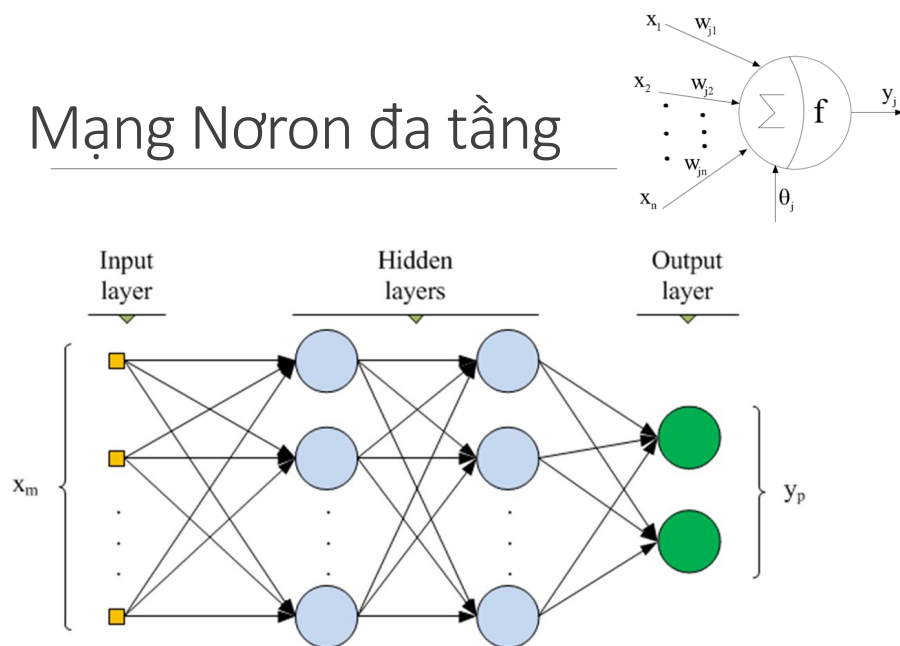


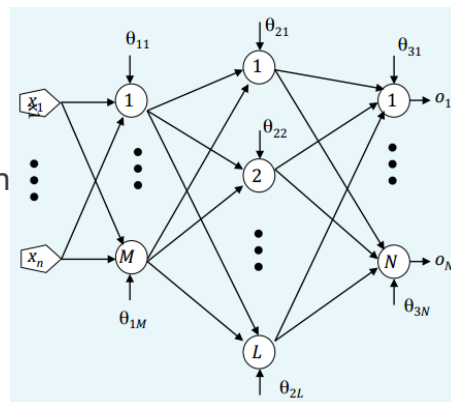
## Mạng Nơ-ron đa tầng



32

## Mạng nơ-ron đa tầng

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



33

## Mạng nơ-ron đa tầng (MLP)

Huấn luyện mạng MLP:

- **Giải thuật lan truyền ngược (back propagation)**

**Yan Le Cun đề xuất năm 1986 - tổng quát hoá luật Delta**

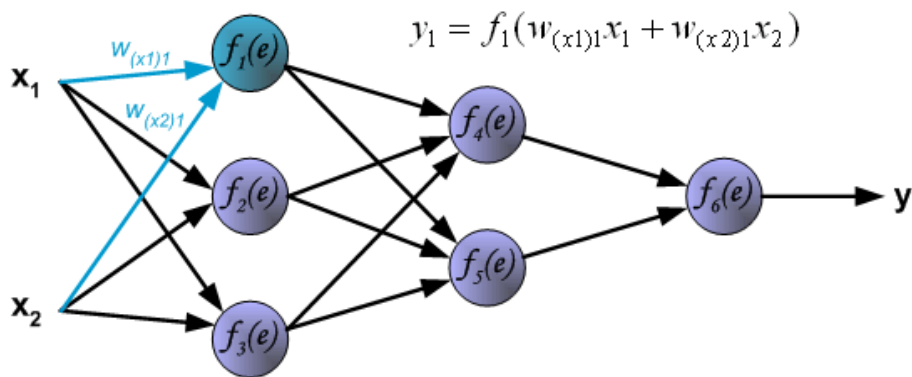
- Định nghĩa hàm lỗi: bình phương sai khác giữa đầu ra và đầu ra mong muốn
- Tính toán lỗi
- Lan truyền lỗi từ đầu ra ngược trở về đầu vào để cập nhật các trọng số  $w$ . Trọng số được cập nhật dựa trên gradient của hàm lỗi.

34

- Giải thuật học lan truyền ngược tìm kiếm một vector các trọng số (weights vector) giúp **cực tiểu hóa lỗi tổng thể** của hệ thống đối với tập học
- Giải thuật BP bao gồm 2 giai đoạn (bước)
  - Giai đoạn **lan truyền tiến tín hiệu (Signal forward)**. Các tín hiệu đầu vào (vector các giá trị đầu vào) được lan truyền tiến từ tầng đầu vào đến tầng đầu ra (đi qua các tầng ẩn)
  - Giai đoạn **lan truyền ngược lỗi (Error backward)**
    - Căn cứ vào giá trị đầu ra mong muốn của vector đầu vào, hệ thống tính toán giá trị lỗi
    - Bắt đầu từ tầng đầu ra, giá trị lỗi được lan truyền ngược qua mạng, từ tầng này qua tầng khác (phía trước), cho đến tầng đầu vào
    - Việc lan truyền ngược lỗi (error back-propagation) được thực hiện thông qua việc tính toán (một cách truy hồi) giá trị gradient cục bộ của mỗi nơ-ron

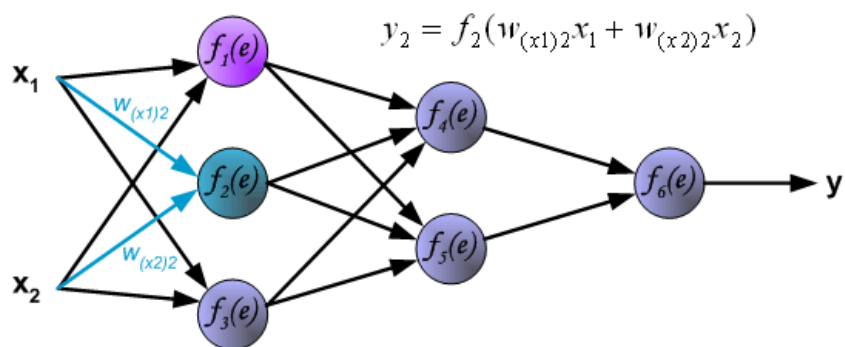
35

## Mạng nơ-ron MLP



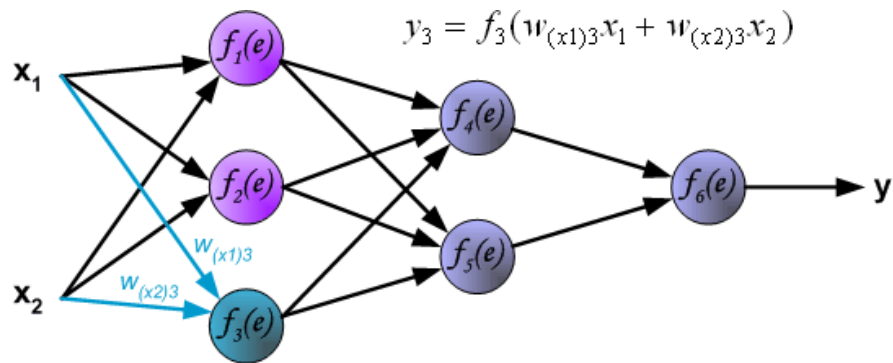
36

## Mạng nơ-ron MLP



37

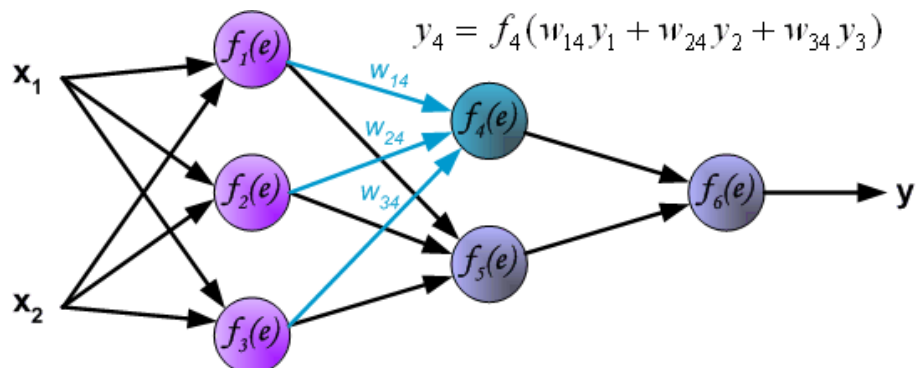
## Mạng nơ-ron MLP



38

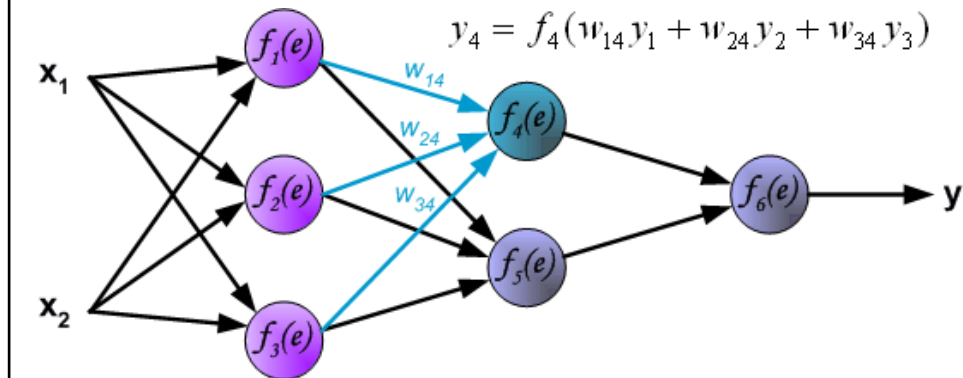
## Mạng nơ-ron MLP

$w_{mn}$  là trọng số nối kết giữa đầu ra của nơron m và đầu vào của nơron n ở tầng kế tiếp



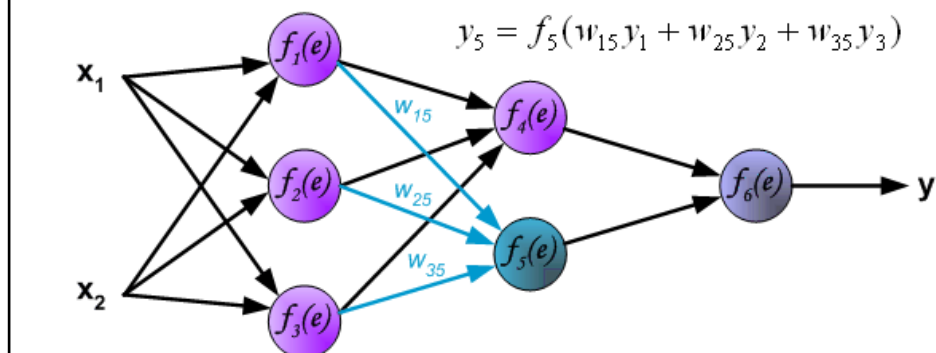
39

## Mạng nơ-ron MLP



40

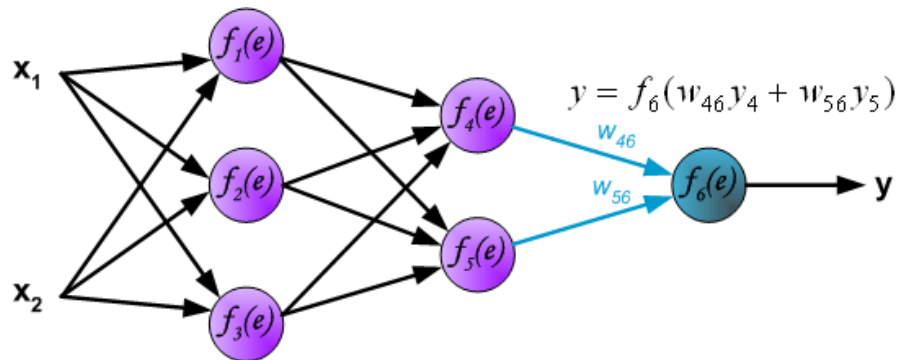
## Mạng nơ-ron MLP



41

## Mạng nơ-ron MLP

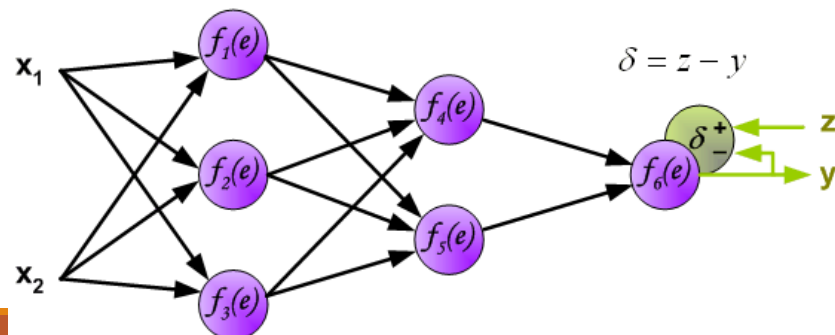
Quá trình lan truyền được thực hiện đến khi gặp đầu ra



42

## Mạng nơ-ron MLP: lan truyền ngược

Tín hiệu đầu ra của mạng  $y$  được so sánh với giá trị đầu ra mong muốn (mục tiêu), được tìm thấy trong tập dữ liệu huấn luyện. Sự khác biệt được gọi là tín hiệu báo lỗi  $d$  của nơ-ron lớp đầu ra

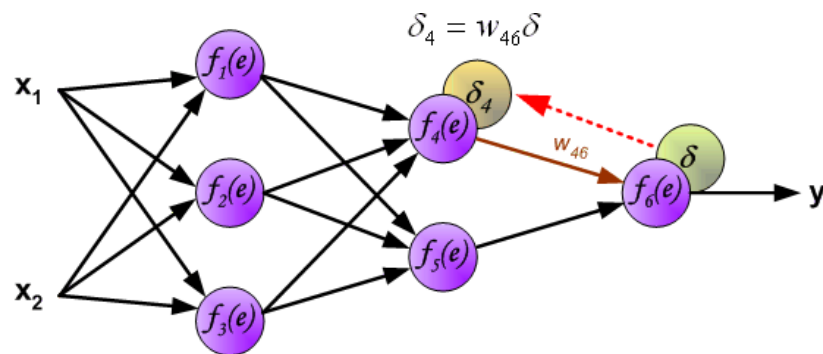


43

## Mạng nơ-ron MLP: lan truyền ngược

Ý tưởng:

truyền tín hiệu báo lỗi d trở lại với tất cả các nơ-ron

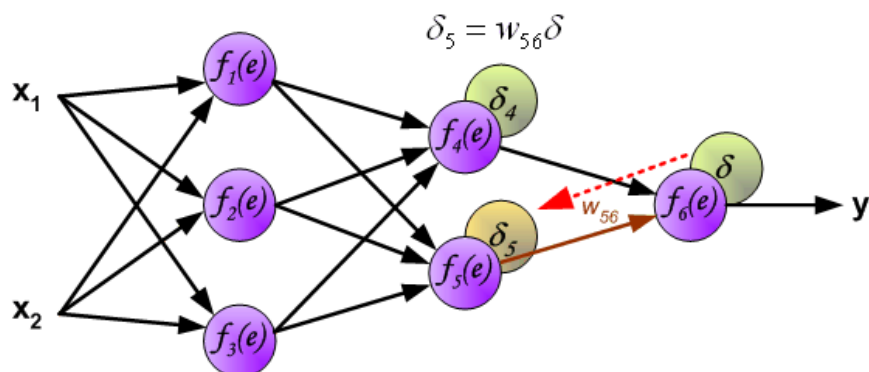


44

## Mạng nơ-ron MLP: lan truyền ngược

Ý tưởng:

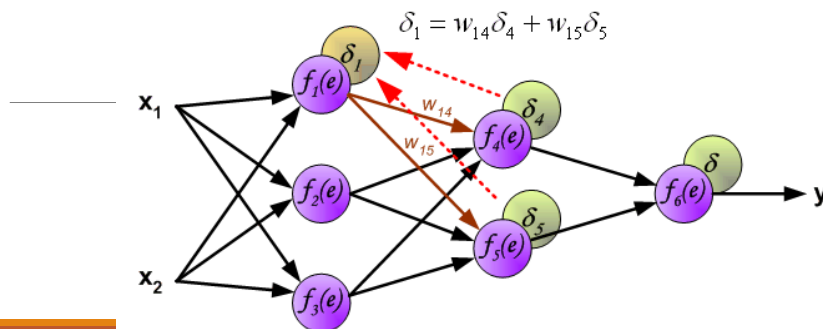
truyền tín hiệu báo lỗi d trở lại với tất cả các nơ-ron



45

## Mạng nơ-ron MLP: lan truyền ngược

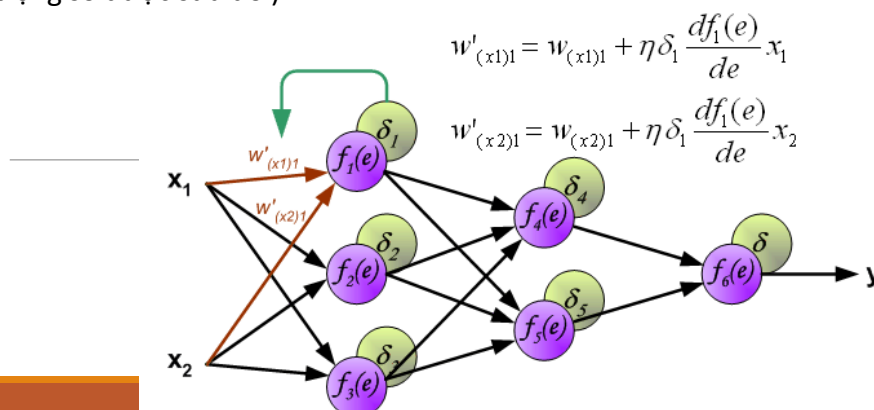
Các hệ số trọng lượng của  $w_{mn}$  được sử dụng để truyền lỗi trở lại. Chỉ có hướng dòng chảy dữ liệu được thay đổi (tín hiệu được truyền từ đầu ra sang đầu vào). Kỹ thuật này được sử dụng cho tất cả các lớp mạng. Nếu các lỗi truyền xuất phát từ vài tế bào thần kinh, chúng được thêm vào. Hình minh họa dưới đây:



46

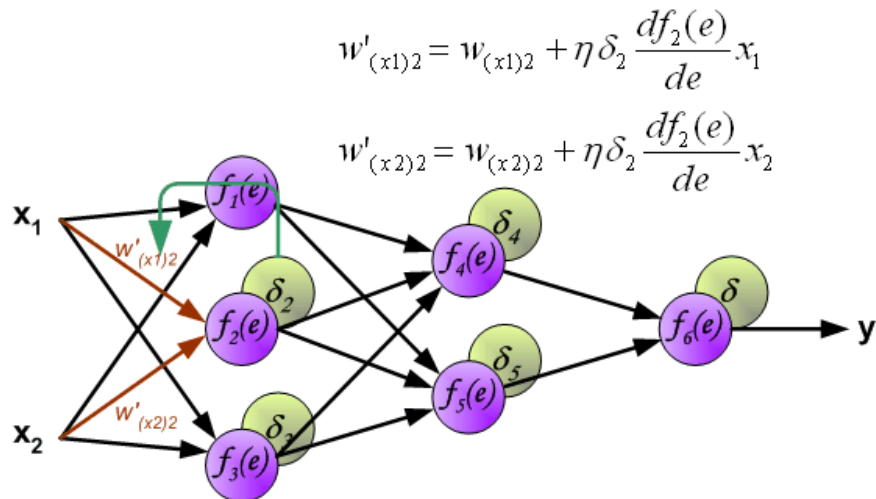
## Mạng nơ-ron MLP: lan truyền ngược

Khi tín hiệu lỗi cho mỗi neuron được tính, các hệ số trọng lượng của mỗi nút đầu vào nơ-ron có thể được sửa đổi. Trong các công thức dưới  $df(e)/de$  biểu thị dẫn xuất của chức năng kích hoạt nơ-ron (trọng số được sửa đổi).



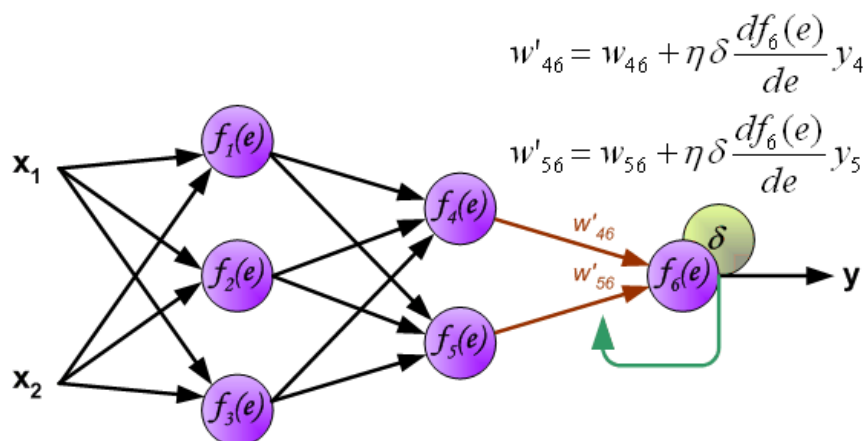


## Mạng nơ-ron MLP: lan truyền ngược



48

## Mạng nơ-ron MLP: lan truyền ngược



49