Класичний і сучасні персептрони

Н.І.Недашківська

Інститут прикладного системного аналізу Національного технічного університету України "Київський політехнічний інститут ім. Ігоря Сікорського"

Київ-2019

Нервова клітина, яка називається **нейроном** – це базовий елемент нервової системи людини.
В нейроні виділяють тіло клітини та два види відростків:
1) за якими нейрон отримує інформацію – **дендрити**,
2) за яким нейрон передає інформацію – **аксон**.
Кожний нейрон людини має лише один вихідний відросток.
Нейрон передає збудження іншим нейронам через нервові стики, які називаються **синапсами**. Мозок людини налічує порядку 10^{11} нейронів і $10^{14} - 10^{15}$ синапсів.

В результаті функціонування синапсів збудження може підсилюватися або послаблюватися. Синаптична вага штучного нейрону може приймати як додатні так і від'ємні значення. В результаті до нейрону приходять збуджуючі сигнали та сигнали, які мають гальмівну дію.

Нейрон додає всі вхідні імпульси.

Якщо їх сума перевищує деяке порогове значення, то вихідний сигнал нейрону пересилається за допомогою аксону до інших нейронів.

Таким чином, нейрон може знаходитися у двох різних станах: у «вимкненому» частота подачі сигналів маленька, у збудженому нейрон «включається» і частота подачі імпульсів сильно збільшується. Активація нейронів залежить від сигналів, що приходять через синапси і потім дендрити від інших нейронів.

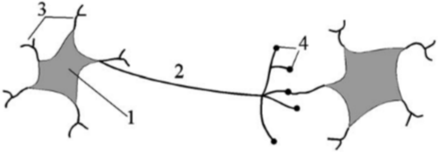


Рис.: Модель нейрона: 1 – тіло клітини, 2 – аксон, 3 – дендрити, 4 – синапси

Нейрони працюють стохастично, вони пересилають сигнали через випадкові проміжки часу. В цьому їх відмінність від транзисторів, з яких складається процесор комп'ютера. Останні працюють синхронізовано.

Риси мозку людини

В основу штучних нейронних мереж покладено наступні риси мозку людини, які дозволяють їм добре справлятися з інтелектуальними задачами:

- простий обчислювальний елемент - нейрон;
- дуже велике число нейронів бере участь в обробці інформації;
- один нейрон пов'язаний з великою кількістю інших нейронів (глобальні зв'язки);
- зв'язки між нейронами змінюються у часі;
- паралельність обробки інформації.

Нейронна мережа - це розподілений паралельний процесор, який складається з великої кількості елементарних одиниць обробки інформації, що накопичують експериментальні знання і надають їх для обробки. Для накопичення знань застосовуються зв'язки між нейронами, які називаються **синаптичними вагами**.

Нейронна мережа - це сукупність великого числа порівняно простих елементів - нейронів, топологія з'єднань між якими залежить від типу мережі. Щоб створити нейронну мережу для вирішення певної задачі потрібно вибрати яким чином з'єднувати нейрони один з одним і підібрати значення вагових параметрів на цих зв'язках. Вплив одного нейрону на інший визначається встановленими зв'язками. Вага зв'язку визначає силу впливу.

Модель МакКаллока-Піттса (1943)

$u_1, u_2, \ldots, u_n$ – вхідні сигнали даного нейрону, які надходять від інших нейронів,
$w_1, w_2, \ldots, w_n$ – синаптичні ваги, v – порогове значення.
Тоді вихідний сигнал нейрону дорівнює

$$y = \begin{cases} 1, \text{ при } \sum_{i=1}^{n} w_i u_i \geq v, \\ 0, \text{ при } \sum_{i=1}^{n} w_i u_i < v. \end{cases}$$

Модель може бути представлена у вигляді:

$$y = f\left(\sum_{i=0}^{n} w_i u_i\right),$$

де

$$f(x) = \begin{cases} 1, \text{ при } x \geq 0, \\ 0, \text{ при } x < 0. \end{cases}$$

і $w_0 = v$, $u_0 = -1$.

Модель нейрона МакКаллока-Піттса (1943)

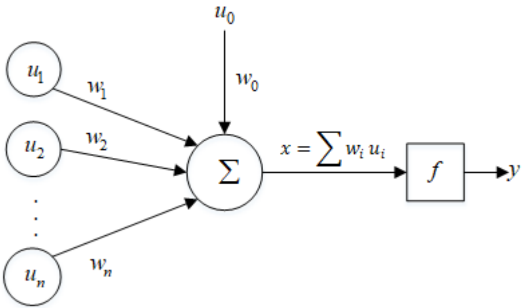


Рис.: Модель нейрона МакКаллока-Піттса

Інші функції активації

Функція активації f обмежує амплітуду вихідного сигналу нейрона.
В якості функції активації в інших моделях нейронних мереж використовуються порогові функції виду

$$f(x) = \begin{cases} 1, \text{ при } x \geq 0, \\ -1, \text{ при } x < 0. \end{cases}$$

або

$$f(x) = \begin{cases} 1, \text{ при } x > 1, \\ -1, \text{ при } x < -1, \\ x, \text{ при } |x| \leq 1. \end{cases}$$

Сигмоїдальна функція:

$$s(x) = \frac{1}{1 + e^{-\beta x}} > 0,$$

де β - параметр нахилу. В граничному випадку $s(x)$ вироджується у порогову. На відміну від порогової, $s(x)$ диференційовна.

Функція *гіперболічного тангенсу*

$$f(x) = tanh(\frac{\alpha x}{2}) = \frac{1 - e^{-\alpha x}}{1 + e^{-\alpha x}} > 0.$$

tanh значно швидше зростає і спадає за s. Для s нуль - точка насичення: при навчанні значення s в нуль, вхід буде наближатися до $-\infty$, а похідна - до нуля, це стабільний стан.

Для *tanh* нуль - це найменш стабільна проміжна точка.

Персептрон – найпростіша однонаправлена нейронна мережа. Її запропонував і дослідив Розенблатт в кінці 1950х – на початку 1960х років.
Сигнал x на виході лінійної частини персептрона задається виразом

$$x = \sum_{i=1}^{n} w_i u_i - v = \sum_{i=0}^{n} w_i u_i,$$

де $w_0 = v, u_0 = -1$.
Функція активації:

$$f(x) = \begin{cases} 1, \text{ при } x \geq 0, \\ -1, \text{ при } x < 0. \end{cases}$$

Персептрон ділить n-вимірний простір вхідних векторів u на два підпростори за допомогою розділяючої гіперплощини

$$\sum_{i=1}^{n} w_i u_i - v = \sum_{i=0}^{n} w_i u_i = 0.$$

Нехай ваги $w_0, w_1, ..., w_n$ в рівнянні розділяючої гіперплощини невідомі.
Є навчальна вибірка $u(t)$, $t = 1, 2, ...$:

$$u(t) = (u_1(t), u_2(t), ..., u_n(t))^T.$$

Нехай навчальна вибірка лінійно роздільна за допомогою гіперплощини на два класи $\{u_1(t)\} \in C_1$ і $\{u_2(t)\} \in C_2$.

Навчання персептрона Розенблатта

В t-й момент часу сигнал на виході лінійної частини персептрона задається виразом

$$x(t) = \sum_{i=0}^{n} w_i(t)u_i(t) = w^T(t)u(t),$$

де

$$u(t) = (-1, u_1(t), u_2(t), ..., u_n(t))^T$$

і

$$w(t) = (v(t), w_1(t), w_2(t), ..., w_n(t))^T.$$

Навчання моделі полягає в тому, що вектор ваг на кроці $t+1$ коригується за формулами

$$w(t+1) = \begin{cases} w(t), \text{ при } u(t) \in C_1 \text{ і } w^T(t)u(t) \geq 0 \\ w(t), \text{ при } u(t) \in C_2 \text{ і } w^T(t)u(t) < 0 \end{cases}$$

і

$$w(t+1) = \begin{cases} w(t) + \eta u(t), \text{ при } u(t) \in C_1 \text{ і } w^T(t)u(t) < 0 \\ w(t) - \eta u(t), \text{ при } u(t) \in C_2 \text{ і } w^T(t)u(t) \geq 0 \end{cases},$$

де параметр $0 < \eta < 1$ – це крок коригування, і початкові значення встановлюються рівними нулю

$$w(0) = 0.$$

Визначимо еталонний або заданий сигнал $d(t)$:

$$d(t) = \begin{cases} 1, \text{ при } u(t) \in C_1 \\ -1, \text{ при } u(t) \in C_2 \end{cases}.$$

Значення вихідного сигналу персептрона дорівнює

$$y(t) = sgn(w^T(t)u(t)).$$

Тоді

$$w(t+1) = w(t) + \eta(d(t) - y(t))u(t).$$

Різниця $d(t) - y(t)$ інтерпретується як похибка між еталоном $d(t)$ і фактичним вихідним сигналом $y(t)$.

Навчання персептрона Розенблатта

За умови лінійної роздільності вхідних сигналів алгоритм збігається, тобто

$$w(t_0) = w(t_0 + 1) = w(t_0 + 2) = \dots .$$

Після виконання навчання роздільна гіперплощина описується рівнянням

$$\sum_{i=0}^{n} w_i(t_0)u_i = 0,$$

і персептрон коректно класифікує як сигнали, які належать навчальній вибірці $\{u(t)\}$, так і ті, що не належать навчальній множині, але є лінійно роздільними.

У перших моделях **багатошарових персептронів** використовувалася *сигмоїдальна функція:*

$$s(x) = \frac{1}{1 + e^{-\beta x}} > 0,$$

де β - параметр нахилу. В граничному випадку $s(x)$ вироджується у порогову. Функція $s(x)$ диференційовна, на відміну від порогової.
Багатошарові персептрони з функцією активації $s(x)$ навчаються методом **зворотного розповсюдження похибки** (error back-propagation), який використовує алгоритм мінімізації середньоквадратичної помилки (LMS) на основі градієнтного спуску.

Функції активації у сучасних перцептронів

В багаторівневій нейронній мережі Neocognitron, розробленій К.Фукусімою в 1980-х, запропоновано **функцію активації ReLU** (rectified linear units):

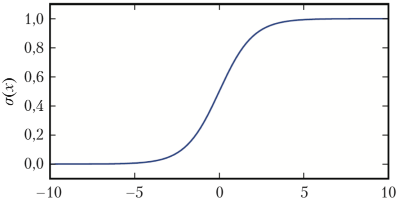
$$ReLU(x) = \begin{cases} 0, \text{ при } x < 0, \\ x, \text{ при } x \geqslant 0 \end{cases}$$

або $\mathrm{ReLU}(x) = \max(0, x)$.

Функції активації у сучасних перцептронів

Причини появи ReLU пов'язані з властивостями сигмоїдальної функції:

- Їй важко відрізнити активацію «з силою 5» ($s(5) = 0,9933$ при $\beta = 1$) від активації «з силою 10» ($s(10) = 0,99995$).
- Вміє відрізняти «достатньо активовані» нейрони, коли результат активації близький до одиниці, від «недостатньо активованих», коли цей результат близький до нуля.



Поняття «сили активації» може бути виражене функцією ReLU. Розглянемо функцію активації, яка є сумою нескінченого ряду сигмоїдальних:

$$S(x) = s(x + 1/2) + s(x - 1/2) + s(x - 3/2) + \ldots,$$

де кожен сигмоїд в цій сумі зміщений вправо відносно попереднього на одиницю, і нехай $\beta = 1$.
Якщо $x \to -\infty$, то $S \to 0$. Справа від нуля сигмоїди, центр яких знаходиться правіше від x, будуть як і раніше близькі до нуля, і ряд буде збігатися в будь-якій кінцевій точці. Сигмоїди, центр яких знаходиться лівіше за x, будуть близькі до одиниці. Тому активація за допомогою такої множини сигмоїдів в результаті дає значення x.

Функції активації у сучасних перцептронів

$$S(x) = \sum_{i=0}^{\infty} s\left(x + 1/2 - i\right) \approx \int_{1/2}^{\infty} s\left(x + 1/2 - y\right) dy =$$

$$= \left(- \ln \left(1 + \exp(x + 1/2 - y) \right) \right)_{y=1/2}^{y=\infty} = \ln(1 + \exp(x)),$$

оскільки $\int s(x)dx = \ln(1 + \exp(x)) + C$.
Сума $S(x)$ нескінченного числа сигмоїдальних функцій краще **виражає поняття «сили активації»** ніж окремий сигмоїд і може бути **наближена функцією** $\ln(1 + \exp(x))$ **і ReLU(x)**.

По-друге, **нейрони з ReLU функцією ефективніші** за ті, що базуються на сигмоїдальній $s(x)$. Похідна $s^{'}(x) = s(x)(1 - s(x))$. Для підрахунку похідної від ReLU(x) потрібно лише виконати одне порівняння x з нулем. Тому мережі з ReLU-нейронів можуть **містити значно більшу кількість нейронів** ніж мережі з більш складними функціями активації за всіх інших одникових умов.

По-третє, активація за допомогою ReLU більш точно відображає процеси, які проходять у мозку людини. Енергоефективність мозку досягається **розріженістю активації** його нейронів. Використовуючи ReLU-активацію з регуляризацією, можна добитися розріженості активації, на відміну від сигмоїдальної функції, для якої в кожний момент часу активується приблизно половина нейронів.

Функції активації у сучасних перцептронів

Leaky ReLU:

$$LReLU(x) = \begin{cases} ax, \text{ при } x < 0, \\ x, \text{ при } x \geqslant 0 \end{cases},$$

де $a > 0$ - невелике, наприклад, $a = 0.1$.
Мета - покращення розв'язку за рахунок можливості навчання нейронів в області $(-\infty, 0)$. Відомі приклади застосувань нейронних мереж в акустиці, де функція LReLU покращує розпізнавання мови.

Параметризована ReLU (Parametric ReLU, PReLU) співпадає з LReLU. Відмінність: параметр $a > 0$ також піддається навчанню.
Експоненційний лінійний нейрон (Exponential Linear Unit, ELU):

$$ELU(x) = \begin{cases} a(\exp(x) - 1), \text{ при } x < 0, \\ x, \text{ при } x \geqslant 0 \end{cases}.$$

- Присутні від'ємні значення, що важливо для навчання.
- Швидке насичення при зменшенні аргументу, що важливо для збереження розрідженості.

Розробляють також інші функції активації, які мають переваги при розв'язанні окремих задач.

Дякую за увагу!