Práctica 7: Métodos Bayesianos

Alejandro Pulido Sánchez

1. *Dadas dos variables aleatorias discretas, X e Y, y dada su distribución de probabilidad conjunta que aparece en la tabla, se pide:*

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

   Para comprobar si la distribución es conjunta se deben cumplir 2 premisas. La primera es que la suma de todas las probabilidades de los posibles resultados sea igual a 1.

   (2/16+1/16+1/16+1/16+1/16+2/16+2/16+1/16+1/16+1/16+1/16+0+0+2/16+0+0) = 16/16=1

   La segunda es que todas las probabilidades sean no negativas. En este caso todas las probabilidades son mayores o igual que 0.

   b. *¿Cuál es la probabilidad de $P(X = x1)$?*

   Sumamos todas las probabilidades asociadas a X1:
   (2/16+1/16+1/16+0) = 4/16 = 1/4

   c. *¿Cuáles son las distribuciones marginales de cada $P(X = x)$ y $P(Y = y)$?*

   Para calcular las distribuciones marginales de cada variable aleatoria, podemos sumar las probabilidades en la tabla sobre las filas para la distribución marginal de $P(X = x)$ y sobre las columnas para la distribución marginal de $P(Y = y)$.

X/Y	**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	6/16
y3	1/16	1/16	1/16	0	3/16
y4	0	2/16	0	0	2/16
P(X)	4/16	6/16	4/16	2/16	1

La distribución marginal de $P(X = x)$ es:

- $P(X =$ x1) = 4/16 = 1/4
- $P(X =$ x2) = 6/16 = 3/8
- $P(X =$ x3) = 4/16 = 1/4
- $P(X =$ x4) = 2/16 = 1/8

La distribución marginal de $P(Y = y)$ es:

- $P(Y =$ y1) = 5/16
- $P(Y =$ y2) = 6/16 = 3/8
- $P(Y =$ y3) = 3/16
- $P(Y =$ y4) = 2/16 = 1/8

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

Podemos verificar si se cumplen estas dos propiedades con las distribuciones marginales calculadas en la pregunta anterior.

Para la propiedad 1, debemos sumar todas las probabilidades de cada distribución marginal y asegurarnos de que sumen 1:

- Para $P(X = x)$: 1/4 + 3/8 + 1/4 + 1/8 = 1
- Para $P(Y = y)$: 5/16 + 3/8 + 3/16 + 1/8 = 1

Entonces, ambas distribuciones marginales cumplen la propiedad 1 y la 2.

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:*
   a. $X1$ = < sunny, cool, normal, false >
   b. $X2$ = < overcast, hot, high, true >

Para obtener la clasificación de la clase mediante naive bayes nos valdremos de la siguiente formula:

$$v_{NB} = \arg\max_{v_j \in V} p'(v_j) \prod_i p'(a_i / v_j)$$

Se inicia el proceso de clasificación para la instancia X1 mediante el cálculo de vNB, el cual se obtiene al maximizar la probabilidad condicional p' de la variable objetivo (play) dada la instancia de los atributos <sunny, cool, normal, false> utilizando el clasificador Naive Bayes. Para esto, se estima la probabilidad de cada atributo condicionado a la variable objetivo (play) utilizando la frecuencia de ocurrencia de cada valor en el conjunto de entrenamiento.

p ’ (play = yes) = 9/14 = 0.64
p ’ (play = no) = 5/14 = 0.36

p ’ (outlook = sunny / play = yes) = 2/9 = 0.22
p ’ (outlook = sunny / play = no) = 3/5 = 0.6
p ’ (temperature = cool / play = yes) = 5/9 = 0.55
p ’ (temperature = cool / play = no) = 1/5 = 0.2
p ’ (humidity = normal / play = yes) = 6/9 = 0.66
p ’ (humidity = normal / play = no) = 1/5 = 0.2
p ’ (windy = false / play = yes) = 5/9 = 0.55
p ’ (windy = false / play = no) = 2/5 = 0.4

Clasificación:
p ’ (play = yes | < sunny, cool, normal, false >) :
p ’ (yes) p’ (sunny / yes) p’ (cool / yes) p’ (normal / yes) p’ (false / yes) = **0.0294**

p ’ (play = no | < sunny, cool, ormal, alse n f >) :
p ’ (no) p’ (sunny / no) p’ (cool / no) p’ ($n$ormal / no) p’ ($f$alse / no) = **0.0034**

La clasificación resultante muestra que la instancia **X1** pertenece a la clase "**yes**", ya que la estimación de la probabilidad para esta clase es mayor que para la clase "no".

Posteriormente, se repite el proceso para la instancia X2, en la cual se utiliza el atributo outlook con el valor overcast, el atributo temperature con el valor hot, el atributo humidity con el valor high y el atributo windy con el valor true. En este caso, al no tener ejemplos de entrenamiento para los atributos outlook con la clase no y temperature con la clase yes, la estimación de probabilidad para ambas clases será nula. Aun así procedemos a calcurlas.

p ’ (play = yes) = 9/14 = 0.64
p ’ (play = no) = 5/14 = 0.36

p ’ (outlook = overcast / play = yes) = 2/9 = 0.22
p ’ (outlook = overcast / play = no) = 0/5 = 0
p ’ (temperature = hot / play = yes) = 0/9 = 0
p ’ (temperature = hot / play = no) = 2/5 = 0.4
p ’ (humidity = high / play = yes) = 3/9 = 0.33
p ’ (humidity = high / play = no) = 4/5 = 0.8
p ’ (windy = true / play = yes) = 4/9 = 0.4
p ’ (windy = true / play = no) = 3/5 = 0.6

Clasificación:
p ’ (play = yes | < overcast, hot, high, true >) :
p ’ (yes) p’ (overcast / yes) p’ (hot / yes) p’ (high / yes) p’ (true / yes) = **0**

p ’ (play = no | < overcast, hot, high, true >) :
p ’ (no) p’ (overcast / no) p’ (hot / no) p’ (high / no) p’ (true / no) = **0**

Al aplicar una corrección de muestreo, se puede obtener una clasificación para la instancia X2. En este caso, la corrección de muestreo aumenta la estimación de la probabilidad de la clase "no" y se obtiene que la instancia X2 pertenece a la clase "no" en lugar de la clase "yes", debido a que la estimación de la probabilidad para la clase "no" es mayor que para la clase "yes" después de la corrección de muestreo (**yes = 0.0205, no = 0.0685**).

3. *Utilizando Weka y el clasificador NaiveBayes determinar la clasificación de los ejemplos anteriores:*

   a. *¿Coindice con la clasificación calculada en el ejercicio anterior?*

   Como vemos en la siguiente imagen, la clasificación dada es inversa a la obtenida. Es decir, para la primera clasificación he obtenido el valor **yes** de la clase **play**, sin embargo, en los **datos** se ha clasificado la instancia como **no**. Lo mismo sucede con la segunda clasificación.

```
@data

sunny,cool,normal,FALSE,no
overcast,hot,high,TRUE,yes
```

   Por tanto, si Weka clasifica erróneamente ambas instancias, querrá decir que obtiene el mismo resultado que hemos obtenido nosotros en el apartado anterior.

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

```
=== Summary ===

Correctly Classified Instances           0               0      %
Incorrectly Classified Instances         2             100      %
Kappa statistic                         -1
Mean absolute error                      0.7973
Root mean squared error                  0.7978
Relative absolute error                159.4562 %
Root relative squared error            154.7908 %
Total Number of Instances                2
```

Se puede observar que ambas instancias han sido clasificadas erróneamente y, por tanto, la clasificación obtenida por Weka es la misma que hemos obtenido nosotros.

4. *Entrenar, con Weka, un clasificador Naive Bayes para el conjunto de datos weather.nominal.practica*
   a. *Estimar la tasa de error cometida por el clasificador utilizando validación cruzada de 10 particiones.*

```
=== Summary ===

Correctly Classified Instances           8               57.1429 %
Incorrectly Classified Instances         6               42.8571 %
Kappa statistic                          0.0667
Mean absolute error                      0.3939
Root mean squared error                  0.4767
Relative absolute error                 82.7157 %
Root relative squared error             96.6215 %
Total Number of Instances               14
```

Como se puede observar, el clasificador Naive Bayes cometió una tasa de error del **42.8571%** al ser evaluado mediante validación cruzada de 10 particiones.

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

La implementación del algoritmo Naive Bayes Simple en WEKA sigue los mismos pasos que la implementación manual descrita en los apuntes. En primer lugar, se calcula la probabilidad a priori de cada valor posible de la clase objetivo en el conjunto de datos de entrenamiento. A continuación, se calculan las probabilidades condicionales de cada valor de los atributos en relación a las diferentes clases objetivo. Para clasificar las instancias, el algoritmo asume la independencia entre los atributos y calcula la probabilidad conjunta de todos los atributos de la instancia en función de las diferentes clases objetivo, y luego clasifica la instancia en la clase que tenga la probabilidad más alta.

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:*
    a. *Estimar la tasa de error cometida por el clasificador utilizando validación cruzada de 10 particiones.*

```
=== Summary ===

Correctly Classified Instances          71               84.5238 %
Incorrectly Classified Instances        13               15.4762 %
Kappa statistic                          0.6553
Mean absolute error                      0.254
Root mean squared error                  0.3689
Relative absolute error                 55.1685 %
Root relative squared error             76.9784 %
Total Number of Instances               84
```

Como se puede observar, el clasificador Naive Bayes cometió una tasa de error del **84.5238%** al ser evaluado mediante validación cruzada de 10 particiones.

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

Al repetir instancias obtenemos una tasa de error menor pues se refuerza la información dada en los datos de entrenamiento. Esto puede producir un sobreajuste ya que se está impulsando el aprendizaje memorístico. Es decir, se ve mejorada la clasificación de datos con los que ya ha sido entrenado, pero perjudica gravemente a la generalización del modelo.