תרגול 6

18/5/20 - ב׳

נדבר על: 1. מודל ה- PAC: ◦ הגדרה
◦ למידות - PAC
◦ דוגמא - איך לקבוע האם מחלקה היא למידת-PAC

2. מימד - VC: ◦ דוגמאות: * Halfspace הומוגני
* סף פולינומיאלי
ע״י שימוש לבדוק האם מחלקה היא למידת-PAC

(1) מודל ה- PAC

① הגדרה:

◦ X - domain set
◦ Y - label set (בלי הגבלת הכלליות): $Y=\{\pm 1\}$ או $Y=\{0,1\}$.
◦ $H\subset Y^X$ - מחלקת ההיפותזות
◦ D - ההתפלגות (לא ידועה)
◦ $S=\{(x_i,y_i)\}_{i=1}^m$ - training data (נדגם מ- D).

הנחת ה-i.i.d: $x_i \overset{iid}{\sim} D$ (נסמן גם: $S\sim D^m$)
הנחת הריאליזביליות: f שקובעת את התוויות מעל X מקיימת $f\in H$.
⇐ אלגו׳ הלמידה יסומן כ- $A(S)$.

2. סוגי שגיאות:
1. שגיאת ההכללה (האמיתית): $L_{D,f}(h)=P_{x\sim D}[h(x)\neq f(x)]$
2. empirical risk (שערוך לשגיאה האמיתית): $L_S(h)=\frac{1}{m}\left|\left\{\begin{matrix} i\in[m] \mid \\ h(x_i)\neq y_i\end{matrix}\right\}\right|$

② למידות-PAC

מחלקה H היא למידת PAC אם קיימת פונק׳ $m_H:(0,1)^2\to\mathbb{N}$ ואלגו׳ למידה A כך שלכל $\epsilon,\delta\in(0,1)$ • התפלגות D מעל X
• פונק׳ תיוג $f:X\to\{0,1\}$. אם הנחת הריאליזביליות עבור H,D,f מתקיימת ואם נריץ את A על $m\geq m_H(\epsilon,\delta)$ דוגמאות שנדגמו iid

מ-D ולמתייג f כך ש- $\forall$ (מתקיים) $A(S)=h_S$ המקיימת:

$$\underset{S\sim D^m}{P}\left[L_{D,f}(h_S)\le \varepsilon\right] \ge 1-\delta$$

אם קיים אלגו' למידה כזה, נאמר ש-A הוא אלגו' למידה למחלקה H.

- ε - פרמטר ה- accurance
- δ - פרמטר ה- confidence (Probability)
- m_H - נקרא ה- sample complexity (סיבוכיות המדגם)

כדי להראות שמחלקה היא למידת-PAC נראה 2 דברים:

1) קיים אלגו' למידה למחלקה.
2) קיימת פונק' $m_H(\varepsilon,\delta)$ שמקיימת את הנדרש בהגדרת למידות-PAC.

☆ ראינו בהרצאה אלגו' כללי: $ERM_H(S)\in \underset{h\in H}{\text{argmin}}\, L_S(h)$ שכל ההחלטה שלו היא כלל ה-ERM.

☆ ראינו בהרצאה שכל מחלקת היפותזות סופית H - למידת PAC. וכי אלגו' שמחליט לפי כלל ה-ERM לומד מחלקה זו, כאשר סיבוכיות המדגם היא: $m_H(\varepsilon,\delta)\le \left\lceil \frac{\log|H|+\log(\frac{1}{\delta})}{\varepsilon}\right\rceil$

- תלות ב-ε - $\frac{1}{\varepsilon}$ = חזקה ישר
- תלות ב-δ - לוג' = חלשה
- תלות ב-$|H|$ - לוג' = חלשה.

③ דוגמא

- H = מחלקת המלבנים המיושרים עם הצירים ב-R^2.
- $X=R^2$, $Y=\{\pm1\}$, f מתייגת נק' כך: אם היא בתוך המלבן (או על שפה): +1, אחרת: -1.

☆ נזכור שאנו פועלים תחת הנחת הריאליזביליות ולכן לכל אוסף המתויגים אכן קיים מלבן שמפריד בין הקבוצות.

נרצה למצוא אלגו' למידה שעבורו ניתן לחסום את $m_H(\varepsilon,\delta)$: (כפונק' של ε,δ).

האלגוריתם:

בהינתן אוסף תיוגים, נקח את המלבן R' שהוא המינימלי שמפריד בין הקבוצות המתויגים.

תמיד יש הפרש חיובי בין R ל- R' (מאחר וב- training data לא נמצאות כל הנק' מבפנים של R)

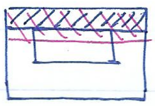
<u>נוכיח שניתן לחסום את m_H :</u>

נראה שלכל $\epsilon,\delta \in (0,1)$ קיימת $m_H(\epsilon,\delta)=m$ כך שלכל התפלגות D מעל $\mathbb{R}^2$, ולכל מלבן R בהסתברות לפחות $(1-\delta)$ מעל הבחירה של דוגמאות ב-S וגודל מדגם $\geq m$ מתקיים ש: $L_{D,R}(R') \leq \epsilon$.

<u>R,R' מסומנים בגלל שהחלקים הללו לא גורמים לשגיאה</u>

$\Leftarrow$ D מישור נחלק ל-3 חלקים: $\mathbb{R}^2/R$, R/R' , R'

אז נחלק את $R \setminus R'$ ל- 4 סטריפים ונניח שלכל אחד תהיה שגיאה $\geq \epsilon/4$.

נגדיר את הסטריפים T, T': T  T' . T' לפי הציור, T הוא בעל שטח של $\frac{\epsilon}{4}$.

<u>נבחין ש:</u> ל-T' יש שטח $\geq \epsilon/4$ $\Leftrightarrow$ $T' \subseteq T$.

$T' \nsubseteq T$ $\Leftrightarrow$ $\exists s \in S$ s.t: $s \in T$.

* אם נקח דגימה מ- D, הסיכוי <u>שלא</u> הגרלנו נק' ב-T הוא $1-\epsilon/4$.

וכיוון שכל הגרלה היא iid, אז הסיכוי לשלוף m דגימות ושאף לא יהיו ב- T הוא: $(1-\epsilon/4)^m$.

נעשה כנ"ל לכל שאר הסטריפים. מאורעות אלו <u>לא בהכרח זרים</u> אז לפי באופן החסם במאורעות שמתקיים: $Pr[A\cup B] \leq Pr(A)+Pr[B]$ נסיק:

שהסיכוי לשלוף m דגימות שלא נמצאות באף אחד מארבעת הסטריפים הוא $\geq \sum_{i=1}^{4}(1-\epsilon/4)^m = 4(1-\epsilon/4)^m$.

* נבחין שזה מאורע לא טוב (בדגימים מחוץ ל-T, כלומר מ- $R \setminus R'$: האזור הבעייתי)

נרצה לחסום הסתברות זו שמעט ע"י δ: $4(1-\epsilon/4)^m \leq \delta$

נמצא את m הדרוש לנו:

<u>נסמן:</u> $\epsilon/4=x$ ונקבל: (I) $4(1-x)^m \leq \delta$

ומכיוון ש- $1-x \leq e^{-x}$, נקבל (II) $4(1-x)^m \leq 4e^{-xm} = 4e^{-\frac{\epsilon m}{4}} \leq \delta$

מספיק רק את (I).

נבודד את m לפי (II): $e^{-\frac{\epsilon m}{4}} \leq \delta/4 \Rightarrow -\frac{\epsilon m}{4} \leq \ln(\frac{\delta}{4})$

$\Rightarrow m \geq \left\lceil \frac{4\ln(\frac{\delta}{4})}{\epsilon} \right\rceil$

■

(2) מימד VC

הראנו שמחלקת סופית למידה - PAC. מה עם מחלקות אינסופיות?
עבור מחלקות אינסופיות נחשב את מימד VC וזה יתן לנו כיוון להבין אם הן למידות-PAC או לא.

- **הגבלה של H ל-C:** יהא $C=\{x_1,..,x_m\}\subset X$, ומחלקה H של סיווג בינארי מעל X. ההגבלה של H ל-C היא:

$$H_C=\{h_C\in\{0,1\}^C \mid \exists h\in H \text{ s.t: } \forall x\in C \ \ h_C(x)=h(x)\}$$

כלומר: $H_C=\{(h(x_1),..,h(x_m)) \mid h\in H\}$

☆ נבחין שמתקיים: $|H_C|\le 2^{|C|}$.
כי ישנן שש x׳ים סה״כ שרק אחת התיוגים אפשרי מבחינתם.

- אם $|H_C|=2^{|C|}$ ל-$C\subset X$ סופית נאמר ש-H **מנפצת את C**.

- $\text{VCdim}(H)=\sup\{m\in\mathbb{N} \mid \exists C \text{ with } |C|=m \text{ and } H \text{ shatters } C\}$

איך מחשבים את $\text{VCdim}(H)$?
נרצה להראות ש- $d=\text{VCdim}(H)$. נצטרך ש:
1. קיים $C\subseteq X$, $|C|=d$ שמנופץ ע״י H ($\text{VCdim}(H)\ge d$)
2. לכל $C\subseteq X$, $|C|>d$, C לא מנופץ ע״י H ($\text{VCdim}(H)\le d$)

למה מימד-VC חשוב?

המשפט היסודי של הלמידה הסטטיסטית: H מחלקת היפותזות של מסווגים בינאריים על X ונניח $\text{VCdim}(H)=d$. אזי: קיימים קבועים C_1, C_2 כך שסיבוכיות המדגם של למידת-PAC את H חסום כך:

$$C_1\frac{d+\log(\frac{1}{\delta})}{\varepsilon}\le m_H(\varepsilon,\delta)\le C_1\frac{d\log(\frac{1}{\varepsilon})+\log(\frac{1}{\delta})}{\varepsilon}$$

בנוסף, סיבוכיות המדגם הנ״ל מושגת ע״י ERM.

VC-dim במחלקות סופיות:

טענה:

- לכל H סופית: $Vcdim(H) \le \log_2|H|$.
- נימוק: $Vcdim(H)=d$. אז: $|H| \ge 2^d \Leftrightarrow \log_2|H| \ge d$

(יש לפחות 2^d פונק' ב-H שמסכימות עם h על $x\in C$, אז יש לפחות 2^d פונק' שונות ב-H). ☒

נראה שאפשר לקבל < ואפשר לקבל = כלומר זה חסם הדוק

- נראה שייתכן שיהיה הפרש שרירותי בין $\log_2|H|$, $Vcdim(H)$:
- נגדיר: $H_{singleton} = \{h_z \mid h_z(x) = \mathbb{1}_{[x=z]},\ z\in X\} \cup \{h^-\}$ (מחזירה 1- לכל x)

<u>מתקיים:</u>

⊛ <u>לכל קבוצה בגודל 1</u>: $\{a\}$ נוכל לבחור את h_a, h_b ל-$b\ne a$.
$h_a(a)=1$, $h_b(a)=-1$ ⇐ לכן $\{a\}$ מנותצת את H.

⊛ <u>לקבוצה גודל > 1</u>: $\{a,b\}$, לכל פונק' שנבחר, במקרה הטוב רק אחת תתייג ב-1, והאחרת ב-1-. (או ששתיהן "יתייגו" ב-1-). לא נוכל לתייג את שתיהן ב-1! לכן לא נקבל את כל האפשרויות לכל התיוגים בנק' בקבוצה, וכל קב' מגודל > 1 כבר לא מנותצת את H.

⇐ $Vcdim(H)=1$. מכיוון ש: $z\in X$ אינסופי, כך גם גודל המחלקה, ולמימד VC קבוע אז הפער אכן שרירותי! ☒

- נראה שקיימת מחלקת היפותזות סופית H עם $Vcdim(H)=\log_2|H|$.
- נבחר $X=\{\pm1\}^d$ (הקוביה ה-d-מימדית), ו- $H=Y^X$.

במקרה זה מתקיים: $|H|=|Y|^{|X|}=2^{2^d}$. נקח את C להיות כל הדגימות האפשריות (כלומר: $|C|=2^d$) ומכיוון שאנו מסתכלים על כל הדגימות האפשריות, ומאחר והמחלקה שאנו מסתכלים על כל התוצאות האפשריות (מתאגדרת H) נקבל לכל תיוג של C היפותזה $h\in H$ שמספקת את התווית הנ"ל. כלומר: C מנותצת ע' H. ☒

☆ בדוגמת המלבנים מהחלק הראשון מתקיים: $Vcdim(H)=4$.

נבחין ש- Vcdim זה לא תמיד מס' הפרמטרים החופשיים במחלקה.

נראה ע' דוג': ל- $x\in\mathbb{Z}$: $H=\{x\mapsto sign(sin(\theta(x))) \mid \theta\in\mathbb{R}\}$

מתייגים נק' $x\in\mathbb{Z}$ לפי האם הפונק' (הנק' בו עוברת) מתחת או מעל לציר ה-X.

באופן אינטואיטיבי לכל דגימה (עם תיוגים) נוכל לייצר θ כך שנחזיר את התיוגים שקיבלנו. (נגרום לגל של פונק' ה-sin להתאים..) θ = לתדירות ⇐ כלומר Vcdim לא חסום והמחלקה לא למידה PAC. ($|Vcdim|=\infty$, פרמטר 1.)

דוגמאות:

1) Halfspace הומוגניים:

מחלקת המפרידים הלינאריים ההומוגניים מעל $X=\mathbb{R}^d$ מוגדרת ע"י:

$$\mathcal{H}=\{h_w \mid h_w(x)=\text{sign}(\langle w,x\rangle),\ w\in\mathbb{R}^d\}$$

טענה: $VCdim(\mathcal{H})=d$, (כאשר $X=\mathbb{R}^d$).

הוכחה:

(1) נקח את $C=\{e_1,\dots,e_d\}$. יהא תיוג $\{y_1,\dots,y_d\}\in\{\pm 1\}^d$, ונגדיר את $w=(y_1,\dots,y_d)^T$. מתקיים: $\langle w,x\rangle = w$

מפני שלכל y_i שנבחר נוכל לקבל תיוג שלו מה שבעצם נותן לקבל את כל התיוגים ל-C. כלומר C מנותצת את $\mathcal{H}$.

(2) נוכיח בשלילה: יהא $C=\{x_1,\dots,x_{d+1}\}$, אזי יש תלות ליניארית בין איברי C כלומר: קיימים $\{a_i\}_{i=1}^{d+1}$ לא כולם אפס שמקיימים: $\sum_{i=1}^{d+1} a_i x_i = 0$

בה"כ $a_{d+1}\neq 0$. אזי נחלק ב: $(-a_{d+1})$:

$$x_{d+1} = \sum_{i=1}^{d} b_i x_i, \quad b_i=\frac{-a_i}{a_{d+1}}.$$

כעת נראה שאי-אפשר להשיג את התיוג הבא (ולכן C לא מנותצת ע"י $\mathcal{H}$):

$$y_i=\begin{cases}\text{sign}(b_i) & : i\in[d]\\ -1 & : i=d+1\end{cases}$$

נניח בשלילה שקיים w שמשיג את התיוג שהגדרנו למעלה. אזי $y_i=\text{sign}(\langle w,x_i\rangle)$.

מתקיים: $\langle w,x_{d+1}\rangle = \langle w,\sum_{i=1}^{d} b_i x_i\rangle = \sum_{i=1}^{d} b_i\langle w,x_i\rangle$ סימנו כסימן b_i $\Rightarrow \sum + > 0$

כלומר: $y_{d+1}=\text{sign}(\langle w,x_{d+1}\rangle)=1$ בסתירה להגדרת התיוג y_{d+1}.

אזי אין w שמשיג את התיוג ו-C אינה מנותצת ע"י $\mathcal{H}$.

$\Rightarrow VCdim(\mathcal{H})=d$ ☐

2) Polynomial Threshold

מונום: תהא $\mathbb{R}^d \supseteq X=(x_1,..,x_d)$. פונק' מונום היא $x \mapsto \prod_{i=1}^{d} x_i^{n_i}$ עבור סט משתנים ש $\{n_1,..,n_d\} \in \mathbb{N}$.

דרגת המונום היא $\sum_{i=1}^{d} n_i$

פולינום: נאמר ש- $p(x)$ הוא פולינום מדרגה r אם הוא סכום לינארי של מונומים מדרגה מרבית r.

$$Pol_r = \{x \mapsto sign(p(x)) \mid p(x): \mathbb{R}^d \to \mathbb{R} \text{ פולינום מדרגה } r\}$$

★ נבחין ש: Pol_1 היא מחלקה ש halfspaces הומוגניים שראינו קודם.

נראה ש: $Vcdim(Pol_r) \leq \binom{d+r-1}{d}$:

הוכחה: יהא $x \in \mathbb{R}^d$, נגדיר $\Psi_r(x)$ כוקטור ש כל המונומים מדרגה r. $\Psi_r: \mathbb{R}^d \to \mathbb{R}^{\binom{d+r-1}{r}}$. (בחירת r כדורים מתוך d כדורים עם החלפה בלי חשיבות לסדר).

אז אפשר לכתוב: $p(x) = \langle \Psi_r(x), w \rangle$.

ועכשיו אנחנו במקרה ש Halfspaces הומוגניים ☺, ולכן $Vcdim(Pol_r) \leq \binom{d+r-1}{d}$ ∎