За попр. Рч., умова $VCdim(\mathcal{H}) < \infty$ є необхідною для PAC навчаності $\mathcal{H}$. Виявляється, що вона є й достатньою, причому і для агностичної навчаності. Ці умови пов'язує теорема:

Th. (Основна теорема теорії статистичного машинного навчання). Нехай $\mathcal{H}$ - клас гіпотез у задачі бінарної класифікації на множині $\mathcal{X}$. Тоді наступні його властивості еквівалентні:

1. $\mathcal{H}$ має властивість рівномірної збіжності.
2. $\mathcal{H}$ агностично PAC навчаний.
3. Будь-який $ERM_{\mathcal{H}}$ - алгоритм задовольняє умові агн. PAC навчаності.
4. $\mathcal{H}$ PAC навчаний.
5. Будь-який $ERM_{\mathcal{H}}$ - алгоритм задовольняє умові PAC навчаності

6. $\mathcal{H}$ має скінченну VC-вимірність.

$\triangleright$ Побудуємо схему залежності між цими твердженнями:

$$\begin{tikzcd}
\underline{1.} \arrow[r, Rightarrow] & \underline{3.} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \underline{2.} \arrow[d, Rightarrow] & \\
& \underline{5.} \arrow[r, Rightarrow] & \underline{4.} \arrow[r, Rightarrow] & \underline{6} \arrow[lllu, Rightarrow, bend left]
\end{tikzcd}$$

1. => 3. => 2. доводили у розділі про рівн. збіжність (при цьому 3. => 2. за def. агн. PAC навч.), 3. => 5. і 2. => 4. – бо умови агн. PAC навч. сильніші, 5. => 4. за def. PAC навч. і 4. => 6. за попереднім Тв. Тому залишилось довести 6. => 1.: з $VCdim(\mathcal{H}) < \infty$ випливає, що він рівн. збіжний. Це випливає з двох наступних тверджень:

def. Функція зростання (growth function) класу гіпотез $\mathcal{H}$ у задачі бін. клас. на мн. X – це

$$\tau_{\mathcal{H}} : \mathbb{N} \to \mathbb{N} : \tau_{\mathcal{H}}(m) := \max_{\substack{C \subset X \\ |C| = m}} |\mathcal{H}_C|.$$

Нагадаємо, що $|\mathcal{H}_C|$ – це кількість різних обмежень ф-цій $h \in \mathcal{H}$

на C і що за <u>def.</u> VC-вим. якщо $\mathrm{VCdim}(\mathcal{H}) = d$, то $\underbrace{\forall m \le d}$ $\exists\, C$:

$|C| = m$ і $|\mathcal{H}_C| = 2^m$ (тобто $\mathcal{H}_C$ - усі ф-ції на C), тому

зокрема, $\Gamma_{\mathcal{H}}$ - асимптотична для $d = \infty$.

$\Gamma_{\mathcal{H}}(m) = 2^m$ для $m \le d$. Виявляється, що для більших m при скінч. d зростання стає поліноміальним:

<u>Th. 1</u> (Лема Зауера, Sauer-Shelah-Perles) Нехай $\mathcal{H}$ - клас гіпотез у задачі бін. клас. на мн. $\mathcal{X}$ VC-вимірності $\mathrm{VCdim}(\mathcal{H}) = d < \infty$. Тоді $\forall\, m \in \mathbb{N}$

$$\Gamma_{\mathcal{H}}(m) \le \sum_{i=0}^{d} C_m^i.$$

Зокрема, $\Gamma_{\mathcal{H}}(m) < \left(\frac{em}{d}\right)^d$ для $m > d+1$.

▷ Спочатку доведемо другу частину:

<u>Lem. 1.</u> $\forall\, m, d \in \mathbb{N}$ таких, що $m > d+1$ $\sum_{i=0}^{d} C_m^i < \left(\frac{em}{d}\right)^d$

▷ Індукція за d:

<u>$d=1$.</u> $\sum_{i=0}^{1} C_m^i = 1 + m < \begin{bmatrix} e > 2 \\ m > 1 \end{bmatrix} < em$, що й потрібно.

<u>$d \to d+1$.</u> Отже, нехай тв. вірне для d. Тоді

$$\sum_{i=0}^{d+1} C_m^i = \sum_{i=0}^{d} C_m^i + C_m^{d+1} \le \begin{bmatrix} \text{гіпотеза} \\ \text{інд.} \end{bmatrix} < \left(\frac{em}{d}\right)^d + \frac{m(m-1)\dots(m-d)}{(d+1)!} =$$

$$= \left(\frac{em}{d}\right)^d \left(1 + \left(\frac{d}{em}\right)^d \frac{m\cdot\ldots\cdot(m-d)}{(d+1)!}\right) \le \begin{bmatrix} m-d+1,\ldots \\ \ldots, m-1<m \end{bmatrix} \le \left(\frac{em}{d}\right)^d \left(1+\left(\frac{d}{e}\right)^d \frac{m-d}{(d+1)!}\right) \le$$

З доведення формули Стірлінга $d! \underset{d\to\infty}{\sim} \left(\frac{d}{e}\right)^d \sqrt{2\pi d}$ випл., що її права частина – це оцінка знизу, що вірна $\forall d$:

$d! \ge \left(\frac{d}{e}\right)^d \sqrt{2\pi d}$, тому

$$\le \left(\frac{em}{d}\right)^d \left(1 + \cancel{\left(\frac{d}{e}\right)^d} \frac{m-d}{\cancel{\left(\frac{d}{e}\right)^d}\sqrt{2\pi d}\cdot(d+1)}\right) = \left(\frac{em}{d}\right)^d \cdot \frac{1}{d+1}\left(d+1+\frac{m-d}{\sqrt{2\pi d}}\right) <$$

$$< [2\pi d > 4] < \left(\frac{em}{d}\right)^d \frac{1}{d+1}\left(d+1+\frac{m-d}{2}\right) \le \begin{bmatrix} \text{в дужках:} \\ \frac{d}{2}+1+\frac{m}{2} \le m, \\ \text{бо } d+2 \le m \end{bmatrix} \le \left(\frac{em}{d}\right)^d \frac{m}{d+1} <$$

З іншого боку,

$$\left(\frac{em}{d+1}\right)^{d+1} \Big/ \left(\frac{em}{d}\right)^d = \frac{em}{d+1}\left(\frac{d}{d+1}\right)^d > \begin{bmatrix} \text{означення числа } e: \\ \left(1+\frac{1}{d}\right)^d \underset{d\to\infty}{\to} e \text{ зліва} \end{bmatrix} > \frac{m}{d+1}\cancel{\left(1+\frac{1}{d}\right)^d} \cdot \cancel{\left(\frac{d}{d+1}\right)^d} = \frac{m}{d+1},$$

тому нарешті $< \left(\frac{em}{d+1}\right)^{d+1}$, що й потрібно (причому ми довели строгу нерівність на $\forall$ кроці) △ (Лем. 1).

Перше тв. Тч. 1. випливає з наступного: $\forall \mathcal{H}$ (тут без прип. про VCdim) $\forall$ $C: |C| = m$ $\quad |\mathcal{H}_C| \le |\{B \subset C \mid \mathcal{H} \text{ розбиває } B\}|$ (*).

Дійсно, $VCdim(\mathcal{H}) = d$ означає, що $\mathcal{H}$ може розбивати лише множини з $\le d$ ел-тів, тому права частина (*)

$\leq |\{B \subset C \mid |B| \leq d\}| = \sum_{i=0}^{d} C_m^i$, що й потрібно.

Доведемо (*) індукцією за m (для усіх $\mathcal{H}$).

<u>$m = 1$.</u> Для $|C| = 1$ зліва і справа у (*) часто одночасно або 1 (коли усі $h \in \mathcal{H}$ на $c = 0$ або усі $= 1$, тобто C не розб. і $\varnothing$ розб. завжди), або 2 (коли C розб.: $\exists h \in \mathcal{H}$, що $= 0$ на цьому і $\exists h$, що $= 1$).

<u>$m-1 \rightsquigarrow m$.</u> Припустимо, що (*) вірне для $|C| = m-1$ і усіх $\mathcal{H}$. Нехай $|C| = m$: $C = \{c_1, \dots, c_m\}$. Покладемо $C' := C \setminus \{c_1\} = \{c_2, \dots, c_m\}$. Тоді за прип. індукції ((*) для C'):

$$|\mathcal{H}_{C'}| \leq |\{B \subset C' \mid \mathcal{H} \text{ розб. } B\}| = |\{B \subset C \mid c_1 \notin B \wedge \mathcal{H} \text{ розб. } B\}|.$$

Тепер покладемо

$$\mathcal{H}' := \{h \in \mathcal{H} \mid \exists h' \in \mathcal{H} : h(c_1) \neq h'(c_1),\ h(c_i) = h'(c_i),\ i = \overline{2,m}\},$$

тобто $\mathcal{H}'$ скл. з пар гіпотез $\mathcal{H}$, що узгоджені на C' і відрізняються на c_1 (з пар, бо h', оч., теж $\in \mathcal{H}'$ у позн. вище).

Знову за прип. інд.,

$$|\mathcal{H}'_{C'}| \le |\{B\subset C' \mid \mathcal{H}' \text{ розб. } B\}| = \left[\begin{array}{l} h\in\mathcal{H}' \text{ приймають усі} \\ \text{можл. зн. на } B \Rightarrow \text{отже} \\ \text{й на } B\cup\{c_1\} \text{ за визн. } \mathcal{H}' \end{array}\right] =$$

$$= |\{B\subset C' \mid \mathcal{H}' \text{ розб. } B\cup\{c_1\}\}| = |\{B\subset C \mid c_1\in B \wedge \mathcal{H}' \text{ розб. } B\}| \le$$

$$\le [\mathcal{H}'\subset\mathcal{H}] \le |\{B\subset C \mid c_1\in B\wedge\mathcal{H} \text{ розб. } B\}|.$$

Тепер помітимо, що $|\mathcal{H}_C| = |\mathcal{H}_{C'}| + |\mathcal{H}'_{C'}|$. Дійсно, розглянемо $\mathcal{H}_C \to \mathcal{H}_{C'}: (h(c_1), h(c_2), \dots, h(c_m)) \mapsto (h(c_2), \dots, h(c_m))$ $\forall\, h\in\mathcal{H}$. Це відобр. сюр, але не ін'. : для $h\in\mathcal{H}\setminus\mathcal{H}'$ $h(c_1)$ однозначно визн. $(h(c_2),\dots,h(c_m))$, а для $h\in\mathcal{H}'$ в $(h(c_2),\dots,h(c_m))$ переходять пари $(0, h(c_2),\dots,h(c_m))$ і $(1, h(c_2),\dots,h(c_m))$. Тому $|\mathcal{H}_C| = |(\mathcal{H}\setminus\mathcal{H}')_{C'}| + 2|\mathcal{H}'_{C'}| = |\mathcal{H}_{C'}| - |\mathcal{H}'_{C'}| + 2|\mathcal{H}'_{C'}| = |\mathcal{H}_{C'}| + |\mathcal{H}'_{C'}|$. З пер. вище:

$$|\mathcal{H}_C| \le |\{B\subset C \mid c_1\notin B \wedge \mathcal{H} \text{ розб. } B\}| + |\{B\subset C \mid c_1\in B\wedge\mathcal{H} \text{ розб. } B\}| =$$

$= \left[\begin{array}{c}\text{диз'юнктне}\\ \text{об'єднання}\end{array}\right] = |\{B\subset C \mid \mathcal{H} \text{ розб. } B\}|$, що й потрібно. △ [<u>Рч.1</u>].

<u>Рч.2.</u> Нехай $\mathcal{H}$ – клас гіпотез для задачі бін. клас. на мн. $\mathcal{X}$ з ф-цією зростання $\tau_{\mathcal{H}}$. Тоді $\forall$ розподілу $\mathcal{D}$ на $\mathcal{X}\times\{0,1\}$ $\forall\,\delta\in(0,1)$ вибірка S з m ел-тів, що i.i.d. відн. до $\mathcal{D}$ зі ймовірністю $\ge 1-\delta$ задовольняє умові

$$\forall h \in \mathcal{H} \quad |L_{\mathcal{D}}(h) - L_S(h)| \leq \frac{8}{\delta\sqrt{2m}}\left(1 + \sqrt{\ln \tau_{\mathcal{H}}(2m)}\right).$$

(за умови $\ln \tau_{\mathcal{H}}(2m) > 0$, тобто $\tau_{\mathcal{H}}(2m) \geq 3$).

▷ Отже, потрібно довести, що (екв. умова до формулювання):

$$\underset{S\sim\mathcal{D}^m}{P}\left(\sup_{h\in\mathcal{H}} |L_{\mathcal{D}}(h) - L_S(h)| > \underbrace{\frac{1}{\delta\sqrt{2m}}\left(4 + \sqrt{\ln \tau_{\mathcal{H}}(2m)}\right)}\right) < \delta$$

← $a > 0$ з нер-сті Маркова

В силу нерівності Маркова, для цього достатньо показати, що

$$\underset{S\sim\mathcal{D}^m}{E}\left(\sup_{h\in\mathcal{H}} |L_{\mathcal{D}}(h) - L_S(h)|\right) < \frac{8}{\sqrt{2m}}\left(1 + \sqrt{\ln \tau_{\mathcal{H}}(2m)}\right).$$

↑ вип. величина на $(\mathcal{X}\times\{0,1\})^m$, невід'ємна

Згадаймо, що $\forall h \quad L_{\mathcal{D}}(h) = \underset{S'\sim\mathcal{D}^m}{E} L_{S'}(h)$ (див. доведення $\underline{\text{P4}}$ про рівн. зб. скінч. класів). Тоді за лінійністю E:

$$\underset{S\sim\mathcal{D}^m}{E}\left(\sup_{h\in\mathcal{H}} |L_{\mathcal{D}}(h) - L_S(h)|\right) = \underset{S\sim\mathcal{D}^m}{E}\left(\sup_{h\in\mathcal{H}} \left|\underset{S'\sim\mathcal{D}^m}{E}\left(L_{S'}(h) - L_S(h)\right)\right|\right) \leq$$

$$\leq \left[E(|X|) \leq |E(X)| \text{ за власт. інтегрування}\right] \leq \underset{S\sim\mathcal{D}^m}{E}\left(\sup_{h\in\mathcal{H}} \underset{S'\sim\mathcal{D}^m}{E} |L_{S'}(h) - L_S(h)|\right) \leq$$

$$\leq \left[\begin{array}{l} E(X_h) \leq E(\sup_h X_h) \ \forall h \Rightarrow \\ \Rightarrow \sup_h E(X_h) \leq E(\sup_h X_h). \end{array}\right] \leq \underset{S\sim\mathcal{D}^m}{E}\left(\underset{S'\sim\mathcal{D}^m}{E}\left(\sup_{h\in\mathcal{H}} |L_{S'}(h) - L_S(h)|\right)\right) =$$

$$= \left[\begin{array}{l} S, S' \text{ i.i.d. відп.} \\ \text{до } \mathcal{D}^m \end{array}\right] = \underset{(S,S')\sim\mathcal{D}^{2m}}{E}\left(\sup_{h\in\mathcal{H}} |L_{S'}(h) - L_S(h)|\right) = \left[\begin{array}{l} \underline{\text{def.}}\ L_S\text{, тут} \\ \ell\text{ - це } \ell_{0-1}: \ell(h,(x,y)) = \\ = 0 \text{ при } h(x) = y; \\ 1 \text{ при } h(x) \neq y \end{array}\right] =$$

$$= \mathop{E}_{(S,S')=(z_1,\dots,z_m,z_1',\dots,z_m')\sim D^{2m}} \left(\sup_{h\in\mathcal{H}} \frac{1}{m} \left| \sum_{i=1}^m \ell(h,z_i') - \ell(h,z_i) \right| \right) \circledeq$$

$\forall\, i$ z_i, z_i' незалежно пробігають усі точки $\mathcal{X}\times\{0,1\}$, тому їх можна змінити місцями, отримавши $\ell(h,z_i)-\ell(h,z_i') = -(\ell(h,z_i') - \ell(h,z_i))$. Тому $\forall\, \sigma=(\sigma_1,\dots,\sigma_m)\in\{\pm1\}^m$

$$\circledeq \mathop{E}_{(S,S')\sim D^{2m}} \left(\sup_{h\in\mathcal{H}} \frac{1}{m} \left| \sum_{i=1}^m \sigma_i (\ell(h,z_i') - \ell(h,z_i)) \right| \right) \circledeq$$

Зокрема, це пост. вип. величина на $\{\pm1\}^m$ з однорідним розподілом $\mathcal{U}_\pm^m$ (коли $\forall$ набір з m ±1 має йм. $\frac{1}{2^m}$):

$$\circledeq \mathop{E}_{\sigma\sim\mathcal{U}_\pm^m} \mathop{E}_{(S,S')\sim D^{2m}} \left(\sup_{h\in\mathcal{H}} \frac{1}{m} \left| \sum_{i=1}^m \sigma_i (\ell(h,z_i') - \ell(h,z_i)) \right| \right) = \left[\begin{array}{l} \mathop{E}_{\sigma} \text{ - лін.} \\ \text{вираз, за} \\ \text{лін. } \mathop{E}_{(S,S')} \end{array} \right] =$$

$$= \mathop{E}_{(S,S')\sim D^{2m}} \mathop{E}_{\sigma\sim\mathcal{U}_\pm^m} \left(\sup_{h\in\mathcal{H}} \frac{1}{m} \left| \sum_{i=1}^m \sigma_i (\ell(h,z_i') - \ell(h,z_i)) \right| \right) \circledless$$

$\forall$ фікс. S, S' нехай C - множина їх ел-тів, $|C| \le 2m$. Тоді суттєвими є лише $h\in\mathcal{H}_C$:

$$\mathop{E}_{\sigma\sim\mathcal{U}_\pm^m} \left(\sup_{h\in\mathcal{H}} \frac{1}{m} \left| \sum_{i=1}^m \sigma_i (\ell(h,z_i') - \ell(h,z_i)) \right| \right) = \mathop{E}_{\sigma\sim\mathcal{U}_\pm^m} \left(\max_{h\in\mathcal{H}_C} \frac{1}{m} \left| \sum_{i=1}^m \sigma_i \right.\right.$$

$|\ell(h,z_i') - \ell(h,z_i))|)$.

$\forall h, i \quad \theta_{h,i} : \sigma_i \mapsto \sigma_i(\ell(h,z_i') - \ell(h,z_i))$ – вип. вел. на $\{\pm 1\}$ з рівн. розп. $\mathcal{U}_\pm$, що, оч., має $E(\theta_{h,i}) = 0$. Оск. ℓ приймає знач. 0 і 1, $\theta_{h,i} \in [-1, 1]$. Тому за нер. Хьоффдінга, $\forall \rho > 0$

$$P\left(\left|\frac{1}{m}\sum_{i=1}^{m}\theta_{h,i} - 0\right| > \rho\right) \le 2e^{\frac{-2m\rho^2}{(1-(-1))^2}} = 2e^{-\frac{m\rho^2}{2}}.$$

При цьому $\{\sigma \mid \max\limits_{h\in\mathcal{H}_C}\left|\frac{1}{m}\sum_{i=1}^{m}\theta_{h,i}\right| > \rho\} \subset \bigcup\limits_{h\in\mathcal{H}_C}\{\sigma \mid \left|\frac{1}{m}\sum_{i=1}^{m}\theta_{h,i}\right| > \rho\}$, тому за субад.

$$P\left(\max_{h\in\mathcal{H}_C}\left|\frac{1}{m}\sum_{i=1}^{m}\theta_{h,i}\right| > \rho\right) \le \sum_{h\in\mathcal{H}_C} P\left(\left|\frac{1}{m}\sum_{i=1}^{m}\theta_{h,i}\right| > \rho\right) \le 2|\mathcal{H}_C|e^{-\frac{m\rho^2}{2}}.$$

(можно вважати, що це так і при $\rho = 0$ за непер.).

Лем. 2. Нехай $X \ge 0$ – вип. вел., $a > 0$, $b \ge e$ і для усіх $t \ge 0$ $P(X > t) \le 2be^{-\frac{t^2}{a^2}}$. Тоді $E(X) \le a(4 + 3\sqrt{\ln b})$.

▷ Для $t_i := a(i + \sqrt{\ln b})$, $i \in \mathbb{Z}_+$ за $\underbrace{\text{власт.}}_{}$ (def.) E, оск. $t_i > t_{i-1}$ $\forall i$:

$$E(X) \le a\sqrt{\ln b}\cdot P(X \le \underbrace{a\sqrt{\ln b}}_{=t_0}) + \sum_{i=1}^{\infty} t_i \cdot P(t_{i-1} < X \le t_i) \le$$

$$\le a\sqrt{\ln b} + \sum_{i=1}^{\infty} t_i P(X > t_{i-1}) \le [\text{умова}] \le a\sqrt{\ln b} + \sum_{i=1}^{\infty} a(i + \sqrt{\ln b})\cdot$$

$$\cdot 2b\, e^{-(i-1+\sqrt{\ln b})^2} \le \left[\begin{array}{l} \forall i \ge 2 \text{ при } x \in [i-1+\sqrt{\ln b},\, i+\sqrt{\ln b}] \\ x e^{-(x-1)^2} \ge (i+\sqrt{\ln b})\, e^{-(i-1+\sqrt{\ln b})^2}, \\ \text{бо } x e^{-(x-1)^2} \downarrow \text{ при } x \ge \frac{1+\sqrt{3}}{2} \text{ (ацо } < 2 < 1+\sqrt{\ln b} \end{array}\right] \le a\sqrt{\ln b} +$$

$$+ \cancel{2ab(1+\sqrt{\ln b})\, e^{(1-1+\sqrt{\ln b})^2}} + \sum_{i=2}^{\infty} 2ab \int_{i-1+\sqrt{\ln b}}^{i+\sqrt{\ln b}} x e^{-(x-1)^2} dx = a(2+3\sqrt{\ln b}) +$$

$$+ 2ab \int_{1+\sqrt{\ln b}}^{+\infty} x e^{-(x-1)^2} dx = \left[\begin{array}{l} \text{заміна} \\ x-1=y \end{array}\right] = a(2+3\sqrt{\ln b}) + 2ab \int_{\sqrt{\ln b}}^{+\infty} (y+1) e^{-y^2} dy \le$$

$$\le \left[\begin{array}{l} b > e \Rightarrow y \ge \\ \ge \sqrt{\ln b} \ge 1 \end{array}\right] \le a(2+3\sqrt{\ln b}) + 4ab \int_{\sqrt{\ln b}}^{+\infty} y e^{-y^2} dy = a(2+3\sqrt{\ln b}) + 2ab \cdot$$

$$\underbrace{\left.\left(-e^{-y^2}\right)\right|_{\sqrt{\ln b}}^{+\infty}}_{\frac{1}{b}=} = a(4+3\sqrt{\ln b}) \; \triangle \; (\text{Лем. 2.}).$$

Т.ч., $\forall S, S'$ $\underset{\sigma \sim \mathcal{H}_\pm^m}{E}\left(\sup_{h \in \mathcal{H}} \left|\frac{1}{m}\sum_{i=1}^{m} \theta_{k,i}\right|\right) = \underset{\sigma \sim \mathcal{H}_\pm^m}{E}\left(\max_{h \in \mathcal{H}_c} \left|\frac{1}{m}\sum_{i=1}^{m} \theta_{k,i}\right|\right) <$

$< \sqrt{\frac{2}{m}}\left(4 + 3\sqrt{\ln |\mathcal{H}_c|}\right)$ $\le$ за умови $|\mathcal{H}_c| > e$ (тобто ≥ 3; можна вважати, що це вик. для майже всіх S, S' для дост. великих

m - чому?). За дов. $\tau_{\mathcal{H}}$, це $\leq \sqrt{\frac{2}{m}}\left(4+3\sqrt{\ln \tau_{\mathcal{H}}(|C|)}\right) \leq$

$\leq \left[\begin{array}{l}\text{Впр. перевірити, що} \\ \tau_{\mathcal{H}} \text{ моном. зростає}\end{array}\right] \leq \sqrt{\frac{2}{m}}\left(4+3\sqrt{\ln \tau_{\mathcal{H}}(2m)}\right) < \frac{8}{\sqrt{2m}}\left(1+\sqrt{\ln \tau_{\mathcal{H}}(2m)}\right).$

Оск. це не залежить від S, S', остаточно отримуємо

$< \frac{8}{\sqrt{2m}}\left(1+\sqrt{\ln \tau_{\mathcal{H}}(2m)}\right)$, що й потрібно. $\blacktriangle$ (Рч. 2.).

Тепер можемо завершити доведення Тн., тобто імпл. 6. => 1.

Отже, нехай $\mathcal{H}$: $VCdim(\mathcal{H}) = d < \infty$. Тоді за Рч. 1. для $2m > d+1$ $\tau_{\mathcal{H}}(2m) < \left(\frac{2em}{d}\right)^d$, отже за Рч. 2. $\forall \mathcal{D}, \delta \in (0,1)$ вибірка $S \sim \mathcal{D}^m$ зі ймов. $1-\delta$ задов. умові $\forall h \in \mathcal{H}$

$\left[\begin{array}{c}\Uparrow \\ (m > d)\end{array}\right]$

$$|L_{\mathcal{D}}(h) - L_S(h)| \leq \frac{8}{\delta\sqrt{2m}}\left(1+\sqrt{\ln \tau_{\mathcal{H}}(2m)}\right) < \frac{8}{\delta\sqrt{2m}}\left(1+\sqrt{d\ln\frac{2em}{d}}\right) \leq$$

Далі вважаємо, що $d\ln\frac{2em}{d} \geq 1$, тобто $m \geq \frac{d}{2e}e^{\frac{1}{d}}$:

$$\leq \frac{16}{\delta\sqrt{2m}}\sqrt{d\ln\frac{2em}{d}} = \frac{1}{\delta}\sqrt{\frac{128d}{m}\ln\frac{2em}{d}}.$$

За деф. рівн. збіжності, треба показати, що це $\leq \varepsilon$. Це виконується при $m \geq \frac{128d}{(\varepsilon\delta)^2}\ln m + \frac{128d}{(\varepsilon\delta)^2}\ln\frac{2e}{d}$. Залишилось використати наступне:

Лем. 3. Нехай $a \geq 1$ і $b > 0$. Тоді з $x \geq 4a \ln 2a + 2b$ випливає $x > a \ln x + b$.

▷ Оск. $a \geq 1$, $x > 2b$ (бо $\ln 2a > 0$). Оск. $b > 0$, $x \geq 4a \ln 2a$, тоді $x > 2a \ln x$ за наст. лемою. Т.ч., $2x > 2a \ln x + 2b$. △ (Лем. 3)

Лем. 4. Нехай $a > 0$. Тоді з $x \geq 2a \ln a$ випливає $x > a \ln x$.

▷ Розглянемо $f(x) := x - a \ln x$. $f'(x) = 1 - \frac{a}{x} > 0$ при $x > a$, і $f(2a \ln a) = 2a \ln a - a \ln(2a \ln a) = 2a \ln a - a \ln a - a \ln(2 \ln a) =$
$= a \ln a - a \ln(2 \ln a) = a \ln \frac{a}{2 \ln a} > 0$ при $a > 1$, бо $a > 2 \ln a$ ($\Leftrightarrow e^a > a^2$). При $\ln a > \frac{1}{2}$ $x \geq 2a \ln a \Rightarrow x > a$, тому $f(x) > 0$, що й потрібно; при $\ln a \leq \frac{1}{2}$, тобто $a \leq \sqrt{e}$
$x > 2 \ln x > \sqrt{e} \ln x \geq a \ln x$ завжди (зокрема, для $0 < x \leq 1$) △ (Лем. 4)

Т.ч., при

$$m \geq \frac{512 d}{(\varepsilon\delta)^2} \ln \frac{256 d}{(\varepsilon\delta)^2} + \frac{256 d}{(\varepsilon\delta)^2} \ln \frac{2e}{d}$$

умова виконана. △ (Тл.)

Зам. У останній частині доведення ми часто оцінку для складності вибірки $m_{\mathcal{H}}^{VC}(\varepsilon, \delta)$ у <u>дов.</u> рівном. збіжності. Кращу можна знайти у наст. теоремі (та її доведенні):

<u>Th*.</u> (Основна теорема теорії статистичного машинного навчання - кількісна версія). Існують такі $C_1, C_2, D_1, D_2, E_1, E_2 > 0$, що для будь-якого класу гіпотез $\mathcal{H}$ у задачі бінарної класифікації VC-вимірності $VCdim(\mathcal{H}) = d < \infty$ виконується наступне:

<u>1.</u> $\mathcal{H}$ має властивість рівномірної збіжності, причому його складність вибірки $m_{\mathcal{H}}^{VC}$ задов. умовам

$$\frac{C_1}{\varepsilon^2}\left(d + \ln\frac{1}{\delta}\right) \leq m_{\mathcal{H}}^{VC}(\varepsilon, \delta) \leq \frac{C_2}{\varepsilon^2}\left(d + \ln\frac{1}{\delta}\right) \quad \forall \varepsilon, \delta, d.$$

<u>2.</u> $\mathcal{H}$ агностично PAC навчаний, причому його складність вибірки $m_{\mathcal{H}}$ задов. умовам

$$\frac{D_1}{\varepsilon^2}\left(d + \ln\frac{1}{\delta}\right) \leq m_{\mathcal{H}}(\varepsilon, \delta) \leq \frac{D_2}{\varepsilon^2}\left(d + \ln\frac{1}{\delta}\right) \quad \forall \varepsilon, \delta, d.$$

<u>3.</u> $\mathcal{H}$ PAC навчаний, причому його складність вибірки $m_{\mathcal{H}}$ задов. умовам

$$\frac{E_1}{\varepsilon}\left(d+\ln\frac{1}{\delta}\right) \le m_{\mathcal{H}}(\varepsilon,\delta) \le \frac{E_2}{\varepsilon}\left(d\ln\frac{1}{\varepsilon}+\ln\frac{1}{\delta}\right)\ \forall\, \varepsilon,\delta,d.$$

<u>Rem.</u> Тобто тут константи не залежать від $\mathcal{H}$ (або d).

▷ Див. [<u>1.</u>, Ch. 28]. $\triangle$ (Перехід від <u>1.</u> до <u>2.</u> був у <u>Cor.</u> про рівн. зб. $\mathcal{H}$)

<u>Впр.</u> Порівняти ці оцінки з тими, що були отримані раніше, у випадку скінченного $\mathcal{H}$ (для $d \le \log_2|\mathcal{H}|$).

<u>Rem.</u> Для загальної задачі класифікації та деяких задач регресії (зокр., для ℓ_{sq}) існують узагальнення VC-вимірності, але, взагалі кажучи, не виникає еквівалентності між її скінченністю та рівномірною збіжністю.