# Лекція № 3

# ПЕРСЕПТРОН — ЛІНІЙНИЙ КЛАСИФІКАТОР

Проф. Куссуль Н.М.

# 3.1. Модель персептрона

**Найпростіша мережа, що складається з одного нейрона, називається персептроном. Модель персептрона має вигляд, показаний на рис. 3.1.**

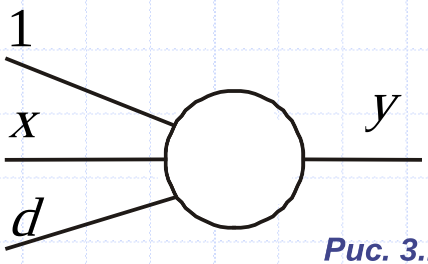



*Рис. 3.1.*

**При цьому** $x \in R^d$, **або** $x \in \{-1, 1\}^d$,
$y \in R$, **або** $y \in \{-1, 1\}$.

**Розглянемо випадок** $x \in R^d$, $y \in \{-1, 1\}$

**Функціонування персептрона описується залежністю:**

$y = \operatorname{sign}(W^{\mathrm{T}} x - \tau) = \eta(x, W)$ **(3.1)** **де** $\tau$ **– деякий поріг.**

**Ця модель персептрона — лінійного порогового елемента (LTU — linear threshold unit) — перша модель нейронної мережі. Вона була запропонована Розенблаттом (Frank Rosenblatt) (рис. 3.1).**

**У геометричній інтерпретації рівняння (3.1) визначає 2 підпростори**

$$\begin{aligned} \{x : y = 1\} \quad &\Leftrightarrow \quad H^{+} = \{x : W^{\mathrm{T}} x \geq \tau\} \\ \{x : y = -1\} \quad &\Leftrightarrow \quad H^{-} = \{x : W^{\mathrm{T}} x < \tau\} \end{aligned} \qquad (3.2)$$

**з розділяючою гіперплощиною (афінний підпростір розмірності** $d-1$**):** $H = \{x : W^{\mathrm{T}} x - \tau = 0\}$ **(3.3)**

**Однорідність рівняння (3.3) дозволяє пронормувати його відносно** $\tau$**, тобто вважати**

$$\tau = \begin{cases} \pm 1, & \text{якщо } \tau \neq 0, \\ 0 \end{cases}$$

**Збільшуючи розмірність простору, отримаємо** $x \in R^d \;\Rightarrow\; \tilde{x} \in R^{d+1}$, **(3.4)**

**де** $\forall i \leq d \quad \tilde{x}_i = x_i, \; \tilde{x}_{d+1} = -1$

$W \in R^d \;\Rightarrow\; \tilde{W} \in R^{d+1}$, **де** $\forall i \leq d \quad \tilde{W}_i = W_i, \tilde{W}_{d+1} = \tau$ **(3.5)**

**Враховуючи (3.4) та (3.5), можна записати** $W^{\mathrm{T}} x - \tau = \tilde{W}^{\mathrm{T}} \tilde{x}$,

**де** $\tilde{x}_{d+1}$ **— Bias-нейрон.**

# 3.2. Можливості персептрона

Оскільки $y \in \{-1, 1\}$ , персептрон дозволяє розділити множину множину вхідних векторів на 2 частини (дихотомізувати) $S \in R^d \to R^2$, тобто розділити простір вхідних векторів на деяку *множину* і її *доповнення*. Однак можливості персептрона обмежені, наприклад, він не може розділити дві множини **(рис. 3.2)**

$\|x\| < 1$ і $\|x\| \geq 1$.

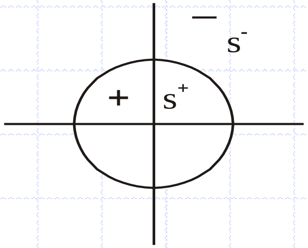
При навчанні нейронної мережі, як правило, математичні вирази длярозділяючих поверхонь відсутні, а навчання виконується тільки на *прикладах*.

Задається навчальна вибірка (скінчена), що складається з пар вхід-вихід

$$T = \{(x_1, t_1), \ldots, (x_n, t_n)\} = \{(x_i, t_i),\ i = \overline{1, n}\ \}. \qquad (3.6)$$

Мета навчання — налаштувати вагові коефіцієнти так, щоб

$$\forall x^* \in R^d,\ x^* \notin T \qquad y = t^*$$ .



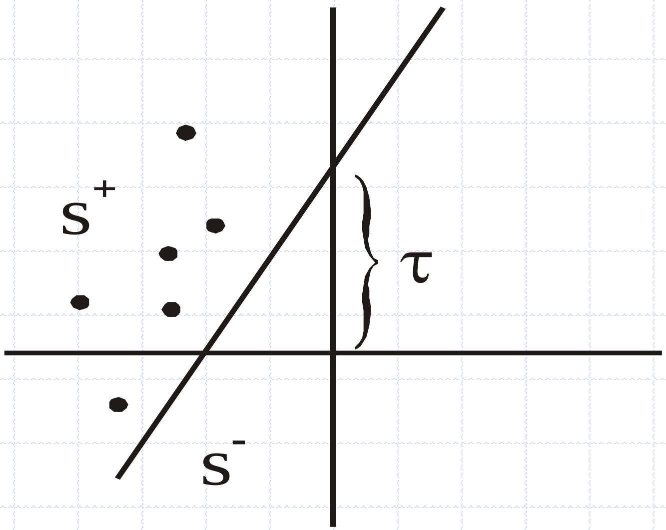
# 3.3. Проблема лінійної роздільності

Чи можна без помилки навчити персептрон на задану навчальну множину?

У $R^2$ розділяюча гіперплощина (3.3) являє собою *пряму*, що поділяє $R^2$ на 2 півплощини **(рис. 3.3).**

Навчальна вибірка (3.6) задає 2 *обмежені* (*скінчені*) множини:

$$S^+ = \{x:\ \exists i:\ t_i = 1,\ \ x = x_i\}$$
$$S^- = \{x:\ \exists i:\ t_i = -1,\ \ x = x_i\} \tag{3.7}$$



# Визначення 3.1.

- Множини $S^+$ і $S^-$ називаються *лінійно роздільними* (linearly separable), якщо існує гіперплощина $H$, обумовлена парамет - рами $\tau$ і $W$ така що

$S^+ \subset H^+, \quad S^- \subset H^-$ або $\forall x \in S^+ \quad W^T x - \tau \geq 0,$ $\forall x \in S^- \quad W^T x - \tau < 0$

Якщо умови (3.8) виконуються для $\tau = 0$, то говорять про однорідну лінійну роздільність.

Визначимо умови лінійної роздільності, використовуючи елементи опуклого аналізу Рокафеллара.

# Визначення 3.2.

- *Опуклою (лінійною)* комбінацією векторів $x_1, \ldots, x_n$ щодо невід’ємних скалярів $\lambda_1, \ldots, \lambda_n$, таких що

$$\sum_{i=1}^{n} \lambda_i = 1$$

називається вектор

$$x^* = \sum_{i=1}^{n} \lambda_i x_i$$

# Приклад.

- Якщо $n = 2,$ то опуклою комбінацією $x_1$ і $x_2$ є вектори, вершини яких розташовані на відрізку, що з'єднує $x_1$ і $x_2$ **(рис. 3.4).**

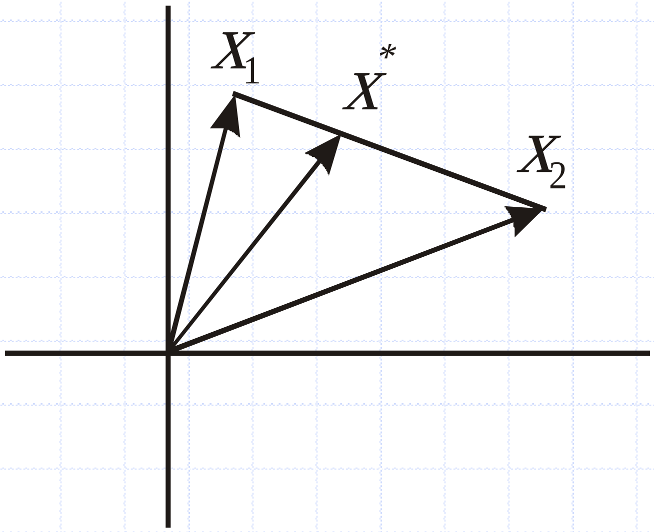



*Рис. 3.4. Лінійна комбінація векторів*

# Визначення 3.3.

- Множина $S \in R^d$ називається ***опуклою***, якщо її можна покрити скінченим числом опуклих комбінацій деяких векторів **(рис. 3.5)**.

Неформально, підмножина лінійного векторного простору є опуклою, якщо для довільної пари точок з цього підпростору, відрізок, що з'єднує ці точки, цілком належить цій підмножині.

- В іншому випадку множина є ***неопуклою*** (рис. **3.6)**.

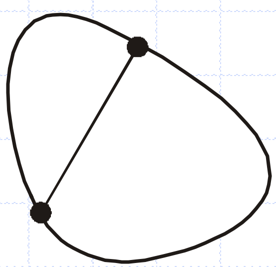


Рис. 3.5. Опукла множина

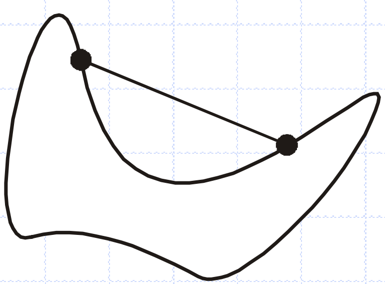


Рис. 3.5. Неопукла множина

# Визначення 3.4.

- *Опуклою оболонкою* $C(S)$ для множини век - торів $S$ називається множина всіх опуклих комбінацій векторів з $S$ :

$$C(S) = \left\{ z : \exists n, \exists x_1, \ldots, x_n \in S, \exists \lambda_1, \ldots \lambda_n : \lambda_i \geq 0, \ \sum_{i=1}^{n} \lambda_i = 1, z = \sum_{i=1}^{n} \lambda_i x_i \right\}$$

Опукла оболонка — це мінімальна (у сенсі включення) опукла множина, що містить множину $S$.

# Теорема Хана-Банаха про лінійну роздільність

Якщо $S$ — скінчена множина, то $C(S)$ — **опуклий багатогранник.**

Необхідну і достатню умову лінійної роздільності скінчених множин $S^+$ і $S^-$ дає теорема 3.1.

**Теорема 3.1 (Хана-Банаха про лінійну роздільність).**
Скінчені множини $S^+$ і $S^-$ є лінійно-раздільними тоді і тільки тоді, коли їх опуклі оболонки $C(S^+)$ і $C(S^-)$ не перетинаються.

# Теорема Хана-Банаха (продовження)

Ця теорема забезпечує необхідну і достатню умову можливості навчання простого персептрона на заданій навчальній множині . Задача *навчання* нейронної мережі зводиться до відповіді на питання: чи існують такі $W$ і $\tau$ , що $S^+$ і $S^-$ є лінійно роздільні.

З'ясуємо умови існування таких $W$ і $\tau$. Для зручності візьмемо розширені множини $\tilde{S}^+$ і $\tilde{S}^- \in R^{d+1}$ виду (3.4). Розглянемо множину

$$\tilde{F} = \tilde{S}^+ \cup \{-\tilde{S}^-\}, \qquad (3.9)$$

де $-\tilde{S}^- = \{\tilde{z} : \forall \tilde{x}_i \in \tilde{S}^-, \ \tilde{z}_i = t_i \tilde{x}_i\},$

тобто в $\tilde{\boldsymbol{z}}_i$ кожен компонент вектора $\tilde{\boldsymbol{x}}_i$ замінений на таке ж число з протилежним знаком.

З урахуванням позначення (3.9) вихідні множини $S^+$ і $S^-$ є лінійно роздільними, якщо

$$\forall \tilde{x} \in \tilde{F} \quad \tilde{W}^T \tilde{x} > 0. \qquad (3.10)$$

# Теорема 3.2 (про альтернативу).

- Або існує $\exists \widetilde{W},$ що задовольняє рівняння (3.10) або існує

$$\exists k : \widetilde{x}_{i_1}, \ldots, \widetilde{x}_{i_k}, \ \lambda_i \geq 0, \ \sum_{j=1}^{k} \lambda_j = 1, \ \sum_{j=1}^{k} \lambda_j \widetilde{x}_{i_j} = 0 \qquad (3.11)$$

Іншими словами, **або** існує гіперплощина, що відокремлює вектори множини $F$ від початку координат, **або** початок координат належить опуклій комбінації таких векторів.

Якщо система (3.11) не має розв'язків, значить, множини є лінійно роздільними і існують розв'язки нерівності (3.10).

Система лінійних нерівностей (3.10) розв'язується з використанням **симплекс-методу.**

Цей метод дозволяє мінімізувати цільову функцію (найчастіше квадратичну) при наявності обмежень у вигляді нерівностей. Симплекс-метод визначає початкову точку, що задовольняє цю нерівність, а потім ітеративно зменшує значення цільової функції.

У нашому випадку цільову функцію можна прийняти тотожньо рівною 0 і використовувати симплекс-метод для знаходження допустимої точки.

# 3.4. Властивості теореми про лінійну роздільність

- Розв’язок нерівності (3.10) не єдиний. Кількість способів розбиття множини $S$ , що містить *n* векторів $S=\{x_i \in R^d, i=\overline{1,n}\}$ ,

$n$

гіперплощинами зі спільним порогом $\tau$ дорівнює числу областей, на які поділяється $R^d$ гіперплощинами з однаковим порогом.

Це випливає з розгляду родини $sign(W^{\mathrm{T}} x_i - \tau) = t_i, \quad i=\overline{1,n}.$

Якщо розглядати $W$ в просторі $R^d$ , то будь-яка гіперплощина з $R^d$ параметрами $W$ $(x_i, \tau)$ поділяє простір вагових коефіцієнтів на 2 підпростори: $\bigcap_n P_i^+$ та $\bigcap_n P_i^-$

Таких гіперплощин $n$ .

Поділяюча гіперплощина (якщо вона існує), може бути обрана у формі

$$\forall i=\overline{1,n} \qquad H=\left\{x : \sum_{j=1}^{d} \alpha_j t_{i_j} x_{i_j} \cdot x - \tau = 0\right\}, \quad \alpha_j > 0 \qquad .(3.12)$$

Тобто ваговий вектор може бути обраний у вигляді

$$W=\sum_{j=1}^{d} \alpha_j t_{i_j} x_{i_j} \qquad \alpha_j > 0.$$

У термінах розширених векторів

$$\tilde{W}=\sum_{j=1}^{d} \alpha_j \tilde{x}_{i_j} \qquad \alpha_j > 0$$

# 3.5. Максимальний розмір навчальної множини

Розмір навчальної множини не повинен перевищувати

$$N^* = 2(d+1) \approx 2d,$$

тобто кожний ваговий коефіцієнт (зв'язок) здатен "запам'ятати" приблизно 2 вхідні образи.

# Алгоритм навчання персептрона за умови існування роздільної гіперплощини

Методи лінійного програмування для розв’язання нерівності (3.10) (симплекс-метод) забезпечують *пакетну обробку*, що не відображає процес навчання.

Для навчання необхідний *інтерак - тивний* або *on-line* алгоритм, що корегує вагові коефіцієнти після пред'явлення мережі кожного образа.

# 3.7. Алгоритм навчання персептрона Розенблатта

- 1. Формуємо множину $\tilde{F} = \tilde{S}^{+} \cup \{-\tilde{S}^{-}\} \in R^{d+1}$

і систему $\forall \tilde{z} \in \tilde{F} \qquad \tilde{W}^{\mathrm{T}} \tilde{z} > 0$

- 2. *Початок*. Вибираємо $\forall \tilde{z} \in \tilde{F}$ як початкове наближення для $\tilde{W}$. Сформуємо випадкову послідовність (циклічну, у якій елементи з'являються з невизначеною частотою) з елементів $\tilde{W}$.

- 3. *Тест*. Вибираємо випадкове значення $\tilde{z}_{i_j} = \mathrm{rand}(\tilde{F})$. Якщо $\tilde{W}^{T} \tilde{z}_{i_j} > 0$ переходимо до п. 3, інакше — до п. 4.

- 4. *Модифікація вагових коефіцієнтів*. Сформуємо обмежену послідовність

$$\varphi_j : 0 < \underline{\varphi} \le \varphi_j \le \overline{\varphi},$$

$$\tilde{w} = \tilde{w} + \varphi_j \tilde{z}_{i_j} \quad .$$

Переходимо до п. 3.

- 5. *Завершення*. Якщо протягом деякого числа ітерацій процедура виконує п. 3 ("зациклюється"), процес навчання завершується.

# Зауваження.

- 1. У базовому алгоритмі навчання персептрона $\varphi_k = 1$ , але найчастіше вибирають

$$\varphi_k = \frac{1}{\left\| \tilde{z}_{i_j} \right\|},$$

для нормування множини $\tilde{F}$ таким чином, щоб усі його век - тори мали одиничну довжину.

Операції 4 обумовлені пошуком розв’язку $\tilde{W}$ у формі

$$\tilde{W} = \sum_j \alpha_j \tilde{z}_{i_j}, \ \ \alpha_j > 0 \cdot$$

Крім того

$$\tilde{W}_j \cdot \tilde{z}_{i_j} = \tilde{W}_{j-1} \cdot \tilde{z}_{i_j} + \varphi_j \left\| \tilde{z}_{i_j} \right\|^2 > \tilde{w}_{j-1} \cdot \tilde{z}_{i_j} \ .$$

Значення $\tilde{W}_j$ — збільшується, щоб після поточного негативного значення на наступному кроці було отримане додатне (п. 4 виконується тільки у випадку негативного добутку).