

ФУНКЦІЯ АКТИВАЦІЇ НЕЙРОНА RELU

Штучна нейронна мережа - це загальна назва для цілого класу моделей. Як правило, вони представляють собою комбінацію нелінійних перетворень вхідних даних. Останнім часом все більшої популярності набуває глибоке навчання (deep learning). Глибоке навчання є галуззю в машинному навчанні і ґрунтується на деякому наборі алгоритмів. За допомогою глибоких нейронних мереж успішно вирішуються різні завдання, пов'язані з класифікацією даних, прогнозування подій, розпізнавання мови, обробки текстів і т.д.

Класичний алгоритм зворотного поширення помилки працює добре з двошаровими та тришаровими нейронними мережами, але при подальшому збільшенні глибини мережі можуть виникнути проблеми. Одна з яких це так зване згасання градієнту. В процесі поширення помилки від початкового шару до вихідного відбувається множення поточного результату на похідну функції активації. Використовуючи традиційну сигмоїдну функцію активації, похідна якої має область визначення менше 5 одиниць, помилка після проходження декількох шарів може бути близькою до нуля. Якщо ж навпаки взяти функцію активації в якій похідна необмежена (як, наприклад, гіперболічний тангенс), то може статися збільшення помилки в процесі навчання, що призведе до нестійкого навчання мережі.

За останні роки великої популярності набула функція активації ReLU (rectified linear unit). Її похідна дорівнює або одиниці, або нулю, і тому не може статися розростання або загасання градієнтів. Більш того, використання даної функції приводить до проріджування ваг.

ReLU має наступну формулу: $\sigma(x) = \max(0, x)$ (рис 1).

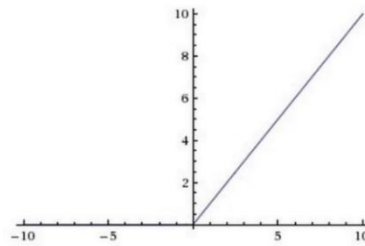


Рис. 1. Графік функції ReLU

Розглянемо позитивні та негативні сторони ReLU.

Позитивні сторони:

1. Сигмоїд і гіперболічний тангенс потребує для виконання операцій велику кількість системних ресурсів, таких як піднесення до степеню, що при великій кількості шарів та нейронів сповільнюють процес навчання, в той час як ReLU може бути реалізований за допомогою простого порогового перетворення матриці активацій в нулі.
2. Застосування ReLU істотно підвищує швидкість збіжності стохастичного градієнтного спуску в порівнянні з сигмоїд і гіперболічним тангенсом. Вважається, що це обумовлено лінійним характером і відсутністю насичення даної функції.

Негативні сторони:

1. На жаль, ReLU не завжди достатньо надійні і в процесі навчання можуть виходити з ладу («вмирати»). Наприклад, великий градієнт, що проходить через ReLU, може привести до такого оновлення ваг, що даний нейрон ніколи більше не активується. Якщо це станеться, то, починаючи з даного моменту, градієнт, що проходить через цей нейрон, завжди буде дорівнює нулю. Відповідно, даний нейрон буде необоротно виведений з ладу. Наприклад, при дуже великій швидкості навчання (learning rate), може виявитися, що до 40% ReLU «мертві» (тобто, ніколи не активуються). Ця проблема вирішується за допомогою вибору належної швидкості навчання.

В даний час існує декілька різновидів даної функції:

1. ReLU з втратами є спробою вирішити проблему, яка пов'язана з втратами звичайної являє собою одну зі спроб вирішити виходу з ладу звичайних. ReLU на інтервалі $x < 0$ на своєму виході дає 0, а LReLU на цьому інтервалі має невелике від'ємне значення (кутовий коефіцієнт близько 0,01). Тобто функція для LReLU має вигляд $f(x) = \alpha x$ при $x < 0$ і $f(x) = x$ при $x \geq 0$, де α - мала константа. Деякі дослідники повідомляють що успішно застосували дану функцію, але результати не завжди стабільні.
2. Для ReLU з параметрами - кутовий коефіцієнт не задається на початку, а обраховується на основі даних які отримуються під час. Процес зворотного поширення помилки і оновлення для PReLU досить простий і подібний до процесу для традиційних ReLU.

В своїй роботі, по розпізнаванню рукописних цифр на базі mnist та багатшарової нейронної мережі, було використано функцію активації ReLU. Вона дозволила збільшити швидкість навчання мережі в порівнянні з сигмоїдом, при цьому помилка навчання була менша.