

Узагальнена PAC навчальності

Задаємо, що у одн. атомній PAC навчальності класу $\mathcal{H}$ ми вимагаємо $L_{\mathcal{D}}(h_S) \leq \min_{h \in \mathcal{H}} L_{\mathcal{D}}(h) + \epsilon$ (зі шв. $> 1 - \delta$), де $h_S = A(S)$ - роз-т роботи алгоритма на вибірці S , $\min$ роз-т ^(оптимальний) який залежить лише від ϵ, δ . Зокрема, це $\Leftrightarrow L_{\mathcal{D}}(h_S) \leq L_{\mathcal{D}}(h) + \epsilon \quad \forall h \in \mathcal{H}$. Тепер покажемо що умову, дозволяючи складності вибірки залежати від h , з якою ми порівнюємо h_S (у цьому випадку це означає, що ці іпотези є (ϵ, δ) -змагальними (competitive)).
def. Для задачі ML на просторі $\mathcal{Z}$ з $\mathcal{F}$ -цією втрат ℓ клас іпотез $\mathcal{H}$ зветься періодично навчальним (periodically learnable), якщо існують $\mathcal{F}$ -ція $m_{\mathcal{H}}^{NOL} : (0, 1)^2 \times \mathcal{H} \rightarrow \mathbb{N}$ і алгоритм ML A такі, що для будь-яких $\epsilon, \delta \in (0, 1)$, $\forall_{h \in \mathcal{H}}$ та розподілу $\mathcal{D}$ на $\mathcal{Z}$ зі шв. імовірністю не меншою ніж $1 - \delta$ A повертає за вибіркою S з $n > m_{\mathcal{H}}^{NOL}(\epsilon, \delta, h)$ ел-тів, що i.i.d. від $\mathcal{D}$,

гіпотезу $A(S)$ з ризиком $L_{\oplus}(A(S)) \leq L_{\oplus}(h) + \epsilon$.

Соч. Будь-який алгоритм PAC навчання клас гіпотез $\mathcal{H}$ нерівномірно навчанням.

В як було зазн. вище, дост. взяти $m_{\mathcal{H}}^{NUL}(\epsilon, \delta, h) := m_{\mathcal{H}}(\epsilon, \delta)$ з лев. алг. PAC навч., тоді умова виконана. $\triangle$

Рем. "Нерівномірно" тут означає саме виникнення залежності $m_{\mathcal{H}}$ від h : для порівняння $A(S)$ з різними гіпотезами $m = |S|$ можуть бути різні. Даві ми побачимо, що це послаблення строю, тобто нерівн. навч. класів гілше.

Структурна мінімізація ризику

Даві розглядаємо класи гіпотез $\mathcal{H}$, що є не гілками гілок, тобто скінченними чи зліченими об'єднаннями класів: $\mathcal{H} = \bigcup_{n=1}^N \mathcal{H}_n$ або $\mathcal{H} = \bigcup_{n=1}^{\infty} \mathcal{H}_n$ відгн. Тут об'єдн. не обов'язково фіз'ичне: класи можуть перетинатися. Для спрощення будемо завжди вважати, що $\mathcal{H} = \bigcup_{n=1}^{\infty} \mathcal{H}_n = \bigcup_{n \in \mathbb{N}} \mathcal{H}_n$

і називати об'єднання злічених. У скінч. випадку це означає просто, що $K_n = K_N \forall n \geq N$. Ідея полягає у тому, що кожній K_n ми приписуємо число $w(n) \in [0, 1]$, що називається вагою (weight) класу K_n і згінтованим так, так, що класи з вагою будуть "вансимілі". Дано точні означення:

дф. Функція ваги (ваговою ф-цією, weight function) на класі ірменез $K = \bigcup_{n \in \mathbb{N}} K_n$ зветься $\forall$ ф-ція $w: \mathbb{N} \rightarrow [0, 1]$ така, що $\sum_{n=1}^{\infty} w(n) \leq 1$.

Ек. Для скінч. об'єднань $K = \bigcup_{n=1}^N K_n$ можна вважати $w(n) = 0$ при $n > N$. Тоді прикладом буде "одноігма" вага $w(n) := \frac{1}{N} \forall n = \overline{1, N}$, що некадає переваги кожній класу.

У (нескінч.) зліч. випадку рівномірно розподілити ваги все не вийде. Тим рідше $w(n) := \frac{a_n}{\sum_{n=1}^{\infty} a_n}$, де $\sum_{n=1}^{\infty} a_n < \infty$.

$\forall$ збіжний невід'ємний числ. ряд, наприклад, $w(n) := \frac{1}{2^n}$
або $w(n) := \frac{6}{(\sqrt{n})^2} \quad \forall n \in \mathbb{N}$.

Планом далі припустимо, що для $\mathcal{H} = \bigcup_n \mathcal{H}_n$
усі $\mathcal{H}_n$ є рівномірно збіжними (для задачі ML з про-
стором $\mathcal{Z}$ і φ -цією витрат ℓ): $\exists m_{\mathcal{H}_n}^{uc} : (0,1)^2 \rightarrow \mathbb{N}$ така,
що $\forall \epsilon, \delta \in (0,1)$ і розподілу $\mathcal{D}$ на $\mathcal{Z}$ зі ім. $1-\delta$ для
вибірки S з $m \geq m_{\mathcal{H}_n}^{uc}(\epsilon, \delta)$ ел-тів, що i.i.d. відносно $\mathcal{D}$,
 $\forall h \in \mathcal{H}_n \quad |L_{\mathcal{D}}(h) - L_S(h)| \leq \epsilon$.

Далі позначатимемо $\forall m \in \mathbb{N}$ і $\delta \in (0,1)$

$$\epsilon_n(m, \delta) := \min \{ \epsilon \in (0,1) \mid m_{\mathcal{H}_n}^{uc}(\epsilon, \delta) \leq m \}, \quad (*)$$

тобто φ -ція $\epsilon_n : \mathbb{N} \times (0,1) \rightarrow (0,1)$ визначає найм. можливу
різн. між $L_{\mathcal{D}}(h)$ і $L_S(h)$ для $h \in \mathcal{H}_n$ і вибірок з
 m ел-тів для фікс. параметра невпевності δ . Вона спа-
дає зі зрост. m , бо кількість під $\min$ збільшується.

Крім того, як завжди, бачимо, що $m_{\mathcal{H}_n}^{uc}(\epsilon, \delta)$ (складність

вибірки) - найменша (для данис ϵ і δ) з такою власт., а
 тому спадат з δ зрост. δ (або ϵ - Впр. або гув. практич.).
 Тому знову не при зрост. δ (для фікс. m) множина
 у правій частині (*) збільшується, а її $\min E_n$ зменшу-
 ється: $E_n(m, \delta_1) \geq E_n(m, \delta_2)$ для $\delta_1 \leq \delta_2$. З деф. рівн.
 зб., що нагадає вище, і (*) випл., що $\forall m \in \mathbb{N}$, $\delta \in (0, 1)$ і
 розн. $\mathcal{D}$ на $\mathbb{Z}$ зі ім. $\geq 1 - \delta$ для S з m -ел-мів, що i.i.d.
 sign. до $\mathcal{D}$ $\forall h \in \mathcal{K}_n$

$$|L_{\mathcal{D}}(h) - L_S(h)| \leq \underbrace{E_n(m, \delta)}_{\text{з задачі}}. \quad (**)$$

Рн. Класи $\mathcal{K} = \bigcup_n \mathcal{K}_n$ - клас інтересу ML на $\mathbb{Z}$ з φ -ією випад.,
 $w: \mathbb{N} \rightarrow [0, 1]$ - вагова φ -ія, усі $\mathcal{K}_n$ мають властивості
 рівн. збінності зі складністю вибірки $m_{\mathcal{K}_n}^{VC}$, і E_n -ак у (*).
 Тоді $\forall \delta \in (0, 1)$, $m \in \mathbb{N}$ і розподіл $\mathcal{D}$ на $\mathbb{Z}$ зі імовірністю
 $\geq 1 - \delta$ для вибірки S з m ел-мів, що i.i.d. sign до $\mathcal{D}$,
 $\forall n \in \mathbb{N}$ і $h \in \mathcal{K}_n$ $|L_{\mathcal{D}}(h) - L_S(h)| \leq E_n(m, w(n) \cdot \delta).$

Закнема, в цих умовах

$$\forall h \in \mathcal{K} \quad L_{\mathcal{D}}(h) \leq L_S(h) + \min_{n: h \in \mathcal{K}_n} \varepsilon_n(m, w(n) \cdot \delta).$$

► Дійсно, з (***) маємо $\forall n$ і $\forall m, \delta, \mathcal{D}$

$$\boxed{\forall h \in \mathcal{K}_n} \xrightarrow{P_{S \sim \mathcal{D}^m}} (|L_{\mathcal{D}}(h) - L_S(h)| \leq \varepsilon_n(m, w(n) \cdot \delta)) \geq 1 - w(n) \cdot \delta,$$

отже $P_{S \sim \mathcal{D}^m} (|L_{\mathcal{D}}(h) - L_S(h)| > \varepsilon_n(m, w(n) \cdot \delta)) \leq w(n) \delta$, магі за

сумарністю

$$P_{S \sim \mathcal{D}^m} (\forall n \forall h \in \mathcal{K}_n \quad |L_{\mathcal{D}}(h) - L_S(h)| > \varepsilon_n(m, w(n) \cdot \delta)) \leq \sum_{n=1}^{\infty} P_{S \sim \mathcal{D}^m}$$

$$(\forall h \in \mathcal{K}_n \quad |L_{\mathcal{D}}(h) - L_S(h)| > \varepsilon_n(m, w(n) \cdot \delta)) \leq \sum_{n=1}^{\infty} w(n) \cdot \delta \leq \delta,$$

звідки і маємо потрібне. $\triangle$

Рез. Дати позначатимемо $n(h) := \min \{n \in \mathbb{N} \mid h \in \mathcal{K}_n\}$ —
перший номер $\mathcal{K}_n$, до якого належить $h \in \mathcal{K} = \bigcup_n \mathcal{K}_n$, з помп.

Рез., магі $\forall m, \delta, \mathcal{D}$ зі ім. $\gamma, 1-\delta$ для $S \sim \mathcal{D}^m$

$$\forall h \in \mathcal{K} \quad L_{\mathcal{D}}(h) \leq L_S(h) + \varepsilon_{n(h)}(m, w(h(h)) \cdot \delta)$$

Це потім буде наступне означення:

дев. Тоборать, що алгоритм A (ML для задачі на $\mathcal{Z}$ з φ -цією φ вразі ℓ) здійснює структурну (структуровану) мінімізацію ризику (structural risk minimization, SRM)

в класі гіпотез $\mathcal{H} = \bigcup_n \mathcal{H}_n$ з ваговим φ -цією $w: \mathcal{N} \rightarrow [0,1]$, де $\forall n \in \mathcal{H}_n$ має власт. рівномірної збіжності зі складовою вибіркою $m_{\mathcal{H}_n}^{uc}$, якщо $\forall m \in \mathcal{N}$ і $\delta \in (0,1)$ за певн. вибіркою S з m ел-тів він повертає гіпотезу

$$h_S = A(S) \in \underset{h \in \mathcal{H}}{\operatorname{argmin}} \left(L_S(h) + \varepsilon_{n(h)}(m, w(n(h)) \cdot \delta) \right),$$

де позн. ε_n і $n(h)$ - як вище.

Рем. Знову не, це не конкретний алгоритм, а гілкий їх для даної клас. На відміну від ERM $_{\mathcal{H}}$ -алгоритмів, усі гіпотези $h \in \mathcal{H}$ "рівноправні" та які лише мінімізують $L_S(h)$, тут виникає упередженість на користь h з певним $\varepsilon_{n(h)}(m, w(n(h)) \cdot \delta)$, тобто (при фікс. m і δ , як і раніше вище) з підкласів $\mathcal{H}_n$ з вагами $w(n)$: ці класи отримують

перевару, навіть якщо n з нас мають більші емп. ризики $L_S(h)$. Виявляється, що це утворює перівн. навчальність!

Th. Нехай $\mathcal{H} = \bigcup_n \mathcal{H}_n$ - клас гіпотез у задачі ML на $\mathbb{Z}$ з φ -гістою ν та ℓ , де $\forall n$ клас $\mathcal{H}_n$ має власт. рівномірної збіжності зі складністю вибірки $m_{\mathcal{H}_n}^{VC}$. Тоді $\mathcal{H}$ перівномірно навчаний для довільного алгоритма, що здійснює SRM в $\mathcal{H}$ з дов. базовою φ -гістою $\omega: N \rightarrow [0, 1]$, і складності вибірки $m_{\mathcal{H}}^{NUL}(\varepsilon, \delta, h) \leq m_{\mathcal{H}_{n(h)}}^{VC} \left(\frac{\varepsilon}{2}, \frac{1}{2} \omega(n(h)) \cdot \delta \right)$.

► Отже, нехай A - SRM-алгоритм для $\mathcal{H}$ з ω . Для довільних $\varepsilon, \delta \in (0, 1)$, $h \in \mathcal{H}$ нехай $m \geq m_{\mathcal{H}_{n(h)}}^{VC} \left(\frac{\varepsilon}{2}, \frac{1}{2} \omega(n(h)) \cdot \delta \right)$. За пореп. Рч., тоді $\forall$ розп. $\mathcal{D}$ на $\mathbb{Z}$ зі ймов. $\geq 1 - \frac{\delta}{2}$ для $S \sim \mathcal{D}^m$ (і ком. нічого нового)

$$\forall h' \in \mathcal{H} \quad L_{\mathcal{D}}(h') \leq L_S(h') + \varepsilon_{n(h')} \left(m \frac{1}{2} \omega(n(h')) \cdot \delta \right)$$

Зокрема, для $h' = A(S)$ за дев. SRM

$$L_{\mathcal{D}}(A(S)) \leq L_S(A(S)) + \varepsilon_{n(A(S))} \left(m \frac{1}{2} \omega(n(A(S))) \cdot \delta \right) = \min_{h' \in \mathcal{H}} (L_S(h') +$$

$$+ \varepsilon_{n(h)} \left(m \frac{1}{2} w(n(h)) \delta \right) \leq L_S(h) + \varepsilon_{n(h)} \left(m \frac{1}{2} w(n(h)) \delta \right) \leq [(*),$$

$$m \geq m_{\kappa_{n(h)}}^{uc} \left(\frac{\varepsilon}{2} \frac{1}{2} w(n(h)) \delta \right) \leq L_S(h) + \frac{\varepsilon}{2}.$$

Итак, $P_{S \sim \mathcal{D}^m} \left(L_{\mathcal{D}}(A(S)) > L_S(h) + \frac{\varepsilon}{2} \right) < \frac{\delta}{2}$. Крим моно, оск.

$$m \geq m_{\kappa_{n(h)}}^{uc} \left(\frac{\varepsilon}{2} \frac{1}{2} w(n(h)) \delta \right), \text{ за } \underline{\text{deb.}} \text{ } \text{період. } 3\delta.$$

$$P_{S \sim \mathcal{D}^m} \left(L_S(h) > L_{\mathcal{D}}(h) + \frac{\varepsilon}{2} \right) < \frac{1}{2} w(n(h)) \delta \leq \frac{\delta}{2},$$

тож за симметричності

$$P_{S \sim \mathcal{D}^m} \left((L_{\mathcal{D}}(A(S)) > L_S(h) + \frac{\varepsilon}{2}) \vee (L_S(h) > L_{\mathcal{D}}(h) + \frac{\varepsilon}{2}) \right) < \frac{\delta}{2} + \frac{\delta}{2} = \delta,$$

отже

$$P_{S \sim \mathcal{D}^m} \left(L_{\mathcal{D}}(A(S)) \leq L_S(h) + \frac{\varepsilon}{2} \leq L_{\mathcal{D}}(h) + \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = L_{\mathcal{D}}(h) + \varepsilon \right) \geq 1 - \delta,$$

що і потріб. за deb. період. павч. $\triangle$

Рем. В якості w тут можна взяти $\forall$ базову p -го, напр. (за Ex. вище) $w(n) = \frac{1}{n}$ для скінч. об'єктів, $w(n) = \frac{1}{2^n}$ або $\frac{6}{(\pi n)^2}$ - для нескінч. Як наслідок нерівн. Тн.

і основної Тн. нерівн. розгляду отриманого критерію період. павч. для біа. класифікації!

Тн. Клас гіпотез $\mathcal{H}$ у задачі бінарної класифікації на множині X є рівномірно навчаним тоді й тільки тоді, коли він є злісним об'єктом (априорно) PAC навчання класів.

► Згідно, що за осн. теоремою ML для бінарної клас. PAC навчання, осн. PAC навч., рівномірна збіжність і умова $VCdim < \infty$ еквівалентні. Отже:

Є. $\mathcal{H} = \bigcup_n \mathcal{H}_n$ де усі $\mathcal{H}_n$ (ан.) PAC навчані, а отже рівн. збіжні, є рівн. навчаним за попередньою Тн.

$\Rightarrow$. Нехай $\mathcal{H}$ неперив. навчаний з осн. A і скл. вимірності $m_{\mathcal{H}}^{NOL}$. Оберемо
 якісь $0 < \varepsilon < \frac{1}{8}$, $0 < \delta < \frac{1}{4}$. $\forall n \in \mathbb{N}$ покладемо $\mathcal{H}_n := \{h \in \mathcal{H} \mid m_{\mathcal{H}}^{NOL}(\varepsilon, \delta, h) \leq n\}$.
 Тоді $\mathcal{H} = \bigcup_{n=1}^{\infty} \mathcal{H}_n$. $\forall n \in \mathbb{N} \forall h \in \mathcal{H}_n$ і $\forall \mathcal{Q}$ на $X \times \{0, 1\}$ за лев. неперив. навчальності

$$P_{S \sim \mathcal{Q}^n} \left(L_{\mathcal{Q}}(A(S)) \leq L_{\mathcal{Q}}(h) + \varepsilon < L_{\mathcal{Q}}(h) + \frac{1}{8} \right) \geq 1 - \delta > \frac{6}{7}.$$

Зокр., якщо $\mathcal{Q}$ загдов. прип. реалізовності для $\mathcal{H}_n$, $L_{\mathcal{Q}}(h) = 0$ для де-
 якого $h \in \mathcal{H}_n$, тоді $P_{S \sim \mathcal{Q}^n} \left(L_{\mathcal{Q}}(A(S)) < \frac{1}{8} \right) > \frac{6}{7}$ $\nRightarrow VCdim(\mathcal{H}_n) = \infty$. Повто-
 рюючи для $\mathcal{H}_n$ гомедення NFL-теорему (з нас $A: (X \times \{0, 1\})^n \rightarrow \mathcal{H}$, але
 для дов. це результево-Впр.), отримуємо: $\exists \mathcal{Q}$ на $X \times \{0, 1\}$, що загдов.
 прип. реалізовності для $\mathcal{H}_n$ і для якого $P_{S \sim \mathcal{Q}^n} \left(L_{\mathcal{Q}}(A(S)) > \frac{1}{8} \right) \geq \frac{1}{7} \downarrow$.
 Т.ч., $\forall n \ VCdim(\mathcal{H}_n) < \infty \Rightarrow \mathcal{H}_n$ (ан.) PAC навчаний за осн. Тн. A
Рем. Зокр., якщо у задачі фин. класифікації $\forall$ зліченного розбиття
 $\mathcal{H} = \bigcup_n \mathcal{H}_n \exists n: VCdim(\mathcal{H}_n) = \infty$ (ув. приклад такого $\mathcal{H}$ дає у аналогії NFL-Тн.),

то за попер. Th. $\mathcal{H}$ не є унів. навч.

У цьому випадку $\text{VCdim}(\mathcal{H}) = \infty$ за def. VC-вим. З іншого боку, $\text{VCdim}(\mathcal{H}) = \infty$ можливе і для об'єднання скінченновимірних (наст. приклад також демонструє існування перив. навчаних PAC-навчаних класів):

Ex. Нехай у задачі $\mathcal{H}$ клас. на $\mathbb{R}$ $\mathcal{H}$ — це клас усіх порогових поліномів, тобто $\mathcal{H} = \left\{ h_p : x \mapsto \begin{cases} 1, & p(x) \geq 0 \\ 0, & p(x) < 0 \end{cases} \mid p \in \mathbb{R}[x] \right\}$.

В одній з Впр. вище (див. практику), ми показали, що $\mathcal{H}$ "схильний до перенавчання", тобто, наприклад, алгоритм з $\text{ERM}_{\mathcal{H}}$ може повернути $h_S : x \mapsto \begin{cases} 1, & (x, y) \in S \text{ і } y = 1 \\ 0, & \text{в інш. випад.} \end{cases}$ з

$L_S(h_S) = 0$ і високого $L_D(h_S)$.] гілшо, він не є
(ан.) РАС навічання, то $VC\text{-dim}(\mathcal{H}) = \infty$: $\forall d \exists C \subset \mathbb{R}$:

$|C| = d$ і $\mathcal{H}$ розб. C . Більш того, $\mathcal{H}$ розбиває будь-яку
скінченну C . Для цього можна використати інтерполяційні
поліноми (Лорентца, Лагранжа...) або показати безпосередньо:

якщо $C = \{x_1, \dots, x_d\}$, де $x_1 < x_2 < \dots < x_d$, то $\forall (y_1, \dots, y_d) \in$
 $\{0, 1\}^d$ можна взяти $p(x) = \sum_{i=1}^d \left(y_i - \frac{1}{2}\right) \underbrace{(x_i - x_1) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_d)}_{\text{степень } d-1}$. Дійсно, тоді $\forall i$
в $p(x_i)$ усі доданки нульові крім i -того, і $p(x_i) = \left(y_i - \frac{1}{2}\right) \underbrace{(x_i - x_1)^2 \dots (x_i - x_{i-1})^2 (x_{i+1} - x_i)^2 \dots (x_i - x_d)^2}_{>0}$, > 0 при $y_i = 1$ і < 0 при $y_i = 0$,

отже $hp(x_i) = y_i$. При $d=1$ це просто $p(x) = y_1 - \frac{1}{2}$ - поліном степеня 0.

З іншого боку, $\mathcal{H} = \bigcup_{n=1}^{\infty} \mathcal{H}_n$, де $\mathcal{H}_n$ - поліноми степеня $\leq n-1$
(це од'єднання функцій; ті ж властивості має і не-
функ. з $\mathcal{H}_n$ - поліномами ст. $\leq n-1$; нагадаємо, що поліноми
ст. 0 - константи). При цьому $VC\text{-dim}(\mathcal{H}_n) = n < \infty \forall n$,

отже $\mathcal{H}_n$ - (апп.) PAC навчаний, а $\mathcal{H}$ - неогорн. навчаний за
 опер. Тн. Дійсно, ми вже показали, що $\forall C \subset \mathbb{R}$ з $|C|=n$
 можна розбити за деп. k_p з p степеня $n-1$, подто $k_p \in \mathcal{H}_n$
 Т.ч., $VCdim(\mathcal{H}_n) \geq n$. Нехай тепер $C \subset \mathbb{R}$: $|C|=n+1$, подто
 $C = \{x_1, \dots, x_{n+1}\}$, $x_1 < \dots < x_{n+1}$. Якщо поліном p приймає
 у цих точках знач. знаків $+, -, +, -, \dots$, подто $k_p(C) = (1, 0, 1, 0, \dots)$,
 то p має $> n$ копарно різних коренів (між x_i і x_{i+1}),
 тому його степінь $> n$, а він $\notin \mathcal{H}_n$. Т.ч., $\mathcal{H}_n$ не розд. C :
 $VCdim(\mathcal{H}_n) \leq n$.

Неодпорідна навчальність у цьому прикладі з'являється через
 те, що SRM-алгоритм, на відміну від ERM, вписує
 вагу $w(n)$ для $\mathcal{H}_n$, де, через збіжність ряду $\sum_{n=1}^{\infty} w(n) \leq 1$,
 $w(n) \xrightarrow{n \rightarrow \infty} 0$: поліноми великих степенів мають **малі**
ваги, а отже рідше зустрічаються як відмови (рез-ти
 роботи) SRM-алгоритма, що заподіє перенавчання.

Вопр. Показати, що VC-вимірність $\forall$ скінченного об'єднання $\bigcup_{n=1}^N K_n$ класів скінч. VC-вимірності нескінченна. Зробити верну оцінку на неї (два варіанти такої оцінки вб. у практиці).

Реш. Порівняємо "зінгу" певної навчання $K = \bigcup_n K_n$ з "зіною" агностичною PAC навчання конкретного K_n (коли ми припускаємо, що потрібна гіпотеза саме там), тобто $VCdim(K_n) = n$, як у поперед. прикладі, і $w(n) = \frac{6}{(\pi n)^2}$. Зокрема, вона спадає зі зрост. n , тобто класи більшої вимірності мають менший пріоритет (взагалі, при SRM порядок класів вивільняє і несе відображає нашу знання про задачу). Під "зіною" розуміємо тут складність вибірки. За мінимого T_n , для $h \in K_n$ і $\forall \epsilon, \delta$

$$m_{K_n}^{NOL}(\epsilon, \delta, h) \leq m_{K_n}^{VC}\left(\frac{\epsilon}{2}, \frac{1}{2} w(n) \delta\right) = m_{K_n}^{VC}\left(\frac{\epsilon}{2}, \frac{3\delta}{(\pi n)^2}\right) \leq \left[\begin{array}{l} \text{кількість} \\ \text{векторів осн.} \\ T_n, \text{ для } C_2, \alpha \end{array} \right] \leq \frac{C_2}{\left(\frac{\epsilon}{2}\right)^2} \left(VCdim(K_n) + \ln \frac{1}{\frac{3\delta}{(\pi n)^2}} \right) = \frac{4C_2}{\epsilon^2} (n + 2 \ln \pi n - \ln 3\delta).$$

З іншого боку, як доводили раніше, для абсолютної PAC навчальності $m_{\mathcal{H}_n}(\varepsilon, \delta) \leq m_{\mathcal{H}_n}^{uc}(\frac{\varepsilon}{2}, \delta)$. Дати для спрощення вважати їх рівними. Знову ж, за кільк. версією осн. Тн, для деякого $0 < C_1 \leq C_2$ $m_{\mathcal{H}_n}^{uc}(\frac{\varepsilon}{2}, \delta) \geq \frac{4C_1}{\varepsilon^2} (n - \ln \delta)$. Тому $m_{\mathcal{H}_n}^{uc}(\varepsilon, \delta, h) - m_{\mathcal{H}_n}(\varepsilon, \delta) \leq \frac{4}{\varepsilon^2} \left((C_2 - C_1)n - (C_2 - C_1)\ln \delta + 2C_2 \ln \frac{\pi n}{\sqrt{3}} \right)$ ⑤. Якщо вважати $C_2 - C_1$ малим (див. деталі про кільк. версію осн. Тн у $[1, \text{Ch. 28}]$), маємо $\leq \frac{8}{\varepsilon^2} C_2 \ln 2n$: ця оцінка не залежить від δ і зростає разом з n : це "непраф" за те, що успішна гіпотеза знайметься саме у $\mathcal{H}_n$, а не у класі з більшими вимірами.

Рем. Залишилось питання про існування $\mathcal{H}$, що не ^{періодично} є навчальним. Для задачі біп. класифікації, як бачили, для цього вже недостатньо $VCdim(\mathcal{H}) = \infty$: потрібно (за Тн.) показати, що $\mathcal{H}$ не можна представити як $\bigcup_n \mathcal{H}_n$, де $\forall n \quad VCdim(\mathcal{H}_n) < \infty$. Виявляється, що клас усіх гіпотез на

нескінченний $\mathcal{H}$ все ще підсюжить:

Тн. (NFL - теорема для нерівномірної навчальності) Для заданої бінарної класифікації на множині X клас усіх функцій $h: X \rightarrow \{0,1\}$ не є нерівномірно навчальним, якщо X нескінченна.

► Отже, $\mathcal{H} = \{h: X \rightarrow \{0,1\}\}$, тобто $\mathcal{H}$ "розділяє" всю нескінченну X . $\mathcal{H}$ нерівн. навчальний. За Тн. вище, може $\exists$ (тобто розділ. у вим. об'єктах) розділення $\mathcal{H} = \bigcup_n \mathcal{H}_n$, де усі $\mathcal{H}_n$ (англ.) PAC навчальні, тобто $VCdim(\mathcal{H}_n) < \infty \quad \forall n$. Нехай $X_1 \subset X$ - якась скінченна підмножина з $|X_1| > VCdim(\mathcal{H}_1)$ ек-мів (бога $\exists$, бо $|X| = \infty$). Тоді само, $|X| = \infty \Rightarrow \exists X_2 \subset X: |X_2| > VCdim(\mathcal{H}_2)$ і $X_1 \cap X_2 = \emptyset$ etc. : $\exists \{X_n \subset X\}_{n=1}^{\infty}$ - сукупність скінч. підмножин таких, що $\forall n \quad |X_n| > VCdim \mathcal{H}_n$ і що $X_n \cap X_m = \emptyset$ для $n \neq m$. Туди цьому $|X_n| > VCdim \mathcal{H}_n$ означає, що $\mathcal{H}_n$ не розділяє X_n : $\exists h_n: X_n \rightarrow \{0,1\}: h_n \notin \mathcal{H}_n$. (Точніше, h_n не є обм. на X_n повної р-ції з $\mathcal{H}_n$).

Парі коректно визначена $h: X \rightarrow \{0, 1\} : x \mapsto \begin{cases} h_n(x), & x \in X_n, \\ 0 & \text{у інш. випад.} \end{cases}$
 $n = \overline{1, \infty}$;

При цьому $h \in \mathcal{H}$ і $\forall n \ h \notin \mathcal{H}_n$, бо $h|_{X_n} = h_n \notin \mathcal{O}_n$.

ор-ції з $\mathcal{H}_n$ за побудовою. $\Downarrow \triangle$