נמשיך לדבר על מודל ה-<u>PAC</u>, נדייק אותו תחת הנחות מסוימות וכו'.

<u>מושגים</u> :

סימונים: • X : Domain Set • Y : Label set
• מחלקת השערות: $H \subseteq Y^X$ • ההתפלגות (לא ידועה): D על X
• training data: S = $(x_1, .., x_m)$ • כלל ההחלטה: A(S).

<u>הנחות</u>: • iid : הדוגמאות ב-S (לדגימות) נדגמו iid מ-D.
• ריאליזביליות: קיים $h\in H$ שעבורו $L_D(h)=0$.

↰ ונזכיר אלגו' שמחזיר כלל ב-ERM, מאחר ומחזיר מחיר = 0 על S כולו.

<u>The Restriction Function</u>

הגדרה כללית לחסם מימד ה-VC: נאפיין H לפני קבוצה סופית C.

הגדרה: סימון קבוצה = $C = \{x_1, .., x_{|C|}\}$.

הגבלת ה h על C-S ($h|_C$): $H_C = \{(h_c = h(x_1),.., h(x_{|C|})) | h\in H\}$

כלומר עבור $h\in H$ ניקח וקטור שואל את איך החיזוי את
התנהגות הפונקציה על C-ה, ונשאל את הפונקציות שמגדירות את C את איך התנהג.

נגדיר: $\tau_H : \mathbb{N} \to \mathbb{N}$ שמחזירה עבור כל הדגימות שבגודל N-X
שלכל גודל, הקף שלו את הגודל H_C הגדולה ביותר
(נעבור על כל הקבוצות שגודלם m בגודל המסויימת וחזרת המדדים
נקבל על גודל המחלקה ונספר את המספר הכולל הקבוצים השונים).
$\tau_H(m) = \{|H_C| \mid C\subseteq X, |C|=m\}$

<u>נקודת דגש להגדרה זו:</u>

(1) ההגדרה של τ_H ש מחזירה המחלקה השערות H בשלב, כיצד נחשוב את
מימד ה-VC של המחלקה H?
← ונראה כל גודל m של C - נחפוק: $H_C \subseteq \{\pm 1\}^m$
ולכן $|H_C| \le 2^m$. את נקודות h החיזויים של הפונק' של גודל, ולכן
את הגודל המקסימלי שנחשוב את המחלקה (נקרא): $\tau_H(m) = 2^m$
את זה כלה יהיה (נכון) הרצוי מן לזה עד מימד ה- VC שלו. פורמ':
$$VC(H) = \max_m (\tau_H(m) = 2^m)$$

Ⓐ

⟸ עבור $m=1$: יש 3 אפשרויות. סכ"ה התיוגים של H ל-c יהיה $\{1,-2\}$

(כיוון ש-1 אם x מעל ה-threshold, או -1 אם מתחת אחרת)

והמסקנה כהב- $\tau_H(1) = 2$

עבור $m=2$: נה"כ הנקודות: $x_1 < x_2$

והתיוגים האפשריים הם (מתחת ל-threshold):

$\pm$ $\pm$
$\pm$ $\pm$

ולכן נוכל לקבל $-$, $+$. בסך יש בסך 3 אפשרויות, אזי: $\tau_H(2) = 3$

והמסקנה הכללית: (נראה) $\tau_H(m) = m+1$:

נניח שתהיינה m נקודות מסודרות לפי (גודל). אזי יש מספר אפשרויות

גדול (מתחת ל-threshold) — אחרת זו m או כל (הראשונות)

או הגבוהה ביותר. יש $(m-1)$ נקודות ביניהן הביניים ולכן סה"כ

$(m+1)$ תיוגים אפשריים.

Agnostic PAC

נרצה 3 הנחות:

(1) הגדרת פונקציית ההפסד: פונקציה נתונה ℓ של H (כה נבחרה לבעיה).

מה שמספר מה טיב ההתאמה פונקציית loss: $\ell(h,(x,y)) \geq 0$

שאינו את הדגימים H ההתאמה שלו.

מהם כללים נפוצים? כך: 1. $\ell^{0-1}(h,(x,y)) = \mathbb{1}_{[h(x) \neq y]}$

2. $\ell^{MSE}(h,(x,y)) = |h(x)-y|^2$

$L_S(h) = \frac{1}{m} \sum_{i=1}^{m} \ell(h,(x_i,y_i))$: סיכון ההפסד האמפירי: בכל

4 הדגימה $S = \{(x_i,y_i)\}_{i \in [m]}$.

סיכון ההפסד של h: $L_D(h) = \mathop{\mathbb{E}}_{(x,y) \sim D}[\ell(h,(x,y))]$

(2) הנחת הריאליזביליות: משווים את ההפסד לפונקציה הטובה

היותר שכנו ימים להבין: $\min_{h \in H} L_D(h)$.

נזכיר את ההגדרות מ- PAC ו- Agnostic PAC :

H מחלקת PAC נלמדת אם קיים $m_H : (0,1)^2 \to \mathbb{N}$ ואלג' למידה A, כך שלכל
כל $\varepsilon, \delta \in (0,1)$ ולכל התפלגות D מעל המרחב המדגם X
ולכל פונק' $f : X \to \{0,1\}$ אם קיים h במחלקה (ריאליזביליות) נתקיים
קיים $f - H, D, f$ אם ש' $m \geq m_H(\varepsilon, \delta)$ אזי:
$$\Pr_{S \sim D^m} [L_{D,f}(A(S)) \leq \varepsilon] \geq 1 - \delta.$$

H מחלקת Agnostic - PAC : אם קיים $m_H : (0,1)^2 \to \mathbb{N}$
ואלג' למידה A, כך שלכל $\varepsilon, \delta \in (0,1)$ ולכל D מעל $X \times Y$
(נזיר) אם ש' $m \geq m_H(\varepsilon, \delta)$ אזי :
$$\Pr_{S \sim D^m} [L_D(A(S)) \leq \min_{h \in H} L_D(h) + \varepsilon] \geq 1 - \delta.$$

קצת לפני המבחן :

משפט (בסיסי בתורת הלמידה הסטטיסטית): מחלקה (סופית) H נלמדת ⟺ H היא נלמדת
(א) H נלמדת - PAC ⟺ (ב) H נלמדת Agnostic-PAC ⟺
מימד ה- VC של H הוא סופי.

(*) נשים לב כי המשפט לא אומר רק על נלמדות!

סוף המשפט: נניח כי $d \leq \infty$ מימד ה-VC של H וַ קבועים C_1, C_2
אוניברסליים (לא תלויים ב- ε, δ, D) כך ש: ⊛H נלמדת - PAC תחת
$$C_1 \frac{d + \log(\frac{1}{\delta})}{\varepsilon} \leq m_H(\varepsilon, \delta) \leq C_2 \frac{d \log(\frac{1}{\varepsilon}) + \log(\frac{1}{\delta})}{\varepsilon}$$

⊛ H נלמדת Agnostic - PAC תחת :
$$C_1 \frac{d + \log(\frac{1}{\delta})}{\varepsilon^2} \leq m_H(\varepsilon, \delta) \leq C_2 \frac{d + \log(\frac{1}{\delta})}{\varepsilon^2}$$

ואם 2 החסמים העליונים מתקיימים, ניתן להשיג ע' ידי אלג' ה- ERM.

דוגמא: נתון SG פונקצ' אינדיקטור לפו חצי מרחב על $\mathbb{R}^d$. נתון מדגם המכיל נקודות חיוביות
ושליל' (תיוג שונה), נרצה פרד בעזרת המחלקה חצי המרחב (נלמד עליו
אם נניח כל לינאריות? (אינדיק' פונקציות S סופי)
התשובה חיובית אם המדגם ריאליזבילי.

$\Leftarrow$ (מסקנה): פתיח $\exists$ בעיית מגזם m (תלוי δ כ בזמן) אלא $h_m = A(s)$

יופ עבור m: $L_D(h_m) \geq \varepsilon + L_D(h)$ $\min_{h \in \mathcal{H}}$

ונרצה שנבחר גם ε. לשם כך p בחר תלויה N-VC של $\mathcal{H}$ הקבוצה.

זו תהיה (h) אלא מחלקת halfspace חיובי (אוסף ל... , חיובי.)

מתקיים כאן שמיד: $VC(\mathcal{H}) = d$ (למשל $\mathcal{H}$ ב-$\mathbb{R}^d$.)

נשמש מהמשפט הס.ע.:
$$m \leq O\left(\frac{d}{\varepsilon^2}\right)$$

וכן $\Leftrightarrow$ $\varepsilon \leq O\left(\sqrt{\frac{d}{m}}\right)$.

ואף: $L_D(h_m) \leq \min_{h \in \mathcal{H}} L_D(h) + \sqrt{\frac{d}{m}} C$ עבור $C \in \mathbb{R}$ קבוע.

(ולכן $m \to \infty$ אלא $\tfrac$ מתכנס δ לערך $(\min_{h \in \mathcal{H}} L_D(h) - \delta$,

___

אך נקפיד הגה הזו לכל תיאור בעזרת ה-uniform convergence

הרעיון הוא אותו m לולאו $\tfrac$, אלא נרצה שבמקרה ודאית תהיה אמ פנים של בעצם.

כי להציג $\mathcal{H}$ פורמית, נשמש התחלה הגדרה:

ε-N-רצ: נאמר ש-D,s הוא "ε-N-רצ" קבוע $l_{\mathcal{H},D}$ אלא מתקיים
$$\forall h \in \mathcal{H}: \quad |L_s(h) - L_D(h)| \leq \varepsilon$$

___

טענה: נאמר ש $\tfrac{\varepsilon}{2}$-N-רצ. (אלא שבמקרה הגבוהה שהודגה ול N-ERM$_{\mathcal{H}}$ (h_s)

מתקיים: $L_D(h_s) \leq \varepsilon + \min_{h \in \mathcal{H}} L_D(h)$

$\Leftarrow$ (מסקנה) אלא h^*-ב התהיה הגבוהה אותה מאת, נסמן: $h^* = \mathrm{argmin}_{h \in \mathcal{H}} L_D(h)$

$h_s = ERM_{\mathcal{H}}(s)$. מאחר s-ל הוא $\tfrac{\varepsilon}{2}$-N-רצ:
$$L_D(h_s) \leq L_s(h_s) + \tfrac{\varepsilon}{2} \leq L_s(h^*) + \tfrac{\varepsilon}{2} \leq \left(L_D(h^*) + \tfrac{\varepsilon}{2}\right) + \tfrac{\varepsilon}{2} = $$
$$= \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} + L_D(h^*) = L_D(h^*) + \varepsilon$$

uniform convergance (התכנסות אחידה). מחלקת היפותיזות $\mathcal{H}$ היא בעלת התכנסות אל ה-uc

אלא קיימת $m_{\mathcal{H}}^{uc}: (0,1)^2 \to \mathbb{N}$ שמתקיים: לכל $\varepsilon, \delta \in (0,1)$ ולכל התפלגות D

על $\mathcal{X} \times \mathcal{Y}$: $\Pr_{s \sim D^m}[s \ \text{הוא} \ \varepsilon\text{-N-רצ}] \geq 1 - \delta$.

___

טענה: $\mathcal{H}$ היא uc $\Leftarrow$ $\mathcal{H}$ היא Agnostic-PAC עם $m_{\mathcal{H}}(\varepsilon, \delta) \leq m_{\mathcal{H}}^{uc}(\tfrac{\varepsilon}{2}, \delta)$:

$\Leftarrow$ צריך להוכיח עם D, ε, δ ולכל D^m $s \sim$ (אלא $m \geq m_{\mathcal{H}}^{uc}(\varepsilon, \delta)$)

מתקיים: $\Pr_{s \sim D^m}[L_D(h_s) \leq \varepsilon + \min_{h \in \mathcal{H}} L_D(h)] \geq 1 - \delta$:

נמשש uc, נקח D, ε, δ: $\Pr_{s \sim D^m}[s \ \text{הוא} \ \varepsilon\text{-N-רצ}] \geq 1 - \delta$

ולפי טענה זה גורר את שהוכחנו: $\Pr_{s \sim D^m}[L_D(h_s) \leq \min_{h \in \mathcal{H}} L_D(h) + \varepsilon] \geq 1 - \delta$.

וכן זה מה שרצינו להוכיח $\textcircled{11}$