**Теоретичний матеріал до лабораторної роботи №4**

**РОЗПІЗНАВАННЯ ДРУКОВАНИХ І РУКОПИСНИХ ЛІТЕР**

На рис.1 представлена схема персептрона, призначеного для розпізнавання букв українського алфавіту. Персептрон має 33 вихідних нейрона: кожній букві алфавіту відповідає свій вихідний нейрон. Вважається, що сигнал першого вихідного нейрона повинен бути рівним одиниці, якщо персептрону пред'явлена буква «А», і нулю для всіх інших букв. Вихід другого нейрона має дорівнювати одиниці, якщо персептрону пред'явлена буква «Б», і нулю в усіх інших випадках. І так далі до букви «Я».

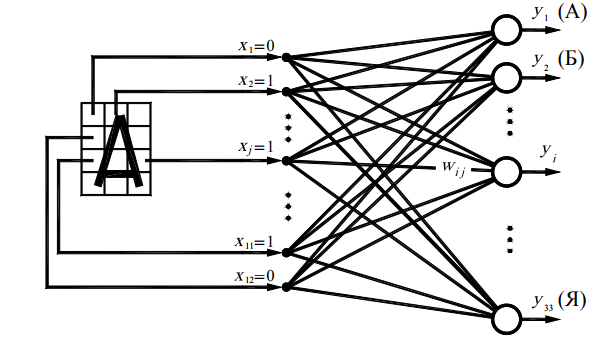


Рис.1 Персептрон, призначений для розпізнавання букв українського алфавіту

При виконанні попередньої лабораторної роботи ми переконалися, що персептрон навчився розпізнавати не тільки букви, на яких його навчали, але і букви, яких в навчальній множині не було, якщо вони не дуже відрізнялися від букв навчальної множини. Властивість розпізнавати нові образи, які персептрон ніколи «не бачив», ми назвали властивістю *узагальнення*.

Подальший розвиток ідеї персептрона був пов'язаний зі спробами розширити коло його застосування і вдосконалити алгоритм навчання. Істотний розвиток персептрона був зроблений американськими вченими *Б.Уідроу і М.Е.Хоффом*, які замість вивченої на першій лабораторній роботі ступінчастої активаційної функції ввели безперервну нелінійну функцію активації

, (1)

графік якої зображено на рис.2.

Цю функцію назвали *сигмоїдою*, через те, що її графічне зображення нагадує латинську букву «S». Інша назва сигмоїди - *логістична функція*.

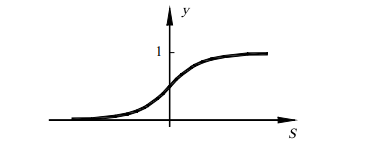


Рис.2 Сигмоїдальна активаційна функція

Подібно звичайній пороговій функції активації, сигмоїда відображає точки області визначення (- ∞, + ∞) в значення з інтервалу (0, +1).

Практично сигмоїда забезпечує безперервну апроксимацію класичної порогової функції.

Для сигмоїд взяли позначення . Персептрони з сигмоїдальними активаційними функціями з одним виходом назвали *адалайн*, з декількома виходами - *мадалайн* (від англійських слів *ADAptive LInear NEuron і Many ADALINE*).

Поява персептронів з безперервними активаційними функціями зумовило появу нових підходів до їхнього навчання. Б.Уідроу і М.Е.Хофф запропонували мінімізувати квадратичну помилку, яка визначається формулою:



в якій - необхідний (бажаний) вихід *i*-го нейрона, а - той, який вийшов в результаті обчислень персептрона.

Розглянемо алгоритм корекції вагових коефіцієнтів персептрона, що має *J* входів і *I* виходів (рис. 3).

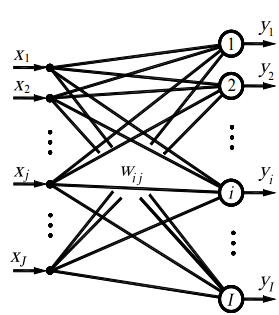


Рис. 3 Персептрон з має *J* входами і *I* виходами

Квадратична помилка навчання персептрона *ε* залежить від того, якими є вагові коефіцієнти . Іншими словами *ε* є функцією від вагових коефіцієнтів . У шкільних курсах зазвичай вивчаються функції тільки від одного аргументу: які на координатної площині *x*, *y* зображуються, як відомо, у вигляді кривих ліній. Якщо функція *z* залежить від двох аргументів: , то вона зображується в тривимірній системі координат *x, y, z* у вигляді поверхні. Функція-похибка персептрона залежить від великої кількості аргументів, тому для її графічного представлення необхідна багатовимірна система координат, яку ми в нашому тривимірному світі уявити собі не можемо. У цій багатовимірної системі координат функція зображується у вигляді багатовимірної поверхні, званої *гіперповерхнею*.

Щоб хоч якось уявити собі гіперповерхню, припустимо, що всі аргументи мають постійні значення за винятком двох, наприклад і , які є змінними. Тоді в тривимірній системі координат , , *ε* гіперповерхня матиме вигляд фігури, що нагадує параболоїд, яку назвемо *псевдопарабалоїдом* (рис.4). Процес навчання персептрона тепер можна уявити як віднайдення такого поєднання вагових коефіцієнтів , якому відповідає найнижча точка *гіперпсевдопараболоїда*. Завдання такого роду називаються *оптимізаційними*. Кажуть, що оптимізаційна задача полягає в *мінімізації функції* в багатовимірному просторі параметрів .

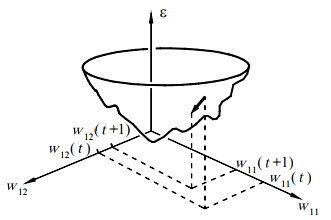


Рис.4. Графічне зображення функції-помилки персептрона в тривимірній системі координат , , *ε*

Таким чином, якщо раніше говорили, що персептрон навчають методом «заохочення - покарання», то тепер стали говорити, що завдання вивчення персептрона - це завдання оптимізації (мінімізації) персептронної помилки (похибки).

Існує безліч методів розв'язання оптимізаційних задач. Найбільш простим методом є перебір вагових коефіцієнтів з подальшими обчисленнями і порівняннями між собою відповідних цим коефіцієнтам значень функції ε. Більш ефективний метод *градієнтного спуску,* згідно з яким зміна (корекція) кожного вагового коефіцієнта *Δ* проводиться в сторону, протилежну *градієнту* функції *ε*. Градієнт функції є дуже важливим математичним поняттям, яке зазвичай проходять на перших курсах вузів. Тут ми не будемо на ньому зупинятися, а тільки зазначимо, що градієнт функції являє собою вектор, проекціями якого на осі координат є похідні від функції *ε* по цих координатах (їх позначають ∂*ε*/∂), і що градієнт функції завжди спрямований у бік її найбільшого зростання. Оскільки наше завдання полягає в знаходженні мінімуму функції , то нам треба опускатися по поверхні помилок, що забезпечується рухом в сторону, протилежну градієнту цієї функції. Звідси і згадана вище назва - *метод градієнтного спуску*.

Рух у бік, протилежний градієнту (тобто протилежний напрямку зростання функції), буде здійснюватися, якщо на кожній ітерації до координат поточної точки ми будемо додавати величину, прямо пропорційну похідній по координаті , взяту з протилежним знаком:



де *η* - деякий коефіцієнт, який зазвичай задається в межах від 0,05 до 1, і зветься, як і раніше, *коефіцієнтом швидкості навчання*.

Зверніть увагу, що згідно з формулою (3) ми рухаємося не тільки в сторону спадання функції, але і зі швидкістю, прямо пропорційною швидкості спадання (крутизні) функції, тому що робимо крок Δ, пропорційний похідній, взятій зі знаком мінус.

Квадратична помилка *ε* є складною функцією, що залежить від вихідних сигналів персептрона , які, в свою чергу, залежать від , тобто . За правилом диференціювання складної функції



Вихідні сигнали нейронів , обчислюються за допомогою сигмоїдальних активаційних функцій , аргументом яких є суми . Отже,



Крім того, якщо продиференціювати (2) по , де n ∈ [1, *I*], то вийде , значить



Підставивши (5) і (6) в (4) і потім отриманий вираз в (3), остаточно будемо мати



Цей вираз отримано для нейронів з активаційними функціями будь-якого виду. Якщо - сигмоїда, задана формулою (1), то



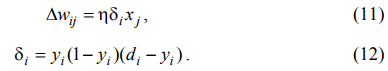
Підставивши цей вираз в (7), отримаємо:



Отже, ми отримали *ітераційну формулу* для навчання персептрона



де



Введену тут за допомогою формули (12) величину надалі будемо називати нейронною помилкою. Алгоритм (10) - (12) називають *узагальненим дельта-правилом*. Його перевага в порівнянні зі звичайним дельта правилом складається в більш швидкій збіжності і в можливості більш точної обробки вхідних і вихідних безперервних сигналів, тобто в розширенні кола задач, які вирішуються персептронами.

Отже, ведення сигмоїдальної функції активації замість функції-сходинки і поява нового алгоритму навчання - узагальненого дельта-правила, розширило сферу застосування персептрона. Тепер він може оперувати не тільки з бінарними (типу «нуль» і «одиниця»), але і з безперервними (аналоговими) вихідними сигналами.