

2) / 3 (на оцінку) Розглянемо ще один клас функцій f -рів на $\{0,1\}^d$ - функції парності (parity functions): для довільної підмножини $I = \{i_1, \dots, i_k\} \subset \{1, \dots, d\}$ (де $i_1 < \dots < i_k$; k тут не фіксоване, а довільне big o of d).

$$h_I(x_1, \dots, x_d) := (x_{i_1} + \dots + x_{i_k}) \bmod 2,$$

тобто застосовуємо до $x_{i_1}, \dots, x_{i_k}$ бінарну функцію (XOR). Зокрема, сюди входить постійна 0 (для $I = \emptyset$), але не постійна 1.

Знайти $VCdim(\mathcal{H})$ для $\mathcal{H} = \{h_I\}_{I \subset \overline{1,d}}$ (точне значення).

9/3. $\mathcal{K}$ має стільки не єл-тів, стільки підмножини у
 $\{1, \dots, d\}$, тобто 2^d . Тому оцінка зверху $V(\dim(\mathcal{K})) \leq$
 $\leq \log_2 |\mathcal{K}| = \underline{d}$. Нагадаємо, як вона поводиться! $\forall C$
 $\exists |C| = d+1 \exists 2^{d+1}$ оп-цій $C \rightarrow \{0, 1\}$, отже C не
 розбивається $\mathcal{K}$ $\exists |\mathcal{K}| = 2^d$.

Покажем, что $VCdim(\mathcal{H}) \geq d$. Для этого достаточно
 мы не мучаясь, что у задачи про линейность -
 $C = \{e_i\}_{i=1}^d$, где $e_i = (0, \dots, 1, \dots, 0)$. Рассмотрим $h: C \rightarrow \{0, 1\}$
 гом. ф-ция: $h(e_{i_1}) = \dots = h(e_{i_k}) = 1$, $h(e_j) = 0$ для остальных
 $j = \overline{1, d}$. Тогда $h_{\{i_1, \dots, i_k\}}: (x_1, \dots, x_d) \mapsto \sum_{l=1}^k x_{i_l} \pmod{2}$ на-
 лаженого по $\mathcal{H}$ и где мы не знаем, $\forall l = \overline{1, k}$
 $e_{i_l} = (0, \dots, 1, \dots, 0) \mapsto 0 + \dots + 1 + \dots + 0 \pmod{2} = 1$ и $\forall j \neq i_1, \dots, i_k$
 $e_j = (0, \dots, 1, \dots, 0) \mapsto 0$. Ит.д., $\mathcal{H}$ разделяет C , что и дает требуемое.
 Отсюда, $VCdim(\mathcal{H}) = d$

1. Представим задачу ERM-проблемы для задачи минимизации
 функции χ на $\mathbb{R}^d$ ф-цией потерь $\ell(h, (x, y)) = |h(x) - y|$ как
 задачу линейного программирования.

Отсюда, если даны пары $S = \{(x_i, y_i)\}_{i=1}^m$ и необходимо найти

$w \in \mathbb{R}^d$, що minimize

$$\sum_{i=1}^m |\langle w, x_i \rangle - y_i|$$

(тут розглядаємо одновимірний варіант; як показано у лекції, загальний можна звести до цього).

Задаємо означення модуля: $\forall i, w$

$$\langle w, x_i \rangle \geq y_i \Rightarrow |\langle w, x_i \rangle - y_i| = \langle w, x_i \rangle - y_i$$

$$\langle w, x_i \rangle \leq y_i \Rightarrow |\langle w, x_i \rangle - y_i| = y_i - \langle w, x_i \rangle$$

$$\Rightarrow [\forall \forall \text{ var.}] \Rightarrow$$

$$\Rightarrow \langle w, x_i \rangle - |\langle w, x_i \rangle - y_i| \leq y_i \leq \langle w, x_i \rangle + |\langle w, x_i \rangle - y_i|.$$

Тому, позначивши $\varepsilon_i := |\langle w, x_i \rangle - y_i| \forall i$, маємо задачу мінімізації $\varepsilon_1 + \dots + \varepsilon_m$ за лін. умов

$$\langle w, x_i \rangle + \varepsilon_i \geq y_i$$

$$\varepsilon_i - \langle w, x_i \rangle \geq -y_i$$

Побудуємо формально: можна розв'язати задачу LP

$\bar{w} = (w_1, \dots, w_d, \varepsilon_1, \dots, \varepsilon_m)$ для лінійної оп-ції $\langle u, \bar{w} \rangle$, де $u =$

на $\mathbb{R}^{d+m}$

Діємо, нехай $d = m$ і $S = \{(e_i, 1)\}_{i=1}^m$, де, як
завжди, $e_i = (0, \dots, 1, \dots, 0)$. Тоді $R = \max \{e_i\} = 1$.
Для $w = (1, \dots, 1)$ $y_i \langle w, e_i \rangle = 1 \cdot 1 = 1 > 1 \quad \forall i$, а
 $|w|^2 = 1 + \dots + 1 = m \Rightarrow |w| = \sqrt{m}$. Напевно, перешли

на цих двох кроках наступний крок: $w^{(1)} = 0$, тоді
 $y_1 \langle w^{(1)}, e_1 \rangle = 0$. Отже, $w^{(2)} = w^{(1)} + y_1 e_1 = (1, 0, \dots, 0)$.
Але тоді $y_2 \langle w^{(2)}, e_2 \rangle = 0$ (при цьому $y_1 \langle w^{(2)}, e_1 \rangle =$
 $= 1 > 0$), тоді $w^{(3)} = w^{(2)} + y_2 e_2 = (1, 1, 0, \dots, 0)$ etc.

Ал. завершити роботу після m кроків з $w^{(m+1)} = (1, \dots, 1)$.

б. Показати те саме для $d=3$: $\forall m \exists S = \{(x_i, y_i) \in$
 $\in \mathbb{R}^3 \times \{\pm 1\}\}_{i=1}^m$, така, що перешли завершити роботу
за m кроків і верна оцінка у теоремі для $m \in \underline{\mathbb{D}(3)}$

3. Покажем, что n -ая $A = \sum_{i=1}^m x_i x_i^T$, где $\{x_i\}_{i=1}^m \subset \mathbb{R}^d$ - векторы-столбцы, невырождена $\Leftrightarrow \text{span } \{x_i\}_{i=1}^m = \mathbb{R}^d$, тогда $\forall y \in \mathbb{R}^d \exists \lambda_1, \dots, \lambda_d \in \mathbb{R} : y = \sum_{i=1}^m \lambda_i x_i$ ($\{x_i\}$ полный).

Отсюда, пусть $x_i = \begin{pmatrix} x_{i,1} \\ \vdots \\ x_{i,d} \end{pmatrix} \forall i$. Тогда $A = \sum_{i=1}^m \begin{pmatrix} x_{i,1} \\ \vdots \\ x_{i,d} \end{pmatrix} (x_{i,1} \dots x_{i,d}) =$

$$= \begin{pmatrix} \sum_i x_{i,1} x_{i,1} & \sum_i x_{i,1} x_{i,2} & \dots & \sum_i x_{i,1} x_{i,d} \\ \sum_i x_{i,2} x_{i,1} & \sum_i x_{i,2} x_{i,2} & \dots & \sum_i x_{i,2} x_{i,d} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_i x_{i,d} x_{i,1} & \sum_i x_{i,d} x_{i,2} & \dots & \sum_i x_{i,d} x_{i,d} \end{pmatrix} = \begin{pmatrix} \langle y_1, y_1 \rangle & \langle y_1, y_2 \rangle & \dots & \langle y_1, y_d \rangle \\ \langle y_2, y_1 \rangle & \langle y_2, y_2 \rangle & \dots & \langle y_2, y_d \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle y_d, y_1 \rangle & \langle y_d, y_2 \rangle & \dots & \langle y_d, y_d \rangle \end{pmatrix},$$

где $\forall \bar{y} = 1, d \quad y_{\bar{y}} = \begin{pmatrix} x_{1,\bar{y}} \\ \vdots \\ x_{m,\bar{y}} \end{pmatrix}$ - вектор $\bar{y}$ -й столб. x_i , $y_{\bar{y}} \in \mathbb{R}^m$.

Тогда за m стр. каноничная матрица

$$A = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,d} \\ x_{2,1} & x_{2,2} & \dots & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \dots & x_{m,d} \end{pmatrix} \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,d} \\ x_{2,1} & x_{2,2} & \dots & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \dots & x_{m,d} \end{pmatrix} = X X^T, \text{ где } X - d \times m \text{-матр.}$$

столбцами x_i - столб. x_i . Как известно, $\text{rank } X X^T = \text{rank } X$, тогда A невр. $\Leftrightarrow \text{rank } A = d \Leftrightarrow \text{rank } X = d \Leftrightarrow$ столбцы X линейно независимы $d \Leftrightarrow \text{span } \{x_i\} = \mathbb{R}^d$.

1) заданы линейные отображения $f: \mathbb{R}^d \rightarrow \mathbb{R}^d$, $g: \mathbb{R}^d \rightarrow \mathbb{R}^m$, $g^*: \mathbb{R}^d \rightarrow \mathbb{R}^m$, $g^* = g^T$.
 A - это нормальная матрица. $g^* = g^T$.
 2) $\text{Im } f = \text{Im } (g \circ g^*) = \text{Im } g = \text{span} \{x_i\}_{i=1}^m$.

Д(3) (не на оцінку)

1. В задані дін. клас. на $\mathbb{R}^d$ оцінати $V(\cdot)$ -випірність класу замкнених куль (можна і з характ. ф-ції).

$$\mathcal{K} = \left\{ h_{\sigma, \epsilon} : x \mapsto \begin{cases} 1, & |x - \sigma| \leq \epsilon \\ 0 & \text{і ин. вип.} \end{cases} \right\} \quad \begin{matrix} \sigma \in \mathbb{R}^d \\ \epsilon \in [0, +\infty) \end{matrix}$$

(раніше ми доводили, що підклас $\{\sigma = 0\}$ має $V(\cdot)$ -вип. 1).

2. Нехай $\eta > 0$ - довільне. Розглянемо η -перцептрон, що діє як звичайний, але замість $w^{(t+1)} := w^{(t)} + \eta \cdot x_i$ робить $w^{(t+1)} := w^{(t)} + \eta \cdot y_i \cdot x_i$. Показати що $\forall$ навч. даних S такий алгоритм закінчує роботу за ту ж кількість кроків, що звичайний перцептрон, і повертає результат w_η , що пропорційний результату звичайного ($w_\eta = \lambda w$ для деякого $\lambda > 0$).

3. Задача 2.6 вище.