

1. [5] מה איננו יתרון של המשואה הנורמלית לעומת GD עבור רגרסיה לינארית?
- a. במשוואה הנורמלית ניתן להשתמש במימדים גבוהים מאוד
 - b. במשוואה הנורמלית אין קצב למידה.
 - c. ב GD רצוי לבצע נירמול
 - d. במשוואה הנורמלית אין צורך באיטרציות רבות
 - e. ב GD יש צורך בבדיקת התכנסות
2. [7] קלסיפייר בינארי (מסווג חיובי או שלילי) נמדד על קבוצת מבחן, והתקבלה טבלת עמימות. נבחנו 1000 מיקרים ומתוכם נחזו 100 חיוביים. ה precision הוא 0.8 וידוע כי מספר החולים בפועל היה 200.
- a. הראה את טבלת העמימות המפורטת
[[80,120],[20,780]]
 - b. חשב Recall
80/200
 - c. חשב F1-score,
 $2(0.4)(0.8)/(0.4+0.8)=64/120$
3. [5] אם נצייר גרף המודד את שגיאת הוולידציה ושגיאת האימון כפונקציה של גודל קבוצת האימון. רשום מה יהיה נכון ככל שנגדיל את קבוצת האימון?
- a. שגיאת הוולידציה תרד ושגיאת האימון גם תרד
 - b. שגיאת הוולידציה תעלה ושגיאת האימון תעלה
 - c. שגיאת הוולידציה תרד ושגיאת האימון תעלה
 - d. שגיאת הוולידציה תעלה ושגיאת האימון גם תעלה
4. [5] בהינתן קלט, קלסיפייר בינארי מחזיר שיערוך של ההסתברות לקטגוריה +. הקלסיפייר נמדד מול קבוצת מבחן וקיבל precision 0.4, Recall 0.999 מה כדאי לעשות כדי לשפר את ה precision על חשבון ה Recall?
- a. להנמיך את סף ההסתברות ל +
 - b. להגדיל את סף ההסתברות ל +
 - c. להגדיל את סף ההסתברות ל -
 - d. להקטין את ה false Negatives
5. [7] דני מעונין לבצע קלסיפיקציה של חמישה סוגי פרחים. לצורך כך יצא דני לשדות ואסף מדגם שלמספר רב של פרחים. לכל פרח שראה, מדד את המאפיינים הנומריים הבאים: אורך העלה, גודל הפרח, צבע עלי הכותרת של הפרח (אותו מדד בעזרת חיישן RGB). ברשימה תיגום נפרדת הוא רשם לכל פרט שאסף את סוג הפרח אותו זיהה מומחה לבוטניקה שהתלווה אליו. כאשר חזר למעבדה הסתבר כי מכשירי המדידה היו רועשים, ורשימת התיגום אבדה.
- a. הסבר בקצרה כיצד יוכל דני לפתור את בעיית הלמידה מבלי שיהיה לו תיגום? באיזה אלגוריתם שלמדנו יוכל לנסות להשתמש?

- i. K-NN עם $K=5$
- ii. PCA ל 5 הרכיבים העיקריים הראשונים
- iii. קלסיפיקציה ל 5 סוגי הפרחים בעזרת עץ החלטה
- iv. K-means עם $k=5$

b. בהמשך לסעיף הקודם, הסבר מה נדרש דני לעשות לנתונים לפני תחילת הפעלת האלגוריתם?

- i. PCA
- ii. Min-Max Scaling
- iii. Std Scaling
- iv. K-NN
- v. רגרסיה לוגיסטית

6. [6] קלסיפייר בינארי מבוסס רגרסיה לוגיסטית, מחשב את הסתברות תמונת קלט להיות של חתול. לאחר אימון שבו נלמד וקטור משקולות (כולל bias) w , נתנו לקסיפייר תמונה (ווקטור x) של חתול, והקסיפייר חישב $w \cdot x = -2$ כי

a. חשב את ההסתברות ש x איננו חתול.

$$1 - g(-2) = 1 - 0.12 = 0.88$$

b. בהנחה ש x אכן איננו חתול ($t=0$), חשב את ה CrossEntropy על דוגמא זו.

$$-(1-t) \ln(1-g(-2)) = -\ln(0.88)$$

c. בהנחה שמשתמשים בGD, וקבוצת האימון D מכילה את הדוגמא הנ"ל שאיננה חתול ($x, t=0$) בלבד, וקצב הלמידה הוא 0.5, כיצד ישתנה הביאס (א.ז. מהו Δw_0)?

$$\Delta w_0 = 0.5(t - g(-2))1 \approx -0.06$$

7. [8] מבצעים אלגוריתם Backward Feature Selection מתחילים ב 5 features אך כאשר מגיעים ל 3 features, מקבלים שגיאות אימון גדולות על ניסיונות להסיר feature נוסף.

a. כמה אימונים עשינו תוך כדי הרצת האלגוריתם? הראה את החישוב!

- i. 3
- ii. 7
- iii. 12
- iv. 4

באיטרציה ראשונה מבצעים 5 אימונים כדי להחליט איזה feature להוריד
באיטרציה השניה 4, ובאיטרציה שלישית 3 :סה"כ $12=5+4+3$

b. כמה פעמים נפעיל פרוצדורת ה Cross-Validation נ כדי לבחור את סט ה features הרצוי?

הראה את החישוב!

- i. 2
- ii. 6
- iii. 3

3: מפעילים CV על מודל עם 5 מאפיינים, על מודל עם 4 מאפיינים ועל מודל עם 3 מאפיינים

8. [10] פונקציות מרחק – dissimilarity וקרנלים: נתונים הנקודות (1,1), (2,2),

a. חשב את מרחק מנהטן בין הנקודות לעיל:

2

b. חשב מרחק גאומטרי בין הנקודות לעיל, כאשר פונקציית המרחק היא $e^{-\gamma d(x,y)}$, $d(x,y)$ הינו מרחק מנהטן ו $\gamma = 1$

$$e^{-2} = 0.135$$

c. סמן נכון/לא נכון לגבי γ (של הסעיף הקודם)

i. ככל ש γ יהיה גדול, המרחק הגאומטרי בין 2 נקודות שונות יתקרב ל 0 (נכון / לא נכון)

ii. ככל ש γ יהיה גדול, רגרסיה ליניארית לוקאלית (LWL) - locally weighted linear regressor - המשתמשת במרחק הגאומטרי לעיל כפונקציית משקל, תקצה משקלים קטנים יותר לנקודות רחוקות והדבר יקטין את שגיאת הוואריאנס. (נכון / לא נכון)

iii. ככל ש γ יהיה גדול, SVM המשתמש בקרנל (ראדיאלי) המתבסס על המרחק הגאומטרי יהיה דומה בביצועיו ל 1NN (נכון / לא נכון)

9. [8] סמן איזה משפט נכון ואיזה אינו נכון:

a. ה capacity ב SVM, נותן אפקט זהה לרגולריזציה מסוג L2. (נכון / לא נכון)

b. קלסיפייר מסוג knn יכול להפריד בעיה לא ליניארית כמו XOR. (נכון / לא נכון)

c. קלסיפייר שנותן 100% accuracy על train ו 70% accuracy על test עדיף מאשר קלסיפייר שנותן 70% accuracy על train ו 100% accuracy על test. (נכון / לא נכון)

d. ב PCA: רכיב ראשי i ורכיב ראשי i+1 הינם וקטורים עצמיים של מטריצת הוואריאנס של D, כך שהערך העצמי של רכיב ראשי i+1 גדול מהערך העצמי של רכיב ראשי i. (נכון / לא נכון)

10. [9] רוצים ללמוד פונקציות בוליאניות שצורתם 1CNF בעלות n משתנים בוליאניים (A, B, C, D, \dots) המקבלים ערכים $\{0, 1\}$ ומסודרים לקסיקוגרפית. כל משתנה זולת A יכול להופיע בצורה של ליטרל חיובי או שלילי או להופיע בכלל. המשתנה הראשון A , יכול להופיע רק בצורה שלילית - $(\neg A, \dots)$ או לא להופיע בכלל.

a. כמה היפותזות יש במרחב דו מימדי $n=2$? רשום אותם. זכור כי גם ההיפותזה הריקה "" קיימת. אין לרשום היפותזות שקולות לוגית.

4: $\neg A \neg B, \neg AB, \neg A, \neg B, B, ""$

b. מה גודלו של מרחב ההיפותזות במקרה ה n מימדי?

$2^{(n-1)}$

c. עבור מרחב היפותזות $n=2$, קיבלו קבוצת דוגמאות $D = \{((10), 1), ((00), 1), ((11), 0)\}$ מפעילים את אלגוריתם הקונסיסטנטיות שלמדנו כדי לקבל היפותזה הקונסיסטנטית עם D .

i. מאיזה ביטוי לוגי יתחיל אלגוריתם הקונסיסטנטיות?

$\neg A \neg B$

ii. רשום את ההיפותזה שתיבחר ע"י האלגוריתם

$\neg B$

iii. מה יחזיר האלגוריתם אם נוסיף ל D גם את $((01), 1)$

היפותזה ריקה

11. [6] נתונה קבוצת אימון $D = \{(0, 1), (1, 2), (2, 3)\}$ ומעונינים לבצע רגרסיה לינארית למשתנה בודד (Single

feature) x , בשיטת GD. מתחילים במשקולות מאופסות (גם הביאס) וקצב הלמידה 0.1.

מהן המשקולות שתתקבלנה לאחר EPOCH אחד של GD?

לשם הבהרה: הדוגמא $(0, 1)$ שב D מורכבת מקלט $x=0$ ותיוג $t=1$

$$W1 = 0.1 \text{mean}(t-y)x = 0.1(1/3)((1-0)0 + (2-0)1 + (3-0)2) = (1/30)(2+6) = 8/30$$

$$W0 = 0.1 \text{mean}(t-y)1 = 0.1(1/3)((1-0) + (2-0) + (3-0)) = (1/30)(1+2+3) = 6/30$$

12. [9] עצי החלטה: נתונה קבוצת אימון עבור קלסיפיקציה בינארית המכילה קלטים דו ממדיים $(F1, F2)$ ומאפיינים

נומרים. $D = \{((1, 2), +), ((2, 3), +), ((3, 2), -), ((3, 4), -)\}$ האלגוריתם עוצר כאשר מושגת הפרדה לקבוצות

הומוגניות ומיצר פיצולים בינאריים מהצורה $F_i < c$ (כך שלכל צומת נוצרים 2 ילדים).

a. צייר את עץ ההחלטה (כולל את הבדיקות $F_1 < c$ בצמתים) שיתקבל ע"י שימוש ב-Information Gain?

צומת יחיד $F_1 < 2.5$

b. מהו ה-Information Gain שנותר בעקבות הפיצול שיבצע השורש (בסעיף הקודם).

$$E(D) - 0.5(0) - 0.5(0) = 1 - 0 = 1$$

c. אילו פיצולים פוטנציאליים יבדקו על מנת לבנות את העץ שבסעיפים הקודמים?
 ז.א. עבור אילו בדיקות $F_1 < c$ יחשב האלגוריתם את ה-Information Gain כדי לבחור את צמתי העץ? (זיכרו כי קימת דרך יעילה לצמצום את מספר הבדיקות תוך שימוש ב- labels של D).

3 בדיקות: $F_1 < 2.5$; $F_2 < 3.5$; $F_2 < 2.5$

13. [9] בהינתן קבוצת אימון $D = \{((1,1),+), ((1,0),+), ((0,1),-)\}$ רוצים לסווג קלט דו מימדי חדש $x = (0,0)$ בעזרת קלסיפייר בינארי מבוסס על Naive Bayes. רשום בקצרה ובצורה מסודרת את החישובים להם נזקק הקלסיפייר.

a. חשב את השערוכים הנדרשים לצורך סיווג $x = (0,0)$ הנדרשים (רק מה שדרוש).
 השערוכים יחושבו בעזרת ספירה פשוטה (וללא תיקוני m-estimate).
 תזכורת: יש לחשב את השערוכים הנדרשים כדי להחליט איזו הסתברות היא גדולה יותר ע"פ MAP: $p(+|(00))$ או $p(-|(00))$

$$p(+)=2/3, p(-)=1/3, p(x_1=0|+)=0, p(x_2=0|+)=1/2, p(x_1=0|-)=1, p(x_2=0|-)=0,$$

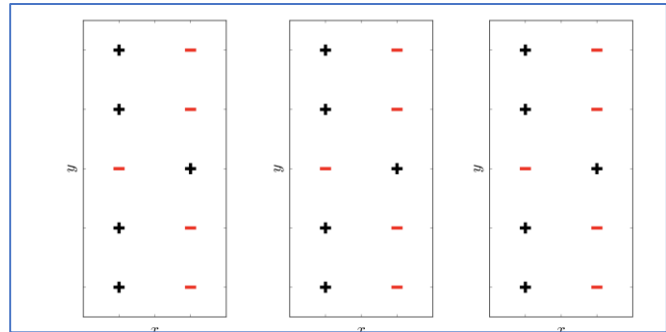
b. על פי השיערוכים שחישבת בסעיף הקודם. מה שיערוך ההסתברויות של $p(+|(00))$, $p(-|(00))$

ללא תיקונים בשיערוך הסתברות 2 הקטגוריות בהינתן x היא 0

c. [בנוס] חשב את השערוכים הנדרשים תוך שימוש ב-Laplace Smoothing. מה יהיה החיזוי עבור $x = (0,0)$? (תיזכורת: אין צורך לעשות תיקון לשערוכי ה-priors).

$$\begin{aligned} p(+)=2/3, p(-)=1/3, \\ p(x_1=0|+)= (0+1)/(2+2)=1/4, p(x_2=0|+)= (1+1)/(2+2)=1/2, \\ p(x_1=0|-)= (1+1)/(1+2)=2/3, p(x_2=0|-)=1/3 \\ p(+,00)=2/3 \cdot 1/4 \cdot 1/2=2/24 \\ p(-,00)=1/3 \cdot 2/3 \cdot 1/3=2/27, \end{aligned}$$

14. [6] צייר את גבול ההפרדה (= גבול החלטה) של קבוצת הנקודות הנתונה בציור (במישור האוקלידי). על פי כל אחד משיטות הקלסיפיקציה הבאות: רגרסיה לוגיסטית, 1-NN, ועץ החלטה בינארי.



רגרסיה לוגיסטית

1NN

עץ החלטה
בינארי

