ЛЕКЦІЯ №1
Поняття штучного інтелекту. Нейрон та нейромережі.

- Штучний інтелект – один з найперспективніших напрямків комп'ютерних наук, який вивчає методи розв'язання задач, для яких не існує способів вирішення. Системи штучного інтелекту можуть оперувати даними та самонавчатися. Сфери застосування таких систем є необмеженими – від створення роботів, які самостійно приймають рішення, до машин з автопілотом чи онлайн-перекладачі в реальному часі.



- Загалом поняття "штучний інтелект" є досить розмитим. Практично вся сучасна техніка обладнується мікрочіпами, а виробники переконують споживачів про наявність ШІ в їх виробах. Але, в більшості це є просте копіювання людиноподібної лінії поведінки на штучно створеному об'єкті для зменшення витрат і часу людини.

Інтелектуальна задача – це процес знаходження алгоритму рішення певного класу задач. Інтелектуальними задачами є задачі, для розв'язання яких немає чітко заданого алгоритму, який завжди приводить до потрібного результату, а інтелектуальною діяльністю можна назвати процес вирішення інтелектуальних задач. Прикладом інтелектуальної задачі є розпізнавання образів, тобто визначення належності об'єкта, що спостерігається, до однієї із заздалегідь визначених категорій.

Інтелектуальним задачам *властиві* неповнота, неточність та суперечливість знань, а також велика розмірність простору рішень, що не дає змоги розв'язувати їх простим перебором. У таких задачах часто немає чітких критеріїв для вибору оптимального рішення, а сама задача не завжди цілком формалізується.

Основними властивостями інтелектуальних задач є:

- символьне подання умов задачі;
- відсутність чіткої постановки задачі;
- відсутність прийнятного для практичного використання алгоритму рішення, який завжди забезпечує правильний результат;
- неповнота, неточність та суперечливість знань;
- відсутність чітких однозначних критеріїв вибору оптимального рішення;
- велика розмірність простору рішень.
- Інтелектуальні задачі часто називають неформалізованими задачами, багато з них які розв'язуються інтелектуальними системами.

Напрямки досліджень в галузі штучного інтелекту

- В дослідженнях у галузі штучного інтелекту склалося два головних напрямки: біонічний і прагматичний.
- Біонічний напрямок досліджень в галузі штучного інтелекту засновано на припущенні про те, що якщо в штучній системі відтворити структури і процеси людського мозку, то й результати вирішення завдань такою системою будуть подібні до результатів, що отримує людина. В цьому напрямку досліджень виділяються:
- Нейромережні алгоритми. В його основі лежать системи елементів, які подібно до нейронів головного мозку здатні відтворювати деякі інтелектуальні функції. Прикладні системи, розроблені на основі цього підходу, називаються нейронними мережами.
- Структурно-евристичний підхід. В його основі лежать знання про поведінку спостережувального об'єкта або групи об'єктів і міркування про ті структури, які могли б забезпечити реалізацію спостережуваних форм поведінки. Прикладом подібних систем служать мультиагентні системи.
- Еволюційні алгоритми. В цьому випадку можна вирішити завдання, що формулюється в термінах еволюціонуючої популяції організмів – сукупності підсистем, що протидіють і співпрацюють, в результаті функціонування яких забезпечується необхідна рівновага (стійкість) всієї системи в умовах постійно змінних впливів середовища. Такого роду підхід реалізовано в прикладних системах на основі генетичних алгоритмів.

- **Нечітка логіка.** Найбільш вражаючим в людському інтелекті є здатність приймати правильні рішення в умовах неповної та нечіткої інформації. Побудова моделей наближених роздумів людини і використання їх в комп'ютерних системах представляє сьогодні одну з найважливіших проблем науки. "Штучний інтелект", який легко вирішує завдання управління складними технічними комплексами, часто є безпорадним в простих ситуаціях повсякденного життя. Для створення інтелектуальних систем, здатних адекватно взаємодіяти з людиною, потрібно застосовувати новий математичний апарат, який переводить неоднозначні життєві твердження в мову чітких і формальних математичних формул.
- **Прагматичний напрямок** ґрунтується на припущенні про те, що розумова діяльність людини є «чорним ящиком». Але, якщо результат функціонування штучної системи збігається із результатом діяльності експерта, то таку систему можна визнати інтелектуальною незалежно від способів отримання цього результату. При такому підході не ставиться питання про адекватність використаних в комп'ютері структур і методів до тих структур чи методів, якими користується в аналогічних ситуаціях людина, а розглядається лише кінцевий результат вирішення конкретних завдань.

Розпізнавання образів

Це напрямок, що сформувався від зародження ШІ, на зараз це самостійна наука. Основним підходом є опис класів об'єктів через значення вагомих ознак. Кожному об'єкту ставиться у відповідність матриця ознак за якою відбувається розпізнавання. Для поділу об'єктів на класи використовують спеціальні математичні процедури і функції.

Нейронні мережі

Принцип створення штучних нейронних мереж запозичено з біології. Вони утворені з елементів, які відтворюють елементарні функції біологічного нейрона. Штучні нейронні мережі відтворюють певні властивості, які притаманні мозку людини. Вони навчаються на основі досвіду, узагальнюють свій досвід, здатні виділяти головне з інформації, що надходить.

Здатність нейронної мережі до навчання вперше була досліджена Дж. Маккалоком і У. Піттом в дослідах 1943 року на створеної ними моделі нейрона. Автори описали принципи побудови нейронних мереж. Пізніше, в 1962 році, Ф. Розенблатт запропонував свою модель нейронної мережі – перцептрон, а в 1986 р. Дж. Хінтон і його колеги опублікували статтю з описом моделі нейронної мережі і алгоритмом її навчання, що дало поштовх до ефективного вивчення нейронних мереж.

Для моделей, побудованих за типом нейронних мереж людського мозку, характерно легке розпаралелювання алгоритмів і висока продуктивність. З людським мозком їх зближує важлива властивість, яка відсутня в простих електронних машинах: нейронні мережі працюють навіть за умови неповної інформації про навколишнє середовище, тобто, як і людина, вони можуть відповідати не тільки "так " або " ні ", але і" не знаю точно, але скоріше так ".

Нейронним мережам сьогодні під силу розпізнавання сигналів, мови, зображень, пошук даних, фінансове прогнозування, шифрування даних. Нейромережний підхід використовується у великій кількості завдань – для кластеризації інформації з Інтернету, для імітації та моделювання складно влаштованого людського мозку, для розпізнавання образів і ін.. Зараз продовжується вдосконалення методів синхронної роботи нейронних мереж на паралельних пристроях.

До переваг нейронних мереж можна віднести самонавчання, само налаштування, гнучкість конфігурації, високу ефективність. Серед найбільш відомих сьогодні нейронних мереж виділяють мережі Хопфілда, нейронні мережі зі зворотним поширенням похибки і самоорганізовані карти.

Наразі НМ розв'язують наступні класи задач.

- *Класифікація образів*: вхідний образ має вигляд вектора ознак, для якого розпізнається відповідний еталонний вектор.
- *Кластеризація* (категоризація), при якій близькі образи розміщуються в один кластер (одну групу).
- *Апроксимація функцій*, при якій по даних вхід-вихід для невідомої функції F(x) знаходиться її оцінка.
- *Прогнозування*, по якому передбачається деяке значення у деякий майбутній момент часу. Фахівці інвестиційного банку за допомогою програмного нейропакета роблять короткострокові прогнози коливань курсів валют.
- *Оптимізація*, завданням якої є знаходження такого рішення, яке задовільняє обмеження системи і максимізує чи мінімізує цільову функцію. *Задача комівояжера* є класичним прикладом задачі оптимізації.

Основні недоліки НМ

- *НМ можуть «помилятись»*, тобто не мають достатнього рівня надійності. Наприклад, не треба нейронні мережі використовувати для нарахування заробітної плати, або при введенні у дію бойових балістичних ракет (у таких випадках кажуть про недостатню надійність нейронних мереж).
- *Не можна гарантувати*, що розроблена мережа є оптимальною мережею.
- У загальному випадку невідома кількість потрібних внутрішніх шарів та кількість нейронів в цих шарах.
- Невідома *кількість еталонних пар* «вхід-вихід» та кількість кроків, які потрібні для навчання.
- Основні успіхи нейронних мереж отримані як результат використання алгоритму зворотного розповсюдження похибок, який має ряд проблем: немає гарантії, що мережа є оптимальною і може бути навчена *за виправданий час*; немає впевненості, що мережа *не попаде у локальний мінімум* і тому може отримати гірше навчання, у порівнянні з оптимумом.

Математична нерівність вирішальної функції – основа роботи штучного нейрону

Відомо, що будь-яка *математична нерівність виконує логічну функцію*. Цю особливість нерівності можна використати для розділу на «худих» та «товстих» множини людей з характеристиками (x_1, x_2), де x_1=0..150 кг – вага людини, а x_2=0..2,0 м – її зріст (рис. 1). Координати найбільш «худої» та найбільш «товстої» людини помічені чорними точками.

Тоді всю множину людей, що розміщена у прямокутнику при x_1=0..150 кг, x_2=0..2 м можна розділити на дві частки («худих» та «товстих») за допомогою прямої лінії (*вхідної вирішальної функції нейрону*), яка знаходиться між вказаними чорними точками у вигляді нерівності

$$u_1 = x_1 - 150x_2 + 45 \geq 0. \qquad (1)$$

Нерівність (1) розділяє площу ознак (x_1, x_2) на дві на півплощини: в одній півплощині нерівність виконується, а в іншій – ні. У цьому можна пересвідчитись, якщо підставити в нерівність (1) координати будь – якої довільно обраної точки. На практиці для перевірки нерівності найчастіше використовують точку $(x_1=0, x_2=0)$. В цій точці ми отримуємо відповідь на питання: «Чи виконується нерівність u_1 в півплощині, в якій знаходиться точка $(x_1=0, x_2=0)$?». У нашому випадку ці координати початку осей координат дають значення $u_1 = 45 \geq 0$, тобто нерівність виконується на тій півплощині. Ця півплощина на прямій u_1 помічається пунктирною лінією. На іншій півплощині нерівність не виконується для всіх її точок.

Таким чином, *вхідна функція нейрону* є вирішальною функцією і виконує логічну умову: якщо $u \geq 0$ (нейрон видає сигнал *«1=Істина»*), то це – людина «повна», якщо ні, то «худа» (нейрон видає сигнал *«0=Хибність»*).

З рис. 1, а видно, що *розміщення вирішальної функції (1) є невизначеним* для розділу класів можна використати безліч подібних вирішальних функцій у залежності від індивідуального для даного експерта тлумачення поняття «повноти».

Відповідний нейрон Adaline згідно вирішальної функції (1) наведений на рис. 2. Нелінійний вихід нейрону рис. 2 призначений для обмеження числових величин NET значеннями y= –1...+1, які виконують логічний висновок нейрона і одночасно запобігають отриманню занадто великих вихідних цифр, які можуть призвести до аварійної зупинки ЕОМ по переповненню.

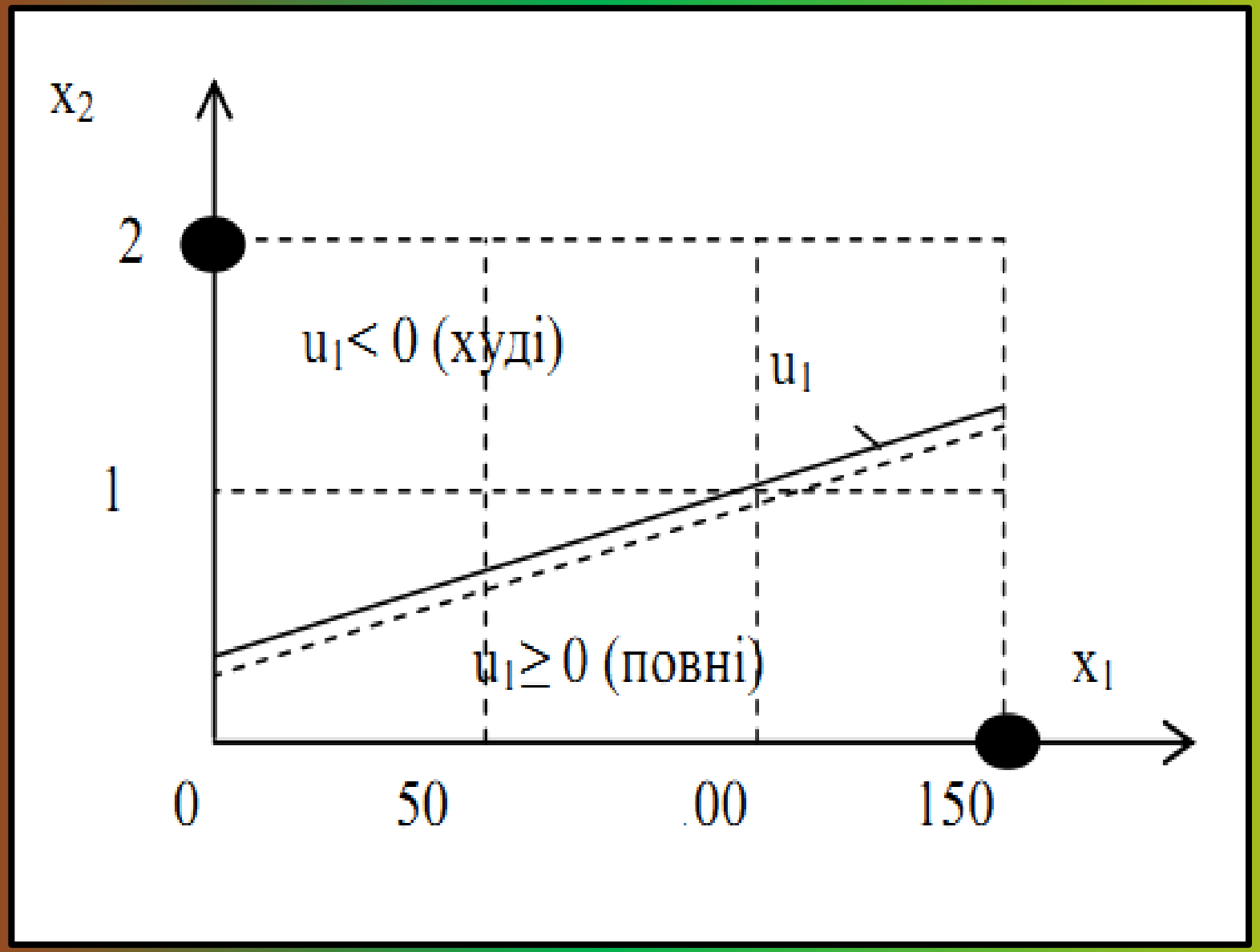


Рис.1. Класифікація об'єктів за двома ознаками без нормалізації

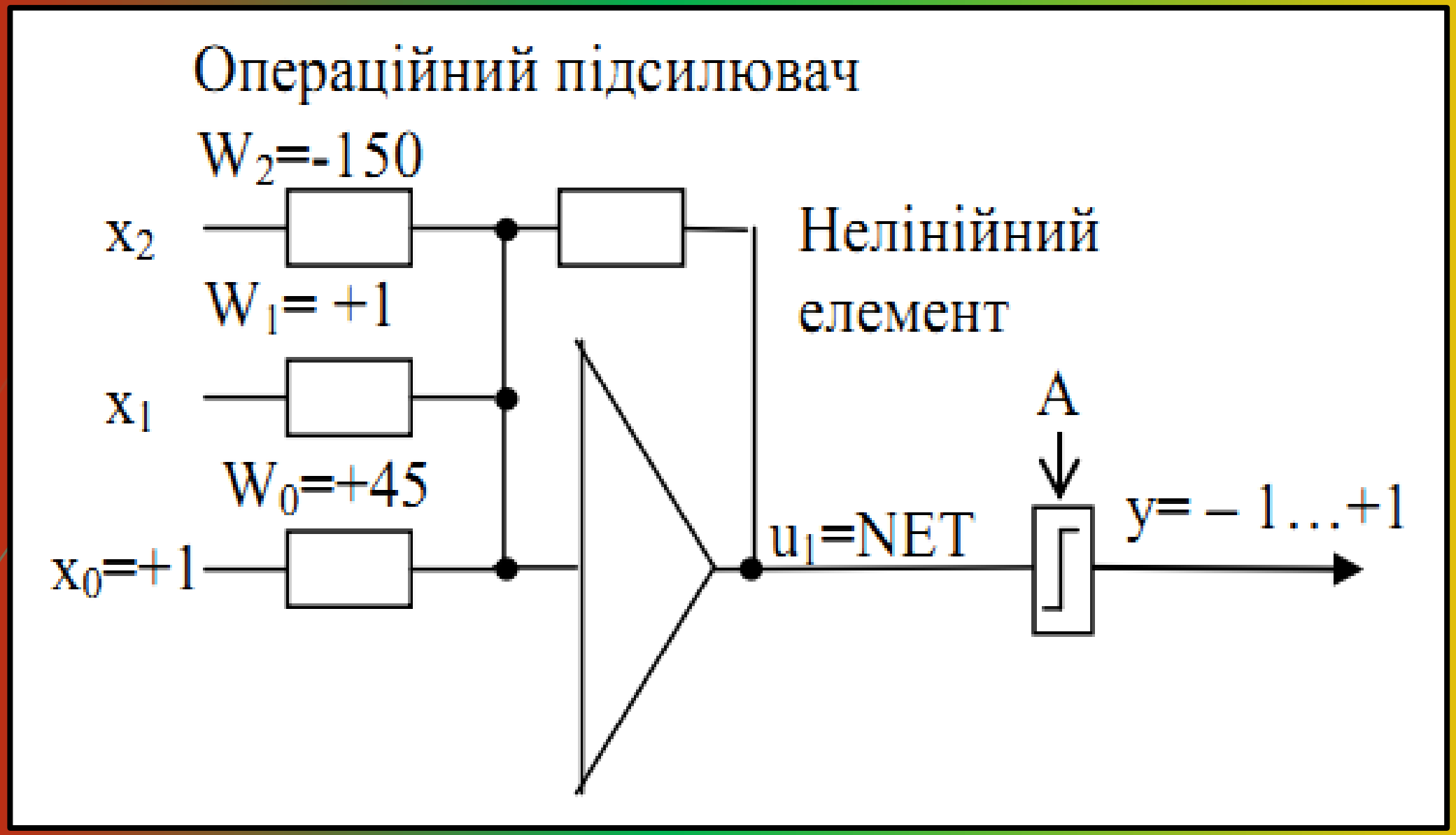


Рис. 2 Апаратна схема штучного нейрона

Таким чином, для об'єктів з кількісними ознаками за допомогою математичних нерівностей ми можемо створювати логічні вирази для їх класифікації.

Цю задачу можна розв'язати двома шляхами:

- *За рахунок створення програми.* Тоді можна казати про *програмну реалізацію* нейронної мережі.
- *За рахунок створення спеціального пристрою*, якій зветься штучним нейроном (наприклад, за рахунок використання опера-ційного підсилювача та нелінійного перетворювача виходу – рис. 2). Тоді можна казати про *апаратну реалізацію* нейронної мережі, маючі на увазі що мережа складається з окремих пристроїв – штучних нейронів.

Операційний підсилювач перемножує вхідні постійні величини та змінні сигнали x_j на коефіцієнти передачі W_j (вагові коефіцієнти) і потім отримує їх підсумок.

Нас у даному випадку цікавить лише знак функції u, а не її величина. Тому додатні значення перетворюють на «+1» (маленький зріст), а *від'ємні значення – на «-1» або «0»* (високий зріст). Це змушує мати на виході операційного підсилювача нелінійний елемент, який дає логічні відповіді (ІСТИНА/TRUE/+1) та (ХИБНІСТЬ/FALSE/–1, «0»). Практично ми перетворюємо математичну нерівність в логічну умову.

Функція *u* іноді зветься *NET* (у матричній формі це має вигляд рівняння U = NET = WX), а вихідна нелінійна функція *y = f(NET)* – зветься OUT. Замість вихідних сигналів «*y*=+1/–1» можуть застосовуватись сигнали «*y*=+1/0».

Функція перетворення значення NET у вихідну функцію y =f(OUT) зветься *активаційною (вихідною) функцією нейрона.*

Розглянемо іншу задачу: розділ літаків на бомбардувальники та винищувачі по значеннях кількісних ознак (x_1, x_2), як показано на рис. 3, а (тут x_1= 0...3 – швидкість (М); x_2= 0...40 – розміри (м)). Ми можемо провести на рис. 2 (а) пряму лінію

$$u_2 = 7x_1 - x_2 + 7 \geq 0, \qquad (2)$$

яка є границею для двох класів об'єктів і знаходиться між граничними значеннями характеристик бомбардувальників та винищувачів.

Нерівність $u_2 \geq 0$ використовується для розділу області ознак (x_1, x_2) на дві півплощини: в одній півплощині з точкою (x_1= 0, x_2 =0) нерівність (2) дотримується (визначаються винищувачі при $u_2 \geq 0$), а в іншій – ні (визначаються бомбардувальники при $u_2 < 0$).

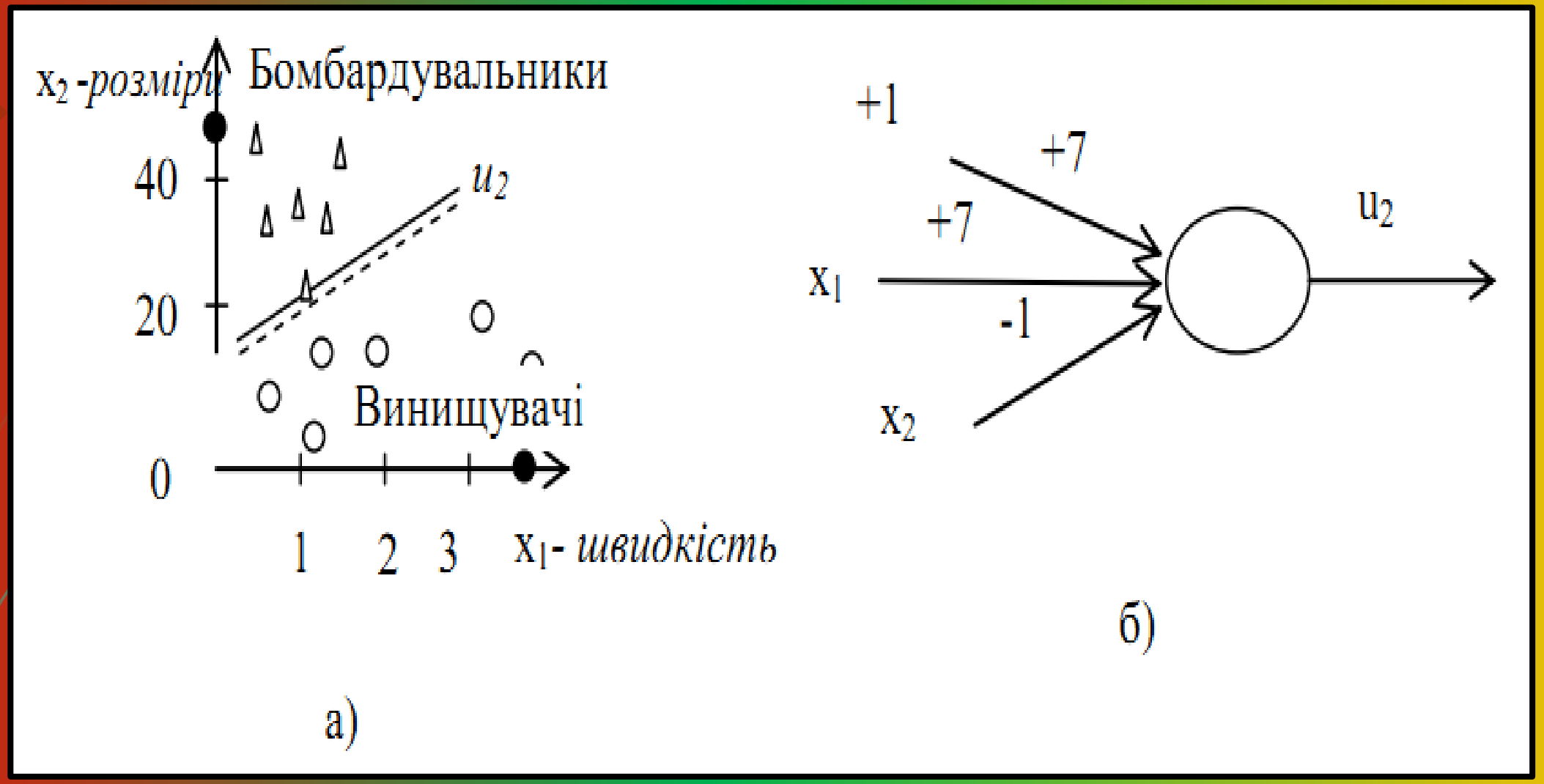


Рис. 3. Класифікація об'єктів (а). Схема нейрона (б)

Знак нерівності («більше або дорівнює» ≥, або «менше або дорівнює» ≤) не має значення для аналізу і довільно обирається експертом (звичайно використовується «≥»). На рис. 3 нерівність (2) виконується на півплощині, яка помічена пунктирною лінією; якщо ми змінимо знак нерівності, то пунктирна лінія переміститься на іншу півплощину, але від цього принцип класифікації не змінюється. Функція u_2 – це лінія у просторі ознак, яка розділяє між собою об'єкти двох класів. Очевидно, що таких функцій можна провести багато і вони можуть мати різний вигляд.

Звичайно нейрон Adaline позначається у спрощеному вигляді (рис. 3, б). Якщо відомі вхідні функції, то це означає відомість вагових коефіцієнтів; тоді для класифікації об'єктів *можна не застосовувати процес навчання*.

Функція активації y = F(NET) = F(u) = OUT може мати різний заданий програмістом вигляд. На рис. 4 вони показані у порядку, який приблизно відображує їх уживаність. Усі вони можуть бути зсунуті вздовж осі u.

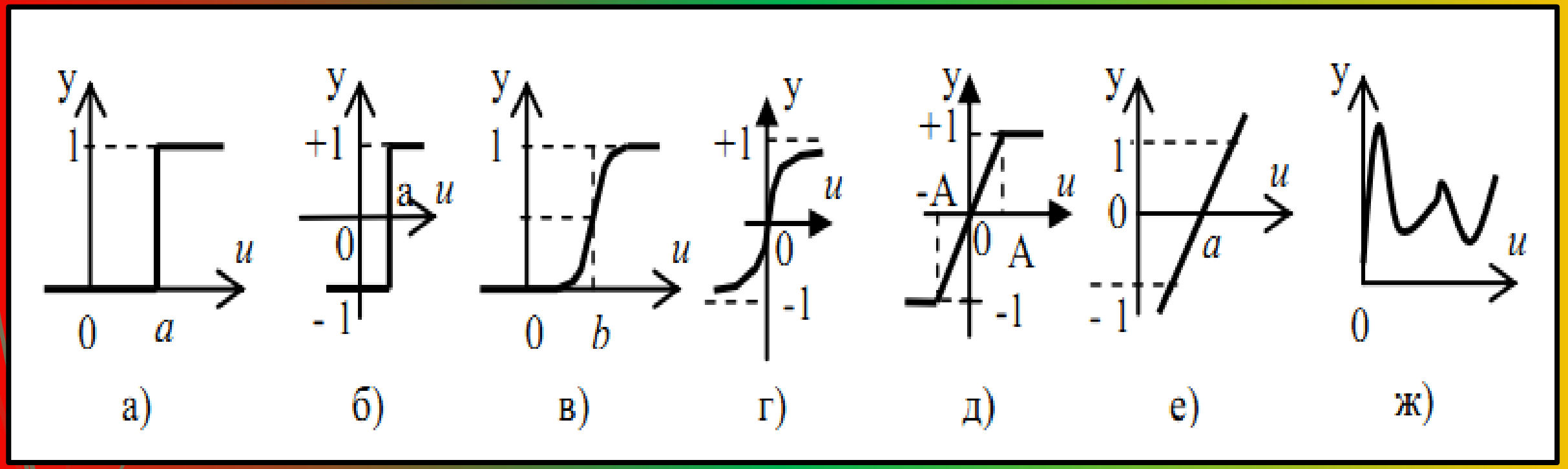


Рис. 4. Функції активації штучного нейрону

Іноді після активаційної функції на виході нейрона використовується додаткова вихідна функція $f_{вих}(y)$, яка призначена для перетворення та уточнення значення виходу нейрону.

Найбільш розповсюдженими функціями активації є: уніполярна стрибкоподібна функція (рис. 4, а); біполярна стрибкоподібна функція.

(рис. 4, б) – сігмоїда.

(рис. 4, в) – тангенс гіперболічний (рис. 4, г); складена з відрізків прямих ліній (рис. 4, д); лінійна залежність (рис. 4, е); довільна задана програмістом функція (рис. 4, ж). Тобто активаційна функція *задається користувачем* і може мати будь-який вигляд, наприклад, при бажанні вона може мати вигляд або прямої лінії, або складатись з 10-ти відрізків прямих ліній, або мати задану користувачем будь-яку нелінійну залежність. Вигляд активаційної функції обирається користувачем у залежності від вимог конкретної задачі.

Звичайно для складної НМ, яка розв'язує конкретну задачу, *всі нейрони мають однакову активаційну функцію*

Головне призначення активаційної функції – обмеження максимального сигналу виходу *значенням модуля «1»*: це обмежує розрахункові числові значення НМ і не дозволяє їм наблизитись до неприпустимо великих величин, які викликають зупинку роботи ЕОМ. Слід сказати, що такої переваги не має активаційна функція рис. 4, е, яка є лінійною функцією *OUT = Ku*, де *K* – константа.

Функції моделей нейронів

Свого часу, *У. Маккалох і У. Пітс* першими опублікували дослідження про роботу штучних нейронів з пороговим виходом «1\0» у вигляді електичної схеми рис. 5 (а)

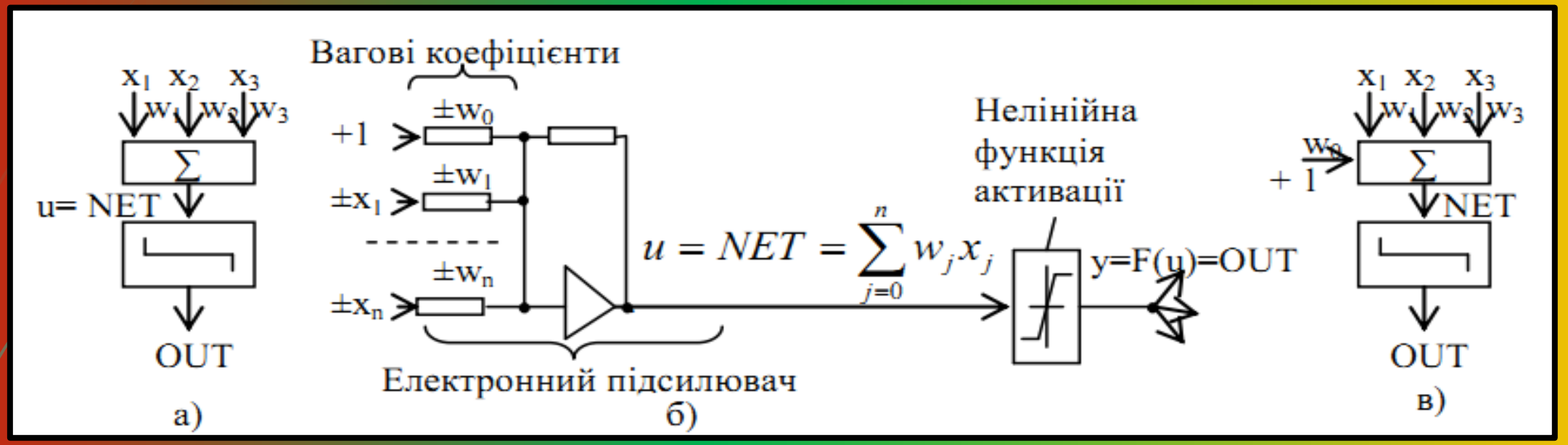


Рис. 5. Схеми штучного нейрону

а) штучний нейрон У. Маккалоха і У. Пітса (1943);

б) штучний нейрон *ADALINE* Б. Відрова та М. Хоффа (1959);

в) персептрон Розенблата (1960)

- *Б. Відроу та М. Хофф* розробили моделі *ADALINE* (*ADAptive Linear NEuron* – Адаптивні Лнійні Нейрони) та мережі з *ADALINE*. Нейрон *ADALINE* показаний на рис. 7.4.1, б. Як бачимо, на вхід нейрону була додана постійна складова. Зараз *ADALINE* є найбільш поширеним елементом нейронних мереж.

Нейрон ADALINE має наступні особливості:

- вхідні сигнали x_j множаться на вагові коефіцієнти w_j, підсумовуються і підсилюється по потужності електронним підсилювачем, в результаті чого на його виході отримуємо значення *вирішальної функції* $u = NET$, яке, у залежності від вхідних координат x_j, може мати додатне або від'ємне значення

$$u = \sum_{j=0}^{n} w_j \cdot x_j \qquad (3)$$

– значення u = NET проходить через нелінійну *функцію активації* у вигляді

$$y = F(u) = F(NET) = OUT = \begin{cases} 1 \text{ – при } NET \geq 0, \\ 0 \text{ – при } NET < 0. \end{cases}$$

і подаються на входи інших нейронів у вигляді логічних сигналів OUT, рівних 1 або 0.

Сигнал NET може бути розрахованим у матричній формі

$$NET[1,1] = X^T[1,n+1]\ W[n+1,1],$$

де $X(x_0=1=const,\ x_1,\ldots,\ x_n)$ – вектор вхідних сигналів (значення $x_0=1=const$ вектора X задається програмістом); $W\ (w_0,\ w_1,\ldots,\ w_n)$ – вектор вагових коефіцієнтів.

Більшість мереж використовують для знаходження зваженої суми входів нейрона формулу (3), але в деяких випадках, якщо цього вимагає поставлена задача, можуть використовуватись і інші формули, наприклад,

$$\text{u} = \sum_{j=0}^{n} w_j x_j^2 \ , \ \sum_{j=0}^{n} w_j (a_1 x_j + a_2 x_j^2) \ .$$

Введення подібних нелінійностей може збільшити обчислювальну потужність НМ, тобто дозволяє розв'язати поставлену проблему з меншою кількістю нейронів.

Математично доведено, що будь-яка нерівність, в тому числі і у вигляді (2), виконує логічні функції: якщо отримана сума «NET = u» більше заданого порогового значення, то вихід дорівнює 1, інакше – 0.

Структура штучного нейрону наведена на рис. 6. Наразі існують різні моделі нейронів. Нейрони можуть реалізуватись апаратно та програмно. Найбільше поширення отримала програмна реалізація НМ – з використанням ЕОМ. Тому для вирішальної функції $u = NET \geq 0$, для функцій активації $z = f_a(u)$ чи виходу

$Y = f_{вих}(z) = OUT$ можуть бути використані довільні *функції будь-якої складності*. Обмеження тут накладаються лише з точки зору зменшення складності аналізу та навчання. При *цьому кількість активаційних та вихідних функцій не обмежується*.

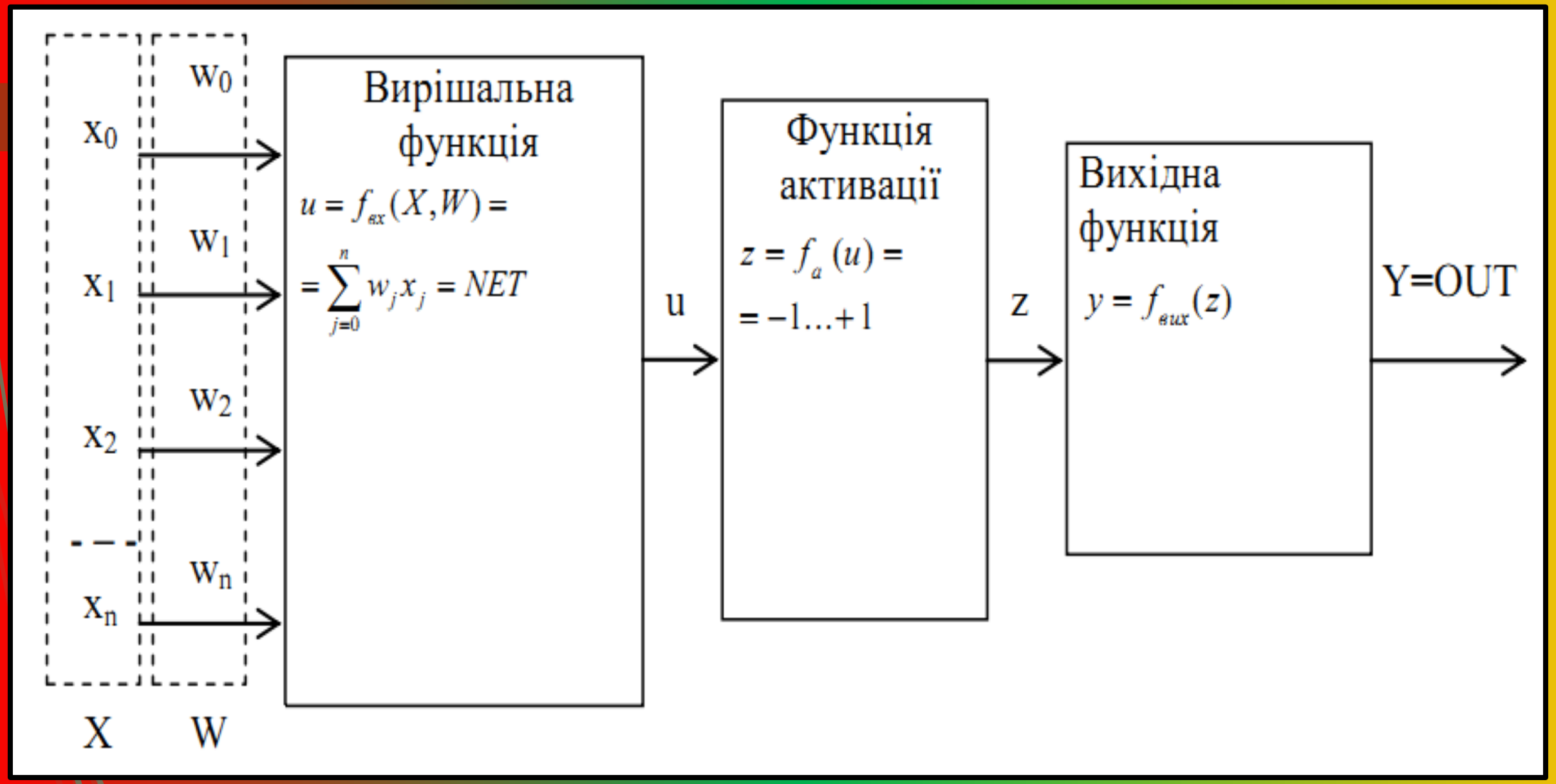


Рис. 6. Структура штучного нейрона

Наприклад, існують моделі нейронів:

- з введенням конденсатора та джерела струму у входи електронного підсилювача, що дозволяє аналізувати динамічні процеси;

  з *добутком* вхідних змінних при визначенні нерівності NET≥0; – з використанням *нелінійної залежності* при визначенні нерівності

- NET≥0 (в тому числі, функції *еліпса*);
- з урахуванням взаємного *відношення* підсумку збуджуючих вхідних сигналів до підсумку гальмуючих вхідних сигналів тощо.

Слід нагадати, що *розміщення вирішальних функцій у просторі ознак є невизначеним*: для розділу класів можна використати безліч вирішальних функцій. Найліпшим вважається її розміщення при рівній відстані від об'єктів різних класів.

Прості функції нейронів є базисом створення потужних мереж.

- *Нейробіолог США Ф. Розенблатт* виконав моделювання *персептрону* – багатошарової НМ з ваговими коефіцієнтами прямих зв'язків між шарами. Кожний шар має довільну кількість нейронів.

- Наразі НМ розглядаються як з'єднані між собою шари, кожний з яких вміщує деяку кількість нейронів, які обмінюються інформацією з використанням прямих та зворотних зв'язків (рис. 7).

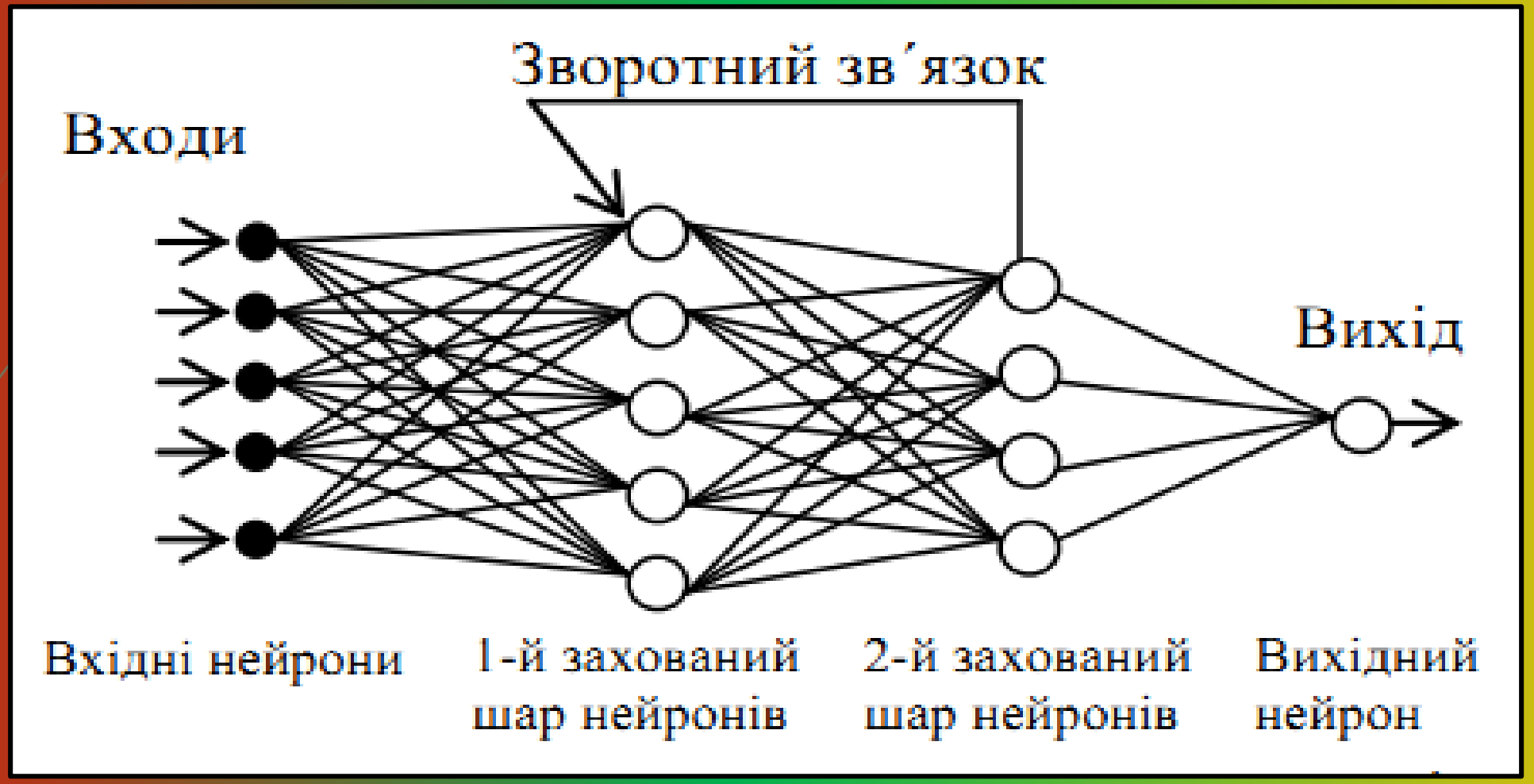


Рис.7. Нейромережа зі зворотніми зв'язками