

# Métodos Bayesianos I

García Prado, Sergio  
 sergio@garciparedes.me

22 de abril de 2017

## Resumen

[TODO]

1. SE SABE QUE UN 1% DE LAS MUJERES DE 40 AÑOS QUE PARTICIPAN EN UN EXAMEN RUTINARIO TIENEN CÁNCER DE MAMA. TAMBIÉN SE SABE QUE UN 80% DE LAS MUJERES QUE TIENEN CÁNCER DE MAMA, DARÁN POSITIVO AL HACERSE UNA MAMOGRAFÍA. SIN EMBARGO, UN 9,6% DE LAS MUJERES QUE NO TIENEN CÁNCER DE MAMA DARÁN POSITIVO EN UNA MAMOGRAFÍA. EN ESTE CONTEXTO UNA MUJER DE 40 AÑOS SE SOMETE A UN EXAMEN RUTINARIO Y SU MAMOGRAFÍA DA POSITIVO. ¿CUÁL ES LA PROBABILIDAD DE QUE REALMENTE TENGA CÁNCER DE MAMA?

[TODO]

$$X = \text{Tener Cancer de mama} \rightarrow \{0, 1\} \quad (1)$$

$$Y = \text{Dar positivo al hacerse una mamografía} \rightarrow \{0, 1\} \quad (2)$$

$$Z = \text{Ser una mujer de 40 años y someterse a examen rutinario} \rightarrow \{0, 1\} \quad (3)$$

$$Pr(Y = 1|X = 1) = 0,800 \quad (4)$$

$$Pr(Y = 1|X = 0) = 0,096 \quad (5)$$

$$Pr(X = 1|Z = 1) = 0,010 \quad (6)$$

$$Pr(X = 1|Z = 1, Y = 1) = Pr(X = 1) \cdot Pr(Y = 1|X = 1) \cdot Pr(Z = 1|X = 1) \quad (7)$$

$$Pr(Y = 1|X = 1) = 0,800 \quad (8)$$

$$Pr(Y = 1|X = 0) = 0,096 \quad (9)$$

$$Pr(X = 1) = 0,010 \quad (10)$$

$$Pr(X = 1|Y = 1) = Pr(X = 1) \cdot Pr(Y = 1|X = 1) = 0,010 \cdot 0,800 = 0,008 \quad (11)$$

2. DADAS DOS VARIABLES ALEATORIAS DISCRETAS,  $X$  E  $Y$ , Y DADA SU DISTRIBUCIÓN DE PROBABILIDAD CONJUNTA QUE APARECE EN LA TABLA, SE PIDE:

- 2.1. ¿CUMPLE LA DISTRIBUCIÓN CONJUNTA LAS PROPIEDADES DE UNA DISTRIBUCIÓN DE PROBABILIDADES?

[TODO]

	$x_1$	$x_2$	$x_3$	$x_4$	$Pr(Y)$
$y_1$	2/16	1/16	1/16	1/16	5/16
$y_2$	1/16	2/16	2/16	1/16	6/16
$y_3$	1/16	1/16	1/16	0	3/16
$y_4$	0	2/16	0	0	2/16
$Pr(X)$	4/16	6/16	4/16	2/16	16/16

**Tabla 1:** Frecuencias relativas de la distribución de probabilidad conjunta de  $X$  e  $Y$ 

2.2. ¿CUÁL ES LA PROBABILIDAD DE  $Pr(X = x_1)$ ?

[TODO]

2.3. ¿CUÁLES SON LAS DISTRIBUCIONES MARGINALES DE  $Pr(X = x)$  Y  $Pr(Y = y)$ ?

[TODO]

2.4. ¿VERIFICAN LAS DISTRIBUCIONES MARGINALES LAS PROPIEDADES DE UNA DISTRIBUCIÓN DE PROBABILIDADES?

[TODO]

3. UTILIZANDO EL CONJUNTO DE DATOS *weather-nominal-practica* QUE SE PROPORCIONA, DETERMINAR LA CLASIFICACIÓN NAIVE BAYES DE LAS SIGUIENTES INSTANCIAS, UTILIZANDO LA ESTIMACIÓN DE MÁXIMA VEROSIMILITUD (FRECUENCIAL)

$$x_1 = \langle \text{sunny}, \text{cool}, \text{normal}, \text{false} \rangle \quad (12)$$

$$x_2 = \langle \text{overcast}, \text{mild}, \text{high}, \text{true} \rangle \quad (13)$$

$$Pr(\text{play} = \text{yes}) = 9/14 \quad (14)$$

$$Pr(\text{play} = \text{no}) = 5/14 \quad (15)$$

$$Pr(\text{outlook} = \text{sunny} | \text{play} = \text{yes}) = 2/9 \quad (16)$$

$$Pr(\text{outlook} = \text{overcast} | \text{play} = \text{yes}) = 2/9 \quad (17)$$

$$Pr(\text{outlook} = \text{rainy} | \text{play} = \text{yes}) = 5/9 \quad (18)$$

$$Pr(\text{outlook} = \text{sunny} | \text{play} = \text{no}) = 3/5 \quad (19)$$

$$Pr(\text{outlook} = \text{overcast} | \text{play} = \text{no}) = 0/5 \quad (20)$$

$$Pr(\text{outlook} = \text{rainy} | \text{play} = \text{no}) = 2/5 \quad (21)$$

$$Pr(\text{temperature} = \text{hot} | \text{play} = \text{yes}) = 1/9 \quad (22)$$

$$Pr(\text{temperature} = \text{mild} | \text{play} = \text{yes}) = 4/9 \quad (23)$$

$$Pr(\text{temperature} = \text{cool} | \text{play} = \text{yes}) = 4/9 \quad (24)$$

$$Pr(\text{temperature} = \text{hot} | \text{play} = \text{no}) = 2/5 \quad (25)$$

$$Pr(\text{temperature} = \text{mild} | \text{play} = \text{no}) = 2/5 \quad (26)$$

$$Pr(\text{temperature} = \text{cool} | \text{play} = \text{no}) = 1/5 \quad (27)$$

$$Pr(humidity = high | play = yes) = 3/9 \quad (28)$$

$$Pr(humidity = normal | play = yes) = 6/9 \quad (29)$$

$$Pr(humidity = high | play = no) = 4/5 \quad (30)$$

$$Pr(humidity = normal | play = no) = 1/5 \quad (31)$$

$$Pr(windy = true | play = yes) = 4/9 \quad (32)$$

$$Pr(windy = false | play = yes) = 5/9 \quad (33)$$

$$Pr(windy = true | play = no) = 3/5 \quad (34)$$

$$Pr(windy = false | play = no) = 2/5 \quad (35)$$

$$Pr(outlook = sunny, temperature = cool, humidity = normal, windy = false | play = yes) = \quad (36)$$

$$Pr(play = yes) \cdot Pr(outlook = sunny | play = yes) \cdot Pr(temperature = cool | play = yes) \cdot \quad (37)$$

$$Pr(humidity = normal | play = yes) \cdot Pr(windy = false | play = yes) = \quad (38)$$

$$9/14 \cdot 2/9 \cdot 4/9 \cdot 6/9 \cdot 5/9 = 40/1701 = 0,02351557 \quad (39)$$

$$Pr(outlook = sunny, temperature = cool, humidity = normal, windy = false | play = no) = \quad (40)$$

$$Pr(play = no) \cdot Pr(outlook = sunny | play = no) \cdot Pr(temperature = cool | play = no) \cdot \quad (41)$$

$$Pr(humidity = normal | play = no) \cdot Pr(windy = false | play = no) = \quad (42)$$

$$5/14 \cdot 3/5 \cdot 1/5 \cdot 1/5 \cdot 2/5 = 3/875 = 0,003428571 \quad (43)$$

$$Pr(outlook = overcast, temperature = mild, humidity = high, windy = true | play = yes) = \quad (44)$$

$$Pr(play = yes) \cdot Pr(outlook = overcast | play = yes) \cdot Pr(temperature = mild | play = yes) \cdot \quad (45)$$

$$Pr(humidity = high | play = yes) \cdot Pr(windy = true | play = yes) = \quad (46)$$

$$9/14 \cdot 2/9 \cdot 4/9 \cdot 3/9 \cdot 4/9 = 16/1701 = 0,00940623 \quad (47)$$

$$Pr(outlook = overcast, temperature = cool, humidity = high, windy = true | play = no) = \quad (48)$$

$$Pr(play = no) \cdot Pr(outlook = overcast | play = no) \cdot Pr(temperature = cool | play = no) \cdot \quad (49)$$

$$Pr(humidity = high | play = no) \cdot Pr(windy = true | play = no) = \quad (50)$$

$$5/14 \cdot 0/5 \cdot 2/5 \cdot 4/5 \cdot 4/5 = 0 \quad (51)$$

[TODO]

4. UTILIZANDO WEKA Y EL CLASIFICADOR NAIVEBAYES DETERMINAR LA CLASIFICACIÓN DE LOS EJEMPLOS ANTERIORES, ¿COINCIDE CON LA CLASIFICACIÓN CALCULADA EN EL EJERCICIO ANTERIOR?

[TODO]

5. ENTRENAR CON WEKA, UN CLASIFICADOR NAIVE BAYES PARA EL CONJUNTO DE DATOS *weather-nominal*

5.1. ESTIMAR LA TASA DE ERROR COMETIDA POR EL CLASIFICADOR UTILIZANDO VALIDACIÓN CRUZADA DE 10 PARTICIONES

[TODO]

5.2. EXAMINAR LA SALIDA PROPORCIONADA POR EL EXPLORER Y DETERMINAR CÓMO ESTÁ ESTIMANDO ESTA IMPLEMENTACIÓN DE *Naive Bayes* LOS PARÁMETROS DEL CLASIFICADOR

[TODO]

6. EL CONJUNTO DE DATOS *weather-nominal-X6* SE HA GENERADO REPITIENDO CADA INSTANCIA DEL CONJUNTO *weather-nominal* SEIS VECES. ENTRENAR CON WEKA UN CLASIFICADOR NAIVE BAYES PARA ESTE CONJUNTO DE DATOS:

6.1. ESTIMAR LA TASA DE ERROR COMETIDA POR EL CLASIFICADOR UTILIZANDO VALIDACIÓN CRUZADA DE 10 PARTICIONES

[TODO]

6.2. COMPARE ESTA TASA DE ERROR CON LA ESTIMADA EN EL EJERCICIO ANTERIOR Y DISCUTA LOS RESULTADOS

[TODO]

REFERENCIAS

[CCAG17] Teodoro Calonge Cano and Carlos Javier Alonso González. Técnicas de Aprendizaje Automático, 2016/17.

[GP17] Sergio García Prado. Métodos bayesianos 1. <https://github.com/garciparedes/machine-learning-bayesian-1>, 2017.

[too] Weka. <http://www.cs.waikato.ac.nz/ml/weka/>.