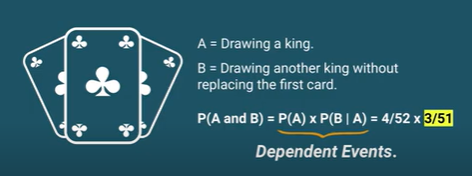
* Xác suất : xác suất chỉ là 1 con số từ 0 đến 1 biểu thị khả năng điều gì đó sẽ xảy ra hoặc không. Đôi khi con số này dựa trên những quan sát lặp đi lặp lại trong 1 thời gian dài hoặc nó có thể chỉ thể hiện niềm tin vào điều gì đó
* Điều gì sẽ xảy ra nếu tung 1 đồng xu thì xác suất nhận được mặt ngửa là ½ và xác suất nhận được mặt sấp cũng là ½ .Nếu bạn được mặt ngửa trong lần đầu tiên thì điều đó không ảnh hưởng đến việc bạn được mặt ngửa hay mặt sấp tsrong lần tung tiếp theo. Vì vậy, nếu chúng ta muốn biết xác suất có được 2 mặt ngửa liên tiếp, chúng ta sẽ nhân xác suất có mặt ngửa ở lân tung đầu tiên với xác suất được mặt ngửa ở lần tung thứ 2 bằng ¼. Giả sử A là sự kiện mặt ngửa trong lần tung thứ nhất, B là sự kiện mặt ngửa trong lần tung thứ 2. Muốn biết xác suất để có 2 mặt ngửa liên tiếp, ta tính theo công thức xác suất của A và B bằng xác suất của A nhân với  xác suất của B khi A đã xảy ra. Trong trường hợp này vì các sự kiện độc lập nên bằng xác suất của A nhân xác suất của B

P(A.B) = P(A) = P(A)P(B) = ¼

* Nhưng ở trong đời thực, mọi thứ không dễ dàng và rõ ràng như vậy về xác suất của các sự kiện xảy ra xung quanh chúng ta thường phụ thuộc vào kết quả của các sự kiện khác. VD xác suất chuyến bay khởi hành đúng giờ phụ thuộc vào thời tiết, nếu có nhiều khả năng xảy ra điều kiện thời tiết bất lợi thì khả năng chuyến bay bị trễ. Tính VD có 1 bộ bài và muốn biết xác suất để rút được 2 quân vua liên tiếp mà không hoàn trả. Gọi A là biến cố rút lần đầu tiên ra quân bài vua và B là biến cố rút được 1 quân vua khác. Ta có xác suất của A và B bằng xác suất của A nhân với xác suất của B khi A đã xảy ra. Do rút 2 lá liên tiếp và không hoàn trả nên đây là xác suất không độc lập



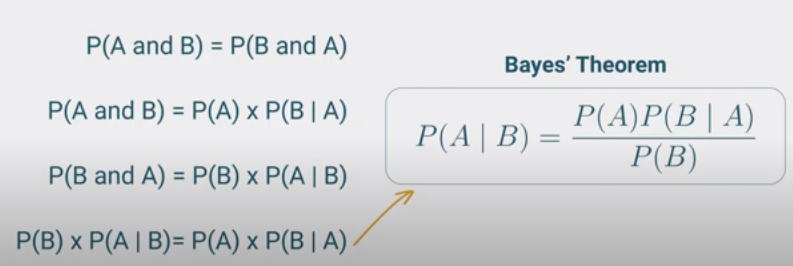
P(A.B) = P(A) = 4/52 3/51

* Vậy chúng ta có công thức xác suất có điều kiện là P(A.B) = P(A)

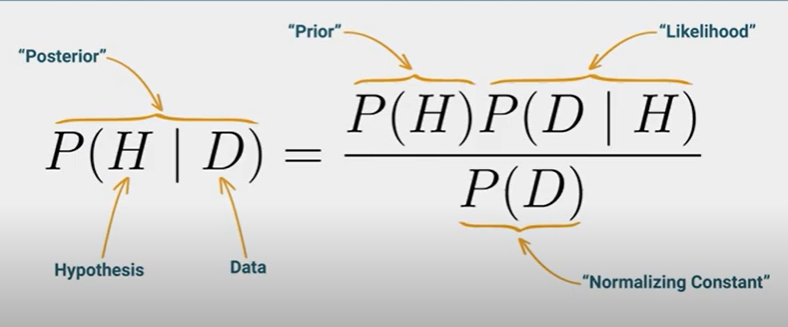
Theo tính chất giao hoán thì P(A.B) = P(B.A) với P(A.B) = P(A)và P(B.A) = P(B)

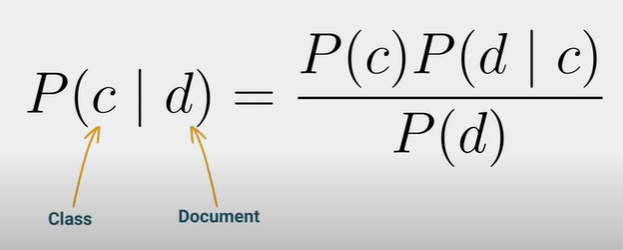
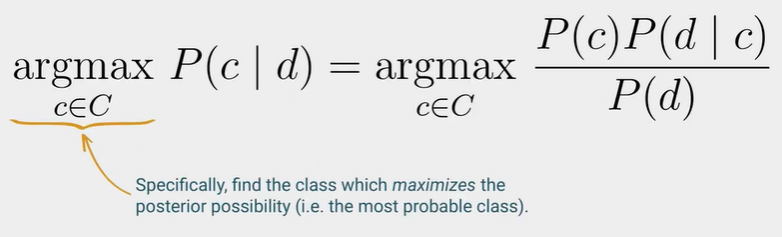
Suy ra P(A)) = P(B). Từ đó ta có công thức Bayes:

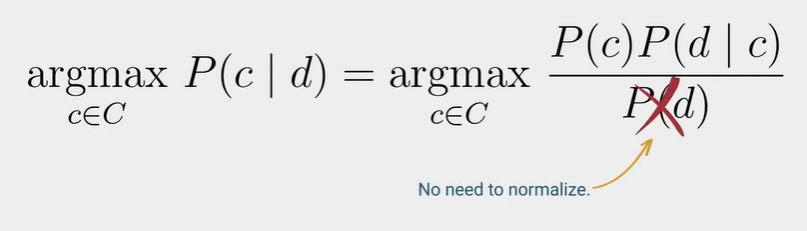
=



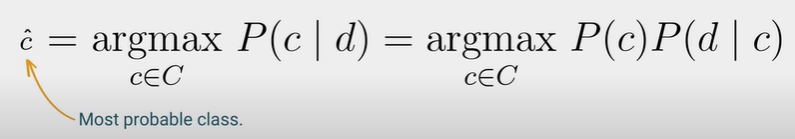
* Thay A và B bằng H (giả thuyết) và D (dữ liệu). Đây là xác suất của 1 giả thuyết là đúng dựa trên 1 số dữ liệu được quan sát. Nhưng theo cách này, định lý bayes cung cấp 1 cách để cập nhật mức độ tin tưởng khi có dữ liệu mới. VD: bạn có thể đang chạy 1 số dịch vụ đăng kí và có 1 spps niềm tin ban đầu về việc liệu khách hàng có đăng ký hay không và khi dữ liệu mới có sẵn thông qua việc khách hàng thực hiện 1 số hành động nhất định mà niềm tin sẽ điều chỉnh như khách hàng dăng ký dùng thử hoặc trả lời khảo sát….tất cả điều đó có thể ảnh hưởn đến niềm tin theo thời gian. Xác suất mà chúng ta đang tidnh toán này được gọi là xác suất hậu nghiệm.
* Để tính hậu nghiệm mà chúng ta cần để định lượng sức mạnh về niềm tin ban đầu trước khi xem xét bất kỳ bằng chứng hoặc dữ liệu nào. P(H) được gọi là dữ liệu trước, ngoài ra còn cần xác suất quan sát dữ liệu dựa trên niềm tin gọi là khả năng nhân với dữ liệu trước và chia cho xác suất của dữ liệi để chuẩn hoá xác suất nằm trong khoảng từ 0 đến 1 thường được gọi là hằng số chuẩn hoá. Có thể tính toán được tất cả các phần này dựa trên tập dữ liệu đang có.



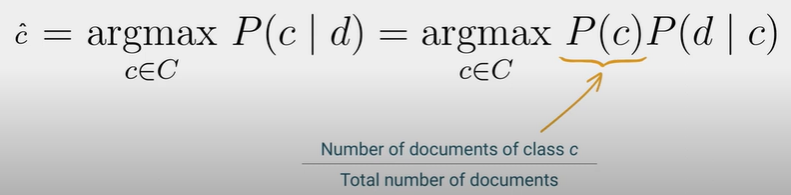
* Định lý bayes khi áp dụng cho phân loại văn bản. Thay A và B thành c và d trong đó c là lớp nhất định được cung cấp d là tài liệu cụ thể là quan tâm hơn đến lớp nào có khả năng thuộc vào loại topic nào. 
* Vì vậy điều mong muốn là lớp trong tập hợp các lớp có thể giúp tối đa hoá xác suất này. 
* Ngoài ra vì là tài liệu nên chúng ta chỉ quan tâm đến việc biết lớp nào có khả năng xảy ra cao nhất nên ta sẽ loại bỏ hằng số chuẩn hoá.



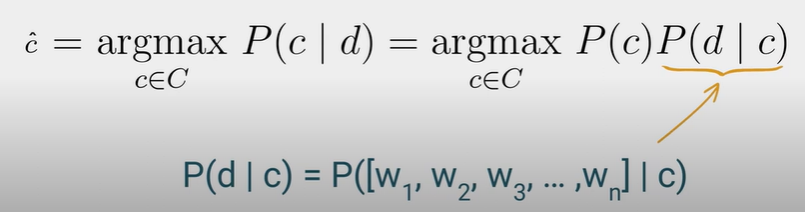
* là lớp có khả năng xảy ra cao nhất



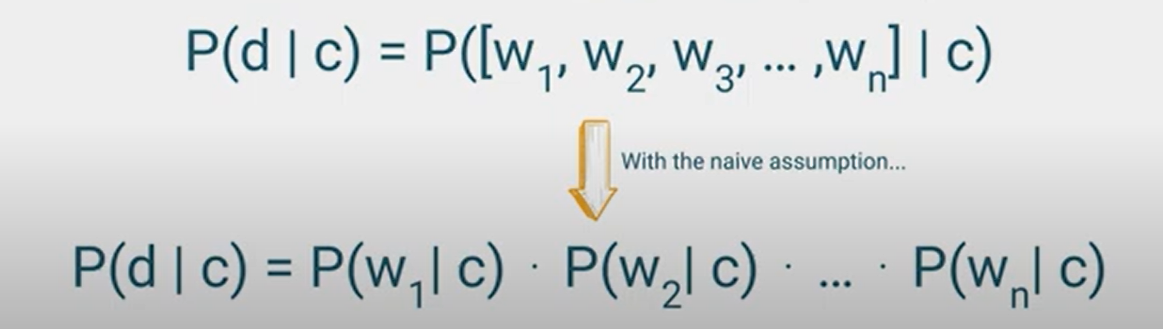
* P(c) là xác suất của số lượng văn bản loại c chia cho tổng số tài liệu



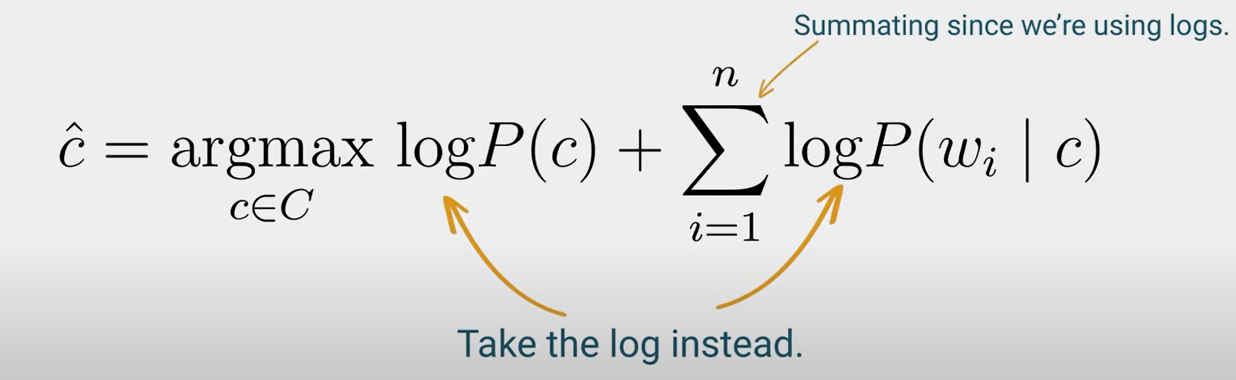
* Khả năng xảy ra là xác suất của 1 chuỗi từ cụ thể cho một lớp nên xác suất của các từ cụ thể này xuất hiện theo thứ tự cụ thể. Điều này rất khó để ước tính xác suất của mỗi chuỗi từ mà chúng ta



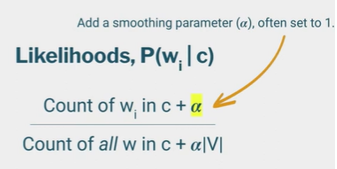
* Xử lý vấn đề: sự xuất hiên của từ thứ 2 là độc lập của nhau trong một lớp cụ thể được gọi là naïve vì giả định của nó khôn đúng trong cuộc sống thực, nhưng trong mục đích phân loại, nó đúng
* Với giả định này, chúng ta đi từ việc tính toán khả năng xảy ra như một chuỗi từ cụ thể cho 1 lớp đến xác suất của mỗi từ riêng lẻ xuất hiện trong 1 lớp và cách tính này dễ dàng hơn nhiều. Để có được kết quả xác suất cho toàn bộ dữ liệu chúng ta nhân xác suất của từng từ riêng lẻ với nhau



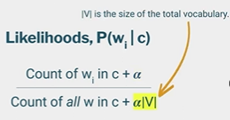
* Nhưng khi chúng ra nhân các xác suất nhỏ với nhau và nếu 1 tài liệu đủ dài với vốn từ đa dạng điều này có thể dẫn đến tình trạng tràn tài liệu. Thay vào đó chúng ta làm việc trong không gian Log, nên chúng ta lấy Log của tài liệu trước đó và ta lấy Log của từng xác suất từ riêng lẻ rồi cộng chúng lại với nhau và đây là trình phân loại naïve bayes



* Khi tính toán khả năng 1 từ nhưng lại nhận được kết quả là 0 do từ đó hoàn toàn không xuất hiện trong lớp, điều này sẽ làm cho toàn bộ xác suất bằng 0 vì như công thức ở trên thì sẽ nhân 0 với chuỗi các từ riêng lẻ khác hoặc nếu lấy Log thì sẽ gây ra lỗi. Vì vậy. nếu tình cờ có 1 tài liệu có trường hợp như thế thì xác suất của lớp đó sẽ về 0. Để giải quyết vấn đề này, ta thêm tham số làm mịn alpha vào tử số, thường được mặc định là 1, vì vậy trong trường hợp từ không xuất hiện, ta sẽ không nhận được kết quả là 0



* Ta cũng thêm tham số làm mịn tương tự và mẫu số nhân với kích thước của tổng số từ vựng



* VD:



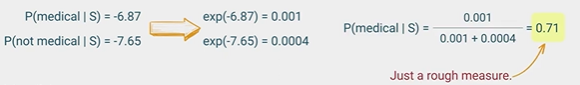
* Ban đầu: P =
* Sau khi thêm tham số làm mịn: P =
* Sau khi giải quyết xong thì làm tương tự với mỗi từ và đó là trình phân loại đã được đào tạo. Giả sử bây giờ muốn phân loại 1 tài liệu mới trong TH này là **“**thử nghiệm vắc xin **”** mà chúng ta sẽ gán là chữ cái S. tất cả những gì chúng ta cần là nhập xác suất trước đó của lớp y tế và cộng cùng với Log của xác suất của mỗi từ trong chuỗi.



Tương tự với not medical:



* Ta đang tìm kiếp lớp tối đa hoá xác suất và do đó trình phân loại trả về medical là lớp có thể xảy ra nhất và đó là cách nó hoạt động
* Nếu chúng ta hài lòng với việc chỉ biết lớp có thể xảy ra nhất nhưng nếu cần, chúng ta cũng có thể truy xuất một số thước đó xác suất. để thực hiện điều đó, chúng ta cần đảo ngược thao tác Log bằng cách nâng số mũ tự nhiên lên giá trị log: e^x



* Xác suất của medical là 71%. Đây là thước đo và sắp xếp sơ bộ . Trong hướng dẫn, trình phân loại naïve bayes được biết đến như 1 trình phân loại tốt nhưng có mức độ tin cậy khaasc nhau về khả năng ước tính của nó do các giả định naïve của nó.