

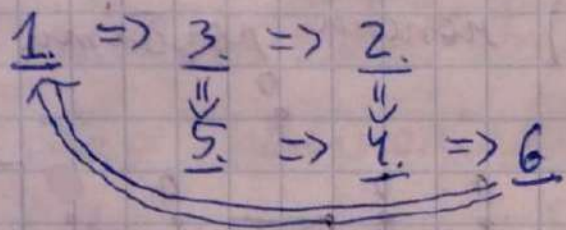
За поперед.  $P_n$ , умова  $VCdim(K) < \infty$  є необхідною для PAC  
навчальності  $K$ . Визначається, що вона є і достатньою, принаймні і  
для априорної навчальності. Ці умови пов'язані з теоремами:

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

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

6.  $\mathcal{K}$  має скінченну VC-вмірність.

► Подуємо схему залежності між цими твердженнями:



$\underline{1} \Rightarrow \underline{3} \Rightarrow \underline{2}$  доводимо у розділі про рівн. зв'язність (при цьому  $\underline{3} \Rightarrow \underline{2}$  за дов. см. PAC навч.),  $\underline{3} \Rightarrow \underline{5}$  і  $\underline{2} \Rightarrow \underline{4}$  - бо умови см. PAC навч. еквівалентні,  $\underline{5} \Rightarrow \underline{4}$  за дов. PAC навч. і  $\underline{4} \Rightarrow \underline{6}$  за попереднім Рс. Тому залишається довести  $\underline{6} \Rightarrow \underline{1}$ : з  $VCdim(\mathcal{K}) < \infty$  випливає, що він рівн. зв'язний. Це випливає з двох наступних тверджень:

дов. Функція зростання (growth function) класу гіпотез  $\mathcal{K}$  у заданій фін. клас. на  $n$ -х,  $X$  - це

$$T_{\mathcal{K}}: \mathbb{N} \rightarrow \mathbb{N} : T_{\mathcal{K}}(m) := \max_{\substack{S \subseteq X \\ |S|=m}} |\mathcal{K}_S|.$$

Копіюємо, що  $|\mathcal{K}_S|$  - це кількість різних об'єктів  $h \in \mathcal{K}$

на  $C$  и что за deb. VC-бук. хитро  $VCdim(\mathcal{H}) = d$ , но  $\exists C$ :  
 $|C| = m$  и  $|\mathcal{H}_C| = 2^m$  (тобто  $\mathcal{H}_C$  - все  $\varphi$ -ы на  $C$ ), тогда  
 $T_{\mathcal{H}}(m) = 2^m$  для  $m \leq d$ . Вспомогат.  $T_{\mathcal{H}}$  - асимптотична для  $d = \infty$ .  
 Вспомогат., что для любого  $m$   
 существует  $n$  и  $d$  такие, что  $n$  и  $d$  удовлетворяют:

Р. 1.

(Лема Загера, Сауер - Шелак - Перлес) Если  $\mathcal{H}$  -

класс  $n$ -омегз у заданн. фин. клас. на  $\mathcal{H}$ .  $VCdim(\mathcal{H}) = d < \infty$ . Тогда  $\forall m \in \mathbb{N}$

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

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

► Стратегия доказательства: рассуждения:

Лем. 1.  $\forall m, d \in \mathbb{N}$  такие, что  $m > d+1$   $\sum_{i=0}^d C_m^i < \left(\frac{em}{d}\right)^d$

► Индукция по  $d$ :

$d=1$ .  $\sum_{i=0}^1 C_m^i = 1+m < \left[\frac{em}{m>1}\right] < em$ , что и требовалось.

$d \rightsquigarrow d+1$ . Опред.  $n$  и  $d$  такие, что  $d$  и  $n$  удовлетворяют:  $\sum_{i=0}^{d-1} C_m^i = \sum_{i=0}^d C_m^i + C_m^{d+1} \leq \left[\frac{em}{d}\right] < \left(\frac{em}{d}\right)^d + \frac{m(m-1)\dots(m-d)}{(d+1)!} =$

$$= \left(\frac{em}{d}\right)^d \left(1 + \left(\frac{d}{em}\right)^d \frac{m \dots (m-d)}{(d+1)!}\right) \leq \left[ \begin{matrix} n-d+1, \dots \\ \dots, m-1 \leq m \end{matrix} \right] \leq \left(\frac{em}{d}\right)^d \left(1 + \left(\frac{d}{e}\right)^d \frac{m-d}{(d+1)!}\right) \leq$$

З габегенна дэпрыжэн Смітска  $d! \sim_{d \rightarrow \infty} \left(\frac{d}{e}\right)^d \sqrt{2\pi d}$  будзе,

уо іі права частка - гл агульна злучы, уо тупа  $\forall d$ :

$$d! \geq \left(\frac{d}{e}\right)^d \sqrt{2\pi d} \text{ мору}$$

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

$$< [2\pi d > 4] < \left(\frac{em}{d}\right)^d \frac{1}{d+1} \left(d+1 + \frac{m-d}{2}\right) \leq \left[ \begin{matrix} \text{э гучнае:} \\ \frac{d}{2} + 1 + \frac{m}{2} \leq m, \\ \text{уо } d+2 \leq m \end{matrix} \right] \leq \left(\frac{em}{d}\right)^d \frac{m}{d+1} \leq$$

З імям дэжы,

$$\left(\frac{em}{d+1}\right)^{d+1} / \left(\frac{em}{d}\right)^d = \frac{em}{d+1} \left(\frac{d}{d+1}\right)^d > \left[ \begin{matrix} \text{агульна частка e:} \\ \left(1 + \frac{1}{d}\right)^d \xrightarrow{d \rightarrow \infty} e \text{ з'ява} \end{matrix} \right] > \frac{m}{d+1} \left(1 + \frac{1}{d}\right)^d$$

$$\cdot \left(\frac{d}{d+1}\right)^d = \frac{m}{d+1}, \text{ мору напярэмі } \leq \left(\frac{em}{d+1}\right)^{d+1}, \text{ уо іі нончідэс}$$

(прачыну іі гавару строгі пэрыфірм на  $\forall$  крозі). Лема 1.

Керне нб. Рл. 1. вандувае з наступнаго:  $\forall K$  (мун дэз нун. нб  $V(\dim)$ )

$$\forall C: |C|=m \quad |K_C| \leq |\{B \subset C \mid K \text{ раздубае } B\}|. \quad (*)$$

Діцна,  $V(\dim(K)) = d$  азначае, уо  $K$  може раздубавацца дубе  
нунчідэс з  $\leq d$  ел-мів, мору права частка  $(*)$

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

Доведемо (\*) індукцією за  $m$  (для всіх  $K$ ).

$m=1$ . Для  $|C|=1$  ліва і справа у (\*) мають одночасно або 1 (коли  $\forall i \in K$  на  $C = 0$  або  $\forall i = 1$ , тобто  $C$  не розд. і  $K$  розд. завжди), або 2 (коли  $C$  розд.:  $\exists h \in K$ , що  $= 0$  на  $C$  і  $\exists h$ , що  $= 1$ ).

$m-1 \rightarrow m$ . Припустимо, що (\*) вірно для  $|C| = m-1$ . Нехай  $|C| = m$ :  $C = \{c_1, \dots, c_m\}$ . Покладемо  $C' := C \setminus \{c_1\} = \{c_2, \dots, c_m\}$ .  
 Поді за прип. індукції (\* для  $C'$ ):

$$|K_{C'}| \leq |\{B \subset C' \mid K \text{ розд. } B\}| = |\{B \subset C \mid c_1 \notin B \wedge K \text{ розд. } B\}|.$$

Тепер покладемо

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

тобто  $K'$  сам. з пар ізометр  $K$ , що узгоджені на  $C'$  і відрізняються на  $c_1$  (з пар, до  $h'$ , оч.,  $h \in K'$  у повн. базисі).

Знову за прип. інд.

$$|\mathcal{K}'_c| \leq |\{B \subset C' \mid \mathcal{K}' \text{ розд. } B\}| = \left[ \begin{array}{l} h \in \mathcal{K}' \text{ приймають } y_{C'} \\ \text{можл. ж. на } B \Rightarrow \text{отримав} \\ \text{її на } B \cup \{C_1\} \text{ за визн. } \mathcal{K}' \end{array} \right] =$$

$$= |\{B \subset C' \mid \mathcal{K}' \text{ розд. } B \cup \{C_1\}\}| = |\{B \subset C \mid C_1 \in B \wedge \mathcal{K}' \text{ розд. } B\}| \leq \\ \leq [\mathcal{K}' \subset \mathcal{K}] \leq |\{B \subset C \mid C_1 \in B \wedge \mathcal{K} \text{ розд. } B\}|.$$

Тепер покажемо, що  $|\mathcal{K}_c| = |\mathcal{K}_c| + |\mathcal{K}'_c|$ . Дійсно, розглянемо  
 $\mathcal{K}_c \rightarrow \mathcal{K}_c' : (h(C_1), h(C_2), \dots, h(C_m)) \mapsto (h(C_2), \dots, h(C_m)) \quad \forall h \in \mathcal{K}.$

Це функція зливання, але не ізої : для  $h \in \mathcal{K} \setminus \mathcal{K}'$   $h(C_1)$  означає значення  $(h(C_2), \dots, h(C_m))$ , а для  $h \in \mathcal{K}'$  в  $(h(C_2), \dots, h(C_m))$  переходимо перш  $(0, h(C_2), \dots, h(C_m))$  і  $(1, h(C_2), \dots, h(C_m))$ . Тоді  $|\mathcal{K}_c| = |(\mathcal{K} \setminus \mathcal{K}')_c| + 2|\mathcal{K}'_c| = |\mathcal{K}_c| - |\mathcal{K}'_c| + 2|\mathcal{K}'_c| = |\mathcal{K}_c| + |\mathcal{K}'_c|$ . З пер. виведе:

$$|\mathcal{K}_c| \leq |\{B \subset C \mid C_1 \notin B \wedge \mathcal{K} \text{ розд. } B\}| + |\{B \subset C \mid C_1 \in B \wedge \mathcal{K} \text{ розд. } B\}| = \\ = \left[ \begin{array}{l} \text{визначення} \\ \text{об'єктів} \end{array} \right] = |\{B \subset C \mid \mathcal{K} \text{ розд. } B\}|, \text{ що і потрібно. } \triangle (\text{Рч.1}).$$

Рч.2. Нехай  $\mathcal{K}$  - клас шпектрів для задані дин. клас. на мн.  $X$  з  $\Phi$ -цією зростання  $T_{\mathcal{K}}$ . Тоді  $\forall$  розподілу  $\mathcal{Q}$  на  $X \times \{0,1\}$   $\forall \delta \in (0,1)$  вибірка  $S$  з  $n$  ел-тів, що і.і.д. визн. го  $\mathcal{Q}$  зі ймовірністю  $> 1 - \delta$  задовольняє умові

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

(за умови  $\ln T_{\mathcal{K}}(2m) > 0$ , маємо  $T_{\mathcal{K}}(2m) \geq 3$ ).

Отже, намірно говоримо, що (єв. умова го формулювання):

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

← ато з лем. Маркова

В силу нерівності Маркова, ми зможемо достатньо показати, що

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

↑  
єв. беремо на  $(\mathcal{X} \times \{0,1\})^m$ ,  
небіз'євна

Згадаймо, що  $\forall h \quad L_{\mathcal{D}}(h) = \mathbb{E}_{S' \sim \mathcal{D}^m} L_{S'}(h)$  (єв. говоримо РЧ. про рівн. зб. класів). Тоді за лінійністю  $\mathbb{E}$ :

$$\begin{aligned} \mathbb{E}_{S \sim \mathcal{D}^m} \left( \sup_{h \in \mathcal{K}} |L_{\mathcal{D}}(h) - L_S(h)| \right) &= \mathbb{E}_{S \sim \mathcal{D}^m} \left( \sup_{h \in \mathcal{K}} \left| \mathbb{E}_{S' \sim \mathcal{D}^m} (L_{S'}(h) - L_S(h)) \right| \right) \leq \\ &\leq \left[ \mathbb{E}(|X|) \leq |\mathbb{E}(X)| \text{ за власт. інтегрування} \right] \leq \mathbb{E}_{S \sim \mathcal{D}^m} \left( \sup_{h \in \mathcal{K}} \mathbb{E}_{S' \sim \mathcal{D}^m} |L_{S'}(h) - L_S(h)| \right) \leq \\ &\leq \left[ \mathbb{E}(X_k) \leq \mathbb{E}(\sup_k X_k) \quad \forall k \Rightarrow \right. \\ &\quad \left. \Rightarrow \sup_k \mathbb{E}(X_k) \leq \mathbb{E}(\sup_k X_k) \right] \leq \mathbb{E}_{S \sim \mathcal{D}^m} \left( \mathbb{E}_{S' \sim \mathcal{D}^m} \left( \sup_{h \in \mathcal{K}} |L_{S'}(h) - L_S(h)| \right) \right) = \\ &= \left[ S, S' \text{ i.i.d. sign.} \right] = \mathbb{E}_{(S, S') \sim \mathcal{D}^{2m}} \left( \sup_{h \in \mathcal{K}} |L_{S'}(h) - L_S(h)| \right) = \left[ \begin{array}{l} \text{def. } L_S; \text{ mym} \\ \text{L-se } L_{0-1}: \ell(h(x, y)) = \\ = 0 \text{ при } h(x) = y; \\ = 1 \text{ при } h(x) \neq y \end{array} \right] = \end{aligned}$$

$$= E_{(S, S') = (z_1, \dots, z_m, z'_1, \dots, z'_m) \sim \mathcal{D}^{2m}} \left( \sup_{h \in \mathcal{H}} \frac{1}{m} \left| \sum_{i=1}^m \ell(h, z'_i) - \ell(h, z_i) \right| \right) \quad (\equiv)$$

$\forall i$   $z_i, z'_i$  незалежно вибирають 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  $\forall \sigma = (\sigma_1, \dots, \sigma_m) \in \{\pm 1\}^m$

$$(\equiv) E_{(S, S') \sim \mathcal{D}^{2m}} \left( \sup_{h \in \mathcal{H}} \frac{1}{m} \left| \sum_{i=1}^m \sigma_i (\ell(h, z'_i) - \ell(h, z_i)) \right| \right) \quad (\equiv)$$

Зокрема, це пост. фін. величина на  $\{\pm 1\}^m$  з однаковою розподілом  $\mathcal{U}_{\pm}^m$  (коли  $\forall$  набір  $z$   $m \pm 1$  має ім.  $\frac{1}{2^m}$ ):

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

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

$\forall$  фікс.  $S, S'$  нехай  $\mathcal{C}$  - множина їх ел-тів,  $|\mathcal{C}| \leq 2m$ ,

Тоді суттєвими є лише  $h \in \mathcal{H}_{\mathcal{C}}$ :

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

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

$$\forall h, i \quad \theta_{h,i} : \mathcal{G}_i \mapsto \mathcal{G}_i(\ell(h, z_i) - \ell(h, z_i)) - \text{вип. вел.}$$

на  $\{\pm 1\}$  з рівн. розп.  $\mathcal{U}_{\pm}$ , що, от., має  $E(\theta_{h,i}) = 0$ .

Оскільки  $\ell$  приймає знач. 0 і 1,  $\theta_{h,i} \in [-1, 1]$ . Тоді за кер.

Хвосторівна,  $\forall p > 0$

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

$$\text{Тоді якщо } \left\{ \mathcal{G} \mid \max_{h \in \mathcal{H}_c} \left| \frac{1}{m} \sum_{i=1}^m \theta_{h,i} \right| > p \right\} \subset \bigcup_{h \in \mathcal{H}_c} \left\{ \mathcal{G} \mid \left| \frac{1}{m} \sum_{i=1}^m \theta_{h,i} \right| > p \right\}$$

$> p$ , тоді за след.

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

(можливо вважати, що це max і при  $p=0$  за керен.).

лем. 2. Нехай  $X \stackrel{7,0}{\sim} \text{вип. вел.}$ ,  $a > 0$ ,  $b \geq e$  і для всіх  $t \geq 0$

$$P(X > t) \leq 2be^{-\frac{t^2}{a^2}}. \text{ Тоді } E(X) \leq a(4 + 3\sqrt{\ln b}).$$

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

$$E(X) \leq a\sqrt{\ln b} \cdot P(X \leq a\sqrt{\ln b}) + \sum_{i=1}^{\infty} t_i \cdot P(t_{i-1} < X \leq t_i) \leq a\sqrt{\ln b} + \sum_{i=1}^{\infty} t_i P(X > t_{i-1}) \stackrel{7,0}{\leq} [y \text{ мова}] \leq a\sqrt{\ln b} + \sum_{i=1}^{\infty} a(i + \sqrt{\ln b}).$$

$$\begin{aligned}
& \cdot 2b e^{-(i-1+\sqrt{\ln b})^2} \leq \left[ \begin{aligned} & \forall i \geq 2 \text{ при } x \in [i-1+\sqrt{\ln b}, i+\sqrt{\ln b}] \\ & x e^{-(x-1)^2} > (i+\sqrt{\ln b}) e^{-(i-1+\sqrt{\ln b})^2} \\ & \text{до } x e^{-(x-1)^2} \downarrow \text{ при } x > \frac{1+\sqrt{3}}{2} \text{ (до } 2 < 1+\sqrt{\ln b}) \end{aligned} \right] \leq a \sqrt{\ln b} + \\
& + 2ab (1+\sqrt{\ln b}) e^{(1-1+\sqrt{\ln b})^2} + \sum_{i=2}^{\infty} 2ab \int_{i-1+\sqrt{\ln b}}^{i+\sqrt{\ln b}} x e^{-(x-1)^2} dx = a(2+3\sqrt{\ln b}) + \\
& + 2ab \int_{1+\sqrt{\ln b}}^{+\infty} x e^{-(x-1)^2} dx = \left[ \begin{aligned} & \text{замена} \\ & x-1=y \end{aligned} \right] = a(2+3\sqrt{\ln b}) + 2ab \int_{\sqrt{\ln b}}^{+\infty} (y+1) e^{-y^2} dy \leq \\
& \leq \left[ \begin{aligned} & b > e \Rightarrow y > 1 \\ & y > \sqrt{\ln b} \geq 1 \end{aligned} \right] \leq a(2+3\sqrt{\ln b}) + 4ab \int_{\sqrt{\ln b}}^{+\infty} y e^{-y^2} dy = a(2+3\sqrt{\ln b}) + 2ab \cdot \\
& \left( -e^{-y^2} \right) \Big|_{\sqrt{\ln b}}^{+\infty} = a(4+3\sqrt{\ln b}) \quad \triangleq \quad (\text{лем. 2.}).
\end{aligned}$$

$\frac{1}{b} =$  и.ч.,  $\forall S, S'$   $E_{G \sim \mathcal{K}_{\pm}^m} \left( \sup_{k \in \mathcal{K}} \left| \frac{1}{m} \sum_{i=1}^m \theta_{k,i} \right| \right) = E_{G \sim \mathcal{K}_{\pm}^m} \left( \max_{k \in \mathcal{K}_c} \left| \frac{1}{m} \sum_{i=1}^m \theta_{k,i} \right| \right) <$   
 $< \sqrt{\frac{2}{m}} (4 + 3 \sqrt{\ln |\mathcal{K}_c|}) \quad \textcircled{5}$  за услови  $|\mathcal{K}_c| > e$  (можно  $\geq 3$ ; конечна  
 ввексання, що це викн. для найменше всіх  $S, S'$  для дост. великих

$m$ -чову? ). За доб.  $T_K$ , че  $(\leq) \sqrt{\frac{2}{m}} (4 + 3 \sqrt{\ln T_K(|C|)}) \leq$   
 $\leq \left[ \begin{smallmatrix} \text{впр. невіримо, що} \\ T_K \text{ монот. зростає} \end{smallmatrix} \right] \leq \sqrt{\frac{2}{m}} (4 + 3 \sqrt{\ln T_K(2m)}) < \frac{8}{\sqrt{2m}} (1 + \sqrt{\ln T_K(2m)}).$

Оскільки не задані  $S, S'$ , остаточно отримувемо

$(\leq) \frac{8}{\sqrt{2m}} (1 + \sqrt{\ln T_K(2m)}),$  що й потрібно.  $\Delta$  (Р.2.).

Тепер можна завершити доведення  $T_K$ , маємо ікла. 6.  $\Rightarrow$  1.

Оскільки, нехай  $K: \forall \dim(K) = d < \infty$ . Тоді за Р.1. для

$2m > d+1$   $T_K(2m) < \left(\frac{2em}{d}\right)^d$ , отже за Р.2.  $\forall D, \delta \in (0,1)$   
 $\underbrace{(m \geq d)}_{\uparrow}$  вибірка  $S \sim D^m$  зі імов.  $1-\delta$  задов. умови  $\forall h \in K$   
 $|L_D(h) - L_S(h)| \leq \frac{8}{\delta \sqrt{2m}} (1 + \sqrt{\ln T_K(2m)}) < \frac{8}{\delta \sqrt{2m}} (1 + \sqrt{d \ln \frac{2em}{d}}) (\leq)$

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

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

За доб. рівн. зв'язності, треба показати, що це  $\leq \epsilon$ . Це  
 виконується при  $m \geq \frac{128d}{(\epsilon\delta)^2} \ln m + \frac{128d}{(\epsilon\delta)^2} \ln \frac{2e}{d}$ . Замикаємо  
 виходячи насправді.

Лем. 3. Если  $a > 1$  и  $b > 0$ , то из  $x \geq 4a \ln 2a + 2b$  вытекает  $x > a \ln x + b$ .

► Если  $a > 1$ ,  $x > 2b$  (то  $\ln 2a > 0$ ). Если  $b > 0$ ,  $x \geq 4a \ln 2a$ , то из  $x \geq 2a \ln x$  по лемме. 1. ч.,  $2x \geq 2a \ln x + 2b$ .  $\Delta$  (Лем. 3.)

Лем. 4. Если  $a > 0$ , то из  $x \geq 2a \ln a$  вытекает  $x > a \ln x$ .

► Рассмотрим  $f(x) := x - a \ln x$ .  $f'(x) = 1 - \frac{a}{x} > 0$  при  $x > a$ , и  $f(2a \ln a) = 2a \ln a - a \ln(2a \ln a) = 2a \ln a - a \ln a - a \ln(2 \ln a) = a \ln a - a \ln(2 \ln a) = a \ln \frac{a}{2 \ln a} > 0$  при  $a > 1$ , то  $a > 2 \ln a$  ( $\Leftrightarrow e^a > a^2$ ). При  $\ln a > \frac{1}{2}$   $x \geq 2a \ln a \Rightarrow x > a$ , тогда  $f(x) > 0$ , что и требуется; при  $\ln a \leq \frac{1}{2}$ , тогда  $a \leq \sqrt{e}$   $x > 2 \ln x > \sqrt{e} \ln x > a \ln x$  завершено (замечая, что  $0 < x \leq 1$ ).  $\Delta$  (Лем. 4.)  
И. ч., при

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

условие выполнено.  $\Delta$  (Тн.)

Рем. В оставшейся части говодеана ми гаетмо

оцінку для складності вибірки  $m_{\mathcal{H}}^{VC}(\epsilon, \delta)$  у дев. рівном.  
збігності. Крайню можна знайти у наст. теоремі (на її доведенні):

Тл\* (Основна теорема теорії статистичного машинного навчання -  
кілкісна 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  $VCdim(\mathcal{H}) = d < \infty$  виконуються наступне:

1.  $\mathcal{H}$  має властивість рівномірної збігності, причому його  
складність вибірки  $m_{\mathcal{H}}^{VC}$  задов. умовам

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

2.  $\mathcal{H}$  асимптотично PAC навчаний, причому його складність  
вибірки  $m_{\mathcal{H}}$  задов. умовам

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

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

$$\frac{E_1}{\varepsilon} \left( d + \ln \frac{1}{\delta} \right) \leq m_K(\varepsilon, \delta) \leq \frac{E_2}{\varepsilon} \left( d \ln \frac{1}{\varepsilon} + \ln \frac{1}{\delta} \right) \forall \varepsilon, \delta, d.$$

Рем. Теорема тут константи не залежать від  $K$  (або  $d$ ).

⇒ Дав. [1, Сн. 28].  $\Delta$  (Порівняв від 1-го до 2-го в у Соч. про рівн. зб.  $K$ )

Вопр. Порівняти ці оцінки з тими, що були отримані раніше, у випадку скінченних  $K$  (для  $d \leq \log_2 |K|$ ).

Рем. Для загальної задачі класифікації та деяких задач регресії (зокр., для  $L_2$ ) існують узагальнення VC-вигірності, але, взагалі кажучи, не викликає ефівальності між її скінченністю та рівномірною збіжністю.

Рем. Хоча за  $T_n$  для бін. класифікації PAC навчальність та асимптотична PAC навчальність екв., на практиці між ними може бути різниця не лише у розмірі  $m_n$ . Так, ERM-алгоритми для парадерінегів на  $\mathbb{R}^d$  і булевих кон'юнкцій на  $\{0, 1\}^d$  поліноміальні за  $d$  і  $|S|=m$  (див. практику) за умови реалізованості, але NP-складні без неї. Дав. обговорення обчислюваності і посилання у [1, 4.8]