

Backpropagation

Derivative

- Đạo hàm thể hiện mức độ thay đổi của hàm mất mát khi thay đổi nhỏ ở tham số.
- Định nghĩa: nếu w tăng ϵ , và $J(w)$ tăng $k \times \epsilon$, thì:

$$\frac{dJ(w)}{dw} = k$$

- Nếu đạo hàm nhỏ \rightarrow cập nhật nhỏ.
- Nếu đạo hàm lớn \rightarrow cập nhật lớn (vì ảnh hưởng lớn đến $J(w)$).

Computation graph

- **Computation graph (đồ thị tính toán)** là một khái niệm cốt lõi trong deep learning.
- Nó là cách các framework như **TensorFlow** tự động tính đạo hàm của mạng nơ-ron bằng cách biểu diễn phép tính thành các bước nhỏ, kết nối với nhau bằng các **node** và **edge** (mũi tên).

Backpropagation

- Là quá trình tính **đạo hàm của J theo các tham số** (w, b).
- Diễn ra **từ phải sang trái**, ngược chiều với forward propagation.
- Backprop rất hiệu quả:
 - Nếu có **n node** và **p tham số**, bạn chỉ cần **$n + p$ phép tính** để tìm tất cả đạo hàm, thay vì **$n \times p$** .
 - Với mạng lớn có hàng chục nghìn node và tham số, sự khác biệt này là rất lớn (ví dụ: $10,000 \times 100,000 = 1$ tỷ phép tính).