

## Práctica 7 - TAA

*Pablo Martín de Benito*

### Índice

<b>1</b>	<b>Ejercicio 1</b>	<b>1</b>
<b>2</b>	<b>Ejercicio 2</b>	<b>2</b>
2.1	a) . . . . .	3
2.2	b) . . . . .	3
<b>3</b>	<b>Ejercicio 3</b>	<b>4</b>
<b>4</b>	<b>Ejercicio 4</b>	<b>5</b>
4.1	Estimar la tasa de error cometida por el clasificador utilizando validación cruzada de 10 particiones .	5
4.2	Examinar la salida proporcionada por el Explorer y determinar cómo está estimando esta implementación de Naive Bayes los parámetros del clasificador. . . . .	5
<b>5</b>	<b>Ejercicio 5</b>	<b>6</b>
5.1	Estimar la tasa de error cometida por el clasificador utilizando validación cruzada de 10 particiones.	6
5.2	Compare esta tasa de error con la estimada en el ejercicio anterior y discuta los resultados. . . . .	6

## 1 Ejercicio 1

	$X_1$	$X_2$	$X_3$	$X_4$	$P(Y)$
$Y_1$	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
$P(X)$	4/16	6/16	4/16	2/16	1

- a. ¿Cumple la distribución conjunta las propiedades de una distribución de probabilidades?

Si la cumple porque todas las probabilidades suman uno y están en el intervalo  $[0,1]$ .

- c. ¿Cuál es la probabilidad de  $P(X = x)$  y  $P(Y = y)$ ?

$$P(X = x) = 4/16X_1 + 6/16X_2 + 4/16X_3 + 2/16X_4$$

$$P(Y = y) = 5/16Y_1 + 6/16Y_2 + 3/16Y_3 + 2/16Y_4$$

- b. ¿Cuál es la probabilidad de  $P(X = x_1)$ ?

$$P(X = x_1) = 4/16X_1$$

- d. ¿Verifican las distribuciones marginales las propiedades de una distribución de probabilidades?

Si se verifica porque las probabilidades suman uno.

## 2 Ejercicio 2

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) y sin utilizar ninguna herramienta de minería de datos:

Obtenemos las siguientes probabilidades del archivo de datos, que serán de utilidad más adelante.

$$P(\textit{play} = \textit{yes}) = 9/14$$

$$P(\textit{play} = \textit{no}) = 5/14$$

$$P(\textit{outlook} = \textit{sunny}/\textit{play} = \textit{yes}) = 2/9$$

$$P(\textit{outlook} = \textit{sunny}/\textit{play} = \textit{no}) = 3/5$$

$$P(\textit{outlook} = \textit{overcast}/\textit{play} = \textit{yes}) = 2/9$$

$$P(\textit{outlook} = \textit{overcast}/\textit{play} = \textit{no}) = 0 \rightarrow \textit{EstimacionBayesiana} = \frac{1}{5+3} = 1/8$$

$$P(\textit{temperature} = \textit{cool}/\textit{play} = \textit{yes}) = 5/9$$

$$P(\textit{temperature} = \textit{cool}/\textit{play} = \textit{no}) = 1/5$$

$$P(\textit{temperature} = \textit{hot}/\textit{play} = \textit{yes}) = 0 \rightarrow \textit{EstimacionBayesiana} = \frac{1}{9+3} = 1/12$$

$$P(\textit{temperature} = \textit{hot}/\textit{play} = \textit{no}) = 1/5$$

$$P(\textit{humidity} = \textit{normal}/\textit{play} = \textit{yes}) = 6/9$$

$$P(\textit{humidity} = \textit{normal}/\textit{play} = \textit{no}) = 1/5$$

$$P(\textit{humidity} = \textit{high}/\textit{play} = \textit{yes}) = 3/9$$

$$P(\textit{humidity} = \textit{high}/\textit{play} = \textit{no}) = 4/5$$

$$P(\textit{windy} = \textit{FALSE}/\textit{play} = \textit{yes}) = 5/9$$

$$P(\textit{windy} = \textit{FALSE}/\textit{play} = \textit{no}) = 2/5$$

$$P(\textit{windy} = \textit{TRUE}/\textit{play} = \textit{yes}) = 3/9$$

$$P(\textit{windy} = \textit{TRUE}/\textit{play} = \textit{no}) = 3/5$$

## 2.1 a)

$$X_1 < \textit{sunny, cool, normal, false} >$$

$$P(\textit{play} = \textit{yes} | < \textit{sunny, cool, normal, false} >) = p(\textit{yes})p(\textit{sunny/yes})p(\textit{cool/yes})p(\textit{normal/yes})p(\textit{false/yes}) =$$

$$9/14 * 2/9 * 5/9 * 6/9 = 0.05291005291$$

$$P(\textit{play} = \textit{no} | < \textit{sunny, cool, normal, false} >) = p(\textit{no})p(\textit{sunny/no})p(\textit{cool/no})p(\textit{normal/no})p(\textit{false/no}) =$$

$$5/14 * 3/5 * 1/5 * 1/5 * 2/5 = 0.00342857142$$

## 2.2 b)

$$X_2 < \textit{overcast, hot, high, true} >$$

$$P(\textit{play} = \textit{yes} | < \textit{overcast, hot, high, true} >) = p(\textit{yes})p(\textit{ovecast/yes})p(\textit{hot/yes})p(\textit{high/yes})p(\textit{true, yes}) =$$

$$9/14 * 2/9 * 1/12 * 3/9 * 3/9 = 0.00132275132$$

$$P(\textit{play} = \textit{no} | < \textit{overcast, hot, high, true} >) = p(\textit{no})p(\textit{ovecast/no})p(\textit{hot/no})p(\textit{high/no})p(\textit{true, no}) =$$

$$5/14 * 1/8 * 1/5 * 4/5 * 3/5 = 0.00428571428$$

### 3 Ejercicio 3

Utilizando Weka y el clasificador Naive Bayes determinar la clasificación de los ejemplos anteriores:

Al utilizar el clasificador Naive Bayes con Weka obtenemos los siguientes resultado:

```
=== Classifier model (full training set) ===
```

Naive Bayes Classifier

	Class	
Attribute yes	no	
	(0.63)	(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

Como podemos ver, las estimaciones de las probabilidades no son las mismas que en el ejercicio anterior, esto es porque Weka, al detectar que una probabilidad requiere del cálculo de la estimación bayesiana, la utiliza para todas las probabilidades.

Correctly Classified Instances	0	0%
--------------------------------	---	----

Incorrectly Classified Instances	2	100%
----------------------------------	---	------

Donde podemos darnos cuenta de que no las clasifica bien dichas instancias.

## 4 Ejercicio 4

Entrenar, con Weka, un clasificador Naive Bayes para el conjunto de datos `weather.nominal.practica`

### 4.1 Estimar la tasa de error cometida por el clasificador utilizando validación cruzada de 10 particiones

Incorrectly Classified Instances      6      42.8571 %

### 4.2 Examinar la salida proporcionada por el Explorer y determinar cómo está estimando esta implementación de Naive Bayes los parámetros del clasificador.

Como hemos explicado en el ejercicio anterior, la estimación de las probabilidades que efectúa Naive Bayes es la estimación Bayesiana que viene dada por la siguiente fórmula:

$$p'(a_i|v_j) = \frac{n_c + mp}{n + m}$$

n: número de ejemplos de entrenamiento con clase  $v_j$

$n_c$ : nº ejemplos clase  $v_j$  con valor  $a_i$  para el atributo a.

p: estimación a priori de  $p'(a_i|v_j)$

m: peso de la estimación a priori (nº de ejemplos virtuales)

## 5 Ejercicio 5

El conjunto de datos `weather.nominalX6` se ha generado repitiendo cada instancia del conjunto `weather.nominal.practica` seis veces. Entrenar con Weka un clasificador Naive Bayes para este conjunto de datos:

### 5.1 Estimar la tasa de error cometida por el clasificador utilizando validación cruzada de 10 particiones.

Incorrectly Classified Instances	13	15.4762 %
----------------------------------	----	-----------

### 5.2 Compare esta tasa de error con la estimada en el ejercicio anterior y discuta los resultados.

Al aumentar el número de instancias, como es evidente, obtenemos una menor tasa de error y por lo tanto un mejor clasificador