1 \wedge x_4 \vee x_2 \wedge x_3$  האם עץ בעומק 2 (4 עלים) יכול לייצג פונקציה זו? אם כן צייר עץ כזה ואם לא, מהו עומק העץ המינימלי שיכול לייצג פונקציה זו?  
**לא, העץ בעומק 4**

b. בהינתן  $y = x_1 \wedge x_4 \vee \neg x_1$  האם עץ בעומק 2 (4 עלים) יכול לייצג פונקציה זו? אם כן צייר עץ כזה ואם לא, מה עומק העץ המינימלי שיכול לייצג פונקציה זו?

**כן, קיים עץ בעומק 2**

c. נתונה קבוצת אימון  $D$  המכילה את טבלת האמת של 5 קלטים בוליאניים המתאימה לביטוי הלוגי  $\neg (x_1 \vee x_2)$ . 4 הקלטים האחרים  $x_3, x_4, x_5$  אינם משפיעים על הסיווג. מהו ה Information Gain של פיצול  $D$  בעזרת  $x_1$  ומהו במקרה שנפצל בעזרת  $x_2$ .

$$IG(D, x_1) = E(D) - 0.5(0) - 0.5(0) = 1$$

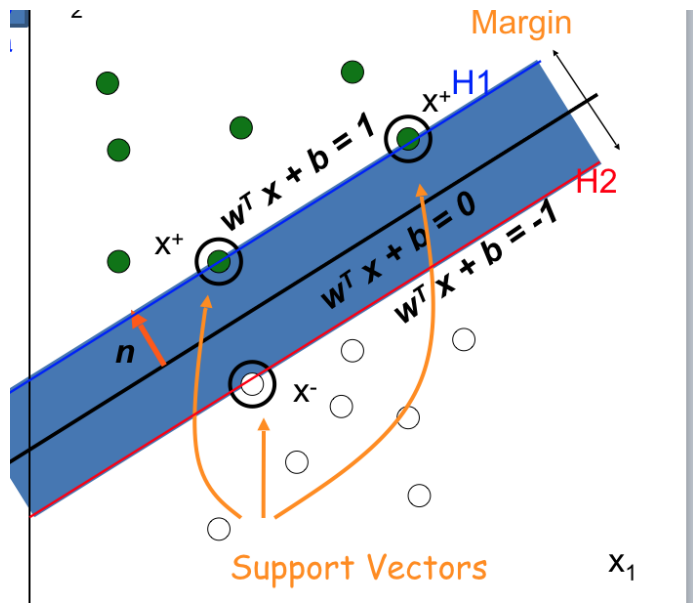
$$IG(D, x_2) = E(D) - 0.5(1) - 0.5(1) = 0$$

13. [6] חישוב דנדוגרמות: חשב את ה Inter Cluster Dissimilarity (IDC) עם  $\text{Linkage} = \text{Max}$  ופונקצית מרחק

$$\{(0,0)\} \cup \{(0,1), (-1,1)\}$$

**$2 \wedge 0.5$**

14. [9] SVM



$$\text{minimize}_w \left\{ \sum_{i=1}^m \max[0, 1 - y_i h(x_i)] + \gamma \sum_{j=1}^n w_j^2 \right\}$$

a. מה מייצג ה-1 ב hinge loss שב-svm?  
את ה Margin

b. אם נשנה את ה-1 ל-2 כיצד שינוי זה ישפיע על האלגוריתם?  
לא ישפיע

c. מה יקרה אם נשנה את ה-1 הנ"ל ל-2 עבור רק מחלקה אחת מסוימת בלבד?

- i. השוליים לא יהיו סימטריים ולמחלקה המסוימת יתאפשר להכניס יותר נקודות כ Support  
וזו יעזור למניעת overfitting כאשר יש מעט נקודות אימון במחלקה המסוימת
- ii. לא ניתן לעשות זאת
- iii. השוליים עדיין יהיו סימטריים וזה לא ישפיע על האלגוריתם
- iv. השוליים לא יהיו סימטריים ולכן פחות נקודות תמך יכנסו מהמחלקה המסוימת וזה יעזור למניעת Underfitting (ביאס)
- v. הוואריאנס יעלה והביאס יקטן עבור המחלקה המסוימת.