



Перцептрон та його різновиди

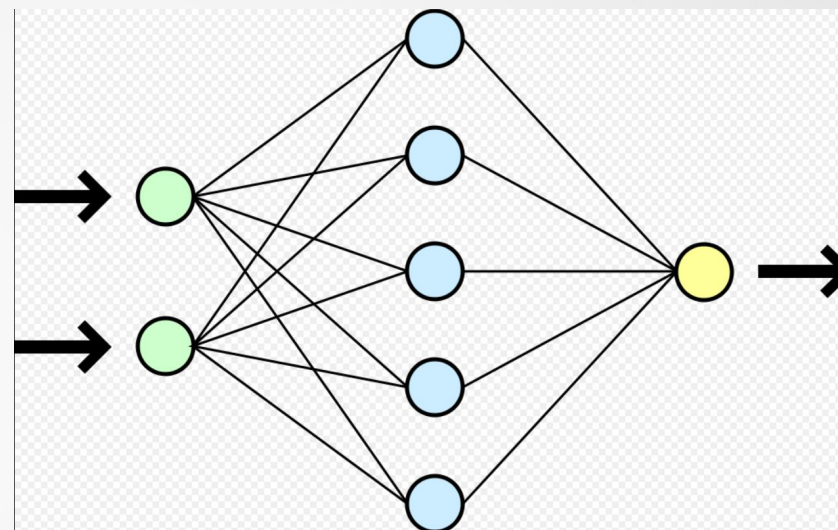
Штучна нейронна мережа

це математична модель, а також її програмна та апаратна реалізація, побудовані за принципом функціонування біологічних нейронних мереж — мереж нервових клітин живого організму. Це поняття виникло при вивченні процесів, які відбуваються в мозку, та при намаганні змоделювати ці процеси.

Після розробки алгоритмів навчання отримувані моделі стали використовуватися в практичних цілях: в задачах прогнозування, для розпізнавання образів, в задачах керування тощо.

Нейронні мережі не програмуються в звичайному розумінні цього слова, вони навчаються. Можливість навчання — одна з головних переваг нейронних мереж перед традиційними алгоритмами.

Технічно, навчання полягає в знаходженні коефіцієнтів зв'язків між нейронами. В процесі навчання нейронна мережа здатна виявляти складні залежності між вхідними даними й вихідними, а також здійснювати узагальнення. Це означає, що в разі успішного навчання мережа зможе повернути правильний результат на підставі даних, які були відсутні в навчальній вибірці



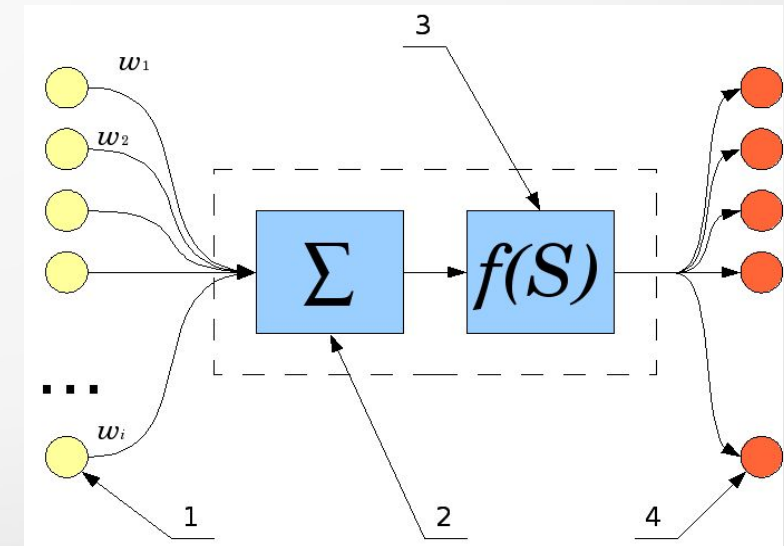
Штучний нейрон

вузол штучної нейронної мережі, що є спрощеною моделлю природного нейрона. Математично, штучний нейрон зазвичай представляють як деяку нелінійну функцію від єдиного аргументу - лінійної комбінації всіх вхідних сигналів. Цю функцію називають функцією активації.

Отриманий результат посилається на єдиний вихід. Такі штучні нейрони об'єднують в мережі - з'єднують виходи одних нейронів з входами інших. Штучні нейрони та мережі є основними елементами ідеального нейрокомп'ютеру.

Схема штучного нейрону

1. Нейрони, вхідні сигнали яких надходять на вхід даного нейрону
2. Суматор вхідних сигналів
3. Обчислювач передавальної функції
4. Нейрони, на входи яких подається сигнал даного нейрону
5. w — ваги вхідних сигналів



Математична

модель

Математично нейрон являє собою ваговий суматор, єдиний вихід якого визначається через його входи і матрицю ваг таким чином:

$$y = f(u), \text{ де } u = \sum_{i=1}^n w_i x_i + w_0 x_0$$

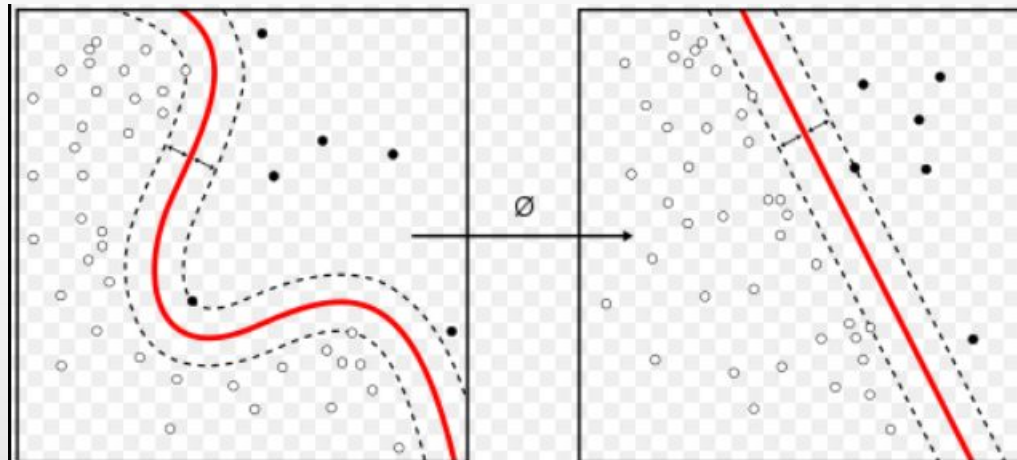
Тут x і w - відповідно сигнали на входах нейрона і ваги входів, функція f називається індукованим локальним полем, а $f(u)$ - передавальною функцією. Можливі значення сигналів на входах нейрона вважають заданими в інтервалі $[0,1]$.

Вони можуть бути або дискретними (0 або 1), або аналоговими. Додатковий вхід x_0 і відповідна йому вага використовується для ініціалізації нейрона. Під ініціалізацією мається на увазі зсув активаційної функції нейрона по горизонтальній осі, тобто формування порогу чутливості нейрона. Крім того, іноді до виходу нейрона спеціально додають якусь випадкову величину, яка називається зсувом. Зсув можна розглядати як сигнал на додатковому, завжди навантаженому, синапсі.

Перцептрон - математична або комп'ютерна модель сприйняття інформації мозком запропонована Френком Розенблатом й реалізована у вигляді електронної машини «Марк-1»

Перцептрон став однією з перших моделей нейромереж, а «Марк-1» — першим у світі нейрокомп'ютером.

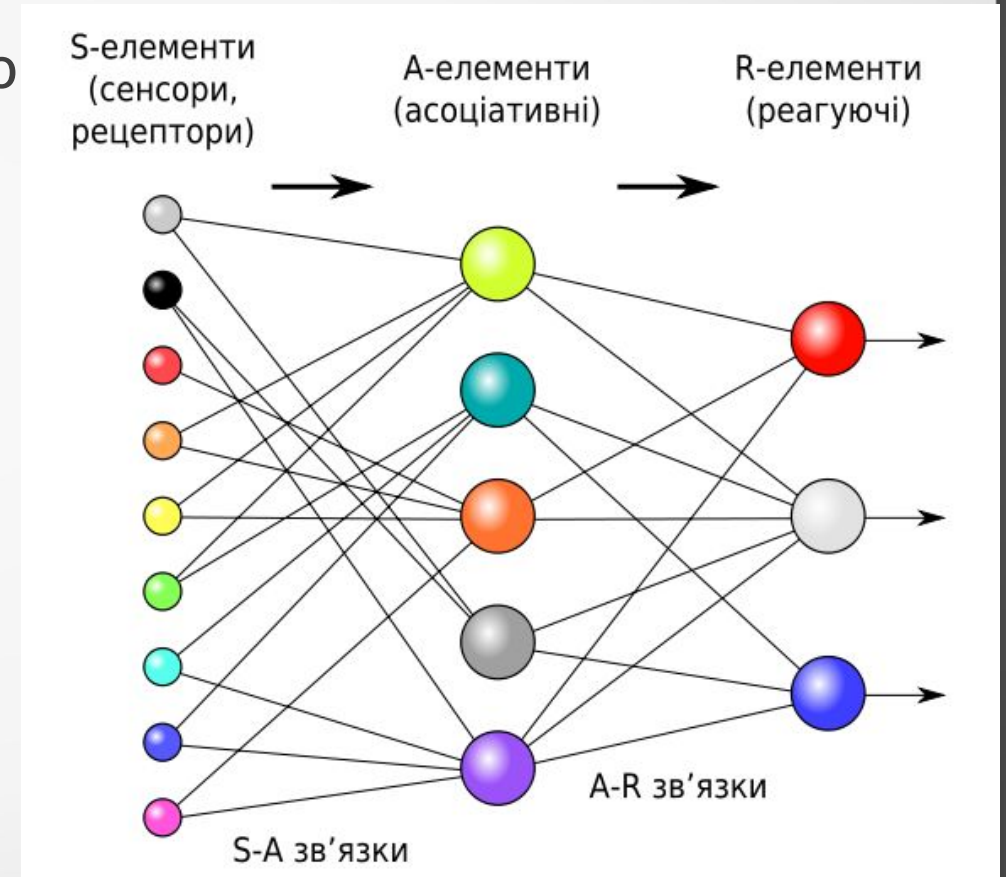
Незважаючи на свою простоту, перцептрон здатен навчатися і розв'язувати досить складні завдання. Основна математична задача, з якою він здатний впоратися — це лінійне розділення довільних нелінійних множин, так зване забезпечення лінійної сепарабельності.



Перцептрон складається з трьох типів елементів, а саме:

сигнали, що надходять від давачів, передаються до асоціативних елементів, а відтак до реагуючих. Таким чином, перцептрони дозволяють створити набір «асоціацій» між вхідними стимулами та необхідною реакцією на виході.

В біологічному плані це відповідає перетворенню, наприклад, зорової інформації у фізіологічну відповідь рухових нейронів.



Опис елементарного

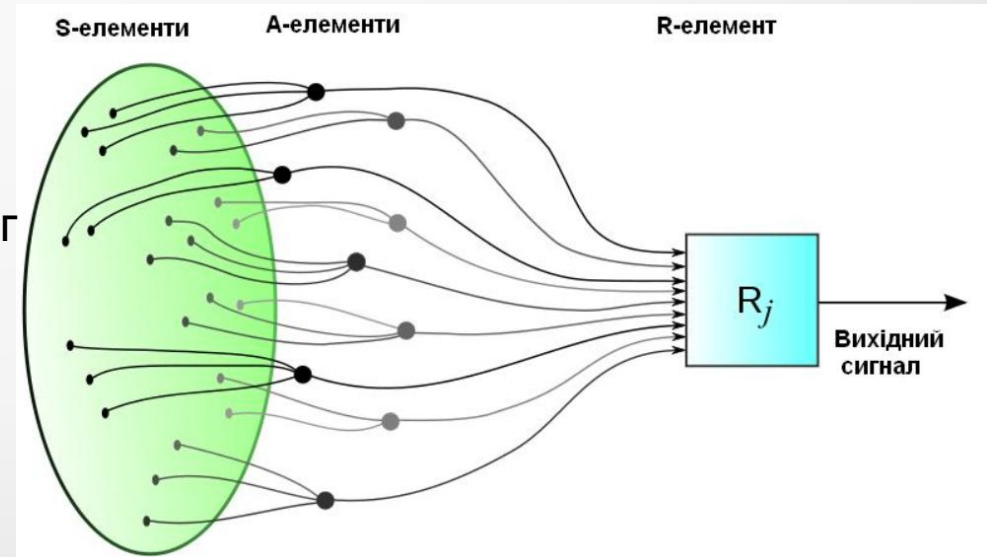
S-елементи — це шар сенсорів, або рецепторів. У фізичному втіленні вони відповідають, наприклад, світлочутливим клітинам сітківки ока або фоторезисторам матриці камери.

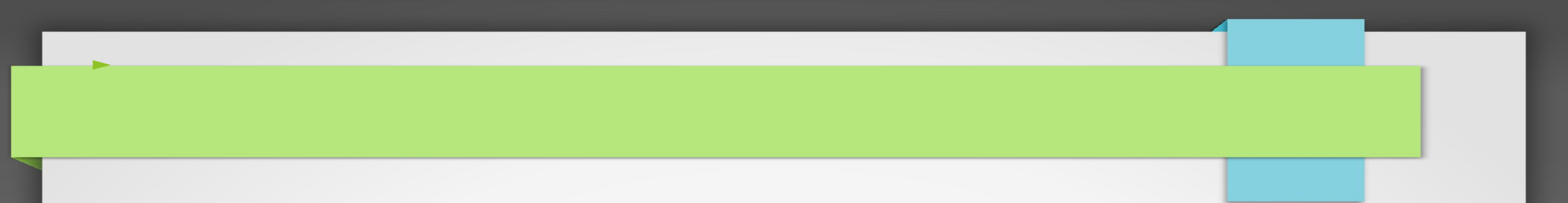
A-елементи називаються асоціативними, тому що кожному такому елементові, як правило, відповідає цілий набір (асоціація) S-елементів. A-елемент активізується, щойно кількість сигналів від S-елементів на його вході перевищує певну величину

Сигнали від збуджених A-елементів, своєю чергою, передаються до суматора R, причому сигнал від i -го асоціативного елемента передається з коефіцієнтом w . Цей коефіцієнт називається вагою A-R зв'язку.

Так само як і A-елементи, R-елемент підраховує суму значень вхідних сигналів, помножених на ваги (лінійну форму). R-елемент, а разом з ним і елементарний перцептрон, видає «1», якщо лінійна форма перевищує поріг θ , інакше на виході буде «-1». Математично, функцію, що реалізує R-елемент, можна записати так:

$$f(x) = \text{sign}\left(\sum_{i=1}^n w_i x_i - \theta\right)$$





Навчання елементарного перцептрона полягає у зміні вагових коефіцієнтів w зв'язків A-R. Ваги зв'язків S-A і значення порогів A-елементів вибираються випадковим чином на самому початку і потім не змінюються.

Після навчання перцептрон готовий працювати в режимі розпізнавання або узагальнення. У цьому режимі перцептрону пред'являються раніше невідомі йому об'єкти, й він повинен встановити, до якого класу вони належать. Робота перцептрона полягає в наступному: при пред'явленні об'єкта, збуджені A-елементи передають сигнал R-елементу, що дорівнює сумі відповідних коефіцієнтів w .

Якщо ця сума позитивна, то ухвалюється рішення, що даний об'єкт належить до першого класу, а якщо вона негативна — то до другого.

Класифікація перцептронів

Перцептрон з одним прихованим шаром

Одношаровий перцептрон

Багатошаровий перцептрон

Розенблатта
Багатошаровий перцептрон

Румельхарта

Перцептрон з одним прихованим шаром

Це класичний перцептрон, у нього є по одному шару S-, A- та R-елементів.

Одношаровий перцептрон

Це модель, у якій вхідні елементи безпосередньо з'єднано з вихідними за допомогою системи ваг. Є найпростішою мережею прямого поширення — лінійним класифікатором, і окремим випадком класичного перцептрона, в якому кожен S-елемент однозначно відповідає одному A-елементові, S-A зв'язку мають вагу +1, і всі A-елементи мають поріг $\theta = 1$. Одношарові перцептрони фактично є формальними нейронами

Багатошаровий перцептрон Розенблатта

перцептрон з додатковими шарами A — елементів, розташованими між S і R елементами. Перцептрон Розенблатта відрізняється від багатошарового перцептрону Румельхарта, і є загальнішим випадком по відношенню до нього. Оскільки елементарний перцептрон вже володів двома шарами зв'язків та трьома шарами елементів (нейронів), то такий перцептрон не вважався багатошаровим, і багатошаровість малася на увазі тільки при наявності мінімум чотирьох шарів елементів. У багатошаровому перцептроні Розенблатта не обов'язково всі зв'язки можна навчати, частина з них може бути випадково обрана і зафіксована. Румельхард ж припускав, що всі зв'язки багатошарового перцептрона повинні мати здатність навчатися. Тому повним еквівалентом багатошарового перцептрону Румельхарта, у Розенблатта є перцептрон зі змінними SA зв'язками.

Багатошаровий перцептрон Румельхарта

окремий випадок перцептрона Розенблатта, в якому один алгоритм зворотного поширення помилки навчає всі шари. На жаль, назва з історичних причин не відображає особливості даного виду перцептрона, тобто не пов'язана з тим, що в ньому кілька шарів. Особливістю є наявність більш ніж одного учня шару. Необхідність у великій кількості шарів-учнів відпадає, оскільки теоретично єдиного прихованого шару достатньо, щоб перекодувати вхідний сигнал таким чином, щоб отримати лінійну карту для вихідного сигналу. Але є припущення, що, використовуючи більше число шарів, можна зменшити число елементів у них, тобто сумарне число елементів у шарах буде менше, ніж при використанні одного прихованого шару.

Відмінності багат шарового перцептрона від перцептрону
Розенблатта
Використання нелінійної функції активації, як правило сигмоїдної.

Число шарів, які навчають, більше одного. Найчастіше використовується не більше трьох.

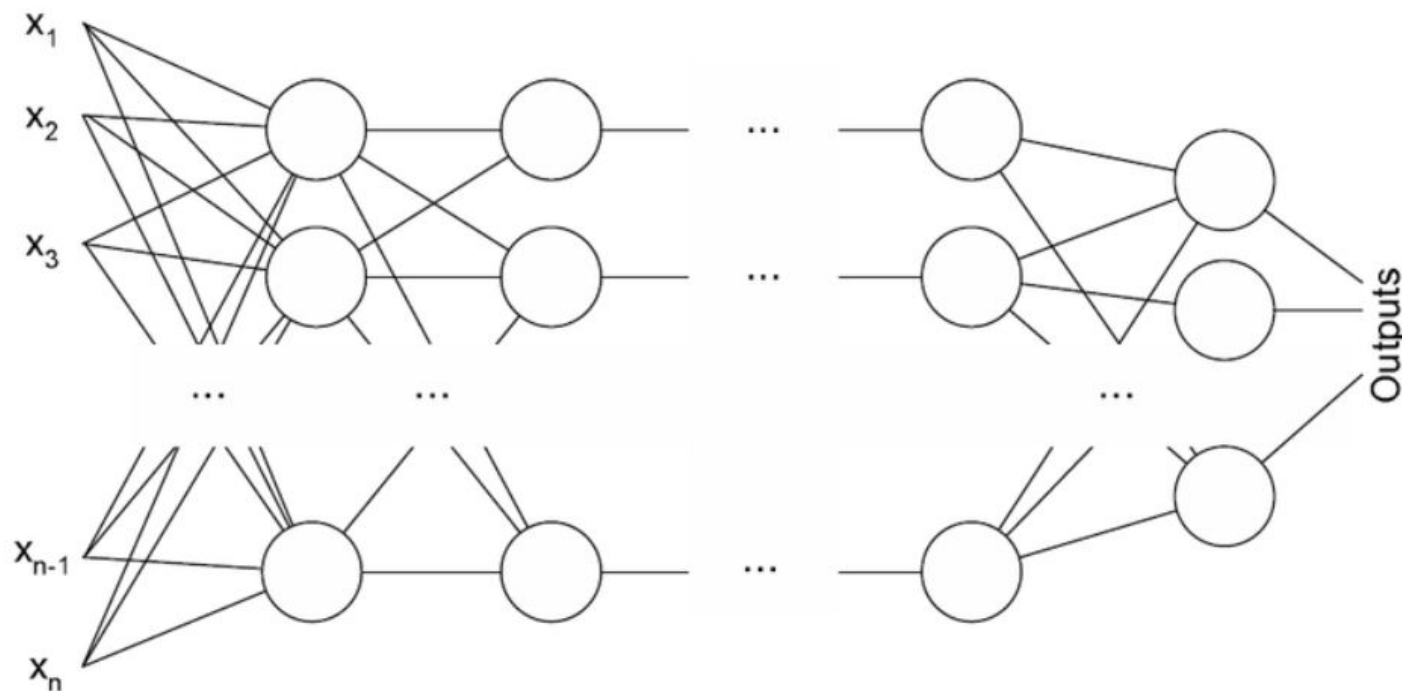
Сигнали, що надходять на вхід, та одержувані з виходу не бінарні, а можуть кодуватися десятковими числами, які потрібно нормалізувати, так щоб значення були на відрізку була від 0 до 1 (нормалізація необхідна як мінімум для вихідних даних, згідно з функцією активації — сигмоїдою).

Допускається довільна архітектура зв'язків (у тому числі, і повнозв'язані мережі).

Помилка мережі обчислюється не як число неправильних образів після ітерації навчання, а як деяка статистична міра нев'язаності між потрібним і одержаним значенням.

Навчання проводиться не до відсутності помилок після навчання, а до стабілізації вагових коефіцієнтів при навчанні або переривається раніше, щоб уникнути перенавчання.

Багатошаровий перцептрон буде володіти функціональними перевагами в порівнянні з перцептроном Розенблатта лише в тому випадку, якщо у відповідь на стимули не просто буде виконана якась реакція (оскільки вже в перцептроні може бути отримана реакція кожного типу), а виразиться у підвищенні ефективності вироблення таких реакцій. Наприклад, покращиться здатність до узагальнення, тобто до правильних реакцій на стимули, яким перцептрон не навчався. Але зараз таких узагальнюючих теорем немає, існує лише маса досліджень різних стандартизованих тестів, на яких порівнюються різні архітектури.



Алгоритми навчання

Важливою властивістю будь-якої нейронної мережі є здатність до навчання. Процес навчання є процедурою налаштування ваг та порогів з метою зменшення різниці між бажаними (цільовими) та отримуваними векторами на виході.

Система підкріплення — це будь-який набір правил, на підставі яких можна змінювати з плином часу матрицю взаємодії (або стан пам'яті) перцептронів.

- Навчання з учителем
- Навчання без учителя
- Метод зворотного поширення помилки

Навчання з

учителем

Класичний метод навчання перцептрон — це метод корекції помилки. Він являє собою такий вид навчання з учителем, при якому вага зв'язку не змінюється до тих пір, поки поточна реакція перцептрона залишається правильною. При появі неправильної реакції вага змінюється на одиницю, а знак (+/-) визначається протилежним від знаку помилки.

Припустимо, ми хочемо навчити перцептрон розділяти два класи об'єктів так, щоби при пред'явленні об'єктів першого класу вихід перцептрона був позитивний (+1), а при пред'явленні об'єктів другого класу — негативним (-1). Для цього виконаємо наступний алгоритм:

1. Випадково вибираємо пороги для А-елементів та встановлюємо зв'язки S-A .
2. Початкові коефіцієнти w вважаємо рівними нулеві.
3. Пред'являємо навчальну вибірку: об'єкти (наприклад, кола або квадрати) із зазначенням класу, до якого вони належать.

Показуємо перцептроніві об'єкт першого класу. При цьому деякі А-елементи збудяться.

Коефіцієнти w , що відповідають цим збудженням елементів, збільшуємо на 1.

Пред'являємо об'єкт другого класу, і коефіцієнти w тих А-елементів, які збудилися при цьому показі, зменшуємо на 1.

4. Обидві частини кроку 3 виконаємо для всієї навчальної вибірки. В результаті навчання сформується значення вагів зв'язків w .

Навчання без

учителя

Крім класичного методу навчання перцептронів, Розенблат також ввів поняття про навчання без учителя, запропонувавши наступний спосіб навчання:

Альфа-система підкріплення — це система підкріплення, за якої ваги всіх активних зв'язків C , що ведуть до елемента U , змінюються на однакову величину r , а ваги неактивних зв'язків за цей час не змінюються.

Пізніше, з розробкою поняття багат шарового перцептронів, альфа-систему було модифіковано, і її стали називати дельта-правилом. Модифікацію було проведено з метою зробити функцію навчання диференційовною (наприклад, сигмоїдною), що в свою чергу потрібно для застосування методу градієнтного спуску, завдяки якому можливе навчання більше ніж одного шару.

Метод зворотного поширення помилки

Для навчання багатошарових мереж ряд учених, було запропоновано градієнтний алгоритм навчання з учителем, що проводить сигнал помилки, обчислений виходами перцептрона, до його входів, шар за шаром. Зараз це є найпопулярніший метод навчання багатошарових перцептронів. Його перевага в тому, що він може навчити всі шари нейронної мережі, і його легко прорахувати локально. Однак цей метод є дуже довгим, до того ж, для його застосування потрібно, щоб передавальна функція нейронів була диференційовною. При цьому в перцептронах довелося відмовитися від бінарного сигналу, і користуватися на вході неперервними значеннями.

Застосування перцептронів

Перцептрон може бути використано, наприклад, для апроксимації функцій, для задачі прогнозування (й еквівалентної їй задачі розпізнавання образів), що вимагає високої точності, та задачі керування агентами, що вимагає високої швидкості навчання.

У практичних задачах від перцептрона вимагатиметься можливість вибору більш ніж з двох варіантів, а отже, на виході в нього має бути більше одного R-елемента.

Апроксимація функцій

Прогнозування та розпізнавання образів

Керування
агентами

Обмеження

Перцептрони мають обмеження в задачах, пов'язаних з інваріантним представленням образів, тобто незалежним від їхнього положення на сенсорному полі та положення щодо інших фігур. Такі задачі виникають, наприклад, якщо нам потрібно побудувати машину для читання друкованих літер або цифр так, щоб ця машина могла розпізнавати їх незалежно від положення на сторінці (тобто щоб на рішення машини не впливали перенесення, обертання, розтяг-стиск символів) або якщо нам потрібно визначити зі скількох частин складається фігура або чи знаходяться дві фігури поруч чи ні.

Перцептрони не мають функціональної переваги над аналітичними методами (наприклад, статистичними) в задачах, пов'язаних із прогнозуванням. Тим не менше, в деяких випадках вони представляють простіший і продуктивніший метод аналізу даних.

Деякі задачі в принципі може бути розв'язано перцептроном, але вони можуть вимагати нереально великого часу або нереально великої оперативної пам'яті.