

# Machine Learning

## Lecture 8 Bayesian Classifier (貝氏分類法)

# Example: 預測下一位客人的行為

- 今有一間沙龍想預測下一位客人會要求的服務
- 假設此沙龍只提供兩種服務：剪髮(0) 跟 美甲(1)

# Example: 預測下一位客人的行為

- 今有一間沙龍想預測下一位客人會要求的服務
- 假設此沙龍只提供兩種服務：剪髮(0) 跟 美甲(1)
- 若去年的統計資料顯示：共有1000名顧客，其中650名要求剪髮，另350名要求美甲
- 請問你預測下一位客人想要求的服務為何？為什麼？

# Example: 預測下一位客人的行為

- 今有一間沙龍想預測下一位客人會要求的服務
- 假設此沙龍只提供兩種服務：剪髮(0) 跟 美甲(1)
- 若去年的統計資料顯示：共有1000名顧客，其中650名要求剪髮，另350名要求美甲
- 請問你預測下一位客人想要求的服務為何？為什麼？

要求剪髮 (0) 的機率  $Pr[0] = \frac{650}{1000} = 0.65$

要求美甲 (1) 的機率  $Pr[1] = \frac{350}{1000} = 0.35$

# Example: 預測下一位客人的行為

- 若進一步研究去年的統計資料，發現1000名顧客：

	男性	女性
剪髮 (0)	400	250
美甲 (1)	0	350

- 若已知下一位顧客是男性，請問你預測他想要求的服務為何？為什麼？

# Example: 預測下一位客人的行為

- 若進一步研究去年的統計資料，發現1000名顧客：

	男性	女性
剪髮 (0)	400	250
美甲 (1)	0	350

- 若已知下一位顧客是男性，請問你預測他想要求的服務為何？為什麼？

已知為男性，要求剪髮 (0) 的機率  $Pr[0 | \text{male}] = \frac{400}{400} = 1$

已知為男性，要求美甲 (1) 的機率  $Pr[1 | \text{male}] = \frac{0}{400} = 0$

# Example: 預測下一位客人的行為

- 若進一步研究去年的統計資料，發現1000名顧客：

	男性	女性
剪髮 (0)	400	250
美甲 (1)	0	350

- 若已知下一位顧客是女性，請問你預測她想要要求的服務為何？為什麼？

# Example: 預測下一位客人的行為

- 若進一步研究去年的統計資料，發現1000名顧客：

	男性	女性
剪髮 (0)	400	250
美甲 (1)	0	350

- 若已知下一位顧客是女性，請問你預測她想要求的服務為何？為什麼？

已知為女性，要求剪髮 (0) 的機率  $Pr[0 | \text{female}] = \frac{250}{600} = 0.42$

已知為女性，要求美甲 (1) 的機率  $Pr[1 | \text{female}] = \frac{350}{600} = 0.58$



# 問題：天氣評估系統

	<b>H<sub>1</sub></b>	<b>H<sub>2</sub></b>	<b>H<sub>3</sub></b>	<b>H<sub>4</sub></b>	<b>B</b>
Day	Outlook	Temp.	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Weak	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cold	Normal	Weak	Yes
D10	Rain	Mild	Normal	Strong	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

**{H<sub>1</sub> = Sunny, H<sub>2</sub> = Cool, H<sub>3</sub> = High, H<sub>4</sub> = Strong, B = ?}**

# 問題：天氣評估系統

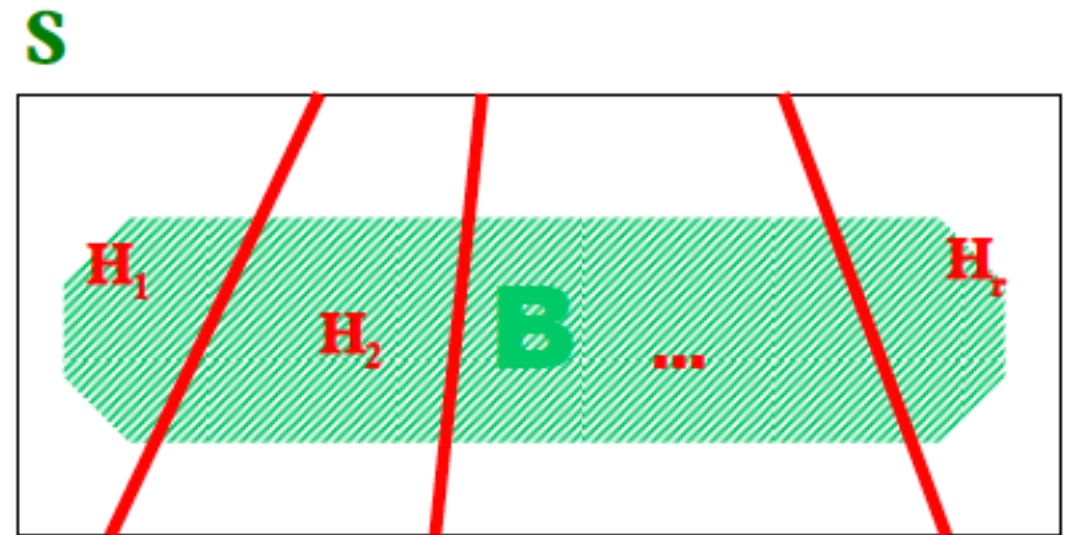
	$H_1$	$H_2$	$H_3$	$H_4$	$B$
Day	Outlook	Temp.	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Weak	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cold	Normal	Weak	Yes
D10	Rain	Mild	Normal	Strong	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Let  $H = (H_1 \cap H_2 \cap H_3 \cap H_4)$        $P(B=Yes | H)$     v.s.     $P(B=No | H)$

# 貝氏定理

- 由結果去追溯某個原因發生的機率，即由後天去推測先天。
- 設  $\{H_1, H_2, \dots, H_r\}$  為樣本空間  $S$  中的分割 ( $r \geq 2$ )， $B$  為  $S$  中的任意事件。若  $P(B) > 0$ ， $P(H_i) > 0$ ， $i = 1, 2, \dots, r$ ， $j = 1, 2, \dots, r$ ，則：

$$\begin{aligned} P(H_j | B) &= \frac{P(H_j \cap B)}{P(B)} \\ &= \frac{P(H_j \cap B)}{\sum_{i=1}^r P(H_i \cap B)} \\ &= \frac{P(H_j)P(B | H_j)}{\sum_{i=1}^r P(H_i)P(B | H_i)} \end{aligned}$$



文氏圖 (Venn Diagram)

- $P(H_j)$ ：事前機率(先天機率)，依據現有資訊所求得的機率。
- $P(H_j | B)$ ：事後機率，根據額外的資訊，經修正求得的機率。

# 貝氏定理

## ● 舉例：依下表，

- ① 請算出一年下雨的機率  $P(\text{雨})$ 。
- ② 假設某一天在下雨，這一天是秋天的機率為何？

	春	夏	秋	冬
雨	0.3	0.4	0.2	0.3
晴	0.7	0.6	0.8	0.7

## ● 解：

- ① 
$$\begin{aligned} P(\text{雨}) &= P(\text{春})P(\text{雨}|\text{春}) + P(\text{夏})P(\text{雨}|\text{夏}) + P(\text{秋})P(\text{雨}|\text{秋}) + P(\text{冬})P(\text{雨}|\text{冬}) \\ &= 0.25 \times 0.3 + 0.25 \times 0.4 + 0.25 \times 0.2 + 0.25 \times 0.3 \\ &= 0.3 \end{aligned}$$
- ② 
$$\begin{aligned} P(\text{秋}|\text{雨}) &= P(\text{雨} \cap \text{秋}) / P(\text{雨}) = P(\text{秋})P(\text{雨}|\text{秋}) / P(\text{雨}) \\ &= 0.25 \times 0.2 / 0.3 \\ &= 0.1666 \end{aligned}$$



# 貝氏分類

❖ 以  $B = \text{Yes}$  來說，其事後機率為：

$$\begin{aligned} P(B = \text{Yes} | H) &= \frac{P(H \cap B = \text{Yes})}{P(H)} = \frac{P(H | B = \text{Yes})P(B = \text{Yes})}{P(H)} \\ &= \frac{P(H_1 \cap H_2 \cap H_3 \cap H_4 | B = \text{Yes})P(B = \text{Yes})}{P(H)} \end{aligned}$$

# 樸素貝氏分類 (naive Bayesian classifier)

- 單純貝氏分類法假定**所有變數(屬性)對分類均是有用的**，且這些變數間是**相互獨立**。
- **條件獨立 (Conditionally Independent)**
  - 假設有三個隨機變數A、B與C。當變數A與變數B在給定條件C的情況下互相獨立，即稱**條件獨立**，可表示成：

$$P(A \cap B | C) = P(A | C) \times P(B | C)$$

# 樸素貝氏分類

❖ 以  $B = \text{Yes}$  來說，其事後機率為：

$$\begin{aligned} P(B = \text{Yes} | H) &= \frac{P(H \cap B = \text{Yes})}{P(H)} = \frac{P(H | B = \text{Yes})P(B = \text{Yes})}{P(H)} \\ &= \frac{P(H_1 \cap H_2 \cap H_3 \cap H_4 | B = \text{Yes})P(B = \text{Yes})}{P(H)} \\ &= \frac{P(B = \text{Yes})P(H_1 | B = \text{Yes})P(H_2 | B = \text{Yes})P(H_3 | B = \text{Yes})P(H_4 | B = \text{Yes})}{P(H)} \\ &= \frac{P(B = \text{Yes}) \prod_{i=1}^4 P(H_i | B = \text{Yes})}{P(H)} \end{aligned} \quad \text{條件獨立} \quad \dots\dots\dots(1)$$

# 樸素貝氏分類

- 以  $B = \text{Yes}$  來說，其事後機率為：

$$\begin{aligned} P(B = \text{Yes} | H) &= \frac{P(H \cap B = \text{Yes})}{P(H)} = \frac{P(H | B = \text{Yes})P(B = \text{Yes})}{P(H)} \\ &= \frac{P(H_1 \cap H_2 \cap H_3 \cap H_4 | B = \text{Yes})P(B = \text{Yes})}{P(H)} \\ &= \frac{P(B = \text{Yes})P(H_1 | B = \text{Yes})P(H_2 | B = \text{Yes})P(H_3 | B = \text{Yes})P(H_4 | B = \text{Yes})}{P(H)} \\ &= \frac{P(B = \text{Yes}) \prod_{i=1}^4 P(H_i | B = \text{Yes})}{P(H)} \end{aligned} \quad \text{條件獨立} \quad \dots\dots\dots(1)$$

- 以  $B = \text{No}$  來說，其事後機率為：

$$\begin{aligned} P(B = \text{No} | H) &= \frac{P(H \cap B = \text{No})}{P(H)} = \frac{P(H | B = \text{No})P(B = \text{No})}{P(H)} \\ &= \frac{P(H_1 \cap H_2 \cap H_3 \cap H_4 | B = \text{No})P(B = \text{No})}{P(H)} \\ &= \frac{P(B = \text{No}) \prod_{i=1}^4 P(H_i | B = \text{No})}{P(H)} \end{aligned} \quad \text{條件獨立} \quad \dots\dots\dots(2)$$



# Example

- 上述公式中的 $P(B=Yes)$ 和 $P(B=No)$ 都是屬於**事前機率**。在本範例中， **$P(B=Yes) = 9/14$ ， $P(B=No)=5/14$** 。
- 以上述公式 (1)、(2)來說，不論是在求解B為Yes或No，分母 $P(H)$ 皆相同，故可視為**固定的常數**，為計算的簡便性可省略它!!因此，公式 (1)、(2) 可修正為概似函數 (Likelihood Function)：

$$P(B = Yes | H) = P(B = Yes) \prod_{i=1}^4 P(H_i | B = Yes) \quad \text{.....(3)}$$

$$P(B = No | H) = P(B = No) \prod_{i=1}^4 P(H_i | B = No) \quad \text{.....(4)}$$

# Example

$$P(H_1 | B = Yes), P(H_2 | B = Yes), P(H_3 | B = Yes), P(H_4 | B = Yes)$$

Day	Outlook	Temp.	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Weak	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cold	Normal	Weak	Yes
D10	Rain	Mild	Normal	Strong	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

	Outlook		
	Sunny	Overcast	Rain
Yes	2	4	3
No	3	0	2
Yes	2/9	4/9	3/9
No	3/5	0/5	2/5

$P(\text{Sunny} | \text{Yes})$

	Temp.		
	Hot	Mild	Cool
Yes	2	4	3
No	2	2	1
Yes	2/9	4/9	3/9
No	2/5	2/5	1/5

	Humidity	
	High	Normal
Yes	3	6
No	4	1
Yes	3/9	6/9
No	4/5	1/5

	Wind	
	Weak	Strong
Yes	6	3
No	2	3
Yes	6/9	3/9
No	2/5	3/5

# Example

	Humidity	
	High	Normal
Yes	3	6
No	4	1
Yes	3/9	6/9
No	4/5	1/5

	Temp.		
	Hot	Mild	Cool
Yes	2	4	3
No	2	2	1
Yes	2/9	4/9	3/9
No	2/5	2/5	1/5

	Outlook		
	Sunny	Overcast	Rain
Yes	2	4	3
No	3	0	2
Yes	2/9	4/9	3/9
No	3/5	0/5	2/5

	Wind	
	Weak	Strong
Yes	6	3
No	2	3
Yes	6/9	3/9
No	2/5	3/5

- 因此，以本範例的待分類資料  $\{H_1 = \text{Sunny}, H_2 = \text{Cool}, H_3 = \text{High}, H_4 = \text{Strong}\}$  來說：

$$P(B=\text{Yes} | H) = 9/14 \times 2/9 \times 3/9 \times 3/9 \times 3/9 = 0.0053$$

$$P(B=\text{No} | H) = 5/14 \times 3/5 \times 1/5 \times 4/5 \times 3/5 = 0.0206$$

- 根據計算結果，此待分類的天氣屬性，被判別為  $B = \text{No}$ 。

$\{H_1 = \text{Sunny}, H_2 = \text{Cool}, H_3 = \text{High}, H_4 = \text{Strong}, B = \text{No}\}$

# 樸素貝氏分類

- 現有一組預測貸款歸還情形的訓練資料如下表，資料類型可分成兩大類：

- 離散型資料欄位

- 資料的值域有限或可數

- 連續型資料欄位

- 資料的值域繁多或不可數

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Figure 5.9. Training set for predicting the loan default problem.



# 朴素貝氏分類

## ● 就離散型資料欄位(屬性)而言：

■ 對於某個離散型資料欄位  $X$  而言，其條件機率  $P(X = x_i | Y = y)$  是指在  $y$  類別中的部份訓練資料，具有特別的屬性值  $x_i$ 。

● 如前圖，有3/7無拖欠貸款的人，是擁有房子的。  $P(\text{Home Owner} = \text{Yes} | \text{Defaulted Borrower} = \text{No})$

● 2/3有拖欠貸款的人，是單身的。  $P(\text{Marital Status} = \text{Single} | \text{Defaulted Borrower} = \text{Yes})$

# 朴素貝氏分類

## ● 就連續型資料欄位(屬性)而言：

■ 可以假設該連續變數符合**某個機率分配**，然後使用訓練資料來估計**該分配的參數**。

■ **常態分配(高斯分配)**最常被用來表示連續變數的類別條件機率。

● 該分配有兩個參數：**平均數  $\mu$** 與**變異數  $\sigma^2$**

● 對於每個類別 $y_j$ 而言，連續型資料欄位  $X$  的類別條件機率如下：

$$P(X = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

● 參數  $\mu$  可以經由連續型資料欄位  $X$  中、屬於類別 $y_j$ 之**樣本平均數  $\bar{X}_j$** 估計而來。而  $\sigma^2$  可以從**樣本變異數  $S_j^2$** 估計而來。

# 朴素貝氏分類

- 以本範例中的**年收入(Annual Income)**而言，在**類別為 no** 時的樣本平均數與變異數為：

$$\bar{X}_{No} = \frac{125 + 100 + \dots + 75}{7} = 110$$

$$S_{No}^2 = \frac{(125 - 110)^2 + (100 - 110)^2 + \dots + (75 - 110)^2}{7} = 2975$$

$$S_{No} = 54.54$$

- 假設現有一測試資料，年收入為120K，那麼可以計算類別為no 時的條件機率 (以常態分配為主)：

$$P(\text{Income} = 120 | Y = \text{No}) = \frac{1}{\sqrt{2\pi}(54.54)} e^{\frac{-(120-110)^2}{2 \times 2975}} = 0.0072$$



# 樸素貝氏分類

- 因此，本範例的所有可能條件機率整理如下圖所示：

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

(a)

$P(\text{Home Owner}=\text{Yes}|\text{No}) = 3/7$   
 $P(\text{Home Owner}=\text{No}|\text{No}) = 4/7$   
 $P(\text{Home Owner}=\text{Yes}|\text{Yes}) = 0$   
 $P(\text{Home Owner}=\text{No}|\text{Yes}) = 1$   
 $P(\text{Marital Status}=\text{Single}|\text{No}) = 2/7$   
 $P(\text{Marital Status}=\text{Divorced}|\text{No}) = 1/7$   
 $P(\text{Marital Status}=\text{Married}|\text{No}) = 4/7$   
 $P(\text{Marital Status}=\text{Single}|\text{Yes}) = 2/3$   
 $P(\text{Marital Status}=\text{Divorced}|\text{Yes}) = 1/3$   
 $P(\text{Marital Status}=\text{Married}|\text{Yes}) = 0$

For Annual Income:  
If class=No: sample mean=110  
                    sample variance=2975  
If class=Yes: sample mean=90  
                    sample variance=25

(b)

Figure 5.10. The naïve Bayes classifier for the loan classification problem.



# 朴素貝氏分類

- 現有一筆測試資料  $H = \{\text{Home Owner} = \text{No}, \text{Marital Status} = \text{Married}, \text{Income} = 120\text{K}\}$ ，要得知其是否會拖欠貸款 ( $\text{Defaulted Borrower} = ?$ )。

- 每個類別的事前機率分別為： $P(\text{Yes}) = 0.3, P(\text{No}) = 0.7$ 。

- 每個類別的條件計算公式如下：

$$\begin{aligned} P(B = \text{No} | H) &= P(B = \text{No}) \prod_{i=1}^3 P(H_i | B = \text{No}) \\ &= P(\text{No}) \times P(\text{Home Owner} = \text{No} | \text{No}) \times P(\text{Status} = \text{Married} | \text{No}) \times P(\text{Annual Income} = 120\text{K} | \text{No}) \\ &= 0.7 \times \frac{4}{7} \times \frac{4}{7} \times 0.0072 \\ &= 0.00165. \end{aligned}$$

$$\begin{aligned} P(B = \text{Yes} | H) &= P(B = \text{Yes}) \prod_{i=1}^3 P(H_i | B = \text{Yes}) \\ &= P(\text{Yes}) \times P(\text{Home Owner} = \text{No} | \text{Yes}) \times P(\text{Status} = \text{Married} | \text{Yes}) \times P(\text{Annual Income} = 120\text{K} | \text{Yes}) \\ &= 0.3 \times 1 \times 0 \times 1.2 \times 10^{-9} \\ &= 0. \end{aligned}$$

- 所以這筆測試資料被判為類別 **No**

# 朴素貝氏分類討論

① 當有資料遺漏時，對分類結果不會造成太大的影響。

■ 當訓練範例有資料遺漏時：

- 在進行頻率計數時，僅需忽略對該遺漏值的計算，同時在機率計算中使用實際出現的個數，而非訓練資料的總數。

# 樸素貝氏分類討論

① 當有資料遺漏時，對分類結果不會造成太大的影響。

■ 當訓練範例有資料遺漏時：

- 在進行頻率計數時，僅需**忽略對該遺漏值的計算**，同時在機率計算中使用**實際出現的個數**，而非訓練資料的總數。

■ 當測試資料有遺漏時：

- 假設有一組待分類的天氣屬性如下 (缺 “**Outlook ( $H_1$ )**”)：

$$\{H_2 = \text{Cool}, H_3 = \text{High}, H_4 = \text{Strong}, B = ?\}$$

在計算時，僅需**簡單忽略這個屬性**：

$$P(B=\text{Yes} | H) = 9/14 \times 3/9 \times 3/9 \times 3/9 = 0.0238$$

$$P(B=\text{No} | H) = 5/14 \times 1/5 \times 4/5 \times 3/5 = 0.0343$$

- 這兩個數值分別比之前的計算值高很多，因為計算時少了一部份的數值，但是**不影響結果**，因為這樣的遺漏值對所有類別的計算都影響到了。
- 標準化後，得到的Yes和No的機率分別為41%和59%。



# 樸素貝氏分類討論

②如果某個屬性的資料，並不是在每個類別都會出現的話，則會發生問題。

- 例如：前述“在何種天氣模式下，會/不會去打網球”的問題中，當  $B = \text{No}$  時， $\text{Outlook} = \text{Overcast}$  的個數為 0，則其條件機率  $P(\text{Outlook}=\text{Overcast} \mid B=\text{No}) = 0$ 。

● 這會使得整個類別的概似函數公式為 0。

- 這是當訓練資料過少、而屬性個數過多時所產生的問題。

	Outlook		
	Sunny	Overcast	Rain
Yes	2	4	3
No	3	0	2
Yes	2/9	4/9	3/9
No	3/5	0/5	2/5

# 樸素貝氏分類討論

	Humidity	
	High	Normal
Yes	3	6
No	4	1
Yes	3/9	6/9
No	4/5	1/5

	Outlook		
	Sunny	Overcast	Rain
Yes	2	4	3
No	3	0	2
Yes	2/9	4/9	3/9
No	3/5	0/5	2/5

	Temp.		
	Hot	Mild	Cool
Yes	2	4	3
No	2	2	1
Yes	2/9	4/9	3/9
No	2/5	2/5	1/5

	Wind	
	Weak	Strong
Yes	6	3
No	2	3
Yes	6/9	3/9
No	2/5	3/5

■ 假設有一組待分類的天氣屬性如下：

$\{H_1 = \text{Overcast}, H_2 = \text{Cool}, H_3 = \text{High}, H_4 = \text{Strong}, B = ?\}$

在計算時：

$$P(B=\text{Yes} | H) = 9/14 \times 4/9 \times 3/9 \times 3/9 \times 3/9 = 0.0106$$

$$P(B=\text{No} | H) = 5/14 \times 0/5 \times 1/5 \times 4/5 \times 3/5 = 0$$

標準化後，得到的Yes和No的機率分別為 1 和 0。

■ 在理論上是正確的，因為只要Outlook為Overcast就可打網球。但這個結果違背了單純貝氏分類法的基本假設：**所有變數對分類均是有用的**。這個範例僅靠Outlook = Overcast來決定。

● 解決：Laplace estimator

# Laplace Estimator

- 由於前述討論 ② 的問題是發生在簡單貝氏分類法中，條件機率連乘時發生乘 0 的情況，使得某類別  $b_j$  的概似函數計算結果為 0。

$$P(B = b_j | H) = P(B = b_j) \prod_{i=1}^r P(H_i | B = b_j), \quad j = 1, \dots, m.$$

- 使用Laplace Estimator評估條件機率的作法如下：
  - 概念：對造成條件機率為0之屬性，將其所屬類別  $b_j$  之機率計算公式的分子、分母皆加上一數值，使該機率不為0
    - 分子：加上1。
    - 分母：加上q，其中 q 為該屬性內的不同資料個數。



# Example

- 在是否打網球之分類問題中，條件機率  $P(\text{Outlook} = \text{Overcast} \mid B = \text{No}) = 0$ 
  - 因為在類別為No中，沒有一筆訓練資料其屬性“Outlook”為陰天(Overcast)。
- 在此使用Laplace Estimator，對類別為No之所有Outlook情況的條件機率做修正：

	Outlook		
	Sunny	Overcast	Rain
Yes	2	4	3
No	3	0	2
Yes	$2/9$	$4/9$	$3/9$
No	$3/5$	$0/5$	$2/5$

Red dashed lines and arrows indicate the correction process for the 'No' row. The original values  $3/5$ ,  $0/5$ , and  $2/5$  are crossed out with red 'X' marks. Red arrows point from these cells to the corrected values  $4/8$ ,  $1/8$ , and  $3/8$  respectively.

# Example

■ 假設有一組待分類的天氣屬性如下：

$\{H_1 = \text{Overcast}, H_2 = \text{Cool}, H_3 = \text{High}, H_4 = \text{Strong}, B = ?\}$

在計算時：

$$P(B=\text{Yes} | H) = 9/14 \times 4/9 \times 3/9 \times 3/9 \times 3/9 = 0.0106$$

$$P(B=\text{No} | H) = 5/14 \times 1/8 \times 1/5 \times 4/5 \times 3/5 = 0.00429$$

標準化後，得到的Yes和No的機率分別為 0.71189 和 0.28811。

■ 根據計算結果，此待分類的天氣屬性，被判別為 **B = Yes**。

	Humidity	
	High	Normal
Yes	3	6
No	4	1
Yes	3/9	6/9
No	4/5	1/5

	Outlook		
	Sunny	Overcast	Rain
Yes	2	4	3
No	3	0	2
Yes	2/9	4/9	3/9
No	4/8	1/8	3/8

	Temp.		
	Hot	Mild	Cool
Yes	2	4	3
No	2	2	1
Yes	2/9	4/9	3/9
No	2/5	2/5	1/5

	Wind	
	Weak	Strong
Yes	6	3
No	2	3
Yes	6/9	3/9
No	2/5	3/5