

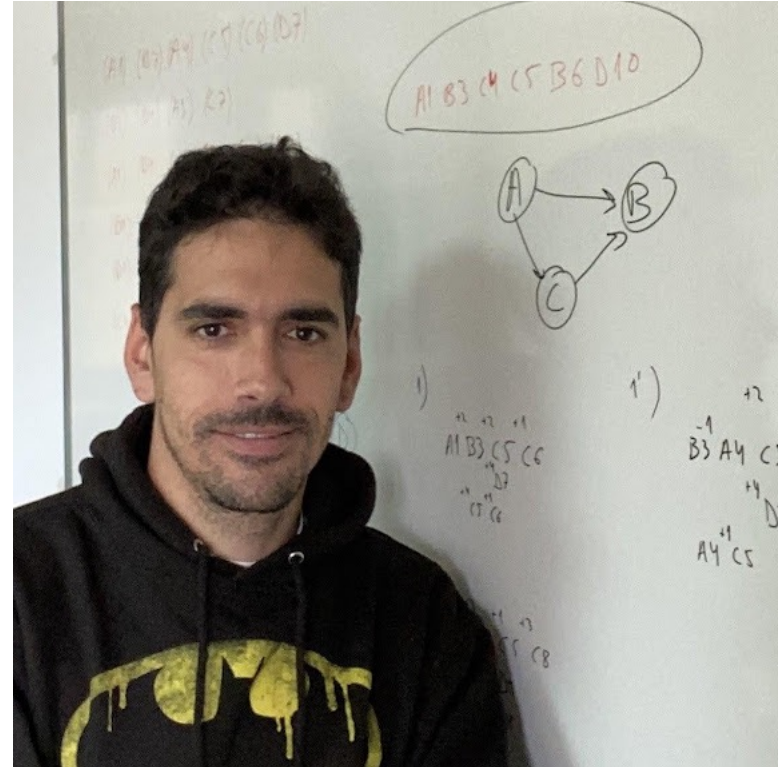
A stylized sunburst graphic in shades of purple and blue, located in the top-left corner of the slide.

# Métodos descriptivos supervisados

Máster en Ciencia de Datos

## José María Luna

recibió el título de Doctor en Ciencias de la Computación en 2014, por la Universidad de Granada. Su carrera investigadora fue inicialmente subvencionada por el Ministerio de Educación de España bajo el programa FPU (predoctoral) y el programa Juan de la Cierva (postdoctoral). Actualmente es profesor de la Universidad de Córdoba en el departamento de Informática y Análisis Numérico. Dr. Luna ha sido autor de los libros monográficos "Pattern Mining with Evolutionary Algorithms" y "Supervised Descriptive Pattern Mining", ambos publicados por la editorial Springer. Además, ha publicado más de 30 artículos en revistas científicas de alto impacto. Actualmente tiene un total de 2271 citas en Google Scholar y un índice H de 25. Su investigación es llevada a cabo en el grupo de investigación *Knowledge Discovery and Intelligent Systems*, donde investiga temas relativos a computación evolutiva, minería de patrones, reglas de asociación y sus aplicaciones.



A stylized sunburst graphic in shades of purple and blue, located in the top-left corner of the slide.

# Métodos descriptivos supervisados (Parte 1)

*Class association rules*

# Class association rules

- Definición formal

- Regla de asociación

- Relación entre *itemsets* de un patrón P de la forma  $X \rightarrow Y$

$$I = \{i_1, \dots, i_n\} \in \Omega$$

$$P \subseteq I$$

$$X \subset P \subseteq I$$

$$Y = P \setminus X$$

- Si el antecedente X de la regla se satisface, entonces es bastante probable que el consecuente Y de la regla también se satisfaga

- Regla de asociación de clase (*class association rule* o CARs)

- Conjunto específico de reglas de asociación donde el consecuente es un único ítem conocido de antemano
    - Relación entre *itemsets* de un patrón P de la forma  $X \rightarrow y$
    - Mejoran la comprensión de las reglas al contener un único elemento en el consecuente

# Class association rules

- Espacio de búsqueda
  - Reglas de asociación clásicas: un conjunto de datos con  $k$  ítems diferentes tiene un total de  $3^k - 2^{k+1} + 1$  reglas diferentes
  - Reglas de asociación de clase: dado  $t$  valores diferentes del consecuente o variable objetivo, existen  $(2^k - 1) \times t$  reglas diferentes

Item1	Item2	Variable objetivo
1	0	$V_1$
1	1	$V_2$
0	1	$V_2$
1	1	$V_1$

Item1  $\rightarrow V_1$

Item2  $\rightarrow V_1$

Item1, Item2  $\rightarrow V_1$

Item1  $\rightarrow V_2$

Item2  $\rightarrow V_2$

Item1, Item2  $\rightarrow V_2$

$k=2, t=2$

Numero reglas =  $(2^2 - 1) \times 2 = 6$

# Class association rules

- Espacio de búsqueda
  - En conjuntos reales, lo que suele ocurrir es que los atributos no son binarios y habría que transformar

Nivel de estudios	Estado civil	Sexo	Superó la prueba
Primaria	Soltero	Hombre	Sí
Primaria	Soltera	Mujer	No
Universitarios	Casado	Hombre	Sí
Universitarios	Casada	Mujer	Sí
Primaria	Divorciado	Hombre	No
Secundaria	Casada	Mujer	Sí

# Class association rules

- Espacio de búsqueda

Primaria	Secundaria	Universitarios	Soltero	Casado	Divorciado	Hombre	Mujer	Si	No
1	0	0	1	0	0	1	0	1	0
1	0	0	1	0	0	0	1	0	1
0	0	1	0	1	0	1	0	1	0
0	0	1	0	1	0	0	1	1	0
1	0	0	0	0	1	1	0	0	1
0	1	0	0	1	0	0	1	1	0

# *Class association rules*

- Objetivo
  - Encontrar un conjunto de ítems que describan adecuadamente la variable objetivo o variable de interés (consecuente de la regla)
- Algoritmos clásicos de minería de patrones
  - Apriori
  - FP-Growth
  - LCM
  - Etc



# *Class association rules*

- Apriori para *class association rules* considerando sólo soporte como métrica:
  1. Dividir el conjunto de datos en subconjuntos según la variable objetivo
  2. Reducir el conjunto si es posible
  3. Para cada subconjunto aplicar el algoritmo Apriori
  4. Cada patrón frecuente será el antecedente que formará una regla junto con la variable objetivo (consecuente)

# Class association rules

- Apriori para *class association rules* considerando sólo soporte como métrica:
  1. Dividir el conjunto de datos en subconjuntos según la variable objetivo

Primaria	Secundaria	Universitarios	Soltero	Casado	Divorciado	Hombre	Mujer	Si
1	0	0	1	0	0	1	0	1
0	0	1	0	1	0	1	0	1
0	0	1	0	1	0	0	1	1
0	1	0	0	1	0	0	1	1

# Class association rules

- Apriori para *class association rules* considerando sólo soporte como métrica:
  1. Dividir el conjunto de datos en subconjuntos según la variable objetivo

Primaria	Secundaria	Universitarios	Soltero	Casado	Divorciado	Hombre	Mujer	No
1	0	0	1	0	0	0	1	1
1	0	0	0	0	1	1	0	1

# Class association rules

- Apriori para *class association rules* considerando sólo soporte como métrica:
  2. Reducir el conjunto si es posible

Primaria	Secundaria	Universitarios	Soltero	Casado	Divorciado	Hombre	Mujer	Si
1	0	0	1	0	0	1	0	1
0	0	1	0	1	0	1	0	1
0	0	1	0	1	0	0	1	1
0	1	0	0	1	0	0	1	1

# Class association rules

- Apriori para *class association rules* considerando sólo soporte como métrica:
  2. Reducir el conjunto si es posible

Primaria	Secundaria	Universitarios	Soltero	Casado	Divorciado	Hombre	Mujer	No
1	0	0	1	0	0	0	1	1
1	0	0	0	0	1	1	0	1

# Class association rules

- Apriori para *class association rules* considerando sólo soporte como métrica:

3. Para cada subconjunto aplicar el algoritmo Apriori

## Items k=1

Primaria (1)  
Secundaria (1)  
Universitarios (2)  
Soltero (1)  
Casado (3)  
Hombre (2)  
Mujer (2)

## Items k=2

Primaria, Soltero (1)  
Primaria, Hombre (1)  
Secundaria, Casado (1)  
Secundaria, Mujer (1)  
Universitarios, Casado (2)  
Universitarios, Hombre (1)  
Universitarios, Mujer (1)

## Items k=3

Primaria, Soltero, Hombre (1)  
Secundaria, Casado, Mujer (1)  
Universitarios, Casado, Hombre (1)  
Universitarios, Casado, Mujer (1)

# Class association rules

- Apriori para *class association rules* considerando sólo soporte como métrica:

3. Para cada subconjunto aplicar el algoritmo Apriori

## Items k=1

Primaria (2)  
Soltero (1)  
Divorciado (1)  
Hombre (1)  
Mujer (1)

## Items k=2

Primaria, Soltero (1)  
Primaria, Divorciado (1)  
Primaria, Mujer (1)  
Primaria, Hombre (1)

## Items k=3

Primaria, Soltero, Mujer (1)  
Primaria, Divorciado, Hombre (1)

# Class association rules

- Apriori para *class association rules* considerando sólo soporte como métrica:

4. Cada patrón frecuente será el antecedente que formará una regla junto con la variable objetivo (consecuente)

## Reglas obtenidas

Primaria -> Si (1)

Secundaria -> Si (1)

Universitarios -> Si (2)

Soltero -> Si (1)

Casado -> Si (3)

Hombre -> Si (2)

Mujer -> Si (2)

Primaria, Soltero -> Si (1)

Primaria, Hombre -> Si (1)

Secundaria, Casado -> Si (1)

Secundaria, Mujer -> Si (1)

Universitarios, Casado -> Si (2)

Universitarios, Hombre -> Si (1)

Universitarios, Mujer -> Si (1)

Primaria, Soltero, Hombre -> Si (1)

Secundaria, Casado, Mujer -> Si (1)

Universitarios, Casado, Hombre -> Si (1)

Universitarios, Casado, Mujer -> Si (1)



# Class association rules

- Apriori para *class association rules* considerando sólo soporte como métrica:

4. Cada patrón frecuente será el antecedente que formará una regla junto con la variable objetivo (consecuente)

## Reglas obtenidas

Primaria -> Si (2)

Soltero -> Si (1)

Divorciado -> Si (1)

Hombre -> Si (1)

Mujer -> Si (1)

Primaria, Soltero -> Si (1)

Primaria, Divorciado -> Si (1)

Primaria, Mujer -> Si (1)

Primaria, Hombre -> Si (1)

Primaria, Soltero, Mujer -> Si (1)

Primaria, Divorciado, Hombre -> Si (1)

# Class association rules

- Uso de *class association rules* para tareas predictivas
  - Este tipo de reglas pueden considerar como variable objetivo la clase o variable a predecir en tareas predictivas
  - La combinación de varias *class association rules* permite formar un clasificador que, dados un conjunto de elementos o variables de entrada, pueda emitir una clasificación acorde al consecuente de la regla
    - SI casado ENTONCES supera la prueba = SI
    - SINO SI soltero y hombre ENTONCES supera la prueba = SI
    - SINO supera la prueba = NO

A stylized sunburst graphic in shades of purple and blue, located in the top-left corner of the slide. It features a semi-circle on the left with several rays extending outwards to the right.

¡Gracias!

**UCO**  
ONLINE

A horizontal bar at the bottom of the slide consisting of three equal-width rectangles of yellow, red, and blue, representing the colors of the Spanish flag.