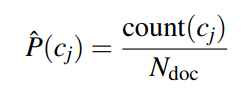
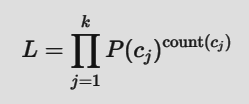
**1. Xác suất tiên nghiệm của lớp Cj:**

Chứng minh công thức:

**Giả sử:**

**MLE (Phương pháp cực đại hợp lý):**

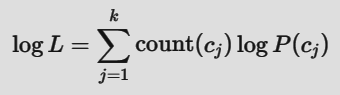
* MLE tìm giá trị tham số xác suất P(cj) sao cho hàm hợp lí (likelihood) là lớn nhất.
* Xác suất để một văn bản thuộc lớp cj là P(cj).
* Vì Ndoc​ văn bản độc lập với nhau, xác suất hợp lý cho toàn bộ tập dữ liệu là:



trong đó count(cj) là số văn bản thuộc lớp cj​.

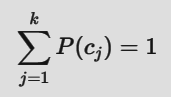
**Hàm log-likelihood:**

Để đơn giản hóa phép tính, ta lấy log của hàm likelihood:

****

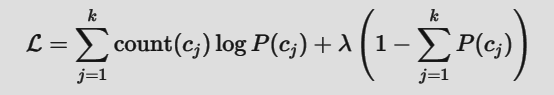
**Ràng buộc xác suất:**

Vì tổng xác suất của tất cả các lớp bằng 1, ta có:

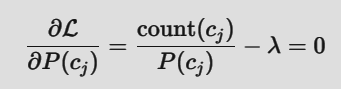
****

**Giải bằng phương pháp Lagrange:**

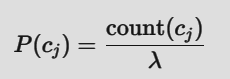
Ta sử dụng phương pháp nhân tử Lagrange để cực đại hóa log L với ràng buộc trên:



Lấy đạo hàm của L theo P(cj):

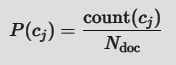


Suy ra:



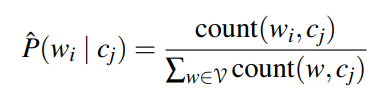
Thay vào ràng buộc :

Do đó:



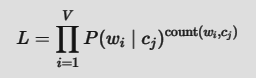
**2. Xác suất có điều kiện của từ wiw\_iwi​ trong lớp cjc\_jcj​:**

Ta cần chứng minh công thức:



**MLE (Phương pháp cực đại hợp lý):**

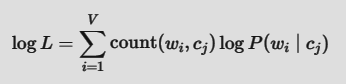
* Xác suất có điều kiện P(wi∣cj) là xác suất để từ wi xuất hiện, với điều kiện văn bản thuộc lớp cj.
* Với giả định các từ trong văn bản độc lập với nhau, xác suất hợp lý của toàn bộ từ trong lớp cj​ là:



trong đó V là số từ trong tập từ vựng V.

**Hàm log-likelihood:**

Lấy log của L:



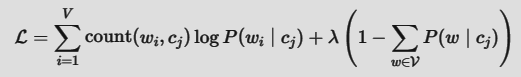
**Ràng buộc xác suất:**

Tổng xác suất của tất cả các từ trong lớp cjc\_jcj​ bằng 1:

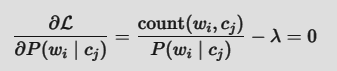


**Giải bằng phương pháp Lagrange:**

Ta cực đại hóa log L với ràng buộc trên:



Lấy đạo hàm của L theo P(wi∣cj):



Suy ra:



Thay vào ràng buộc :



Do đó:

