

Deep Learning

Đào Thành Mạnh

January 18th, 2025

Contents

1	Phần 1: Bộ dữ liệu huấn luyện $\mathcal{D} = \{(x_i, y_i)\}_{i=1}$	3
2	Phần 2: Hàm dự đoán $f(w, b)$ thực hiện ánh xạ $x \rightarrow y$	3
3	Phần 3: Tối ưu hàm mục tiêu $\mathcal{L}(w, b)$ định nghĩa trên \mathcal{D}	3
4	Phần 4: Sử dụng hàm dự đoán $f(w^*, b^*)$ để dự đoán trên tập kiểm tra \mathcal{T}	4
5	PHẦN CÁC CÂU HỎI	4
5.1	Phần 1 (2 điểm) : Bộ dữ liệu huấn luyện $\mathcal{D} = \{(x_i, y_i)\}_{i=1}$	
5.2	Phần 2 (3 điểm) : Hàm dự đoán $f(w, b)$ thực hiện ánh xạ $x \rightarrow y$. .	
5.3	Phần 3 (3 điểm) : Tối ưu hàm mục tiêu $\mathcal{L}(w, b)$ định nghĩa trên \mathcal{D} .	
5.4	Phần 4 (1.5 điểm) : Sử dụng hàm dự đoán $f(w^*, b^*)$ để dự đoán trên tập kiểm tra \mathcal{T}	
6	LỜI GIẢI	5
6.1	Phần 1: Bộ dữ liệu huấn luyện	
6.1.1	Câu a) Số lượng mẫu tập huấn luyện sau khi cắt ảnh	
6.1.2	Câu b) Số lượng mẫu tập huấn luyện sau khi cắt ảnh và tăng cường	
6.1.3	Câu c) Kích thước toàn bộ tập huấn luyện (byte) sau khi cắt ảnh và tăng cường	
6.2	Phần 2: Hàm dự đoán $f(w, b)$	
6.2.1	Câu a) Kích thước của các lớp mạng nơ-ron tích chập $i=1,2,3,4$	
6.2.2	Câu b) Tổng số tham số w và b của tất cả các lớp $i=1,2,...,6$	
6.3	Phần 3: Tối ưu hàm mục tiêu $\mathcal{L}(w, b)$	
6.3.1	Câu a) Tổng chi phí tính toán (flops) của tất cả các lớp $i=1,2,...,6$ cho một lần tính xuôi (forward)	
6.3.2	Câu b) Số lượng các tính toán (flops) trên một bó con (mini-batch) trên một vòng lặp đơn của giải thuật	

- 6.3.3 Câu c) Tổng chi phí tính toán của epochs = 100 vòng lặp của giải thuật, trên toàn bộ dữ liệu \mathcal{D}
- 6.3.4 Câu d) Giá trị đạo hàm lan truyền ngược đầu ra softmax δ_{softmax} của một ảnh mô tả Hồ Gươm
- 6.3.5 Câu e) Giá trị tiếp theo của o_{HT} theo giải thuật lặp
- 6.4 Phần 4: Sử dụng hàm dự đoán $f(w^*, b^*)$ để dự đoán trên tập kiểm tra \mathcal{T}
- 6.4.1 Câu a) Tổng chi phí bộ nhớ (byte) của toàn bộ các lớp $i=0,1,2,\dots,6$ đầu vào và giá trị trung gian các tầng ẩn khi dự đoán
- 6.4.2 Câu b) Tổng chi phí bộ nhớ (byte) của toàn bộ mạng nơ-ron gồm cả chi phí lưu trữ tham số, biểu diễn tầng ẩn và hình ảnh đầu vào khi dự đoán một mẫu
- 6.4.3 Câu c) Tổng chi phí tính toán (flops) dự đoán của mô hình mạng nơ-ron cho ảnh đầu vào kích thước 256×1024 điểm ảnh

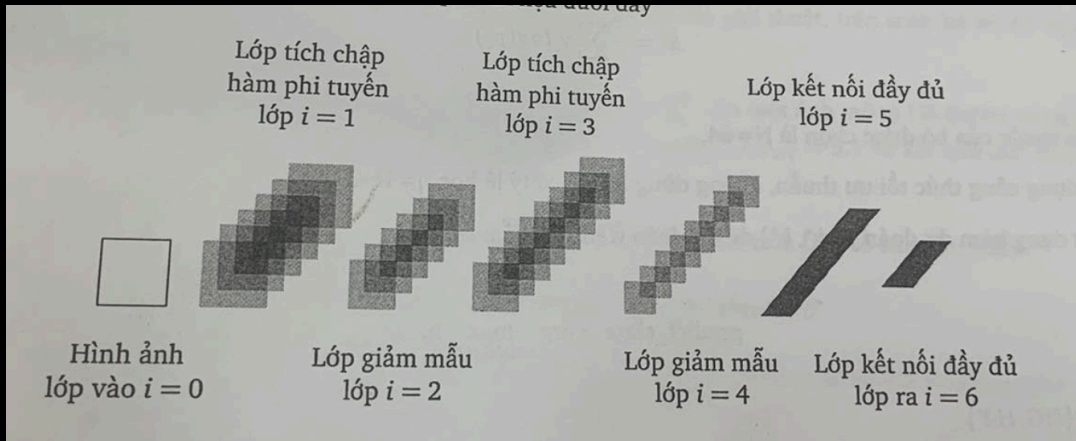
1 Phần 1: Bộ dữ liệu huấn luyện $\mathcal{D} = \{(x_i, y_i)\}_{i=1}$

Bộ dữ liệu 200 ảnh ba lớp mẫu RGB được dùng để xây dựng bộ dữ liệu huấn luyện đầu ra là dự đoán hai nhãn đầu tiên Hồ Gươm (HG), Hồ Tây (HT).

1. Mỗi ảnh được cắt (crop) ngẫu nhiên 256 hình vuông kích thước 128×128 điểm ảnh (pixel).
2. Các hình ảnh cũng được tăng cường (augmentation) bằng cách lật đối xứng trái/phải, trên/dưới, xoay phải 45° , xoay trái 45° .
3. Tiền xử lý trước bằng cách chia các giá trị nguyên không âm RGB cho 255 thành số thực.

2 Phần 2: Hàm dự đoán $f(w, b)$ thực hiện ánh xạ $x \rightarrow y$

Hàm dự đoán sử dụng nơ-ron học sâu gồm 6 lớp minh họa dưới đây:



Hình 1: Mạng nơ-ron gần giống với mạng LeNet-5 của LeCun et al. (1989).

Mạng nơ-ron minh họa trên có các thông số chi tiết như sau

1. Lớp tích chập ($i=1$) phi tuyến hàm truyền ReLU, kích thước nhân $k = (7 \times 7)$, số bộ lọc $f = 16$, bước nhảy $s = 1$, kích thước viền $p = 1$.
2. Lớp giảm mẫu - pooling ($i=2$) sử dụng hàm cực đại, kích thước của số $p = (3 \times 3)$, bước nhảy $s = 2$, kích thước viền $p = 0$.
3. Lớp tích chập ($i=3$) phi tuyến hàm truyền ReLU, kích thước nhân $k = (5 \times 5)$, số bộ lọc $f = 48$, bước nhảy $s = 1$, kích thước viền $p = 0$.
4. Lớp giảm mẫu ($i=4$) sử dụng hàm trung bình, kích thước của số $p = (2 \times 2)$, bước nhảy $s = 2$, kích thước viền $p = 2$.
5. Lớp đầy đủ ($i=5$) có kích thước 4096 nơ-ron dùng hàm truyền tanh.
6. Lớp ra đầy đủ ($i=6$) dùng hàm softmax.

3 Phần 3: Tối ưu hàm mục tiêu $\mathcal{L}(w, b)$ định nghĩa trên \mathcal{D}

$$w^*, b^* = \arg \min \mathcal{L}(w, b)_{\mathcal{D}}$$

Giải thuật tự động ngược hướng vectơ tiếp tuyến (Stochastic Gradient Descent):

1. $t \leftarrow 0$
2. $w_t, b_t \leftarrow$ initial values
3. do

$$4. w_{t+1} \leftarrow w_t - \eta \frac{\partial \mathcal{L}(w, b)}{\partial w}$$

$$5. b_{t+1} \leftarrow b_t - \eta \frac{\partial \mathcal{L}(w, b)}{\partial b}$$

$$6. t \leftarrow t + 1$$

7. **while** (not converge)

- Biết rằng với đầu ra o_c với c được gán hai nhãn đầu tiên HG, HT thì công thức hàm softmax sẽ là:

$$p_c = \text{softmax}(o_c) = \frac{e^{o_c}}{\sum e^{o_c}}$$

- Hàm mất mát (cross-entropy) có công thức như sau:

$$\mathcal{L} = - \sum_c y_c \log(p_c)$$

- Kích thước của bó được chọn là $N = 64$. - Sử dụng công thức tối ưu thuần, không dùng Adam có tỷ lệ học $\eta = 1e - 2$.

4 Phần 4: Sử dụng hàm dự đoán $f(w^*, b^*)$ để dự đoán trên tập kiểm tra \mathcal{T}

$$\text{predict_class} = \arg \max_{c \in [HG, HT]} p_c$$

trong đó $c \in [HG, HT]$.

5 PHẦN CÁC CÂU HỎI

Từ minh họa và mô tả chi tiết trên hãy thực hiện trả lời các câu hỏi. Điền chữ số vào các câu hỏi trên đường liên kết FORMS tại trang 4. Chú ý các phần chỉ dùng để giúp các bạn điền từng số sau đó tính tổng để điền kết quả cuối cùng trên FORMS.

5.1 Phần 1 (2 điểm) : Bộ dữ liệu huấn luyện $\mathcal{D} = \{(x_i, y_i)\}_{i=1}$

- Hãy cho biết số lượng mẫu tập huấn luyện sau khi cắt ảnh?
- Hãy cho biết số lượng mẫu tập huấn luyện sau khi cắt ảnh và tăng cường?
- Hãy cho biết kích thước toàn bộ tập huấn luyện (byte) sau khi cắt ảnh và tăng cường, biết rằng giá trị RGB mỗi điểm ảnh biểu diễn số nguyên không âm $[0, 255]$?

5.2 Phần 2 (3 điểm) : Hàm dự đoán $f(w, b)$ thực hiện ánh xạ $x \rightarrow y$

- Hãy cho biết kích thước của các lớp mạng nơ-ron tích chập $i = 1, 2, \dots, 4$ trên sử dụng quy ước biểu diễn ma trận ba chiều (H, W, D) cao, rộng, sâu với lớp tích chập?
- $i = 1$ - $i = 2$ - $i = 3$ - $i = 4$
- Hãy cho biết tổng số tham số w và b của tất cả các lớp $i = 1, 2, \dots, 6$?

5.3 Phần 3 (3 điểm) : Tối ưu hàm mục tiêu $\mathcal{L}(w, b)$ định nghĩa trên \mathcal{D}

- Hãy cho biết tổng chi phí tính toán o của tất cả các lớp $i = 1, 2, \dots, 6$ đơn vị (flops) cho một lần tính xuôi - forward?
- Giả thiết chi phí lan truyền ngược - backpropagation - là bằng tính xuôi. Hãy cho biết số lượng các tính toán o trên một bó con (mini-batch) trên một vòng lặp đơn của giải thuật?
- Hãy cho biết tổng chi phí tính toán của epochs = 100 vòng lặp của giải thuật, trên toàn bộ bộ dữ liệu \mathcal{D} đã xử lý ở Phần 1?
- Hãy cho biết giá trị đạo hàm lan truyền ngược đầu ra softmax δ_{softmax} của một ảnh mô tả Hồ Gươm trong tập dữ liệu huấn luyện \mathcal{D} là bao nhiêu? Biết rằng giá trị của đầu ra $o_{HG} = 3.2$, $o_{HT} = 5.1$ và kết quả làm tròn hai số sau dấu phẩy.
- Hãy cho biết giá trị tiếp theo của o_{HT} theo giải thuật lặp? Kết quả làm tròn bốn số sau dấu phẩy.

5.4 Phần 4 (1.5 điểm) : Sử dụng hàm dự đoán $f(w^*, b^*)$ để dự đoán trên tập kiểm tra \mathcal{T}

a) Hãy cho biết tổng chi phí bộ nhớ (byte) của toàn bộ các lớp $i = 0, 1, 2, \dots, 6$ đầu vào và giá trị trung gian các tầng ẩn khi dự đoán, biết rằng biểu diễn số thực kích thước tiêu chuẩn (4 bytes)?

b) Hãy cho biết tổng chi phí bộ nhớ (byte) của toàn bộ mạng nơ-ron trên gồm cả chi phí lưu trữ tham số, biểu diễn tầng ẩn và hình ảnh đầu vào khi dự đoán một mẫu?

c) Để dự đoán ảnh đầu vào kích thước 256×1024 điểm ảnh. Hãy cho biết tổng chi phí tính toán (flops) dự đoán của mô hình mạng nơ-ron là bao nhiêu?

6 LỜI GIẢI

6.1 Phần 1: Bộ dữ liệu huấn luyện

6.1.1 Câu a) Số lượng mẫu tập huấn luyện sau khi cắt ảnh

Mỗi ảnh được cắt ngẫu nhiên 256 hình vuông kích thước 128×128 pixel. Tổng số ảnh ban đầu là 200 ảnh. Số lượng mẫu sau khi cắt ảnh:

$$200 \text{ ảnh} \times 256 \text{ mẫu/ảnh} = 51,200 \text{ mẫu}$$

6.1.2 Câu b) Số lượng mẫu tập huấn luyện sau khi cắt ảnh và tăng cường

Mỗi ảnh sau khi cắt được tăng cường bằng cách lật đối xứng trái/phải, trên/dưới, xoay phải 45° , xoay trái 45° . Tổng số phép biến đổi tăng cường: 6 phép. Số lượng mẫu sau khi tăng cường:

$$51,200 \text{ mẫu} \times 6 = 307,200 \text{ mẫu}$$

6.1.3 Câu c) Kích thước toàn bộ tập huấn luyện (byte) sau khi cắt ảnh và tăng cường

Mỗi ảnh cắt có kích thước 128×128 pixel, mỗi pixel có 3 kênh RGB. Mỗi pixel được biểu diễn bằng 3 byte. Kích thước của một ảnh cắt:

$$128 \times 128 \times 3 = 49,152 \text{ byte}$$

Kích thước toàn bộ tập huấn luyện sau khi tăng cường:

$$307,200 \text{ mẫu} \times 49,152 \text{ byte/mẫu} = 15,099,494,400 \text{ byte}$$

6.2 Phần 2: Hàm dự đoán $f(w, b)$

6.2.1 Câu a) Kích thước của các lớp mạng nơ-ron tích chập $i=1,2,3,4$

- Lớp $i=1$ (Conv1): $124 \times 124 \times 16$
- Lớp $i=2$ (Pool1): $61 \times 61 \times 16$
- Lớp $i=3$ (Conv2): $57 \times 57 \times 48$
- Lớp $i=4$ (Pool2): $29 \times 29 \times 48$

6.2.2 Câu b) Tổng số tham số w và b của tất cả các lớp $i=1,2,\dots,6$

Tổng số tham số:

$$2,368 + 19,248 + 165,195,264 + 8,194 = 165,225,074 \text{ tham số}$$

6.3 Phần 3: Tối ưu hàm mục tiêu $\mathcal{L}(w, b)$

6.3.1 Câu a) Tổng chi phí tính toán (flops) của tất cả các lớp $i=1,2,\dots,6$ cho một lần tính xuôi (forward)

Tổng chi phí tính toán:

$$36,019,200 + 536,544 + 62,438,400 + 161,472 + 165,191,168 + 8,192 = 264,355,976 \text{ flops}$$

6.3.2 Câu b) Số lượng các tính toán (flops) trên một bó con (mini-batch) trên một vòng lặp đơn của giải thuật

Chi phí tính toán trên một bó con:

$$264,355,976 \text{ flops} \times 64 = 16,918,782,464 \text{ flops}$$

6.3.3 Câu c) Tổng chi phí tính toán của epochs = 100 vòng lặp của giải thuật, trên toàn bộ bộ dữ liệu \mathcal{D}

Chi phí tính toán trên 100 epochs:

$$81,210,155,827,200 \text{ flops} \times 100 = 8,121,015,582,720,000 \text{ flops}$$

6.3.4 Câu d) Giá trị đạo hàm lan truyền ngược đầu ra softmax δ_{softmax} của một ảnh mô tả Hồ Gươm

Đạo hàm lan truyền ngược:

$$\delta_{\text{softmax}} = p_{HG} - y_{HG} = 0.1301 - 1 = -0.8699$$

6.3.5 Câu e) Giá trị tiếp theo của o_{HT} theo giải thuật lặp

Cập nhật giá trị:

$$o_{HT}^{\text{new}} = o_{HT} - \eta \times \delta_{\text{softmax}} = 5.1 - 0.01 \times (-0.8699) = 5.1 + 0.0087 = 5.1087$$

6.4 Phần 4: Sử dụng hàm dự đoán $f(w^*, b^*)$ để dự đoán trên tập kiểm tra \mathcal{T}

6.4.1 Câu a) Tổng chi phí bộ nhớ (byte) của toàn bộ các lớp $i=0,1,2,\dots,6$ đầu vào và giá trị trung gian các tầng ẩn khi dự đoán

Tổng chi phí bộ nhớ:

$$49,152 + 246,016 + 59,536 + 155,952 + 40,368 + 4,096 + 2 = 555,122 \text{ byte}$$

6.4.2 Câu b) Tổng chi phí bộ nhớ (byte) của toàn bộ mạng nơ-ron gồm cả chi phí lưu trữ tham số, biểu diễn tầng ẩn và hình ảnh đầu vào khi dự đoán một mẫu

Tổng chi phí bộ nhớ:

$$660,900,296 + 555,122 = 661,455,418 \text{ byte}$$

6.4.3 Câu c) Tổng chi phí tính toán (flops) dự đoán của mô hình mạng nơ-ron cho ảnh đầu vào kích thước 256×1024 điểm ảnh

Tổng chi phí tính toán:

$$264,355,976 \text{ flops}$$