

ENTREGA 7.

PRÁCTICA METODOS BAYESIANOS I.

Patricia Aguado Labrador

Ejercicio 1:

- a) Sí, ya que la probabilidad de un par (x,y) es ≥ 0
 b) $P(X = x_1) = 1/4 = 0.25$
 c)

	X1	X2	X3	X4	P(Y)
Y1	2/16	1/16	1/16	1/16	5/16
Y2	1/16	2/16	2/16	1/16	3/8
Y3	1/16	1/16	1/16	0	3/16
Y4	0	2/16	0	0	2/16
P(X)	1/4	3/8	1/4	2/16	1

- d) Sí, ya que $\sum_{(x \in X)} \sum_{(y \in Y)} P(x,y) = 1$

Ejercicio 2:

Datos weather.nominal.practica:

Outlook	Temperature	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	cool	high	true	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
rainy	mild	high	true	yes
rainy	cool	normal	false	yes
rainy	mild	high	true	no

A. X1 = <sunny, cool, normal, false>

$$P'(\text{Play} = \text{no}) = 5/14 = 0.357$$

$$P'(\text{Play} = \text{yes}) = 9/14 = 0.643$$

$$P'(\text{Outlook} = \text{sunny} \mid \text{Play} = \text{no}) = 3/5 = 0.6$$

$$P'(\text{Outlook} = \text{sunny} \mid \text{Play} = \text{yes}) = 2/9 = 0.22$$

$$P'(\text{Temperature} = \text{cool} \mid \text{Play} = \text{no}) = 1/5 = 0.2$$

$$P'(\text{Temperature} = \text{cool} \mid \text{Play} = \text{yes}) = 5/9 = 0.55$$

$$P'(\text{Humidity} = \text{normal} \mid \text{Play} = \text{no}) = 1/5 = 0.2$$

$$P'(\text{Humidity} = \text{normal} \mid \text{Play} = \text{yes}) = 6/9 = 0.66$$

$$P'(\text{Windy} = \text{false} \mid \text{Play} = \text{no}) = 2/5 = 0.4$$

$$P'(\text{Windy} = \text{false} \mid \text{Play} = \text{yes}) = 5/9 = 0.55$$

$$P'(\text{Play} = \text{no} \mid \langle \text{sunny, cool, normal, false} \rangle) = \\ P'(\text{no}) * P'(\text{sunny} \mid \text{no}) * P'(\text{cool} \mid \text{no}) * P'(\text{normal} \mid \text{no}) * P'(\text{false} \mid \text{no}) = 0.0034 \\ (0.1075)$$

$$P'(\text{Play} = \text{yes} \mid \langle \text{sunny, cool, normal, false} \rangle) = \\ P'(\text{yes}) * P'(\text{sunny} \mid \text{yes}) * P'(\text{cool} \mid \text{yes}) * P'(\text{normal} \mid \text{yes}) * P'(\text{false} \mid \text{yes}) = \\ 0.0282 (0.8924)$$

$V_{NB} = \text{yes}$,ya que tiene una probabilidad del 89.24%

B. $X2 = \langle \text{overcast, hot, high, true} \rangle$

$$P'(\text{Play} = \text{no}) = 5/14 = 0.357$$

$$P'(\text{Play} = \text{yes}) = 9/14 = 0.643$$

$$P'(\text{Outlook} = \text{overcast} \mid \text{Play} = \text{no}) = 0/5 = 0$$

$$P'(\text{Outlook} = \text{overcast} \mid \text{Play} = \text{yes}) = 2/9 = 0.22$$

$$P'(\text{Temperature} = \text{hot} \mid \text{Play} = \text{no}) = 2/5 = 0.4$$

$$P'(\text{Temperature} = \text{hot} \mid \text{Play} = \text{yes}) = 0/9 = 0$$

$$P'(\text{Humidity} = \text{high} \mid \text{Play} = \text{no}) = 4/5 = 0.8$$

$$P'(\text{Humidity} = \text{high} \mid \text{Play} = \text{yes}) = 3/9 = 0.33$$

$$P'(\text{Windy} = \text{true} \mid \text{Play} = \text{no}) = 3/5 = 0.6$$

$$P'(\text{Windy} = \text{true} \mid \text{Play} = \text{yes}) = 4/9 = 0.44$$

$$P'(\text{Play} = \text{no} \mid \langle \text{overcast, hot, high, true} \rangle) =$$

$$P'(\text{no}) * P'(\text{overcast} \mid \text{no}) * P'(\text{hot} \mid \text{no}) * P'(\text{high} \mid \text{no}) * P'(\text{true} \mid \text{no}) = 0$$

$$P'(\text{Play} = \text{yes} \mid \langle \text{overcast, hot, high, true} \rangle) =$$

$$P'(\text{yes}) * P'(\text{overcast} \mid \text{yes}) * P'(\text{hot} \mid \text{yes}) * P'(\text{high} \mid \text{yes}) * P'(\text{true} \mid \text{yes}) = 0$$

La estimación de la probabilidad para los dos valores de la clase nos da 0, ya que no tenemos ningún ejemplo para el atributo Outlook con la clase no, y para el atributo Temperature con la clase yes.

En el caso de realizar una corrección de muestreo y dejar a un lado la probabilidad nula de esos atributos con el valor correspondiente de clase, podríamos clasificar la instancia como no, ya que tendría una mayor estimación de la probabilidad (no = 0.0685 y yes = 0.0205).

Ejercicio 3:

Creamos un clasificador Naive Bayes y entrenamos con las intencias del archivo weather.nominal.practica. Obtenemos los siguientes resultados para las instancias del archivo weather.nominal.prueba que utilizamos para realizar el test :

Predice yes para **$\langle \text{sunny, cool, normal, false, no} \rangle$**

Predice no para **$\langle \text{overcast, hot, high, true, yes} \rangle$**

```
=== Confusion Matrix ===
```

```
a b    <-- classified as
0 1 | a = yes
1 0 | b = no
```

a) Observando como están clasificadas las instancias del conjunto de prueba y la matriz de confusión, vemos que los resultados de Weka coinciden con lo calculado en el ejercicio anterior.

Ejercicio 4:

a) La tasa de error que comete el clasificador Naive Bayes utilizando validación cruzada de 10 particiones es de 0.4286 (42.86%).

b) Naive Bayes Classifier

Attribute	Class	
	yes (0.63)	no (0.38)
=====		
outlook		
sunny	3.0	4.0
overcast	3.0	1.0
rainy	6.0	3.0
[total]	12.0	8.0
temperature		
hot	1.0	3.0
mild	5.0	3.0
cool	6.0	2.0
[total]	12.0	8.0
humidity		
high	4.0	5.0
normal	7.0	2.0
[total]	11.0	7.0
windy		
TRUE	5.0	4.0
FALSE	6.0	3.0
[total]	11.0	7.0

Weka utiliza la estimación Bayesiana con el estimador de Laplace para el cálculo de los parámetros del clasificador Naive Bayes.

$$p'(a_j | v_j) = \frac{n_c + mp}{n + m} \quad p = \frac{1}{val(a_i)} \text{ y } m = Val(a_i) \text{ (Estimador de Laplace)}$$

Ejercicio 5:

a) La tasa de error que comete el clasificador Naive Bayes utilizando validación cruzada de 10 particiones es de 0.1548 (15.48%).

b) La tasa de error disminuye considerablemente en comparación con el ejercicio anterior, por ello podemos decir que Naive Bayes funciona mejor cuánto mayor sea el espacio de instancias.