



# Mô hình Naïve Bayes

Lê Thành Sách: [LTSACH@hcmut.edu.vn](mailto:LTSACH@hcmut.edu.vn)

Khoa Khoa học và Kỹ thuật Máy tính

Trường Đại học Bách Khoa Tp.HCM

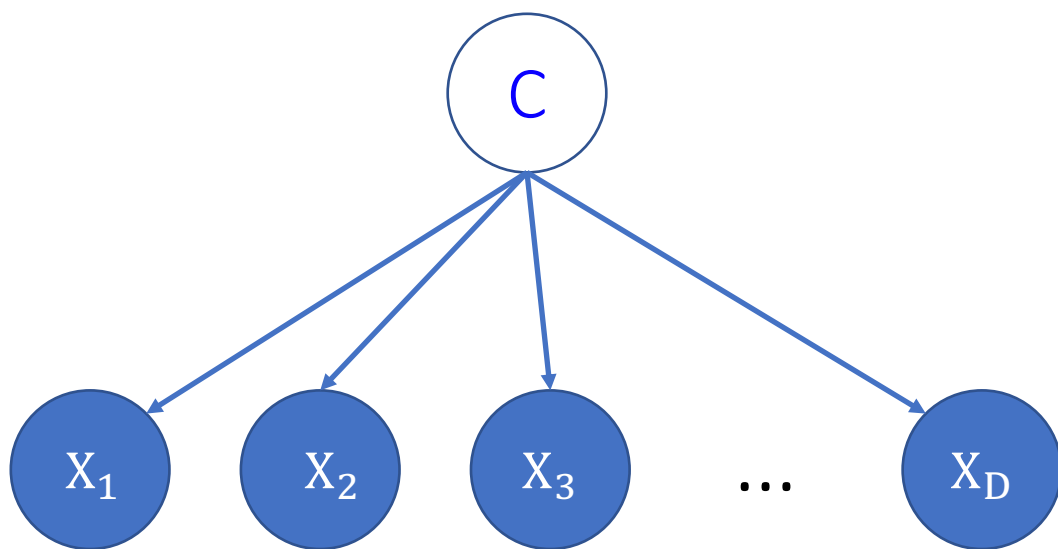


# Nội dung

- ❖ Mô hình Naïve Bayes là gì?
- ❖ Ước lượng các tham số của mô hình Bayes
- ❖ Ứng dụng



# Mô hình Naïve Bayes là gì?



Naïve Bayes là trường hợp đơn giản của mạng Bayes

$\mathcal{X} = \{X_1, X_2, \dots, X_D\}$ : tập các biến sẽ được quan sát trong mô hình

Mô hình Naïve Bayes giả thiết:

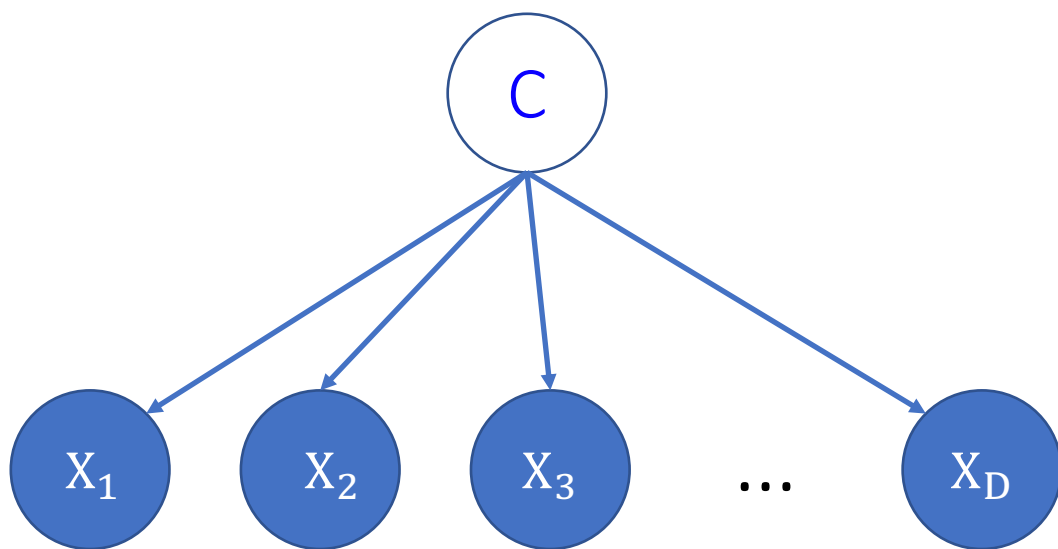
$$X_i \perp \mathcal{X}_{-i} \mid C$$

$$\text{Với: } \mathcal{X}_{-i} = \mathcal{X} - \{X_i\}$$

(Các biến  $X_i$  độc lập xác suất với tất cả các biến còn lại khi cho biết  $C$ )



# Mô hình Naïve Bayes là gì?



Naïve Bayes là trường hợp đơn giản của mạng Bayes

$\mathcal{X} = \{X_1, X_2, \dots, X_D\}$ : tập các biến sẽ được quan sát trong mô hình

Mô hình Naïve Bayes giả thiết:

$$X_i \perp \mathcal{X}_{-i} \mid C$$

Với:  $\mathcal{X}_{-i} = \mathcal{X} - \{X_i\}$

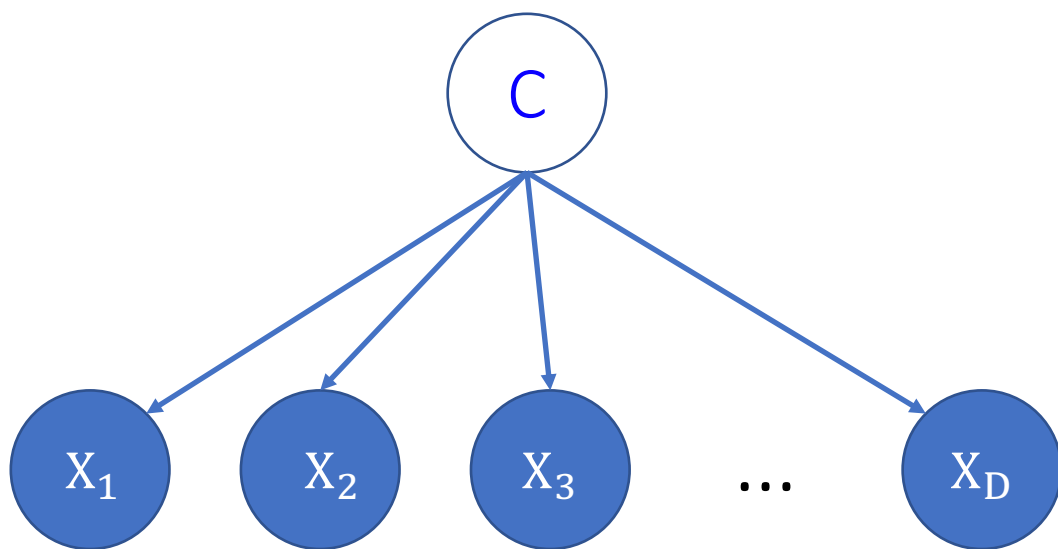
(Các biến  $X_i$  độc lập xác suất với tất cả các biến còn lại khi cho biết C)



Lý do bị gọi là “naïve” (ngây thơ)  
– vì giả thiết quá chặt/mạnh



# Mô hình Naïve Bayes là gì?



Naïve Bayes là trường hợp đơn giản của mạng Bayes

$\mathcal{X} = \{X_1, X_2, \dots, X_D\}$ : tập các biến sẽ được quan sát trong mô hình

Mô hình Naïve Bayes giả thiết:

$$X_i \perp \mathcal{X}_{-i} \mid C$$

Với:  $\mathcal{X}_{-i} = \mathcal{X} - \{X_i\}$

(Các biến  $X_i$  độc lập xác suất với tất cả các biến còn lại khi cho biết C)

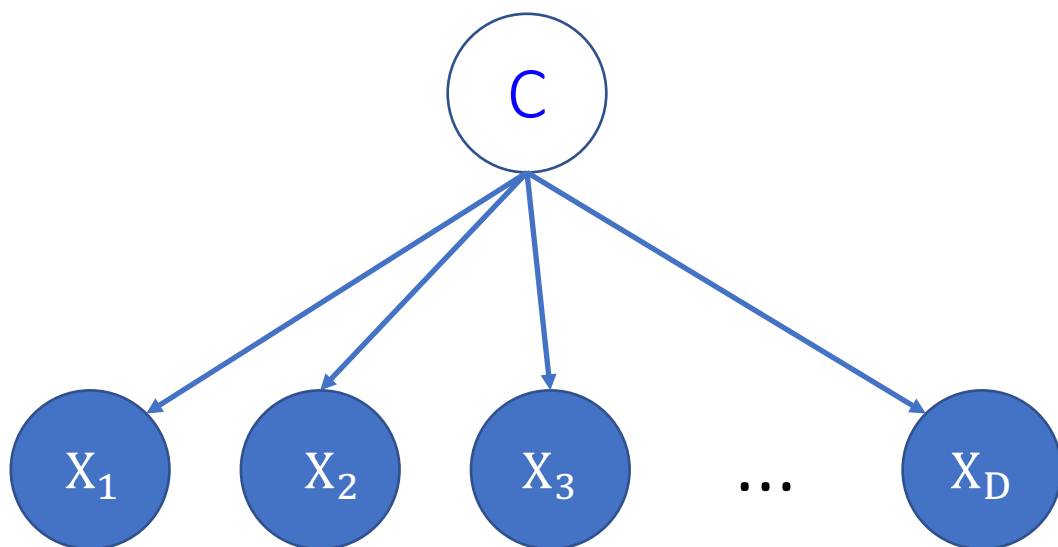


$$P(C, X_1, X_2, \dots, X_D) = P(C) \prod_{i=1}^D P(X_i \mid C)$$



# Mô hình Naïve Bayes là gì?

## Các tham số của mô hình



Tham số của mô hình = tập các giá trị mô tả các phân phối xác suất:

- $P(C)$
- $P(X_i | C)$  for  $i=1,2,\dots, D$

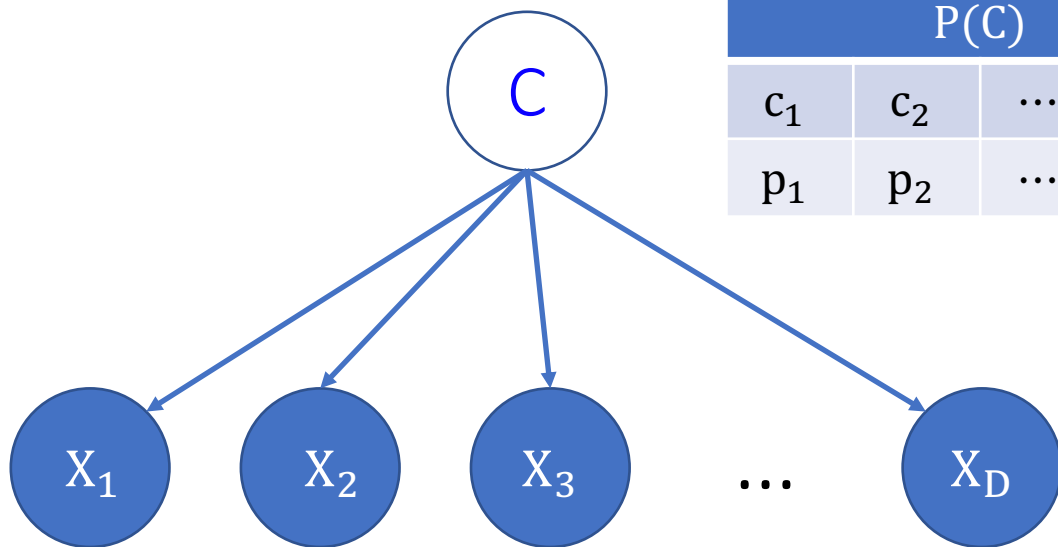
Naïve Bayes là trường hợp đơn giản của mạng Bayes

$\mathcal{X} = \{X_1, X_2, \dots, X_D\}$ : tập các biến sẽ được quan sát trong mô hình



# Mô hình Naïve Bayes là gì?

## Các tham số của mô hình: biến X rời rạc



P(C)			
$c_1$	$c_2$	...	$c_K$
$p_1$	$p_2$	...	$p_K$

$\Sigma=1$

P( $X_1$   C)		
C	$X_1$	P
$c_1$	$x_{1,1}$	$p_{1,1}$
$c_1$	$x_{1,2}$	$p_{1,2}$
$c_1$	...	...
$c_1$	$x_{1,F_1}$	$p_{1,F_1}$
$c_2$	$x_{2,1}$	$p_{2,1}$
$c_2$	$x_{2,2}$	$p_{2,2}$
$c_2$	...	...
$c_2$	$x_{2,F_2}$	$p_{2,F_2}$
...	...	...

$\Sigma=1$

P( $X_D$   C)		
C	$X_D$	P
$c_1$	$x_{D,1}$	$p_{D,1}$
$c_1$	$x_{D,2}$	$p_{D,2}$
$c_1$	...	...
$c_1$	$x_{D,F_D}$	$p_{D,F_D}$
$c_2$	$x_{D,1}$	$p_{D,1}$
$c_2$	$x_{D,2}$	$p_{D,2}$
$c_2$	...	...
$c_2$	$x_{D,F_D}$	$p_{D,F_D}$
...	...	...

$\Sigma=1$

Naïve Bayes là trường hợp đơn giản của mạng Bayes

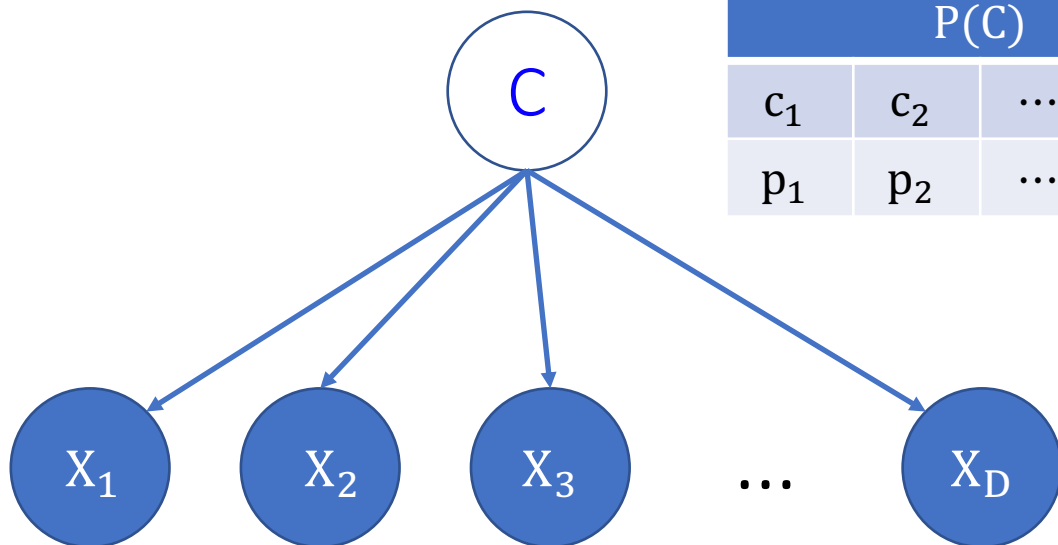
Phân phối xác suất có điều kiện ➡

- K: kích thước tập giá trị của biến C (số lớp trong bài toán phân loại)
- D: số biến phụ thuộc vào C (số đặc trưng trong vectơ đặc trưng)
- $F_i$ : kích thước tập giá trị của biến  $X_i$  (giả sử  $X_i$ : rời rạc)



# Mô hình Naïve Bayes là gì?

## Các tham số của mô hình: biến $X$ liên tục



P(C)			
$c_1$	$c_2$	...	$c_K$
$p_1$	$p_2$	...	$p_K$

$\Sigma=1$

P( $X_1$   C)		
C	$X_1$	P
$c_1$	$\mathcal{P}(\theta_{1,1})$	
$c_2$	$\mathcal{P}(\theta_{2,1})$	
...	...	...
$c_K$	$\mathcal{P}(\theta_{K,1})$	

...

P( $X_D$   C)		
C	$X_D$	P
$c_1$	$\mathcal{P}(\theta_{1,D})$	
$c_2$	$\mathcal{P}(\theta_{2,D})$	
...	...	...
$c_K$	$\mathcal{P}(\theta_{K,D})$	

Naïve Bayes là trường hợp đơn giản của mạng Bayes

$\mathcal{P}(\theta_{c,d})$ : Phân phối xác suất có điều kiện của biến  $X_d$  khi cho biết  $C=c$

- $\theta_{c,d}$ : bộ tham số của mô hình xác suất; ví dụ:  $(\mu, \sigma)$  nếu dùng mô hình Gaussian





# Naïve Bayes cho bài toán phân loại

Huấn luyện:

$X_1$	$X_2$	$\dots$	$X_D$	$C$
$x_1^{(1)}$	$x_2^{(1)}$	$\dots$	$x_D^{(1)}$	$c^{(1)}$
$x_1^{(2)}$	$x_2^{(2)}$	$\dots$	$x_D^{(2)}$	$c^{(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_1^{(N)}$	$x_2^{(N)}$	$\dots$	$x_D^{(N)}$	$c^{(N)}$

$$c^{(i)} \in \{c_1, c_2, \dots, c_K\}$$



Các tham số của mô hình

Làm việc:

$X_1$	$X_2$	$\dots$	$X_D$	$C$
$x_1$	$x_2$	$\dots$	$x_D$	?



Dự báo nhãn  $c$  cho vectơ đặc trưng:

$$\mathbf{x} = [x_1, x_2, \dots, x_D]$$



# Naïve Bayes cho bài toán phân loại

Làm việc:

$X_1$	$X_2$	...	$X_D$	$C$
$x_1$	$x_2$	...	$x_D$	?



Dự báo nhãn  $c$  cho véctơ đặc trưng:

$$\mathbf{x} = [x_1, x_2, \dots, x_D]$$

||

1. Tính:  $P(C | \mathbf{x})$

2. Xác định nhãn:  $c = \operatorname{argmax}_C P(C | \mathbf{x})$



# Naïve Bayes cho bài toán phân loại

Công thức Bayes:

$$P(C | \mathbf{x}) = \frac{P(C) \times P(\mathbf{x} | C)}{P(\mathbf{x})}$$

Posterior probability  
(xác suất hậu nghiệm)

Prior probability  
(xác suất tiên nghiệm)

Likelihood  
(Hệ số khả dĩ/hợp lý)

Evidence probability  
(Xác suất bằng chứng)



# Naïve Bayes cho bài toán phân loại

Công thức Bayes cho mô hình Naïve Bayes:

$$P(C | \mathbf{x}) = \frac{P(C) \times \prod_{i=1}^D P(x_i | C)}{P(\mathbf{x})}$$



$P(\mathbf{x})$  : không phụ thuộc C

$$P(C | \mathbf{x}) \propto \left[ P(C) \times \prod_{i=1}^D P(x_i | C) \right]$$



Xác định nhãn cho  $\mathbf{x}$ :

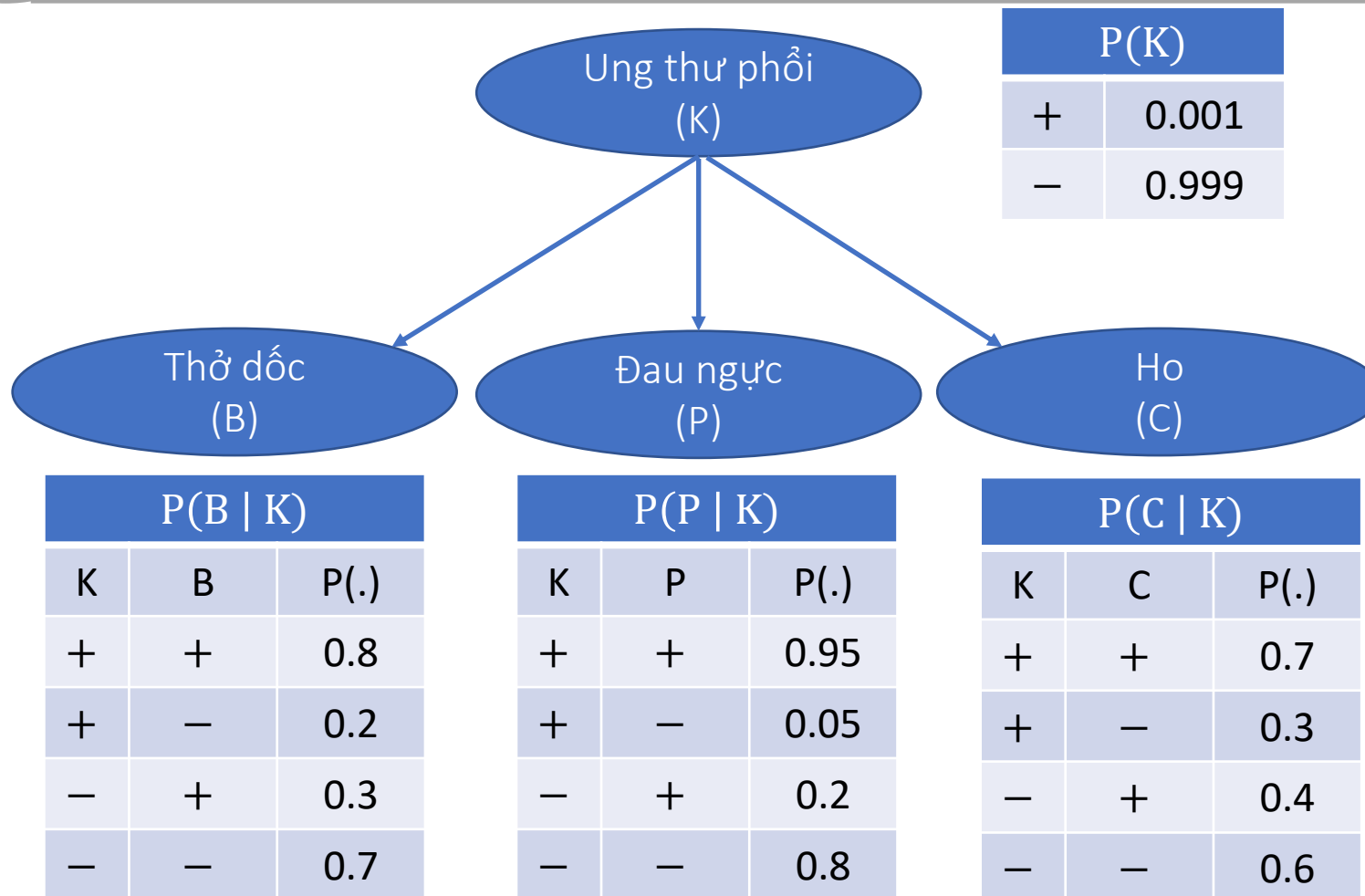
$$c = \operatorname{argmax}_C P(C | \mathbf{x}) = \operatorname{argmax}_C \left[ P(C) \times \prod_{i=1}^D P(x_i | C) \right]$$

$$c = \operatorname{argmax}_C \left[ \log_2 P(C) + \sum_{i=1}^D \log_2 P(x_i | C) \right]$$



# Naïve Bayes cho bài toán phân loại

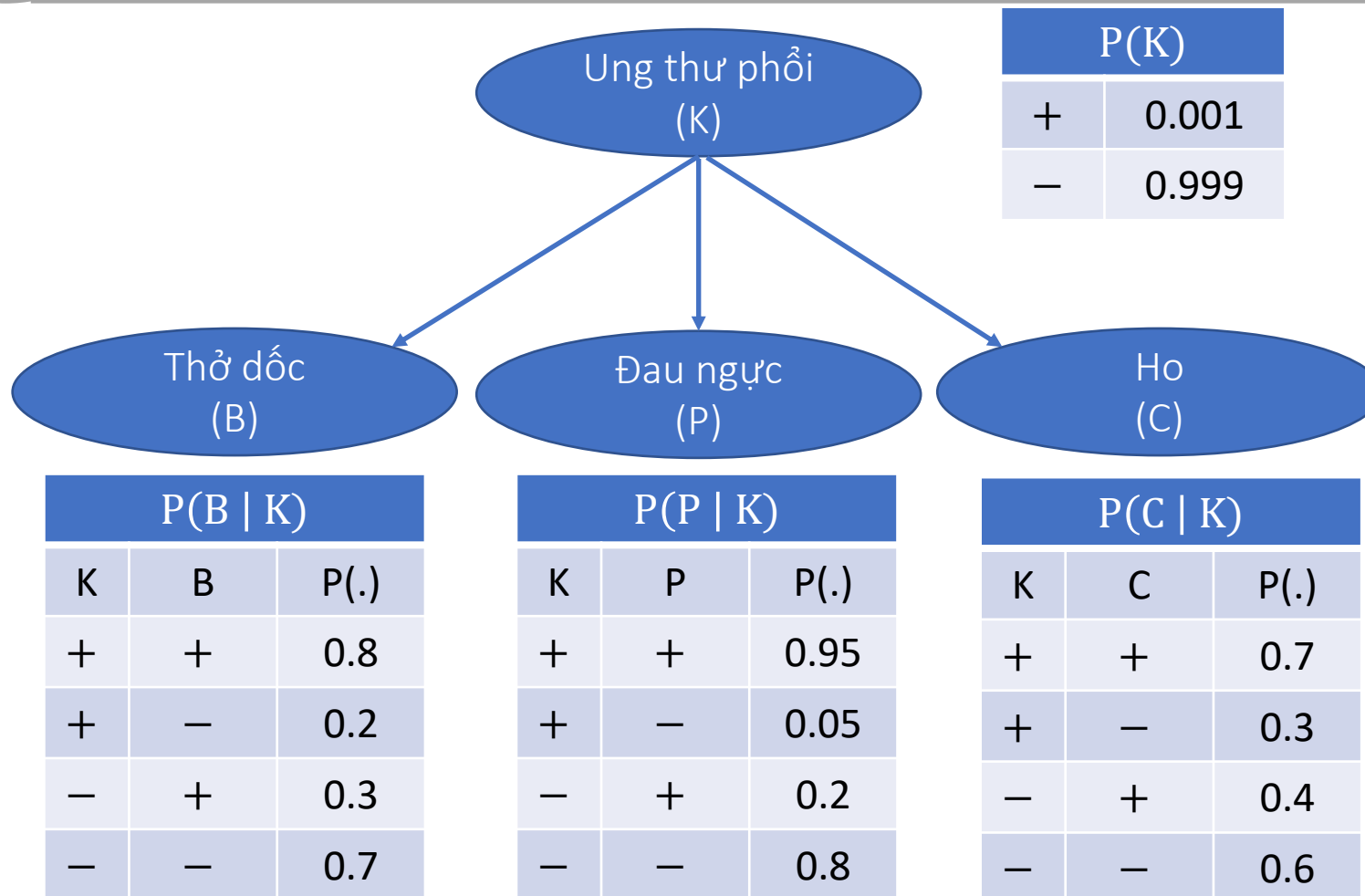
## Ví dụ 1





# Naïve Bayes cho bài toán phân loại

## Ví dụ 1



### Triệu chứng:

- Thở dốc (B): +
- Đau ngực (P): +
- Ho (C) : -
- => Véc tơ đặc trưng:  $x = [+ , + , -]$

Hỏi: có khả năng ung thư phổi?



# Naïve Bayes cho bài toán phân loại

## Ví dụ 1

P(K)	
+	0.001
-	0.999

P(P   K)		
K	P	P(.)
+	+	0.95
+	-	0.05
-	+	0.2
-	-	0.8

Triệu chứng:

- Thở dốc (B): +
- Đau ngực (P): +
- Ho (C) : -
- => Véc tơ đặc trưng:  $x = [+ , + , -]$

Hỏi: có khả năng ung thư phổi?

P(B   K)			P(C   K)		
K	B	P(.)	K	C	P(.)
+	+	0.8	+	+	0.7
+	-	0.2	+	-	0.3
-	+	0.3	-	+	0.4
-	-	0.7	-	-	0.6

$$\begin{aligned}P(C = +|x) &= P(C = +) \times P(B = +|+) \times P(P = +|+) \times P(C = -|+) \\&= 0.001 \times 0.8 \times 0.95 \times 0.3 = 0.000228\end{aligned}$$

$$\begin{aligned}P(C = -|x) &= P(C = -) \times P(B = +|-) \times P(P = +|-) \times P(C = -|-) \\&= 0.999 \times 0.3 \times 0.2 \times 0.6 = 0.035964\end{aligned}$$

$P(C = +|x) < P(C = -|x)$  ➡ Không ung thư phổi!



# Naïve Bayes cho bài toán phân loại

## Ví dụ 2

P(K)	
+	0.001
-	0.999

Ung thư phổi  
(K)

Thở dốc  
(B)

Đau ngực  
(P)

Ho  
(C)

X-Quang  
(X)

P(B   K)		
K	B	P(.)
+	+	0.8
+	-	0.2
-	+	0.3
-	-	0.7

P(P   K)		
K	P	P(.)
+	+	0.95
+	-	0.05
-	+	0.2
-	-	0.8

P(C   K)		
K	C	P(.)
+	+	0.7
+	-	0.3
-	+	0.4
-	-	0.6

P(X   K)		
K	C	P(.)
+	+	0.999
+	-	0.001
-	+	0.002
-	-	0.998

**Triệu chứng:** Thở dốc (B): +; Đau ngực (P): +; Ho (C) : -; X-Quang (X): +

• => Véc tơ đặc trưng:  $x = [+ , + , - , +]$

Hỏi: có khả năng ung thư phổi?





# Naïve Bayes cho bài toán phân loại

## Ví dụ 2

P(K)	
+	0.001
-	0.999

P(P   K)		
K	P	P(.)
+	+	0.95
+	-	0.05
-	+	0.2
-	-	0.8

P(X  K)		
K	C	P(.)
+	+	0.999
+	-	0.001
-	+	0.002
-	-	0.998

**Triệu chứng:** Thở dốc (B): +; Đau ngực (P): +; Ho (C) : -; X-Quang (X): +

• => Véc tơ đặc trưng:  $x = [+ , + , - , +]$

Hỏi: có khả năng ung thư phổi?

P(B   K)		
K	B	P(.)
+	+	0.8
+	-	0.2
-	+	0.3
-	-	0.7

P(C   K)		
K	C	P(.)
+	+	0.7
+	-	0.3
-	+	0.4
-	-	0.6

$$\begin{aligned}P(C = +|x) &= P(C = +) \times P(B = +|+) \times P(P = +|+) \times \\ &\quad P(C = -|+) \times P(X = +|+) \\ &= 0.001 \times 0.8 \times 0.95 \times 0.3 \times 0.999 = 0.00023\end{aligned}$$

$$\begin{aligned}P(C = -|x) &= P(C = -) \times P(B = +|-) \times P(P = +|-) \times \\ &\quad P(C = -|-) \times P(X = +|-) \\ &= 0.999 \times 0.3 \times 0.2 \times 0.6 \times 0.002 = 0.000072\end{aligned}$$





# Ước lượng tham số

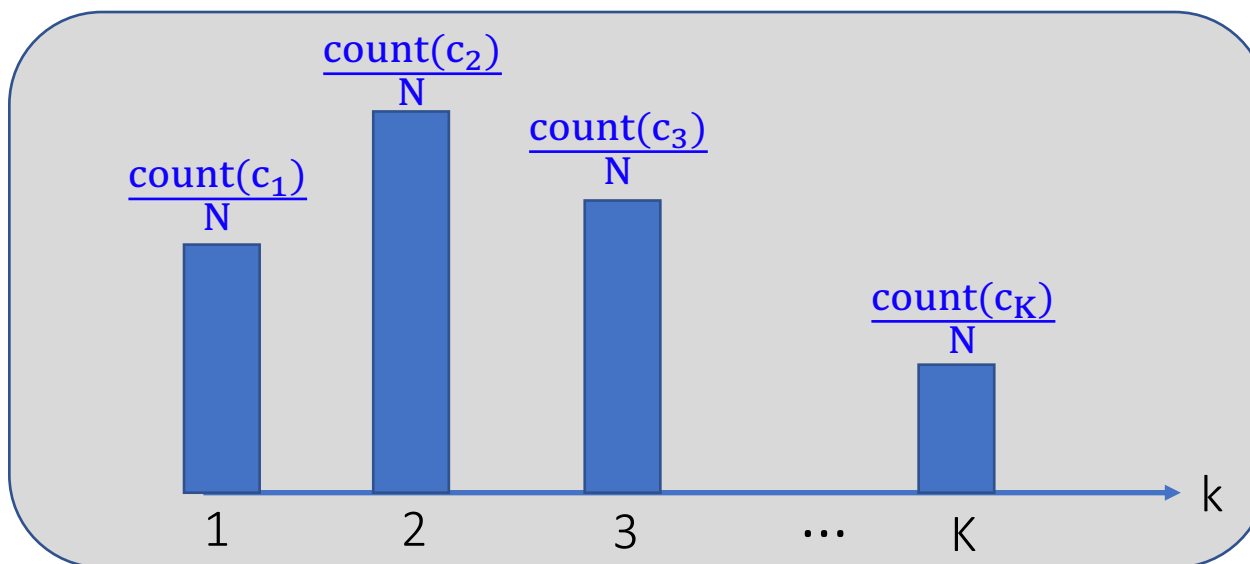
Huấn luyện:

$X_1$	$X_2$	$\dots$	$X_D$	$C$
$x_1^{(1)}$	$x_2^{(1)}$	$\dots$	$x_D^{(1)}$	$c^{(1)}$
$x_1^{(2)}$	$x_2^{(2)}$	$\dots$	$x_D^{(2)}$	$c^{(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_1^{(N)}$	$x_2^{(N)}$	$\dots$	$x_D^{(N)}$	$c^{(N)}$

$c^{(i)} \in \{c_1, c_2, \dots, c_K\}$

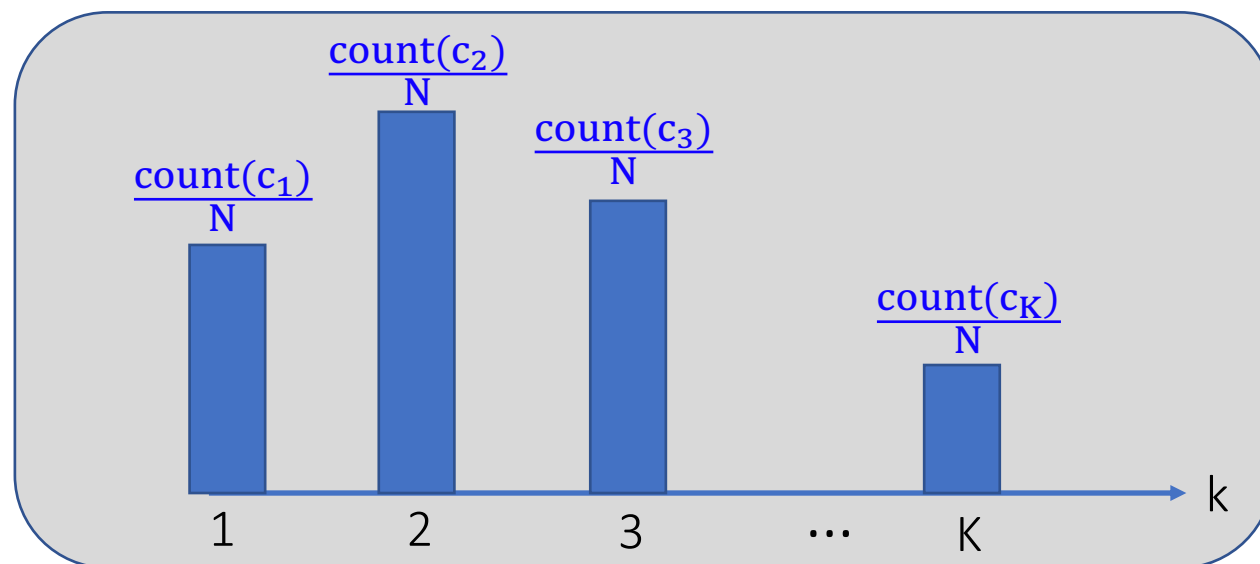


Tính histogram cho cột C





# Ước lượng tham số



$P(C)$			
$c_1$	$c_2$	$\dots$	$c_K$
$\frac{\text{count}(c_1)}{N}$	$\frac{\text{count}(c_2)}{N}$		$\frac{\text{count}(c_K)}{N}$

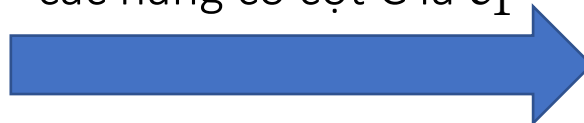


# Ước lượng tham số

Huấn luyện:

$X_1$	$X_2$	...	$X_D$	$C$
$x_1^{(1)}$	$x_2^{(1)}$	...	$x_D^{(1)}$	$c^{(1)}$
$x_1^{(2)}$	$x_2^{(2)}$	...	$x_D^{(2)}$	$c^{(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_1^{(N)}$	$x_2^{(N)}$	...	$x_D^{(N)}$	$c^{(N)}$

(1) Lựa chọn  
các hàng có cột  $C$  là  $c_1$



$X_1$	$X_2$	...	$X_D$	$C$
...	...	...	...	$c_1$
...	...	...	...	$c_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
...	...	...	...	$c_1$



(2) Tính  $P(X_i | C = c_1) = \text{histogram tính trên cột } X_i$

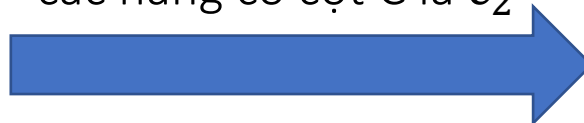


# Ước lượng tham số

Huấn luyện:

$X_1$	$X_2$	...	$X_D$	$C$
$x_1^{(1)}$	$x_2^{(1)}$	...	$x_D^{(1)}$	$c^{(1)}$
$x_1^{(2)}$	$x_2^{(2)}$	...	$x_D^{(2)}$	$c^{(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_1^{(N)}$	$x_2^{(N)}$	...	$x_D^{(N)}$	$c^{(N)}$

(1) Lựa chọn  
các hàng có cột  $C$  là  $c_2$



$X_1$	$X_2$	...	$X_D$	$C$
...	...	...	...	$c_2$
...	...	...	...	$c_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
...	...	...	...	$c_2$



(2) Tính  $P(X_i | C = c_2) = \text{histogram tính trên cột } X_i$

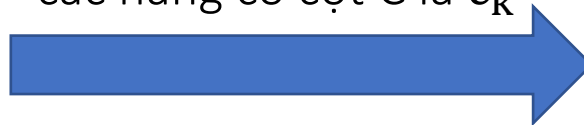


# Ước lượng tham số

Huấn luyện:

$X_1$	$X_2$	...	$X_D$	$C$
$x_1^{(1)}$	$x_2^{(1)}$	...	$x_D^{(1)}$	$c^{(1)}$
$x_1^{(2)}$	$x_2^{(2)}$	...	$x_D^{(2)}$	$c^{(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_1^{(N)}$	$x_2^{(N)}$	...	$x_D^{(N)}$	$c^{(N)}$

(1) Lựa chọn  
các hàng có cột  $C$  là  $c_k$



$X_1$	$X_2$	...	$X_D$	$C$
...	...	...	...	$c_k$
...	...	...	...	$c_k$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
...	...	...	...	$c_k$



(2) Tính  $P(X_i | C = c_k) = \text{histogram tính trên cột } X_i$



# Ước lượng tham số

## Trường hợp biến $X$ liên tục

$X_1$	$X_2$	...	$X_D$	$C$
$x_1^{(1)}$	$x_2^{(1)}$	...	$x_D^{(1)}$	$c^{(1)}$
$x_1^{(2)}$	$x_2^{(2)}$	...	$x_D^{(2)}$	$c^{(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_1^{(N)}$	$x_2^{(N)}$	...	$x_D^{(N)}$	$c^{(N)}$

(1) Lựa chọn các hàng có cột  $C$  là  $c_k$

$X_1$	$X_2$	...	$X_D$	$C$
...	...	...	...	$c_k$
...	...	...	...	$c_k$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
...	...	...	...	$c_k$

$X_1$	$X_2$	...	$X_D$	$C$
...	...	...	...	$c_1$
...	...	...	...	$c_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
...	...	...	...	$c_1$

$X_1$	$X_2$	...	$X_D$	$C$
...	...	...	...	$c_2$
...	...	...	...	$c_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
...	...	...	...	$c_2$

Lưu ý:  $N_1 + N_2 + \dots + N_K = N$



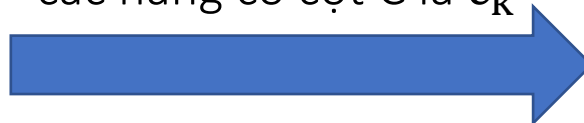
# Ước lượng tham số

## Trường hợp biến $X$ liên tục

Huấn luyện:

$X_1$	$X_2$	...	$X_D$	$C$
$x_1^{(1)}$	$x_2^{(1)}$	...	$x_D^{(1)}$	$c^{(1)}$
$x_1^{(2)}$	$x_2^{(2)}$	...	$x_D^{(2)}$	$c^{(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_1^{(N)}$	$x_2^{(N)}$	...	$x_D^{(N)}$	$c^{(N)}$

(1) Lựa chọn  
các hàng có cột  $C$  là  $c_k$



$X_1$	$X_2$	...	$X_D$	$C$
...	...	...	...	$c_k$
...	...	...	...	$c_k$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
...	...	...	...	$c_k$

Giả thiết: mô hình xác suất của biến  $x_d$  khi biết  $C = c_k$  là Gaussian (single model):

$$p(x; \mu_{k,d}, \sigma_{k,d}) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left[ \frac{x - \mu_{k,d}}{\sigma_{k,d}} \right]^2 \right\}$$





# Ước lượng tham số

## Trường hợp biến X liên tục

Huấn luyện:

$X_1$	$X_2$	...	$X_D$	C
$x_1^{(1)}$	$x_2^{(1)}$	...	$x_D^{(1)}$	$c^{(1)}$
$x_1^{(2)}$	$x_2^{(2)}$	...	$x_D^{(2)}$	$c^{(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_1^{(N)}$	$x_2^{(N)}$	...	$x_D^{(N)}$	$c^{(N)}$

(1) Lựa chọn  
các hàng có cột C là  $c_k$



$X_1$	$X_2$	...	$X_D$	C
...	...	...	...	$c_k$
...	...	...	...	$c_k$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
...	...	...	...	$c_k$

**Nghĩa là:** Lấy mẫu từ mô hình Gaussian

$$p(\mathbf{x}; \mu_{k,d}, \sigma_{k,d}) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left[ \frac{x - \mu_{k,d}}{\sigma_{k,d}} \right]^2 \right\}$$



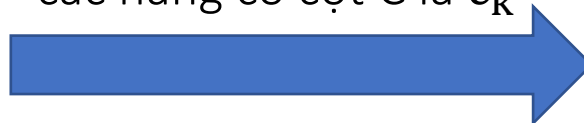
# Ước lượng tham số

## Trường hợp biến X liên tục

Huấn luyện:

$X_1$	$X_2$	...	$X_D$	C
$x_1^{(1)}$	$x_2^{(1)}$	...	$x_D^{(1)}$	$c^{(1)}$
$x_1^{(2)}$	$x_2^{(2)}$	...	$x_D^{(2)}$	$c^{(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_1^{(N)}$	$x_2^{(N)}$	...	$x_D^{(N)}$	$c^{(N)}$

(1) Lựa chọn  
các hàng có cột C là  $c_k$



$X_1$	$X_2$	...	$X_D$	C
...	...	...	...	$c_k$
...	...	...	...	$c_k$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
...	...	...	...	$c_k$



Giả sử dãy giá trị là:  $x_d^{[1]}, x_d^{[2]}, \dots, x_d^{[N_k]}$

Lưu ý:  $N_1 + N_2 + \dots + N_K = N$



$$\mu_{k,d} = \frac{1}{N_k} \sum_{i=1}^{N_k} x_d^{[i]}$$

$$\sigma_{k,d} = \frac{1}{N_k - 1} \sum_{i=1}^{N_k} [x_d^{[i]} - \mu_{k,d}]^2$$

$$p(x; \mu_{k,d}, \sigma_{k,d}) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left[ \frac{x - \mu_{k,d}}{\sigma_{k,d}} \right]^2 \right\}$$

Với:  $d=1,2,..D; k=1,2,..,K$



# Ước lượng tham số

## Ví dụ

### Features :

Outlook (O): {Sunny, Overcast, Rain}

Temperature(T): {Hot, Mild, Cool}

Humidity(H): {High, Normal}

Wind(W): {Strong, Weak}

### Target:

Play(Y): {Yes (+), No(-)}

STT	Outlook	Temperature	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No



# Ước lượng tham số

## Ví dụ

### Features :

Outlook (O): {Sunny, Overcast, Rain}

Temperature(T): {Hot, Mild, Cool}

Humidity(H): {High, Normal}

Wind(W): {Strong, Weak}

### Target:

Play(Y): {Yes, No}

Naïve Bayes Model:

$$P(Y|\mathbf{x}) = \operatorname{argmax}_Y [P(Y) \times P(O|Y) \times P(T|Y) \times P(H|Y) \times P(W|Y)]$$

$(\mathbf{x} = [O = ?, T = ?, H = ?, W = ?])$



# Ước lượng tham số

## Ví dụ

Histogram trên cột “Play”: {Yes: 9, No:5}



P(Y)	
Yes	$9/14=0.64$
No	$5/14=0.36$

STT	Outlook	Temperature	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No



# Ước lượng tham số

## Ví dụ

STT	Outlook	Temperature	Humidity	Wind	Play
1	Overcast	Hot	High	Weak	Yes
2	Rain	Mild	High	Weak	Yes
3	Rain	Cool	Normal	Weak	Yes
4	Overcast	Cool	Normal	Strong	Yes
5	Sunny	Cool	Normal	Weak	Yes
6	Rain	Mild	Normal	Weak	Yes
7	Sunny	Mild	Normal	Strong	Yes
8	Overcast	Mild	High	Strong	Yes
9	Overcast	Hot	Normal	Weak	Yes

Lựa chọn **Play=Yes**: có 9 hàng ( $N_1 = 9$ )



STT	Outlook	Temperature	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No



# Ước lượng tham số

## Ví dụ

STT	Outlook	Temperature	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Rain	Cool	Normal	Strong	No
4	Sunny	Mild	High	Weak	No
5	Rain	Mild	High	Strong	No



STT	Outlook	Temperature	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Lựa chọn **Play=No**: có 5 hàng ( $N_1 = 5$ )



# Ước lượng tham số

## Ví dụ

STT	Outlook	Temperature	Humidity	Wind	Play
1	Overcast	Hot	High	Weak	Yes
2	Rain	Mild	High	Weak	Yes
3	Rain	Cool	Normal	Weak	Yes
4	Overcast	Cool	Normal	Strong	Yes
5	Sunny	Cool	Normal	Weak	Yes
6	Rain	Mild	Normal	Weak	Yes
7	Sunny	Mild	Normal	Strong	Yes
8	Overcast	Mild	High	Strong	Yes
9	Overcast	Hot	Normal	Weak	Yes



### Histogram trên các cột:

- Outlook: {Overcast: 4; Rain:3; Sunny:2}
- Temperature: {Hot: 2; Mild: 4; Cool: 3}
- Humidity: {High: 3; Normal: 6}
- Wind: {Weak:6; Strong: 3}

Lựa chọn **Play=Yes**: có 9 hàng ( $N_1 = 9$ )





# Ước lượng tham số

## Ví dụ

Histogram trên các cột:

- Outlook: {Overcast: 4; Rain:3; Sunny:2}
- Temperature: {Hot: 2; Mild: 4; Cool: 3}
- Humidity: {High: 3; Normal: 6}
- Wind: {Weak:6; Strong: 3}

P(W   Y=Yes)		
Y	W	P(W   Y=Yes)
Yes	Weak	6/9=0.67
Yes	Strong	3/9=0.33

P(O   Y=Yes)		
Y	O	P(O   Y=Yes)
Yes	Overcast	4/9=0.44
Yes	Rain	3/9=0.33
Yes	Sunny	2/9=0.23

P(T   Y=Yes)		
Y	T	P(T   Y=Yes)
Yes	Hot	2/9=0.23
Yes	Mild	4/9=0.44
Yes	Cool	3/9=0.33

P(H   Y=Yes)		
Y	H	P(H   Y=Yes)
Yes	High	3/9=0.33
Yes	Normal	6/9=0.67



# Ước lượng tham số

## Ví dụ

STT	Outlook	Temperature	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Rain	Cool	Normal	Strong	No
4	Sunny	Mild	High	Weak	No
5	Rain	Mild	High	Strong	No



Historgam trên các cột:

- Outlook: {Overcast: 0; Rain:2; Sunny:3}
- Temperature: {Hot: 2; Mild: 2; Cool: 1}
- Humidity: {High: 4; Normal: 1}
- Wind: {Weak:2; Strong: 3}

Lựa chọn **Play=No**: có 9 hàng ( $N_1 = 5$ )



# Ước lượng tham số

## Ví dụ

Histogram trên các cột:

- Outlook: {Overcast: 0; Rain:2; Sunny:3}
- Temperature: {Hot: 2; Mild: 2; Cool: 1}
- Humidity: {High: 4; Normal: 1}
- Wind: {Weak:2; Strong: 3}

P(W   Y=No)		
Y	W	P(W   Y=No)
No	Weak	2/5=0.4
No	Strong	3/5=0.6

P(O   Y=No)		
Y	O	P(O   Y=No)
No	Overcast	0/5=0.0
No	Rain	2/5=0.4
No	Sunny	3/5=0.6

P(T   Y=No)		
Y	T	P(T   Y=No)
No	Hot	2/5=0.4
No	Mild	2/5=0.4
No	Cool	1/5=0.2

P(H   Y=No)		
Y	T	P(H   Y=No)
No	High	4/5=0.8
No	Normal	1/5=0.2



# Ước lượng tham số

## Ví dụ

Historgam trên các cột:

- Outlook: {Overcast: 4; Rain:3; Sunny:2}
- Temperature: {Hot: 2; Mild: 4; Cool: 3}
- Humidity: {High: 3; Normal: 6}
- Wind: {Weak:6; Strong: 3}

Historgam trên các cột:

- Outlook: {Overcast: 0; Rain:2; Sunny:3}
- Temperature: {Hot: 2; Mild: 2; Cool: 1}
- Humidity: {High: 4; Normal: 1}
- Wind: {Weak:2; Strong: 3}



P(O   Y)		
Y	O	P(O   Y)
Yes	Overcast	$4/9=0.44$
Yes	Rain	$3/9=0.33$
Yes	Sunny	$2/9=0.23$
No	Overcast	$0/5=0.0$
No	Rain	$2/5=0.4$
No	Sunny	$3/5=0.6$



# Ước lượng tham số

## Ví dụ

Historgam trên các cột:

- Outlook: {Overcast: 4; Rain:3; Sunny:2}
- Temperature: {Hot: 2; Mild: 4; Cool: 3}
- Humidity: {High: 3; Normal: 6}
- Wind: {Weak:6; Strong: 3}

Historgam trên các cột:

- Outlook: {Overcast: 0; Rain:2; Sunny:3}
- Temperature: {Hot: 2; Mild: 2; Cool: 1}
- Humidity: {High: 4; Normal: 1}
- Wind: {Weak:2; Strong: 3}



P(T   Y)		
Y	T	P(T   Y)
Yes	Hot	2/9=0.23
Yes	Mild	4/9=0.44
Yes	Cool	3/9=0.33
No	Hot	2/5=0.4
No	Mild	2/5=0.4
No	Cool	1/5=0.2



# Ước lượng tham số

## Ví dụ

Histogram trên các cột:

- Outlook: {Overcast: 4; Rain:3; Sunny:2}
- Temperature: {Hot: 2; Mild: 4; Cool: 3}
- Humidity: {High: 3; Normal: 6}
- Wind: {Weak:6; Strong: 3}

Histogram trên các cột:

- Outlook: {Overcast: 0; Rain:2; Sunny:3}
- Temperature: {Hot: 2; Mild: 2; Cool: 1}
- Humidity: {High: 4; Normal: 1}
- Wind: {Weak:2; Strong: 3}



P(H   Y)		
Y	H	P(H   Y)
Yes	High	3/9=0.33
Yes	Normal	6/9=0.67
No	High	4/5=0.8
No	Normal	1/5=0.2



# Ước lượng tham số

## Ví dụ

Histogram trên các cột:

- Outlook: {Overcast: 4; Rain:3; Sunny:2}
- Temperature: {Hot: 2; Mild: 4; Cool: 3}
- Humidity: {High: 3; Normal: 6}
- Wind: {Weak:6; Strong: 3}

Histogram trên các cột:

- Outlook: {Overcast: 0; Rain:2; Sunny:3}
- Temperature: {Hot: 2; Mild: 2; Cool: 1}
- Humidity: {High: 4; Normal: 1}
- Wind: {Weak:2; Strong: 3}



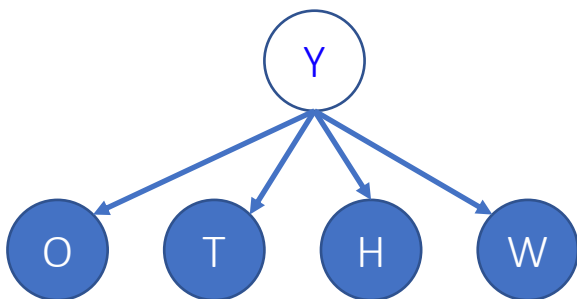
P(W   Y)		
Y	H	P(W   Y)
Yes	Weak	6/9=0.67
Yes	Strong	3/9=0.33
No	Weak	2/5=0.4
No	Strong	3/5=0.6



# Ước lượng tham số

## Ví dụ

<https://medium.com/@arunm8489/naive-bayes-for-machine-learning-238e8039edee>



P(Y)	
Yes	9/14=0.64
No	5/14=0.36

P(O   Y)		
Y	O	P(O   Y)
Yes	Overcast	4/9=0.44
Yes	Rain	3/9=0.33
Yes	Sunny	2/9=0.23
No	Overcast	0/5=0.0
No	Rain	2/5=0.4
No	Sunny	3/5=0.6

P(T   Y)		
Y	T	P(T   Y)
Yes	Hot	2/9=0.23
Yes	Mild	4/9=0.44
Yes	Cool	3/9=0.33
No	Hot	2/5=0.4
No	Mild	2/5=0.4
No	Cool	1/5=0.2

P(H   Y)		
Y	H	P(H   Y)
Yes	High	3/9=0.33
Yes	Normal	6/9=0.67
No	High	4/5=0.8
No	Normal	1/5=0.2

P(W   Y)		
Y	H	P(W   Y)
Yes	Weak	6/9=0.67
Yes	Strong	3/9=0.33
No	Weak	2/5=0.4
No	Strong	3/5=0.6