

PRÁCTICA 7 EXTENSIÓN

NAIVE BAYES EXTENSION



TÉCNICAS DE APRENDIZAJE AUTOMÁTICO
JAVIER ABAD HERNÁNDEZ

1. Considerar el conjunto de datos weather.nominal.practica proporcionado y la siguiente red TAN:

Obtener los parámetros de la red para el nodo Windy utilizando la estimación de máxima

verosimilitud y la corrección de Laplace con $m = 0,5$ (caso particular de la estimación Bayesiana). Hacer el ejercicio sin utilizar software de aprendizaje.

$$\hat{\theta}_{x_i^j|u} = \frac{M[u, x_i^j] + mp}{M[u] + m}$$

Con $M[u]$ el número de ejemplos del conjunto de entrenamiento con $pa_{X_i} = u$

Con $M[u, x_i^j]$ el número de ejemplos de $M[u]$ con $X_i = x_i^j$

$p = 0,5$: Windy unicamente toma dos valores, TRUE o FALSE.

$m = 0,5$.

Cálculos:

1. (Sunny, yes) → Total de instancias: 2 ($M[u]$)
De esas 2 instancias, Windy toma el valor TRUE en 1 y FALSE en 1.
Para TRUE → $(1 + 0.5 * 0.5) / (2 + 0.5) = 0.5$
Para FALSE → $(1 + 0.5 * 0.5) / (2 + 0.5) = 0.5$
2. (Overcast, yes) → Total de instancias: 2 ($M[u]$)
De esas 2 instancias, Windy toma el valor TRUE en 2 y FALSE en 0.
Para TRUE → $(2 + 0.5 * 0.5) / (2 + 0.5) = 0.9$
Para FALSE → $(0 + 0.5 * 0.5) / (2 + 0.5) = 0.1$
3. (Rainy, yes) → Total de instancias: 5 ($M[u]$)
De esas 5 instancias, Windy toma el valor TRUE en 1 y FALSE en 4.
Para TRUE → $(1 + 0.5 * 0.5) / (5 + 0.5) = 0.2272$
Para FALSE → $(4 + 0.5 * 0.5) / (5 + 0.5) = 0.7727$
4. (Sunny, no) → Total de instancias: 3 ($M[u]$)
De esas 3 instancias, Windy toma el valor TRUE en 1 y FALSE en 2.
Para TRUE → $(1 + 0.5 * 0.5) / (3 + 0.5) = 0.3571$
Para FALSE → $(2 + 0.5 * 0.5) / (3 + 0.5) = 0.6428$
5. (Overcast, no) → Total de instancias: 0 ($M[u]$)
Para TRUE → $(0 + 0.5 * 0.5) / (0 + 0.5) = 0.5$
Para FALSE → $(0 + 0.5 * 0.5) / (0 + 0.5) = 0.5$
6. (Rainy, no) → Total de instancias: 2 ($M[u]$)
De esas 2 instancias, Windy toma el valor TRUE en 2 y FALSE en 0.

Para TRUE $\rightarrow (2 + 0.5 * 0.5) / (2 + 0.5) = 0.9$

Para FALSE $\rightarrow (0 + 0.5 * 0.5) / (2 + 0.5) = 0.1$

P(W O,P)	(sunny,yes)	(overcast,yes)	(rainy,yes)	(sunny,no)	(overcast,no)	(rainy,no)
TRUE	0.5	0.9	0.3571	0.2272	0.5	0.9
FALSE	0.5	0.1	0.6428	0.7727	0.5	0.1

2. Considerar el conjunto de datos weather.nominal.practica proporcionado y Entrenar con Weka una red TAN. Comparar ambas redes. Si la estructura es la misma que la proporcionada en el ejercicio anterior, comparar también los parámetros de la red para el nodo Windy.

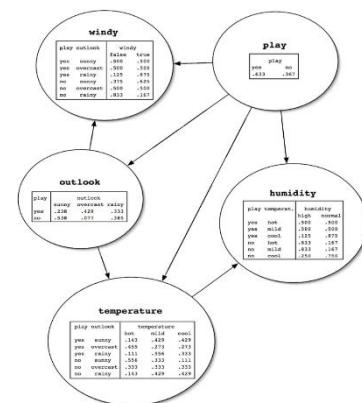
El resultado en todos los casos es el mismo, con la clase play, al no tener en cuenta la dirección, vemos que en windy y Outlook cambian respecto a la representada en el enunciado, ya que cambian los padres. Pero el resto de tablas se mantienen exactamente igual.

La tabla para windy en weka es la siguiente:

play	TRUE	FALSE
yes	0,45	0,55
no	0,583	0,417

La tabla para outlook en weka es la siguiente:

play	windy	sunny	overcast	rainy
yes	TRUE	0,273	0,455	0,273
yes	FAL...	0,231	0,077	0,692
no	TRUE	0,333	0,111	0,556
no	FAL...	0,714	0,143	0,143



Como podemos ver, como he mencionado anteriormente, se ha intercambiado la estructura de las tablas entre windy y outlook y también, al tener distintos padres, cambian los valores.

3. Considerar la estructura la red Bayesiana Credit, en la que todas las variables son discretas:

Clasificación con redes Bayesianas (Tasa error %)				
	Naive Bayes		TAN	
N.º Instancias (X)	100	10000	100	10000
Datos_Credit_X	36%	29,71%	35%	28,27%
Test_Credit_1000	30%	27,5%	33,3%	25%

La tasa de error de ambos algoritmos de clasificación disminuye a medida que aumenta el tamaño del conjunto de datos de entrenamiento.

En cuanto a los resultados, ambos algoritmos presentan tasas de error similares y no se puede determinar con certeza si uno es mejor que el otro en los conjuntos de datos utilizados debido a pequeñas variaciones en las tasas de error.