

INTELIGENCIA ARTIFICIAL Y SISTEMAS
EXPERTOS
Probabilidad Condicionada

Ulises C. Ramirez

8 de Octubre, 2018

Versionado

Para el corriente documento se está llevando un versionado a fin de mantener un respaldo del trabajo y además proveer a la cátedra o a cualquier interesado la posibilidad de leer el material en la última versión disponible.

REPOSITORIO: *<https://github.com/ulisescolina/UC-IA/>*

–ULISES

Probabilidad Condicionada

En este documento se provee un ejemplo de la aplicación de lo visto en clase a un ejemplo inventado por el alumno tomando como base el ejemplo proveído por la cátedra [IA y SE - 2018].

Definición del escenario

Se ha recibido una embarcacion con contenedores que poseen placas madre para ordenadores, en total la cantidad de *items* que se tienen es de 123456, luego de un análisis del inventario se encontró que los items pertenecian a las marcas M1, M2, M3 y M4. Otra de las cuestiones que fueron de interés en el análisis del cargamento es determinar con que *features* dentro del abanico de características de placas madre contaba cada placa, se encontraron las siguientes C1, C2 y C3.

Se debe poder determinar lo siguiente:

1. Probabilidad a priori de cada **Marca** si se presenta un item al azar
2. La probabilidad que se presente una **Marca** sabiendo que el **item** tiene todos los *features*.

Se tiene que la cantidad de placas según cada fabricante es la siguiente:

$$\begin{aligned}Total &= 18800 \\M1 &= 3500 \\M2 &= 5500 \\M3 &= 2500 \\M4 &= 7300\end{aligned}$$

Table 1: Datos tabulados para el contenido recibido

	C1	C2	C3
M1	3130	502	1244
M2	2325	5351	124
M3	273	2312	14
M4	135	7281	4214

Probabilidades para M_i

Para realizar éstos cálculos se realizará la operación presentada en la ecuación 1

$$P(M_i) = \frac{\#M_i}{Total} \quad (1)$$

A continuación, en los **Desarrollos 2, 3, 4 y 5** se exponen los diferentes resultados obtenidos del cálculo de probabilidades a priori para M_i

$$\begin{aligned}
P(M_1) &= \frac{3500}{18800} \\
&= 0,186170
\end{aligned} \tag{2}$$

$$\begin{aligned}
P(M_2) &= \frac{5500}{18800} \\
&= 0,292553
\end{aligned} \tag{3}$$

$$\begin{aligned}
P(M_3) &= \frac{2500}{18800} \\
&= 0,132978
\end{aligned} \tag{4}$$

$$\begin{aligned}
P(M_4) &= \frac{7300}{18800} \\
&= 0,388297
\end{aligned} \tag{5}$$

Probabilidades para C_i

En esta sección se presenta el cálculo realizado para poder definir las diferentes probabilidades a priori de encontrar la característica C_i en el total de placas madre y lo adjuntamos a la **Tabla 1**, estas probabilidades están adheridas a la **Tabla 2**.

El cálculo para las probabilidades a priori para las características las podemos calcular según [IA y SE - 2018], como presenta la ecuación 6

$$P(C_i) = \sum_{j=1}^N P(M_i)P(C_i|M_j) \tag{6}$$

Siendo N la cantidad de, en este caso, **marcas** se tenga en el inventario, para nuestro caso particular se tiene que $N = 4$

Otra forma de verlo, es la siguiente

$$P(C_i) = \frac{\sum_{j=1}^N \#C_i \text{ en } M_j}{\#Total} \tag{7}$$

conceptualmente, lo que expresa la ecuación 7 es que se deben sumar los cardinales para cada columna y este resultado dividirlo por el total de items.

Los calculos para las probabilidades de C_i se encuentran en los **Desarrollos 8, 9 y 10**

$$\begin{aligned}
P(C_1) &= \frac{3130 + 2325 + 273 + 135}{18800} \\
&= 0,311861
\end{aligned} \tag{8}$$

$$P(C_2) = \frac{502 + 5351 + 2312 + 7281}{18800} = 0,821595 \quad (9)$$

$$P(C_3) = \frac{1244 + 124 + 14 + 4214}{18800} = 0,297659 \quad (10)$$

Table 2: Probabilidades para C_i

	C1	C2	C3
M1	3130	502	1244
M2	2325	5351	124
M3	273	2312	14
M4	135	7281	4214
Probabilidad	0,311861	0,821595	0,297659

Cálculo de verosimilitudes

Luego de este cálculo se tendrá una nueva tabla la cual representará la verosimilitud de cada característica C_j dada una marca M_i , para realizar el cálculo se procede a realizar lo siguiente

$$P(C_j|M_i) = \frac{\#C_j}{\#M_i} \quad (11)$$

Esto es, la cardinalidad de un C_j sobre la cardinalidad de un M_i . A continuación se presenta la tabla nueva de la que se habló.

Table 3: Verosimilitudes

	C1	C2	C3
M1	0,894285	0,143428	0,355428
M2	0,422727	0,972909	0,022545
M3	0,109200	0,924800	0,005600
M4	0,018493	0,997397	0,577260

Cálculo de probabilidades condicionales

Este se llevará a cabo mediante la siguiente ecuación:

$$P(M_i|C_1, C_2, \dots, C_j) = P(M_i) \prod_{j=1}^N P(C_j|M_i) \quad (12)$$

La ecuación 12 se puede interpretar conceptualmente como, la probabilidad de que ocurra que se seleccione la marca M_i dado a que esta marca cuente con los features C_1, C_2, \dots, C_j .

$$\begin{aligned} P(M_1|C_1, C_2, C_3) &= 0,186170 \cdot 0,894285 \cdot 0,143428 \cdot 0,355428 \\ &= 0,008487 \end{aligned} \quad (13)$$

$$\begin{aligned} P(M_2|C_1, C_2, C_3) &= 0,292553 \cdot 0,422727 \cdot 0,972909 \cdot 0,022545 \\ &= 0,002712 \end{aligned} \quad (14)$$

$$\begin{aligned} P(M_3|C_1, C_2, C_3) &= 0,132978 \cdot 0,109200 \cdot 0,924800 \cdot 0,005600 \\ &= 0,000075 \end{aligned} \quad (15)$$

$$\begin{aligned} P(M_4|C_1, C_2, C_3) &= 0,388297 \cdot 0,018443 \cdot 0,997397 \cdot 0,577260 \\ &= 0,004123 \end{aligned} \quad (16)$$

Ahora si realizamos la sumatoria de todos los resultados obtenidos anteriormente, tendremos que, $\sum_{i=1}^{N=4} P(M_i|C_{1,2,3}) = 0,015397$ esto es un indicio de que la normalización es necesaria en el ejercicio, para esto, procedemos a dividir a cada resultado de $P(M_i|C_{1,2,3})$ por la suma obtenida anteriormente.

Finalizamos entonces con lo siguiente:

$$\frac{P(M_1|C_1, C_2, C_3)}{0,015397} = 0,551211 \quad (17)$$

$$\frac{P(M_2|C_1, C_2, C_3)}{0,015397} = 0,176138 \quad (18)$$

$$\frac{P(M_3|C_1, C_2, C_3)}{0,015397} = 0,004871 \quad (19)$$

$$\frac{P(M_4|C_1, C_2, C_3)}{0,015397} = 0,267779 \quad (20)$$

La sumatoria de estos nuevos valores normalizados es de 0,999999, lo cual basicamente tiende a 1, es decir, el 100%, el resultado que se aprecia es consecuencia de lo conocido como *error de truncamiento*.

Referencias

[IA y SE - 2018] INTELIGENCIA ARTIFICIAL Y SISTEMAS EXPERTOS. *Sistemas Basados en Probabilidades - Ejemplo. Clase 7.*