

# Bài tập quá trình 1 - NLP

Minh Nhật - 22521060

Ngày 6 tháng 10 năm 2024

## BÀI TẬP 1

Chứng minh rằng tổng xác suất của tất cả các chuỗi có thể từ tập từ vựng  $\mathcal{V}$  trong hai trường hợp sau:

- Không có từ kết thúc  $\langle /s \rangle$ : Mô hình ngôn ngữ có thể sinh chuỗi dài vô hạn.

$$\sum_{n=1}^{\infty} \sum_{x_{1:n}} P(x_{1:n}) = \infty \quad (1)$$

- Có từ kết thúc  $\langle /s \rangle$ : Mô hình ngôn ngữ phải dừng lại việc sinh chuỗi bằng từ kết thúc câu  $\langle /s \rangle$ .

$$\sum_{n=1}^{\infty} \sum_{x_{1:n}} P(x_{1:n} \langle /s \rangle) = 1 \quad (2)$$

## BÀI TẬP 1A.

Ta có:

$$\begin{aligned} \sum_{n=1}^1 \sum_{x_{1:n}} P(x_{1:n}) &= 1, \\ \sum_{n=2}^2 \sum_{x_{1:n}} P(x_{1:n}) &= 1, \\ \Rightarrow \sum_{n=1}^2 \sum_{x_{1:n}} P(x_{1:n}) &= 1 + 1 = 2, \end{aligned}$$

Tương tự, ta được:

$$\sum_{n=1}^3 \sum_{x_{1:n}} P(x_{1:n}) = 1 + 1 + 1 = 3,$$

.....

Giả sử:

---


$$\sum_{n=1}^k \sum_{x_{1:n}} P(x_{1:n}) = k,$$

Với một số  $k \geq 1$ , ta cần chứng minh:

$$\sum_{n=1}^{k+1} \sum_{x_{1:n}} P(x_{1:n}) = k + 1.$$

Ta có:

$$\begin{aligned} \sum_{n=1}^{k+1} \sum_{x_{1:n}} P(x_{1:n}) &= \sum_{n=1}^k \sum_{x_{1:n}} P(x_{1:n}) + \sum_{n=k+1}^{k+1} \sum_{x_{1:n}} P(x_{1:n}) \\ &\Rightarrow \sum_{n=1}^{k+1} \sum_{x_{1:n}} P(x_{1:n}) = k + 1 \text{ (ĐPCM)} \end{aligned}$$

Theo quy nạp, ta được:

$$\sum_{n=1}^{\infty} \sum_{x_{1:n}} P(x_{1:n}) = \infty.$$

## BÀI TẬP 1B.

Ta có:

$$\begin{aligned} \sum_{n=1}^{\infty} \sum_{x_{1:n}} P(x_{1:n}, </s>) &= 1 \\ \Leftrightarrow \sum_{n=1}^{\infty} P(</s>) \sum_{x_{1:n}} P(x_{1:n}) &= 1 \\ \text{hoặc} \\ \Leftrightarrow \sum_{n=1}^{\infty} P(</s> | x_{1:n}) \sum_{x_{1:n}} P(x_{1:n}) &= 1 \end{aligned}$$

(Trong trường hợp  $P(</s>)$  và  $P(x_{1:n})$  không độc lập)

Xét

$$\begin{aligned} &\sum_{x_{1:n}} P(x_{1:n}) \\ &= \sum_{x_{1:n}} \prod_i P(x_i) \\ &= \sum_{x_1} \sum_{x_2} \dots \sum_{x_n} \prod_i P(x_i) \\ &= \sum_{x_1} \sum_{x_2} \dots \sum_{x_n} P(x_1) P(x_2) \dots P(x_n) = 1 \quad (3) \end{aligned}$$

Mà ta lại có:

$$\begin{aligned} \sum_{n=1}^{\infty} P(</s>) &= 1 \quad (4) \\ \sum_{n=1}^{\infty} P(</s> | x_{1:n}) &= 1 \end{aligned}$$

---

(Trong trường hợp  $P(</s>)$  và  $P(x_{1:n})$  không độc lập)

Từ (3) và (4):

$$\Rightarrow \sum_{n=1}^{\infty} \sum_{x_{1:n}} P(x_{1:n} </s>) = 1$$

## BÀI TẬP 2

Cho tập dữ liệu gồm nhiều văn bản thuộc các lớp  $C = \{c_1, c_2, \dots, c_k\}$  và mỗi văn bản chứa các từ từ tập từ vựng  $\mathcal{V}$ . Hãy sử dụng phương pháp **MLE** để tính:

- **Xác suất tiên nghiệm của lớp  $c_j$ :**

$$\hat{P}(c_j) = \frac{\text{count}(c_j)}{N_{doc}}$$

Trong đó:

- $\text{count}(c_j)$ : số văn bản thuộc lớp  $c_j$ .
- $N_{doc}$ : tổng số văn bản.
- **Xác suất có điều kiện của từ  $w_i$  trong lớp  $c_j$ :**

$$\hat{P}(w_i|c_j) = \frac{\text{count}(w_i, c_j)}{\sum_{w \in \mathcal{V}} \text{count}(w, c_j)}$$

Trong đó:

- $\text{count}(w_i, c_j)$ : số lần từ  $w_i$  xuất hiện trong lớp  $c_j$ .
- $\sum_{w \in \mathcal{V}} \text{count}(w, c_j)$ : tổng số lần xuất hiện của tất cả các từ trong lớp  $c_j$ .

## BÀI TẬP 2A

- Đặt  $\theta_j = P(c_j)$  là xác suất của lớp  $c_j$  mà chúng ta muốn ước lượng.

### Likelihood Function

Hàm Likelihood cho số lượng tài liệu trong  $k$  lớp theo phân phối đa thức:

$$L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{j=1}^k \theta_j^{\text{count}(c_j)}$$

---

## Hàm Log-Likelihood

Lấy logarit của hàm Likelihood, ta được hàm Log-likelihood:

$$\log L(\theta_1, \theta_2, \dots, \theta_k) = \sum_{j=1}^k \text{count}(c_j) \log \theta_j$$

## Ràng buộc xác suất

Vì  $\theta_j$  đại diện cho xác suất, ta có ràng buộc:

$$\sum_{j=1}^k \theta_j = 1$$

## Sử dụng Lagrange Multipliers

Chúng ta sử dụng một hệ số Lagrange  $\lambda$  để đưa ràng buộc vào bài toán cực đại hóa. Hàm Lagrangian là:

$$\mathcal{L} = \sum_{j=1}^k \text{count}(c_j) \log \theta_j + \lambda \left( 1 - \sum_{j=1}^k \theta_j \right)$$

## Cực đại hóa Log-likelihood

Để cực đại hóa  $\mathcal{L}$ , ta lấy đạo hàm riêng theo từng  $\theta_j$  và đặt bằng 0:

$$\frac{\partial \mathcal{L}}{\partial \theta_j} = \frac{\text{count}(c_j)}{\theta_j} - \lambda = 0$$

Giải tìm  $\theta_j$ :

$$\theta_j = \frac{\text{count}(c_j)}{\lambda}$$

## Giải cho $\lambda$

Sử dụng ràng buộc  $\sum_{j=1}^k \theta_j = 1$ , ta có thể giải cho  $\lambda$ :

$$\sum_{j=1}^k \frac{\text{count}(c_j)}{\lambda} = 1$$
$$\frac{1}{\lambda} \sum_{j=1}^k \text{count}(c_j) = 1$$

Vì  $\sum_{j=1}^k \text{count}(c_j) = N_{doc}$ , ta có:

---


$$\frac{N_{doc}}{\lambda} = 1 \quad \Rightarrow \quad \lambda = N_{doc}$$

**Ước lượng cuối cùng cho  $\theta_j$**

Thay  $\lambda = N_{doc}$  vào biểu thức của  $\theta_j$ :

$$\theta_j = \frac{\text{count}(c_j)}{N_{doc}}$$

**Kết luận**

Vì vậy, Maximum Likelihood Estimate (MLE) cho xác suất của lớp  $c_j$  là:

$$\hat{P}(c_j) = \frac{\text{count}(c_j)}{N_{doc}}$$

## BÀI TẬP 2B

Chúng ta muốn ước lượng xác suất có điều kiện  $P(w_i | c_j)$ , xác suất của từ  $w_i$  khi biết lớp  $c_j$ .

### Likelihood Function

Hàm Likelihood là:

$$L(\theta_{w_i, c_j}) = \prod_{w \in V} \theta_{w, c_j}^{\text{count}(w, c_j)}$$

### Hàm Log-Likelihood

Lấy log của hàm Likelihood, ta được hàm Log-likelihood:

$$\log L(\theta_{w_i, c_j}) = \sum_{w \in V} \text{count}(w, c_j) \log \theta_{w, c_j}$$

### Cực đại hóa Log-Likelihood

Để cực đại hóa hàm Log-likelihood, ta áp dụng nguyên tắc MLE dưới ràng buộc:

$$\sum_{w \in V} \theta_{w, c_j} = 1$$

Sử dụng Lagrange multipliers, ta định nghĩa hàm Lagrangian:

---


$$\mathcal{L} = \sum_{w \in V} \text{count}(w, c_j) \log \theta_{w, c_j} + \lambda \left( 1 - \sum_{w \in V} \theta_{w, c_j} \right)$$

Lấy đạo hàm riêng theo  $\theta_{w_i, c_j}$  và đặt bằng 0:

$$\frac{\partial \mathcal{L}}{\partial \theta_{w_i, c_j}} = \frac{\text{count}(w_i, c_j)}{\theta_{w_i, c_j}} - \lambda = 0$$

Giải ra  $\theta_{w_i, c_j}$ :

$$\theta_{w_i, c_j} = \frac{\text{count}(w_i, c_j)}{\lambda}$$

### Giải tìm $\lambda$

Sử dụng ràng buộc  $\sum_{w \in V} \theta_{w, c_j} = 1$ :

$$\begin{aligned} \sum_{w \in V} \frac{\text{count}(w, c_j)}{\lambda} &= 1 \\ \frac{1}{\lambda} \sum_{w \in V} \text{count}(w, c_j) &= 1 \end{aligned}$$

Do đó,  $\lambda = \sum_{w \in V} \text{count}(w, c_j)$ .

### Ước lượng cuối cùng cho $\theta_j$

Thay  $\lambda = \sum_{w \in V} \text{count}(w, c_j)$  vào biểu thức của  $\theta_j$ :

$$\theta_j = \frac{\text{count}(c_j)}{\sum_{w \in V} \text{count}(w, c_j)}$$

### Kết luận

Vì vậy, Maximum Likelihood Estimate (MLE) cho xác suất có điều kiện của lớp  $c_j$  là:

$$\hat{P}(c_j) = \frac{\text{count}(c_j)}{\sum_{w \in V} \text{count}(w, c_j)}$$