(X838) מבוא לבינה מלאכותית | פרוייקט סיום תיאורטי

שם : גלעד X | ת"ז :X

שאלה 1 - PAC

(א) צ"ל ממד VC של על מישורים מפרידים $\geq d+1$ ובעצם להראות כי אף קבוצה של $d+2$ נקודות אינה מנופצת:

הוכחה : נניח בשלילה כי קיימת קבוצה של $d+2$ אשר נופצה ע"י $\{h \cap c \mid h \in H\}$. $H_c \in H, H_c =$ ⇐ $H_C = 2^{d+2}$.

ע"פ משפט ראדון , כל קבוצת נקודות מגודל $d+2$ ב R^d ניתנת לחלוקה ל2 תתי קבוצות זרות אשר הקמור שלהן נחתך.

תהי X_1, X_2 חלוקה כזאת של אוסף הנקודות שלנו. לפי הגדרה מקבילה לקמור - אוסף כל הצ"ל של איברים מהקבוצה עם מקדמים חיוביים שסכומם הוא 1.

$X_1 = \sum_1^n a_i p_i, s.t : \sum_1^n a_i = 1$, $X_2 = \sum_1^n b_i q_i, s.t : \sum_1^n b_i = -1$. מכיוון שכל קבוצת היפותזות שמנפצת את C בפרט מנפצת כל תת קבוצה שלה.

הרי ש 2 הקבוצות שלנו מנופצות גם הן . אולם , מחיתוך הקמורים של הקבוצות שלנו , נקבל כי קיימת נקודה, בה"כ p_0כלשהי , אשר נמצאת בחיתוך הנ"ל.

נקודה זו לא ניתנת להפרדה ע"י על מישור , שכן היא שייכת ל2 הקבוצות בו זמנית ⇐קיימת תלות לינארית של P_0 שלנו בנקודות $x_1 \in X_1$ו $x_2 \in X_2$כלשהן.

ברור כי כל על מישור מפריד בין 3 נקודות על אותו קו יצטרך לחצות את הקו הזה בנקודה כלשהי , אך לא קיימת דרך לחצות את הקו כך שהנקודה האמצעית תהיה שייכת

לחצי מישור מסויים וה2 האחרות (החיצוניות יותר) יהיו מצידו השני. ⇐ $H_C \neq 2^{d+2}$ ⇐סתירה לניפוץ $d+2$ נקודות.

(ב) צ"ל ממד VCשל היפר מישורים מפרידים הוא בדיוק $d+1$

להוכחה , נחסום מלמטה את ממד VC ע"י $d+1$ $\iff$ נראה כי קיים סידור כלשהו של $d+1$ נקודות שניתן לניפוץ ע"י על מישורים מפרידים.

נבחר את אוסף הנקודות : 0 , ו הבסיס הסטנדרטי על R^n : $\{(e_1 = (1,0,0,0,...), (e_2 = (0,1,0,0,...)....\}$.

עבור כל חלוקה של הנקודות הנ"ל ל2 קבוצות , נוכל למצוא על מישור מהצורה $(\sum_n w_i e_i) + b > 1$ כך שיפריד בין הנקודות מכיוון שאין תלות לינארית ביניהן,

ע"י w_i חיובי או שלילי בהתאם לצורך שלנו וbבהתאם לקבוצה שנרצה ש 0 יהיה בה .

ולכן קבוצת העל מישורים הללו שווה ל2^{d+1} . ולכן , $d+1 \leq vc_{dim} HYPERPLANES \leq d+1$ ושווה ל$d+1$ בדיוק .

(ג) הראה כי מס' הפונקציות השונות על m דוגמאות שנוירון בינארי עם כלל החלטה קשיח וkכניסות יכול לחשב, חסום ע"י $m^{k+1} + 1 \leq m^{2k}$:

בפשטות: כל דוגמה צריכה לקבל תיוג מתוך 2 תיוגים אפשריים (בינארי) , בהינתן $k=1, m=1$ נקבל סה"כ 2 תיוגים .

כל m שנוסיף , יקבל עוד 2 תיוגים אפשריים . כל k שנוסיף , יתן לנו את כל האפשרויות לשילובים בין ערכים שונים של k ובסופם 2 תיוגים אפשריים, אזי נקבל

שעבור כל m , 2 אפשרויות עם כל שילוב של k'ים כאשר בסה"כ במקרה המחמיר ביותר : $h \leq m^{2k} = powerset(m^{2k})$,

במידה ו ניתן לנפץ כמות גדולה מזה , נקבל כי $d< vcdim$, בסתירה להגדרה וכמובן בסתירה לכך שתיווצר לנו תלות לינארית שלא ניתן לנפץ ע"י חצאי מישורים.

(ד) נתונה רשת נוירונים עם Wמשקלים סה"כ , צ"ל חסם למספר פונקציות שונות הניתנות לחישוב ע"י הרשת:

נחלק ל2 : א. נראה כי מס הפונקציות בשכבה ראשונה כלשהי חסום , ב. נראה כי שאר השכבות חסומות ע"י המספר בשכבה הראשונה.

א. n נוירונים עם משקלים כמספר הכניסות , צריכים לפלוט אחת מ2 אפשרויות לכל דגימה נכנסת , לכל אחת זה על מישור המיוצג ע"י צירוף לינארי של המשקולות כפול התצפית ותוספת של ביאס .

ובמקרה הקיצון נמצא מפריד לכל אחד מהתצפיות משאר התצפיות , אזי נקבל כי מדובר בקבוצת החזקה של 2 עבור 2 מצבים אפשריים לכל תצפית בהתייחסות לכל אחד מהמשקלים ,

וסה"כ גודל קבוצת ההיפותזות האפשריות ללמידה $m^{2*|w|}$.

ב. נב"ש כי מס הפונקציות לא חסום ע"י השכבה הקודמת $\leftarrow$ בהנתן החסם הקודם , נניח כי חישבנו עוד פונקציה חדשה :

נמשיך להסתכל על הנוירונים כמפרידים לינאריים , הפונקציה החדשה שחישבנו מהנוירון הנוסף בשכבת השניה , היא צ"ל של של פונקציה שכבר חושבה.

לכן הלמידה לגביה תסתיים בסיווג של נקודה שכבר סיווגנו - במפריד לינארי שכבר קיים , בסתירה לכך שזו פונקציה חדשה .

(ה) מצא חסם מלעיל למס הדוגמאות הנדרשות לרשת עבור למידת PAC עם $\epsilon = \delta$:

נניח W גדול מאוד , בהתאם למשפט היסודי ללמידת PAC :

נידרש לm דוגמאות שהן $m \geqslant c * \frac{d+log(1/\epsilon)}{\epsilon}$ אבל d גדול ממש לכן נקבל כי נידרש ל $c * \epsilon * VCdim(h)$ דוגמאות .

שאלה 2 - רשתות נוירונים ולמידה לאחור

(א) מימדים של משקולות בהנחת וקטור עמודה כקלט:

בהנתן קלט מגודל $d1$, נקבל כי מימדW_1הוא $transpose$ של גודל הקלט $\Longleftarrow W_1 = Matrix^{dim(d1)*1}$ וb_1בהתאם - אותו גודל .

נשים לב כי הקלט עבור השכבה השניה מורכב מצירוף לינארי של השכבה הראשונה לאחר פונקציית $ReLU$ שבמקרה הקיצון לא תאפס כלום

$$\Leftarrow input(z3) = \sum_{dim1}(w_1x_1 + b_1) - (w_1x_2 + b_1) = \sum_{dim1}(w_1(x_1 - x_2))$$

ונקבל כי מימד W_2וb_2היא כגודל $W1$ ו b_1. סה"כ נראה שכולם וקטורים באורך $d1$.

(ב) הרבה חישובים ידניים , רושם תשובות סופיות מוסיף דף חישובים ידני בסוף .

1. $\frac{\partial L}{\partial(z_3)} = \frac{1}{NUM(P_i)} * (\tilde{y} - y)$
2. $\frac{\partial z_3}{\partial(a)} = w_2$
3. $\frac{\partial a}{\partial(z_2)} = \begin{cases} if z_2 > 0 & 1 \\ else & 0 \end{cases}$
4. $\frac{\partial a}{\partial(z_1)} = \begin{cases} if z_1 > 0 & 1 \\ else & 0 \end{cases}$
5. $\frac{\partial Z_1}{\partial(w_1)} = \overrightarrow{x_1}$
6. נשלב : $\frac{\partial z_1}{\partial(w_1)} = \frac{\partial L}{\partial(z_3)} * \frac{\partial z_3}{\partial(a)} * \frac{\partial a}{\partial(z_1)} * \frac{\partial z_1}{\partial(w_1)} = \frac{1}{N}(\tilde{y} - y) * w_2 * (1/0) * \overrightarrow{x_1}$

(ג) כללי העדכון של GD:

$$w_1 = w_{old} - \alpha(=1) * \frac{\partial L}{\partial w_1}$$

$$w_2 = w_2 - \frac{\partial L}{\partial w_2}$$

$$b_1 = b_1 - 1 * \frac{\partial L}{\partial b1}$$

$$b_2 = b_2 - 1 * \frac{\partial L}{\partial b2}$$

שאלה 3 - אנטרופיה

(א) מכיוון שמדובר ב9 "בינאריים" , נקבל כי $2^9 = 512$, והאנטרופיה יוצאת 9 בהתפלגות אחידה.

(ב) נחשב ונגיע ל $9 * (\frac{1}{4} * log_2(4) + \frac{3}{4} * log_2(\frac{4}{3})) = 7.301$

(ג) חסם עליון יהיה - 1.698

(ד) אנטרופיה של Y- 3.32

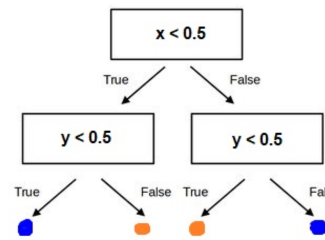
(ה) חסם תחתון - 1.623

(ו) הסתברות לשגיאה גדולה מ0.187

(ז) אני בספק גדול אם האלגוריתם יוכל להגיע לחסם , מספיקים 2 פיקסלים שפספסו ואין דרך להבדיל בין ספרות , נצטרך יותר פיקסלים .

שאלה 4 - עצי החלטה

(א) עץ אופטימלי:



(ב) האלגוריתם לא יצליח להגיע לעץ המושלם שכן אין פה צעד מועדף , כל צעד לא מקטין את אי הוודאות שלנו , ובהצבעת רוב בין מספר זוגי של אפשרויות - לא נקבל רוב .

(ג) נצטרך להוסיף הטייה כלשהי , לדוגמה $x <= k$ במקום קטן בלבד או להוסיף נקודה מלאכותית רנדומלית כך שיהיה יתרון מספרי להעדפה אחת על השניה .

שאלה 5 - התכנסות $KMEANS$

בהנתן A - סכום ריבועי המרחקים של כל אחת מהנקודות למרכז שאליו הן משתייכות , נסמן $A = \sum |x_i - c_{clust}|^2$

כמו כן נזכור : בכל צעד , האלגוריתם מחפש מרכז חדש ע"י מיצוע סכום ריבועי המרחקים של הנקודות באשכול , ובחירת הנקודות הקרובות ביותר לשיוך לאשכול הזה .

(א) צ"ל צעד 1 באלגוריתם מקטין את A אם שינינו את המרכז ולא משנה את Aבמידה ולא :

נוכיח: יהי A כלשהו , $c_{A-new} = \frac{1}{|I|_{i \in I}} \sum |x_i - c_{clust}|^2$ כעת יש 2 אפשרויות: אם לא שינינו מרכז -

$c_{new} = c_{old}$, אזי$A_{new} = \sum |x_i - c_{clust-new}|^2 = \sum |x_i - c_{clust-old}|^2 = A_{old}$כי לא שינינו נקודות , והכי קרובות במדגם עדיין הכי קרובות לאותו מרכז.

אם שינינו מרכזים : עלינו לחשב מחדש את השייכות לאשכולות ע"פ הכלל $C_j^{(t)} = \{x_i | \forall m \in \{1, ..., k\} : d(x_i, c_j^{(t-1)}) \leq d(x_i, c_m^{(t-1)})\}$, כעת :

תהי v נקודה שלא היתה בAכלשהי שלנו , נבדוק האם A_1 קטנה מ A :

$$A = \sum d(x_i - c)^2 = \sum ||x_i - c||^2 = \sum |x_i - v + v - c|^2 = \sum |x_i - v|^2 + |v - c|^2 + 2 * ((x_i - v)(v - c)) > \sum d(x_i - v)^2 = A_1$$

ואכן הוכחנו לוקטור כללי , ולכן נכון לכל צעד שיש בו הזזה של המרכז.

(ב) אותה הוכחה עם שינוי מ1 ל2 באינדקסים .

(ג) נראה כי אין אפשרות ללולאה של חזרה לתיוגים קודמים:

נב"ש כי הגענו לחלוקה כלשהי , ולאחר t צעדים חזרנו לאותה חלוקה :

ע"פ ההוכחה מסעיף א' , Aקטנה בכל שלב , נקבל כי : $A_{old} = \sum |x_i - c|^2 > \sum |x_i - c|^2 = A_{new}$,

אם חזרנו לאותה חלוקה , כל x_iבאשכול הוא הכי קרוב לc מרכז האשכול , ואחד מהם לא היה בשלב הקודם , מה שאומר שלפני הזזת המרכז , הוא היה באשכול אחר ,

אזי הזזת המרכז היתה ב"כיוון" הנקודה שהתווספה , אבל אם כבר היינו בתיוג הנ"ל , אז כבר התחשבנו בנקודה הזו בהזזת מרכז האשכול , בסתירה לכך שהמרחק של הנקודה המדוברת

ממרכז האשכול גדל .

(ד) נוכיח כי האלגו מתכנס :

מכיוון שמדובר במספר סופי של נקודות , בכל שלב אנו מקטינים את A לכל קלאסטר , וממצעים את מרכז האשכול , אזי בערך מחלקים ל2 את הערכים מסביב ,

מכיוון שמרחקים בריבוע הם אי שליליים - קיים אוסף של נקודות כך שהמרחק מכל מרכז הוא מינימלי אזי קיים סידור נקודות כך שAמינימלי לכל אשכול.

ונקבל כי האלגו מתכנס למינימום , לאו דווקא האופטימלי(גלובלי) אבל כן נעצר אחרי מספר צעדים סופי .

נחסום את מספר האיטרציות הנדרש להתכנסות ע"י k^n כאשר מספר זה הוא פשוט כל האפשרויות לסידור כל הנקודות בכל הקלאסטרים .

שאלה 6 - שרשרת מרקוב

כל החישובים נעשו ידני , אז רושם מה ואיך חישבתי ותשובה סופית.

(א) התפלגות סטציונרית - נשאיף לאינסוף n ,נעלה את מטריצת המעבר בין המצבים בחזקה ונכפול בהסתברות למצב התחלתי

$$A = \begin{array}{ccc} & happy & upset \\ happy & 0.8 & 0.2 \\ upset & 0.4 & 0.6 \end{array} \longrightarrow_{lim(n)} \Pi * A^n = \Pi * \begin{array}{ccc} & happy & upset \\ happy & 2/3 & 1/3 \\ upset & 2/3 & 1/3 \end{array} = [2/3, 1/3] = \Pi_0$$

(ב) בהנתן וקטור מצב עבור יום ראשון - נחשב את ההסתברות לכך שהוא שמח ביום רביעי (נכפול מצב התחלה במטריצת מעבר בחזקת ימים:

$\pi_1 = [0, 1] * A^3 = [0.624, 0.376]$ ונקבל כי ההסתברות לכך שהמרצה שמח ביום רביעי הינה 62.4% .

(ג) ידוע כי נתן בחינה קלה ביום רביעי , נחשב הסתברות מותנית עם כל המידע שהתקבל לנו:

יש לנו וקטור הסתברות לגבי מצב הרוח של המרצה בהנחת מטריצת המעבר שחישבנו , כמו כן נתון מבחן קל ויש לנו הסתברות מותנית , נשתמש בנוסחת בייס :

$P(happy|easy - test) = 0.624 * 1 + 0.376 * 0.4 = 77.44\%$ וקיבלנו מעל $3/4$ הסתברות לכך שהמרצה שמח.

(ד) מה התוחלת לערעור בהנתן המידע הקודם ?

נחשב קודם כל את ההסתברות לכך שהמרצה יהיה שמח ביום שאחרי (יום חמישי) ואת התוחלת לניקוד , כנ"ל למצב שהוא עצבני :

יש לנו הסתברות של $2/3$ שהמרצה שמח בחמישי לכן 80% להעלאת 5 נק' ו20% לאי העלאת ציון , נקבל שינוי ממוצע של 4+ נק אם המרצה שמח.

יש $1/3$שהוא עצבני ולכן 70% שיוריד 4 נק' ו30% שלא ישתנה הציון , נקבל שינוי ממוצע של -2.8נק אם המרצה עצבני.

נסכום על תוצאות המאורעות בהנתן ההסתברות לשמח או כועס מהסעיף הקודם : $E[grade - change] = 0.6666 * (0.8 * (+5) + 0.2 * (0)) + 0.3333(0.7 * (-4) + 0)$

ונקבל כי התוחלת לשינוי בציון היא $E[change] = 2.664 - 0.933 = +1.73$, מכאן שמומלץ לערער תמיד , אלא אם הציון בין 64 ל 60 כי אז מסתכנים בנכשל. אבל ברוב המקרים מומלץ לערער.