



UNIVERSIDAD DE
GUADALAJARA
Red Universitaria de Jalisco

CUCEI

Minería de datos

DR. ISRAEL ROMÁN GODÍNEZ

ISRAEL.ROMAN@CUCEI.UDG.MX

Algoritmos de aprendizaje

Algoritmos de aprendizaje automático

- ❖ Zero-R (Clasificación)
- ❖ One-R (Clasificación)
- ❖ Naïve Bayes (Clasificación)
- ❖ k-NN (Clasificación y regresión)
- ❖ K-means (Agrupamiento)
- ❖ A priori (Asociación)

Zero-R

CLASIFICACIÓN

Contexto

- ❖ Holte, en 1993, mostró que “Los algoritmos simples trabajan bien” en un estudio de desempeño de diferentes algoritmos con diferentes complejidades.
- ❖ Existen muchas estructuras simples:
 - Un solo atributo hace todo el trabajo.
 - Todos los atributos contribuyen independientes y de igual forma.
 - Una combinación lineal podría servir.
 - Simples reglas lógicas funcionan.
- ❖ El éxito del método depende del dominio.

Contexto

- ❖ **No es un modelo predictivo** ya que todo el peso de la decisión recae en la clase y no en los atributos o características del conjunto de ejemplos.
- ❖ Se utiliza, generalmente, para **establecer una línea base**. Es decir, un valor de clasificación mínimo que tienen que cumplir los algoritmos para ser considerados.

Ejemplo

id	Outlook	Temp	Humidity	Windy	Play
1	Sunny	Hot	High	False	No
2	Sunny	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	Rainy	Mild	High	False	Yes
5	Rainy	Cool	Normal	False	Yes
6	Rainy	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes
8	Sunny	Mild	High	False	No
9	Sunny	Cool	Normal	False	Yes
10	Rainy	Mild	Normal	False	Yes
11	Sunny	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Rainy	Mild	High	True	No

Algoritmo

Construye una *tabla de frecuencias* y escoge, como la clase, el valor más repetido.

Descripción del modelo

PLAY	
SI	NO
9	5

One-R

CLASIFICACIÓN

Contexto

- ❖ One-R es por “One Rule” (una regla). Es simple y con buen desempeño.
- ❖ Se basa en la premisa de que las estructuras de los datos representan, en general, relaciones simples.
- ❖ El éxito del algoritmo depende fuertemente del dominio.

Algoritmo

- ❖ Genera un conjunto de reglas por cada atributo o predictor en los datos, luego selecciona el conjunto con el error más pequeño.
- ❖ Para construir una regla para el atributo, construimos una tabla de frecuencia para cada atributo contra los valores de las clases objetivo.

Algoritmo

1. Para cada atributo:

1. Para cada valor del atributo, construye reglas como sigue:

1. Calcula la tabla frecuencia para cada valor de clase posible.
2. Encuentra la relación de atributo-clase más frecuente.
 1. Construye reglas que asignen la clases más frecuente a un valor dado en el atributo.
3. Calcula el error total de las reglas para cada atributo.

2. Escoge el atributo con el error total más pequeño

Ejemplo

id	Outlook	Temp	Humidity	Windy	Play
1	Sunny	Hot	High	False	No
2	Sunny	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	Rainy	Mild	High	False	Yes
5	Rainy	Cool	Normal	False	Yes
6	Rainy	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes
8	Sunny	Mild	High	False	No
9	Sunny	Cool	Normal	False	Yes
10	Rainy	Mild	Normal	False	Yes
11	Sunny	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Rainy	Mild	High	True	No

Tablas de frecuencia

		Play	
		Yes	No
Outlook	Sunny	2	3
	Overcast	4	0
	Rainy	3	2

		Play	
		Yes	No
Humidity	High	3	4
	Normal	6	1

		Play	
		Yes	No
Temp	Hot	2	2
	Mild	4	2
	Cool	3	1

		Play	
		Yes	No
Windy	False	6	2
	True	3	3

Reglas

Atributo	Reglas	Errores	Error Total
Outlook	Sunny -> No	2/5	4/14
	Overcast -> Yes	0/4	
	Rainy -> Yes	2/5	
Temp	Hot -> No*	2/4	5/14
	Mild -> Yes	2/6	
	Cool -> Yes	1/4	
Humidity	High -> No	3/7	4/14
	Normal -> Yes	1/7	
Windy	False -> Yes	2/8	5/14
	True -> No*	3/6	

Descripción del modelo

Outlook	Sunny → No
---------	------------

	Overcast → Yes
--	----------------

	Rainy → Yes
--	-------------

Orange
