

שאלון X

1. [7] רגרסיה לוגיסטית פולינומיאלית: מעוניינים לבצע קלסיפיקציה לקלט מממד 2 (שני פיצ'רים; x_1, x_2). רוצים למצוא גבול החלטה שהוא פולינום מדרגה 3.
a. אילו Features תוסיפו?

b. מה יהיה מימד הקלט החדש?

2. [7] טרנספורמציה למרחב הקלטים: מבקשים לעשות רגרסיה למחירי דירות. בדאטהסט יש לנו שלושה פיצ'רים לכל דירה; אורך הדירה (L), רוחבה (W) ומספר הקומה (F).
ידוע כי מחיר הדירה תלוי לוגריתמית בסכום של הקומה ושטח הדירה. מחיר הדירה אינו תלוי בשום דבר אחר.

a. איזו טרנספורמציה ל-3 המאפיינים שבקלט הייתם עושים?

b. בכמה מאפיינים תשתמשו אם אתם חוששים מהתאמת יתר (Overfitting) ולכן רוצים לחסוך במשקולות?

c. בעזרת איזה class של Scikit-learn ניתן לבדוק באיזה בסיס של הלוג יש להשתמש?
i. Pipeline
ii. GridSearchCV
iii. StandardScaler
iv. VotingClassifier

3. [8] רגרסיה לוגיסטית: לאחר ביצוע אימון נמצא מישור הפרדה $w = (0.6, 0, -0.4, 0.693)$.
a. מה המרחק של נקודה $x = (1, 2, -1)$ ממישור ההפרדה?

b. מה תהיה ההסתברות של רגרסיה לוגיסטית על נקודה זו?

c. מה יהיה הסיווג של SVM על אותה נקודה ואותו מישור הפרדה?

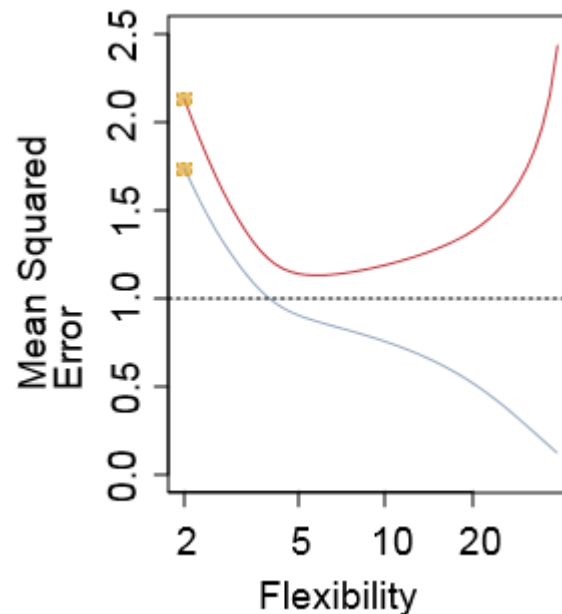
4. [5] רגרסיה לוגיסטית בשיטת GD: בהנחה שקבוצת האימון בממד קלט 1 (מאפיין אחד), מכילה דוגמא אחת שהמאפיין היחיד שלה (x_1) שווה ל-0 והלייבל שלה (t) שווה ל-1 [כלומר $(x_1, t) = (0, 1)$]. בהינתן שבתחילת האיפוק וקטור המשקלים הוא $w = (0, -1)$ (זכרו כי האיבר הראשון בווקטור הוא w_0) וקצב הלמידה (λ) הינו 1. מה יהיה w לאחר איפוק אחד?

5. [8] CE: ביצעו רגרסיה לוגיסטית על תמונות חתולים וכלבים במטרה לזהות תמונות של כלבים (ה-Target הוא תמונות כלבים). בקבוצת המבחן (Test) היו רק 2 תמונות של חתולים. החיזוי על תמונת חתול א' היה 0.3 והחיזוי על תמונת חתול ב' היה 0.6. [השתמשו ב- \ln (לוג עם בסיס טבעי - e)]
 a. מהי שגיאת ה-CE שתימדד על 2 דוגמאות החתולים (ממוצע השגיאה על כל קבוצת המבחן)?

b. מה (שיערוך) ההסתברות הממוצעת לחיזוי נכון של הקלסיפיייר?

6. [5] עצים רנדומליים: בונים יער של עצים רנדומליים בעומק מקסימלי 1 (פיצול 1) כאשר מספר המאפיינים הוא 16 ו- $m' = 4$
 כמה עצים שונים יכולים להיבנות ביער (הסבירו את תשובתכם)?

7. [10] מבצעים רגרסיה בעזרת thin board splines. בדרגות גמישות שונות, הריצו את הרגרסיה על קבוצת האימון וגם על קבוצת וולידציה וצירו את הגרף להלן.



a. מהי שגיאת הוולידציה ומהי שגיאת האימון בדרגת גמישות 20 (עגלו את השגיאה לחצי הקרוב ביותר)

b. כיצד תאבחנו את שגיאות מודל עם דרגת גמישות 20 ומודל עם גמישות 2:

- i. דרגה 20 הינה High variance ואילו דרגה 2 היא High Bias
- ii. דרגה 20 הינה High bias ואילו דרגה 2 היא Low Bias
- iii. דרגה 20 הינה High variance ואילו דרגה 2 היא Low Variances
- iv. דרגה 20 הינה High Bias ואילו דרגה 2 היא High Variance

c. באיזו דרגת גמישות תשתמשו כמודל החיזוי עבור קבוצת המבחן (Test Set)?

d. אם תרצו להוריד את שגיאת הוואריאנס. האם תשתמשו בלוח דק יותר או עבה יותר?
רמז: ככל שהלוח דק יותר כך יותר קל לכופף אותו.

e. ישנן 2 נקודות בהן שגיאת ההכללה (על ה- Validation Set) מקבלת ערך 1.4. בנקודה אחת שגיאת הוואריאנס (Variance) היא 1 ובשניה שגיאת הוואריאנס קטנה מ-1. האם הנקודה הראשונה היא השמאלית או הימנית?

f. בהנחה שהשגיאה הבלתי ניתנת להפחתה מתמצעת ל-0, מהי שגיאת הביאס בנקודה זו (תוחלת ההפרשים בין ההיפותזה לפונקציה שיצרה את הדוגמאות)?

8. [8] מבצעים רגרסיה בשיטת LWLR (משקול גאומטרי ע"פ הנוסחה הנתונה) כדי לחזות את ערך המטרה של דוגמת מבחן $x=(0,0)$ בהינתן קבוצת דוגמאות אימון [בצורה $((x_1, x_2), t)$]

$$D = \{((1,1), 1), ((-1,-1), 0), ((2,2), 0)\}$$

זו הנוסחה של משקל הדוגמאות (במונה יש נורמה 2 בריבוע):

$$\beta_i = e^{\frac{-\|x_i - x\|_2^2}{2\tau^2}}$$

נורמה 2:

$$\|x\|_2 = \sqrt{\sum_{j=1}^n w_j^2}$$

a. משתמשים ב- $\tau=10$ (טאו שווה ל- 10)

מה יהיו המשקלים (β) של 3 הנקודות ב- D?

b. בהנחה שמבצעים LWLR עבור $x=(0,0)$ באיזה טאו (τ) נקבל את הגמישות הגדולה ביותר (ז.א. את ההתאמה הגדולה ביותר לקבוצת האימון)?

i. טאו 10

ii. טאו 1

iii. טאו 0.2

iv. טאו 0.1

9. [8] בקלסיפיקציה בינארית מבצעים מקסימיזציה של המכפלה:

$$l_D(w) = \prod_{p:t_p=1} y_p \prod_{p:t_p=0} (1 - y_p)$$

בהנחה שמספר הדוגמאות בקבוצת האימון הוא m ,

a. כיצד נחשב את ההסתברות הממוצעת לחיזוי t ?

i. נחבר את ההסתברויות y_p (עבור דוגמאות $t_p = 1$) ואת ההסתברויות

$1 - y_p$ (עבור דוגמאות $t_p = 0$) ונעלה את הסכום בחזקת $\frac{1}{m}$

ii. נחבר את ההסתברויות y_p (עבור דוגמאות $t_p = 1$) ואת ההסתברויות

$1 - y_p$ (עבור דוגמאות $t_p = 0$) ונחלק ל- m

iii. נעלה את המכפלה בחזקת $\frac{1}{m}$

iv. נבצע \log של המכפלה

b. כיצד מגיעים מתוך מקסום מכפלת ההסתברויות לעיל למזעור ה-Cross Entropy?

i. מבצעים \log של המכפלה לעיל, מחלקים במספר הדוגמאות והופכים את

הסימן

ii. מבצעים \log של המכפלה לעיל, מעלים בחזקת $1/m$ והופכים את הסימן

iii. מחלקים את המכפלה לעיל במספר הדוגמאות והופכים את הסימן

iv. מבצעים \log של המכפלה לעיל והופכים את הסימן

10. [6] בקבוצת מבחן היו 200 נבדקים. חציים חולים (Label 1) וחציים בריאים (Label 0).

קלסיפייר סיווג נכון 90% מהחולים ושגה ב- 90% מהבריאים.

חשב:

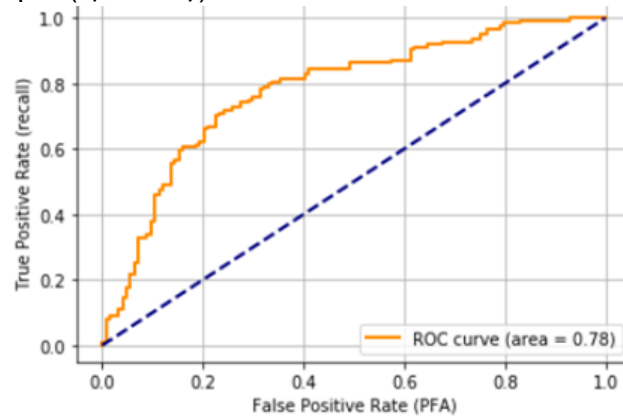
a. מלאו את מטריצת העמימות:

	Predicted 0	Predicted 1
Actual 0		
Actual 1		

b. מהו ה-F1?

c. מהו ה-Balanced Accuracy?

11. [4] נתונה דיאגרמת ROC. מה הספציפיות (Specificity) בנקודה בה ה- Recall=0.6? (בערך)



12. [4] Feature Selection, PCA: נתון קבוצת אימון (Dataset) עם מספר גדול של פיצ'רים. אנו מעוניינים להקטין את המספר הפיצ'רים כדי למנוע Overfitting. אנו יכולים לעשות זאת בעזרת PCA או בעזרת Feature Selection. סמנו נכון או לא נכון

a. קרוב לוודאי שבשימוש ב- PCA נאבד פחות מידע. (נכון/לא נכון)

b. קרוב לוודאי שבשימוש ב- Feature Selection נסנן יותר טוב פיצ'רים רועשים. (נכון/לא נכון)

13. [4] Laplace Smoothing: נתונה קבוצת אימון D של 1000 דוגמאות, חציין מתויגות כ- "1" וחציין מתויגות כ- "0". מבצעים קלסיפיקציה בינארית בשיטת Naive Bayes Classifier. לאחד המאפיינים x_1 יש 100 ערכים המופיעים ב- D . ואולם הערך v כלל אינו קיים בקטגוריה "1". מה יהיה שיערוך ההסתברות של $P(x_1 = v | "1")$ בשיטת הספירה שלמדנו כולל החלקת לאפלאס?

14. [7] CV, Bootstrapping:

- נתונה קבוצת אימון T של 1000 דוגמאות.
 מבצעים Bootstrapping כדי לבחור קבוצת אימון וולידציה.
 מבצעים מספר גדול של מדגמים. כל מדגם מתקבל ע"י 1000 דגימות אקראיות (עם חזרות).
 לכל מדגם, הדוגמאות שנבחרו במדגם (M) נלקחות כקבוצת האימון (Train) ואילו שאר
 הדוגמאות $(T-M)$ נלקחות כוולידציה (Validation).
 a. בהנחה שנלקחו הרבה מדגמים כאלו, מה גודל קבוצת הוולידציה הממוצעת?

- b. כמה קבוצות וולידציה שונות של 20% אפשר לדגום מקבוצה של 100 דוגמאות?

i. 100^{20}

ii. 20^{100}

iii. $\binom{100}{20}$

iv. $\frac{100}{20}$

- c. כמה אימונים נעשה בשיטת LOOCV?

15. [4] מבצעים רגולריזציית לאסו ($L1$). מה יהיה ערך משקולת $w=1$ בסיום איפוק כאשר קצב
 הלמידה (λ) הוא 1, מקדם הרגולריזציה (γ) הוא 1 והגראדיינט על w ($\nabla loss$) הוא 0.5?

16. [5] SVM: בנו מודל SVM עם קרנל לינארי לקבוצת אימון D עם $C=2$ וקיבלו שבנקודה בה ה-
 Margin מתמקסם, ה- Slack Variables הם כולם 0. מה המשמעות של זה?

- i. אין פתרון לבעיית האופטימיזציה
 ii. הדוגמאות ב- D ניתנות להפרדה לינארית
 iii. ה Margin -שנבחר הוא הרחב ביותר האפשרי
 iv. כל נקודות האימון נמצאים בתוך היפר-צינור ה- Margin