



MÁY HỌC NÂNG CAO (Mạng nơ-ron nhân tạo)

Phạm Nguyên Khang
pnkhang@cit.ctu.edu.vn



CANTHO UNIVERSITY

Nội dung trình bày

- Giới thiệu
- Mô hình nơ-ron McCulloch & Pitts
- Mô hình perceptron đa tầng (MLP)
- Một số vấn đề cần chú ý khi huấn luyện mạng nơ-ron
- Bài tập



CANTHO UNIVERSITY

Giới thiệu

- Mạng nơ-ron nhân tạo (Artificial neural network - ANN)
 - Mô hình hoá hoạt động của hệ thần kinh con người
 - Được nghiên cứu lần đầu vào năm 1943 (McCulloch và Pitts, 1943)
 - Perceptron: thể hệ đầu tiên của mạng nơ-ron (Rosenblatt, 1958)
 - Mô phỏng quá trình hoạt động của thị giác con người



CANTHO UNIVERSITY

Giới thiệu

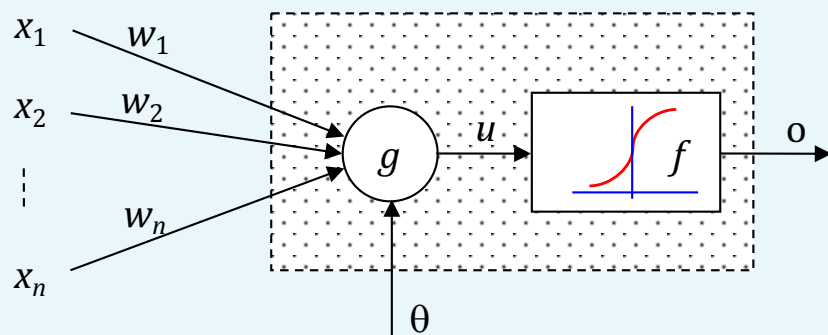
- Lịch sử
 - 1943, McCulloch & Pitts đưa ra mô hình nơ-ron đầu tiên
 - 1982, Mô hình mạng nơ-ron hồi tiếp của Hopfield
 - 1984, Mạng nơ-ron Kohonen hay còn gọi là Mảnh đồ tự tổ chức (SOM)
 - 1985, Mạng nơ-ron đa tầng (MLP)
- Mô hình mạng nơ-ron khác
 - Mạng nơ-ron tích chập (Convolutional NN) của Yan LeCun.



CANTHO UNIVERSITY

Mô hình nơ-ron McCulloch & Pitts

- Nơ-ron: đơn vị tính toán cơ bản/đơn vị của tất cả các mạng nơ-ron:
 - n đầu vào, 1 tham số, 1 đầu ra
 - Hàm mạng/hàm kết hợp
 - Hàm kích hoạt/hàm truyền



$$u = g(x) = \sum_{i=1} w_i x_i + \theta$$



CANTHO UNIVERSITY

Mô hình nơ-ron McCulloch & Pitts

- w_i : trọng số
- θ : ngưỡng (threshold), độ lệch (bias)
- Đầu ra:

$$o = f(u) = f(g(x))$$

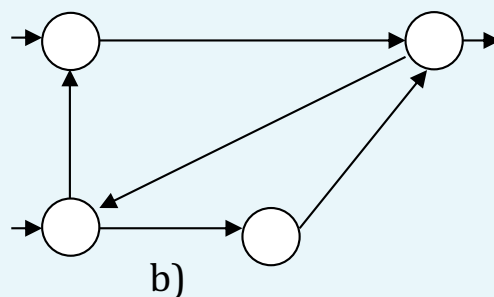
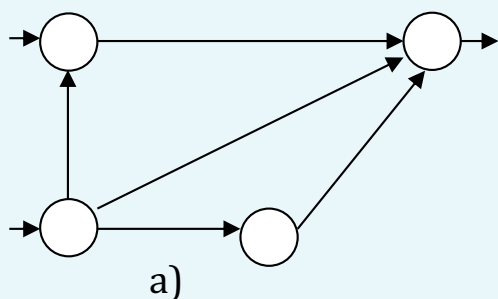
$$f(u) = \frac{1}{1 + e^{-u/T}} \quad f(u)[1 - f(u)] / T$$



CANTHO UNIVERSITY

Kiến trúc mạng

- Mạng truyền thẳng (forward)
- Mạng hồi tiếp (feedback)

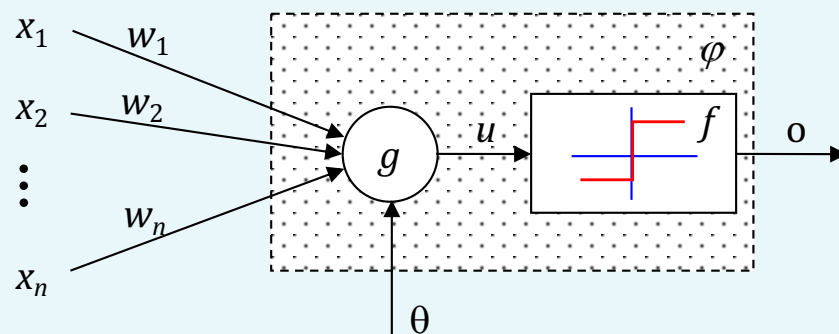




CANTHO UNIVERSITY

Mô hình perceptron đa tầng

- Mô hình perceptron
 - Do Rosenblatt đề xuất năm 1958
 - Tương tự như mô hình nơ-ron của McCulloch&Pitts
 - Perceptron tuyến tính có ngưỡng
 - n đầu vào, 1 ngưỡng, 1 đầu ra





CANTHO UNIVERSITY

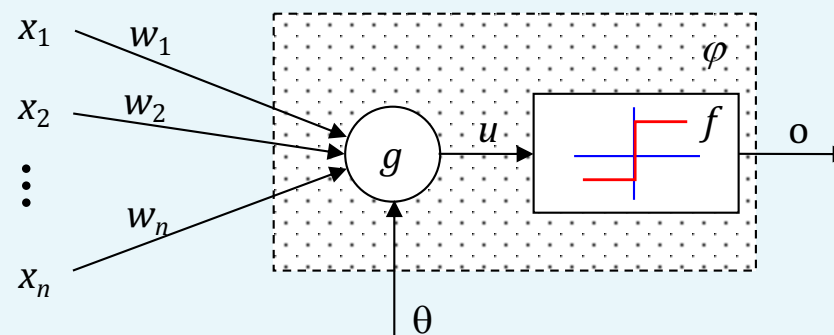
Mô hình perceptron

– Perceptron tuyến tính có ngưỡng

- n đầu vào, 1 ngưỡng, 1 đầu ra
- Hàm mạng tuyến tính

$$u = g(x) = \theta + \sum_{i=1}^n w_i x_i$$

- Hàm kích hoạt là hàm ngưỡng



$$o = f(u) = f(g(x)) = \begin{cases} 1 & g(x) \geq 0 \\ 0 & g(x) < 0 \end{cases}$$



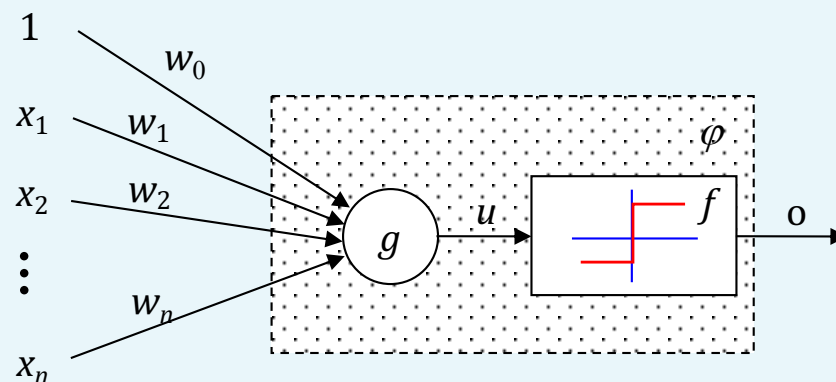
CANTHO UNIVERSITY

Mô hình perceptron

– Perceptron tuyến tính không ngưỡng

- $n + 1$ đầu vào, 1 đầu ra
- Đầu vào giả x_0 luôn có giá trị 1, $w_0 = \theta$
- Hàm mạng tuyến tính

$$u = g(x) = \sum_{i=0}^n w_i x_i$$

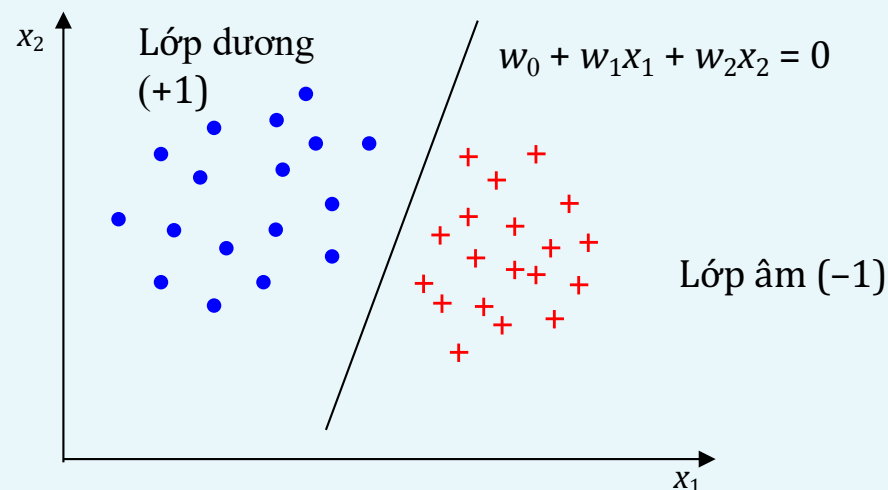




Mô hình perceptron

- Ý nghĩa hình học
 - Ranh giới của hàm kích hoạt là $u = 0$.
 - Phương trình $u = g(x) = 0$ là phương trình của 1 siêu phẳng trong không gian n chiều.

$$u = g(x) = \sum_{i=0}^n w_i x_i = 0$$





Huấn luyện perceptron

- Huấn luyện:
 - Dạy cho perceptron
 - Tìm kiếm n tham số: $w_0, w_1, w_2, \dots, w_n$ sao cho đầu ra của nơ-ron phù hợp với giá trị mong muốn của tất cả dữ liệu học nhất.
- Về mặt hình học:
 - Tìm siêu phẳng tách dữ liệu thành 2 lớp sao cho mỗi lớp về 1 phía của siêu phẳng này.
- Dữ liệu đầu vào:
 - Tập các mẫu huấn luyện
 - Mỗi mẫu huấn luyện gồm:
 - véc-tơ đặc trưng x và nhãn y .
- Tham số:
 - Tốc độ học: η (đọc là eta)



Huấn luyện perceptron

- Trường hợp dữ liệu khả tách tuyến tính
 - Khởi tạo ngẫu nhiên các w
 - Đưa từng mẫu học qua perceptron và quan sát giá trị đầu ra
 - Nếu ra trị đầu ra khác với giá trị mong muốn, cập nhật lại các trọng số theo công thức:

$$w_j = w_j + \eta \cdot (y_i - o_i) \cdot x_{ij}, \forall j = 0..n$$

- Giải thuật: xem giáo trình trang 29



Huấn luyện perceptron

- Trường hợp dữ liệu không khả tách tuyến tính
 - Cố gắng tìm một siêu phẳng “tốt” nhất
 - Tối = lỗi (trên tập học) nhỏ nhất có thể
 - Định nghĩa hàm lỗi theo $E(w)$ theo các trọng số w trên tất cả các phần tử của tập học:

$$E(w) \equiv E(w_0, w_1, \dots, w_n) = \frac{1}{2} \sum_{i=1}^m (y_i - \varphi(x_i))^2$$

- Bài toán huấn luyện trở thành tìm w sao cho $E(w)$ nhỏ nhất



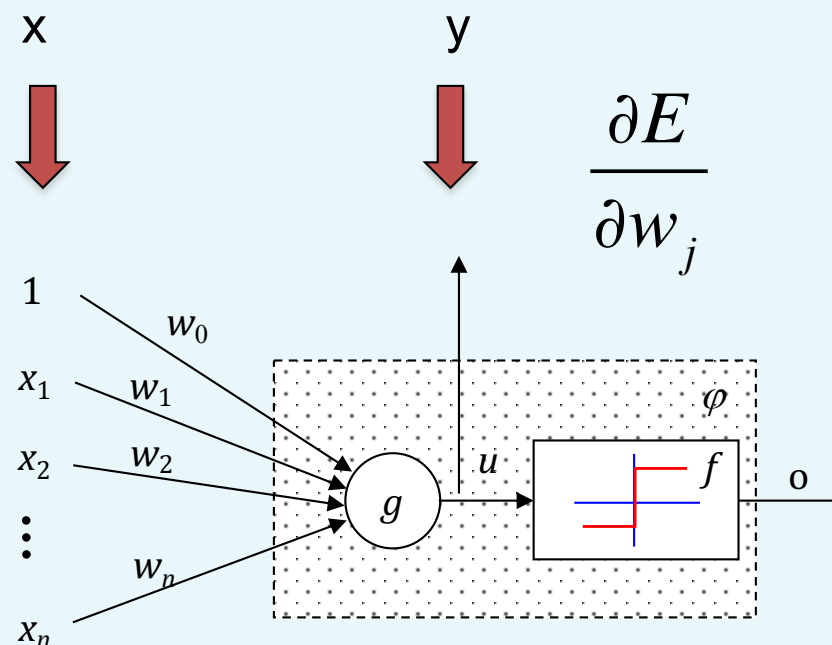
CANTHO UNIVERSITY

Huấn luyện perceptron

- Chú ý:
 - Hàm kích hoạt trong trường hợp này được thay bằng hàm tuyến tính $f(u) = u$ hay:

$$\varphi(x_i) = g(x_i) = \sum_{j=0}^n w_j \cdot x_{ij}$$

- Tương tự như trường hợp khả tách tuyến tính ta tối ưu $E(w)$ bằng phương pháp gradient descent





CANTHO UNIVERSITY

Huấn luyện perceptron

- Giải thuật:
 - Khởi động ngẫu nhiên w
 - Lặp cho đến khi điều kiện dừng thoả mãn:
 - $w = w + \Delta w = w - \eta \Delta E(w)$

- Gradient của $E(w)$:

$$\frac{\partial E}{\partial w_j} = - \sum_{i=1}^m (y_i - g(x_i)) \cdot x_{ij}$$

$$\nabla E(w) = \begin{pmatrix} \frac{\partial E}{\partial w_0} \\ \frac{\partial E}{\partial w_1} \\ \vdots \\ \frac{\partial E}{\partial w_n} \end{pmatrix}$$



CANTHO UNIVERSITY

Huấn luyện perceptron

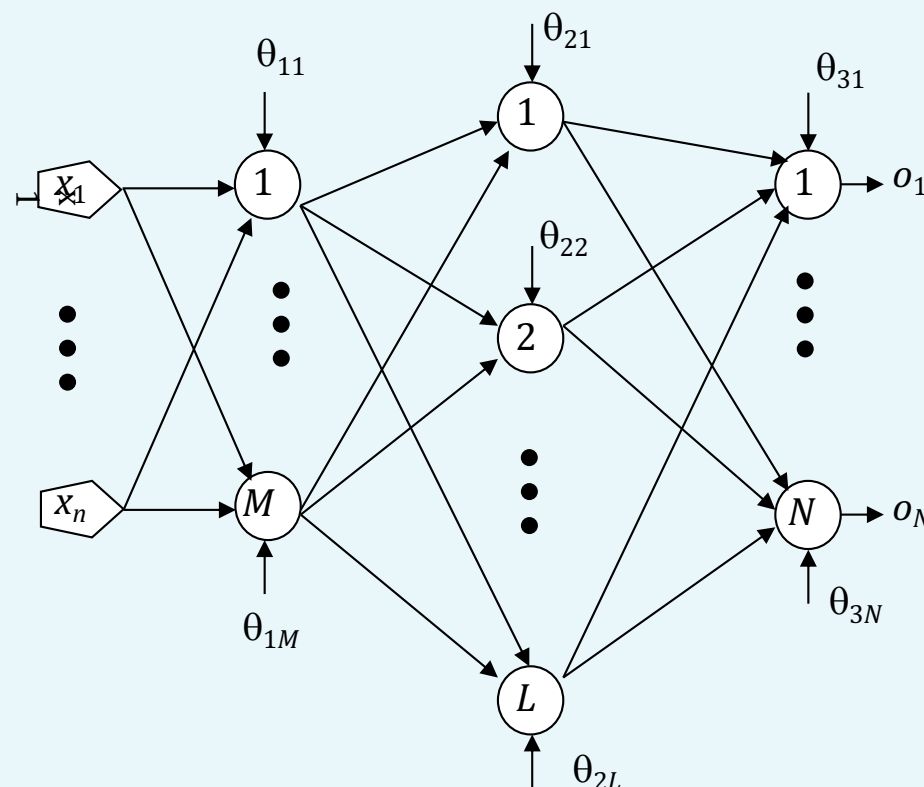
- Luật cập nhật w :
 - Luật gradient chuẩn: cập nhật sau khi xem xét tất cả các phần tử của tập học, cộng dồn Δw
 - Luật Delta: cập nhật sau khi xét mỗi phần tử của tập học
- Giải thuật: trang 36, 37



CANTHO UNIVERSITY

Mạng nơ-ron đa tầng (MLP)

- MLP:
 - Mạng nơ-ron truyền thẳng
 - Nơ-ron: hàm mạng tuyến tính, hàm kích hoạt phi tuyến, liên tục và khả vi (có thể lấy vi phân được)
 - Sigmoid
 - Hyperpolic tangent





CANTHO UNIVERSITY

Mạng nơ-ron đa tầng (MLP)

- Huấn luyện mạng MLP:
 - Giải thuật lan truyền ngược (back propagation)
 - Định nghĩa hàm lỗi: bình phương sai khác giữa đầu ra và đầu ra mong muốn
 - Tính toán lỗi
 - Lan truyền lỗi từ đầu ra ngược trở về đầu vào để cập nhật các trọng số w . Trọng số được cập nhật dựa trên gradient của hàm lỗi.



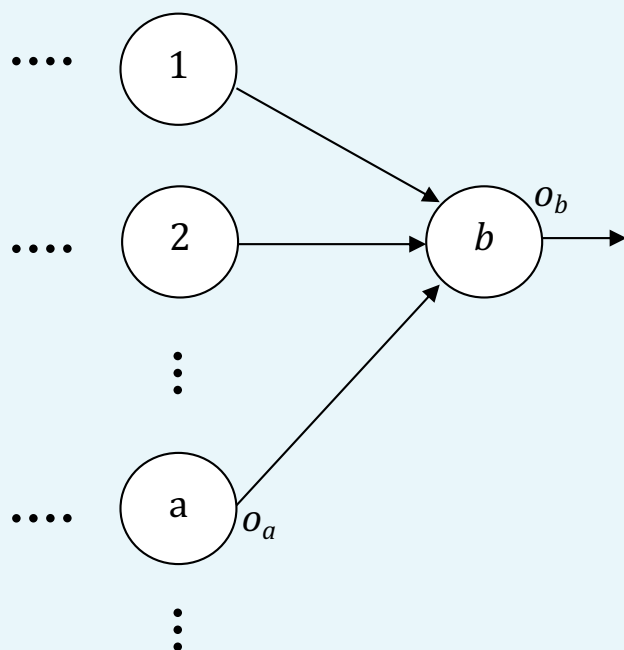
CANTHO UNIVERSITY

Mạng nơ-ron đa tầng (MLP)

- Xét các nơ-ron ở tầng đầu ra:

$$u_b = \sum_a (w_{a \rightarrow b} o_a)$$

$$o_b = f_b(u_b)$$



$$E = \frac{1}{2} (y_b - o_b)^2$$

$$\frac{\partial E}{\partial o_b} = \frac{\partial}{\partial o_b} \left(\frac{1}{2} (y_b - o_b)^2 \right) = -(y_b - o_b)$$



Mạng nơ-ron đa tầng (MLP)

- Xét các nơ-ron ở tầng đầu ra:

$$u_b = \sum_a (w_{a \rightarrow b} o_a) \quad o_b = f(u_b)$$

- Để cập nhật được các trọng số $w_{a \rightarrow b}$ ta cần tính đạo hàm riêng của E theo nó.
- Áp dụng quy tắc chuỗi để tính đạo hàm ta được (xem giáo trình trang 41, công thức 2.17 – 2.20):

$$\frac{\partial E}{\partial w_{a \rightarrow b}} = -(y_b - o_b) \cdot f'_b(u_b) \cdot o_a$$



CANTHO UNIVERSITY

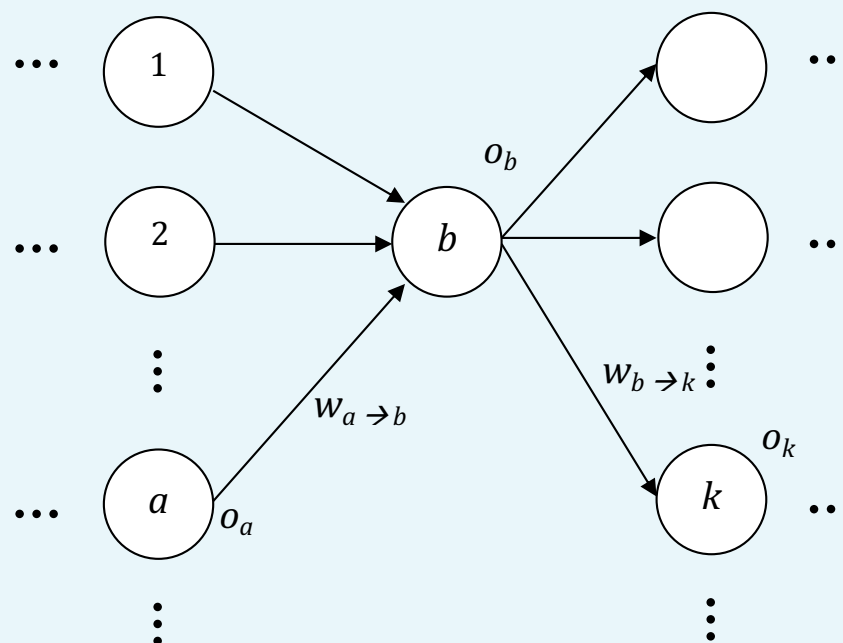
Mạng nơ-ron đa tầng (MLP)

- Xét các nơ-ron ở tầng giữa:

$$\frac{\partial E}{\partial w_{a \rightarrow b}} = \frac{\partial E}{\partial o_b} \cdot f'_b(u_b) \cdot o_a$$

- Trong đó $\frac{\partial E}{\partial o_b}$ được tính bằng:

$$\frac{\partial E}{\partial o_b} = \sum_k \left(\frac{\partial E}{\partial o_k} \cdot f'_k(u_k) \cdot w_{b \rightarrow k} \right)$$





CANTHO UNIVERSITY

Mạng nơ-ron đa tầng (MLP)

- Giải thuật:
 - Khởi tạo ngẫu nhiên w
 - Lặp cho đến khi điều kiện dừng thỏa:
 - Với mỗi mẫu huấn luyện, tính đầu ra o
 - Tính đạo hàm riêng của E theo từng trọng số w
 - Cập nhật trọng số theo đạo hàm riêng này (chú ý dấu trừ vì đi ngược hướng của gradient)
- Bài tập 7 trang 48



CẢM ƠN