

1. $\mathcal{X}$ — множество $X = \{x_i\}_{i=1}^n$ — дискретный и конечный
пространство, тогда $\mathcal{P}(X) = \{P_i\}_{i=1}^n$ — вероятностный

$Y = \{0, 1\}$. Клас $\mathcal{K}$ складається з "синхронних" і "асинхронних" $h_x = 1_{\{x\}} : Y \mapsto \begin{cases} 1, & y=x \\ 0, & y \neq x \end{cases}$ та "негативної" $h^- \equiv 0$. Показати, що $\mathcal{K}$ PASC повнотан y "звичайною" сенсі (з припущенням реалізованості), оцінити $m_{\mathcal{K}}$ та описати алгоритм.

Питання $|\mathcal{K}| = n+1$, тобто можемо використати сен. для скінч. класів: $m_{\mathcal{K}} \leq \left\lceil \frac{1}{\varepsilon} \ln \frac{n+1}{\delta} \right\rceil$. Потрібно знайти кращу оцінку для цього випадку і авно описати алгоритм.

Принципова реалізованості означає, що одна з q -їх h_x, h^- повертає правильні класи (= "правильні" f)

Алгоритм: Якщо серед $S = \{(x_{\bar{y}}, y_{\bar{y}})\}_{\bar{y}=1}^m \in$ позитивна точка, то ~~повертає~~ $y: y_{\bar{y}} = 1$, то повернути $h_{x_{\bar{y}}}$, інакше h^- (для цього треба 1 раз перебрати S). Для цього завжди $L_S = 0 \quad ; \quad L_{\mathcal{D}, f} = 0$ для гіпотези h_x .

$L_{\mathcal{D}, f}(h_S) \neq 0$ можливо лише якщо $h_S = h^-$, але h^- завжди $\in h_{x_c}$. Вони відрізняються лише у x , тому помилкою $\in P := \mathcal{D}(\{x\})$. Якщо для даного ϵ $p \leq \epsilon$, то $L_{\mathcal{D}, f}(h_S) = p \leq \epsilon$. Якщо ж $p > \epsilon$, одержимо й навпаки обрати S , що дає $h_S = h^-$, тоді не міститиме x :

$$\begin{aligned} \mathcal{D}^m(\{S \mid L_{\mathcal{D}, f}(h_S) > \epsilon\}) &= \mathcal{D}^m(\{S = \{(x_{\bar{y}}, y_{\bar{y}})\}_{\bar{y}=1}^m \mid \forall \bar{y} \ x_{\bar{y}} \neq x\}) = \\ &= \mathcal{D}^m((X \setminus \{x\})^m) = (\mathcal{D}(X \setminus \{x\}))^m = (1-p)^m < (1-\epsilon)^m < e^{-\epsilon m}. \end{aligned}$$

Тоді при $m \geq \frac{1}{\delta} \ln \frac{1}{\epsilon}$ це $\leq \delta$, тому $\mathcal{D}^m(\{S \mid L_{\mathcal{D}, f}(h_S) \leq \epsilon\}) \geq 1 - \delta$: маємо PAC навч., і $m_{\mathcal{H}}(\epsilon, \delta) \leq \lceil \frac{1}{\epsilon} \ln \frac{1}{\delta} \rceil$.

1. Показати, що якщо клас $\mathcal{K}$ еквівалентно PASC навчаний, то він PASC навчаний з тим же алгоритмом.

Отже, тут задача класифікації, тобто $Z = X \times Y$,
де Y скінченна, $l = l_{0-1}$, і розподіл на $X \times Y$ задається
 $\mathbb{D}$ на X і $f: X \rightarrow Y$, потім $L_{\mathbb{D}, f}$. Щоб

перевірити "звичайну" PAC навчальність, зрештою
припущення реалізованості: $\exists h^* \in \mathcal{H} : L_{\mathcal{D}, f}(h^*) = 0$.

Але може $\min_{h \in \mathcal{H}} L_{\mathcal{D}, f}(h) = 0$, тому за дов. асп. PAC
набв. для $m_{\mathcal{H}}$ і алгоритма з кодо $\forall \epsilon, \delta$ для вибірок S

з $m \geq m_{\mathcal{H}}(\epsilon, \delta)$ ~~можливо~~ маємо
 $\mathbb{D}^m(\{S \mid L_{\mathcal{D}, f}(h_S) \leq \min_{h \in \mathcal{H}} L_{\mathcal{D}, f}(h) + \epsilon = \epsilon\}) > 1 - \delta$,

що є неможливо.

2. Нехай $\mathcal{X} = \{0, 1\}^d$, $\mathcal{Y} = \{0, 1\}$, і $\mathcal{H}$ - клас булевих
функцій, тобто логічних ф-цій вигляду

$$h: (x_1, \dots, x_d) \mapsto x_{i_1} \wedge \dots \wedge x_{i_k} \wedge \overline{x_{j_1}} \wedge \dots \wedge \overline{x_{j_\ell}}$$

для деяких $i_1, \dots, i_k, j_1, \dots, j_\ell = 1, d$. Наприклад, $x_1 \wedge x_2 \wedge x_4$,
 $x_2 \wedge \overline{x_3} \wedge \overline{x_5}$ etc. Також доцільно - нульову ф-цію: такою
буде $\forall$, що містить пару $x_i \wedge \overline{x_i}$; та монотонна 1,
що sign. порожньому набору індексів.

Отже, задача: дано $S = \{((x_{j_1}, \dots, x_{j_d}), y)\}_{j=1}^m$, треба перевірити

$h_5 \in \mathcal{K}$.

— Чи $\in \mathcal{K}$ (апострично) РАС навчання?

Так, адже він симетричний. Перелічимо усі $h \in \mathcal{K}$: кожен з аргументів x_i може входити до h як x_i , $\bar{x}_i$ або не входити. Тому разом маємо 3^d варіантів, де одинач-на ор-ція - коли h не входить. Додамо ще нульову:

$$|\mathcal{K}| = 3^d + 1$$

Навч. з ^{припущення} ~~належності~~ реалізованості означає, що існує "правильна" $f \in \mathcal{K}$, відповідно до якої розставлені мітки, зокрема, у S . Наприклад, для $f(x_1, x_2, x_3) = \bar{x}_1 \wedge x_2$ можливі $((0, 0, 0), 0)$, $((0, 1, 0), 1)$ etc., але не $((1, 1, 1), 1)$. У цьому випадку ~~для~~ за заг. ор-ною

$$m_{\mathcal{K}} \leq \left\lceil \frac{1}{\epsilon} \ln \frac{|\mathcal{K}|}{\delta} \right\rceil = \left\lceil \frac{1}{\epsilon} \ln \frac{3^d + 1}{\delta} \right\rceil$$

Апостричне навчання означає, що ми можемо

таким образом S , что не violates погрешности
кон'юнкції, наприклад, містить $((0,0,0),0)$ і $((0,0,0),1)$
одночасно. Тоді з заг. ф-ли маємо навічність для
 $m_H \leq \left\lceil \frac{2}{\epsilon^2} \ln \frac{2|H|}{\delta} \right\rceil = \left\lceil \frac{2}{\epsilon^2} \ln \frac{2(3^d+1)}{\delta} \right\rceil$ (і, можливо, певному
 ϵ_{app} для деяких розподілів).

— Знайти алгоритм, що здійснює ERM_H і працює
поліноміально за md .

Просто перебрати усі $h \in H$ і підставити в кон'юнкції з S дає $O(m^d(3^d+1))$, не годиться.

Можна спростувати гіпотезу: спочатку припустимо
 $h=0$. Прокладимо усі ел-ти S , пропускаючи ті, що з
міткою 0. Для першої пари $((x_1, \dots, x_d), 1)$ знаємо h
на добуток x_i для $x_i=1$ і $\bar{x}_i$ для $x_i=0$ (так, що
 $h(x_1, \dots, x_d)=1$). Зазначимо, що для потім. пар за умови
унік. реалізованості ця ф-ція все ще дає 0 (бо саме

такі аргументи не зустрічаються). Дані знову пропускаємо
 нулі, а для пар $((x_1, \dots, x_d), 1)$ вадресовано з h аргументи,
 що "завансають" отримати 1: x_i при $x_i = 0$ і $\bar{x}_i$ при $x_i = 1$.

Наприклад:

Дані	Гіпотеза
$((0, 0, 0), 0)$	0
$((0, 1, 0), 1)$	$\bar{x}_1 \wedge x_2 \wedge \bar{x}_3$
$((0, 0, 1), 0)$	$\bar{x}_1 \wedge x_2 \wedge \bar{x}_3$
$((0, 1, 1), 1)$	$\bar{x}_1 \wedge x_2$ $\uparrow$ викреслюємо
....	

Пл. 1., гіпно отримавмо час $O(md)$. За умови реалізов-
 ності цей алг. $g \in L_s(h_S) = 0$.

3 Розв'язано задачу ML на $\mathbb{Z}$ з розподілом $\mathcal{D}$ і ф-цією
 втрач $\ell: \mathcal{X} \times \mathbb{Z} \rightarrow [0, 1]$. Показати, що $\forall$ d, h_2 . А умова
 з деб. ~~можливо~~ PAC набув.

$$\forall \varepsilon, \delta \in (0, 1) \exists m(\varepsilon, \delta) : \forall m \geq m(\varepsilon, \delta) \mathbb{P}_{S \sim \mathcal{D}^m} (L_{\oplus}(A(S)) \leq \varepsilon) \geq 1 - \delta$$

exhibierenma zjincami

$$\lim_{m \rightarrow \infty} \mathbb{E}_{S \sim \mathcal{D}^m} L_{\mathcal{D}}(A(S)) = 0.$$

$\Rightarrow$. Нехай $\varepsilon > 0$. Можливо вбачаючи $\varepsilon < 2$, моги бути.

$$m \quad \left(\frac{\varepsilon}{2}, \frac{\varepsilon}{2}\right), \quad \exists \quad m > m \quad \left(\frac{\varepsilon}{2}, \frac{\varepsilon}{2}\right)$$

$$\mathbb{E}_{S \sim \mathcal{D}^m} L_{\mathcal{D}}(A(S)) = \int_{\{S \mid L_{\mathcal{D}}(A(S)) \leq \frac{\varepsilon}{2}\}} L_{\mathcal{D}}(A(S)) d\mathcal{D}^m + \int_{\{S \mid L_{\mathcal{D}}(A(S)) > \frac{\varepsilon}{2}\}} L_{\mathcal{D}}(A(S)) d\mathcal{D}^m \leq \left[\begin{array}{l} \ell \leq 1 \Rightarrow \\ \Rightarrow L_{\mathcal{D}} \leq 1 \end{array} \right] \leq$$

$$\leq \int_{\{S \mid L_{\mathcal{D}}(A(S)) \leq \frac{\varepsilon}{2}\}} \frac{\varepsilon}{2} d\mathcal{D}^m + \int_{\{S \mid L_{\mathcal{D}}(A(S)) > \frac{\varepsilon}{2}\}} 1 \cdot d\mathcal{D}^m = \frac{\varepsilon}{2} \mathbb{P}_{S \sim \mathcal{D}^m} \left(L_{\mathcal{D}}(A(S)) \leq \frac{\varepsilon}{2} \right) + \mathbb{P}_{S \sim \mathcal{D}^m} \left(L_{\mathcal{D}}(A(S)) > \frac{\varepsilon}{2} \right) \leq$$

$$\leq \left[P \leq 1, \text{ з умови!} \right] < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \text{ що } \bar{a} \text{ динамічно збігається.}$$

$$\Leftrightarrow \text{За дов. збігності, } \forall \varepsilon, \delta > 0 \exists m_0: \forall m > m_0 \mathbb{E}_{S \sim \mathcal{D}^m} L_{\mathcal{D}}(A(S)) < \varepsilon \delta.$$

Лема про пер. Маркова

$$P_{S \sim \mathcal{D}^m} (L_{\mathcal{D}}(A(S)) > \varepsilon) \leq \frac{\mathbb{E}(L_{\mathcal{D}}(A(S)))}{\varepsilon} < \delta, \text{ що } \bar{a} \text{ наближено.}$$

4. Нести в умові NFL-теорем $m < \frac{|X|}{k}$. Як зміниться і твердження?

Повторно доведення теорем і відбулося, де $m < \frac{|X|}{2}$ було суттєвим:

- На початку отримано скін. $C: |C| = 2m$. Тепер можна обрати $C: |C| = km$.

- Записуємо ел-ти C , що не належать до m -ел-мної вибірки S_j , як $\{y_1, \dots, y_p\}$. Тепер $p \geq km - m = (k-1)m$ (тут $\geq$, а не $=$, бо в посл-сті S_j елементи можуть повторюватися).

- Маємо (для y , для якого цей вираз min)

$$\max_i E_{S \sim \mathcal{D}_i^m} (L_{\mathcal{D}_i}(A(S))) \geq \dots \geq \min_y \frac{1}{T} \sum_{i=1}^T L_{\mathcal{D}_i}(A(S_y^i)) = \frac{1}{T} \sum_{i=1}^T \frac{1}{km} \sum_{x \in C}$$

$$\begin{aligned} & \frac{1}{T} \sum_{i=1}^T \frac{1}{km} \sum_{x \in C} \mathbb{1}_{\{A(S_y^i)(x) \neq f_i(x)\}} \geq \frac{1}{T} \sum_{i=1}^T \frac{1}{km} \sum_{\ell=1}^p \mathbb{1}_{\{A(S_y^i)(y_\ell) \neq f_i(y_\ell)\}} \geq \left[\begin{array}{l} \text{змінило пор. } m \\ \text{забав. } km = \frac{k}{k-1}(k-1) \\ \leq \frac{k}{k-1} p \end{array} \right] \\ & \geq \frac{k-1}{k} \frac{1}{p} \sum_{\ell=1}^p \frac{1}{T} \sum_{i=1}^T \mathbb{1}_{\{A(S_y^i)(y_\ell) \neq f_i(y_\ell)\}} \geq \left[\begin{array}{l} \text{сер. ар. } \geq \text{min} \\ \text{як раніше} \end{array} \right] \geq \frac{k-1}{k} \cdot \frac{1}{T} \cdot \frac{T}{2} = \frac{k-1}{2k} \end{aligned}$$

Поэтому $\exists i: \mathbb{E}_{S \sim \mathcal{D}_i^m} (L_{\mathcal{D}_i}(A(S))) \geq \frac{k-1}{2k}$ (у опр. гребенки
 было $k=2$ и $\geq \frac{1}{4}$). При этом, как правило, $L_{\mathcal{D}_i}(f) = 0$,
 где f - функция-жама, что $f|_C = f_i$. Взаимостно такое
 посл. з пер. Маркова для $\alpha = \frac{k-1}{4k}$ ($\in (0,1)$): $P(L_{\mathcal{D}}(A(S)) \geq \frac{k-1}{4k}) \geq$
 $\geq \frac{\frac{k-1}{2k} - \frac{k-1}{4k}}{1 - \frac{k-1}{4k}} = \frac{k-1}{1+3k}$ (можно спробувати і інші α). Отже, уточнена
Th.: $\forall A, m < \frac{|X|}{k} \exists$ розн. $\mathcal{D}$ на $X \times \{0,1\}$ і $f: X \rightarrow \{0,1\}$ такі, що:
 - $L_{\mathcal{D}}(f) = 0$;
 - $\mathbb{E}_{S \sim \mathcal{D}^m} (L_{\mathcal{D}}(A(S))) \geq \frac{k-1}{2k}$, отже $P_{S \sim \mathcal{D}^m} (L_{\mathcal{D}}(A(S)) \geq \frac{k-1}{4k}) \geq \frac{k-1}{3k+1}$.

5. Покажати, що VC-вирізність має вадмативність монотонності: $\mathcal{H} \subset \mathcal{H}' \Rightarrow VCdim(\mathcal{H}) \leq VCdim(\mathcal{H}')$.

Дійсно, $VCdim(\mathcal{H}) = d$ означає, зокрема, що $\exists C \subset X$:
 $|C| = d$ і $\mathcal{H}$ розбиває C : $\mathcal{H}_C = \{0,1\}^{|C|}$. Але

$\mathcal{H} \subset \mathcal{H}' \Rightarrow \mathcal{H}_C \subset \mathcal{H}'_C$, тодіто менше $\mathcal{H}'_C = \{0, 1\}^{|C|}$ -
 $\mathcal{H}'$ розбиває C . Тому за дов. $VCdim(\mathcal{H}') \geq d$.

Д/з. 1. Формально довести $\forall d$ твердження з лекції
про те, що VC -вмірність класу (х.ф.) d -вимірних
паралелепіпедів дор. $2d$.

2. Знайти VC -вмірності:

- класу $\mathcal{H} = \{h_x\}_{x \in X} \cup \{h^-\}$ з нотер. Д/з;

- класу булевих кон'юнкцій з 2.

Обидва вони скінч., але треба спробувати знайти точне значення,
що краще за оцінку $VCdim(\mathcal{H}) \leq \log_2 |\mathcal{H}|$ з лекції.