

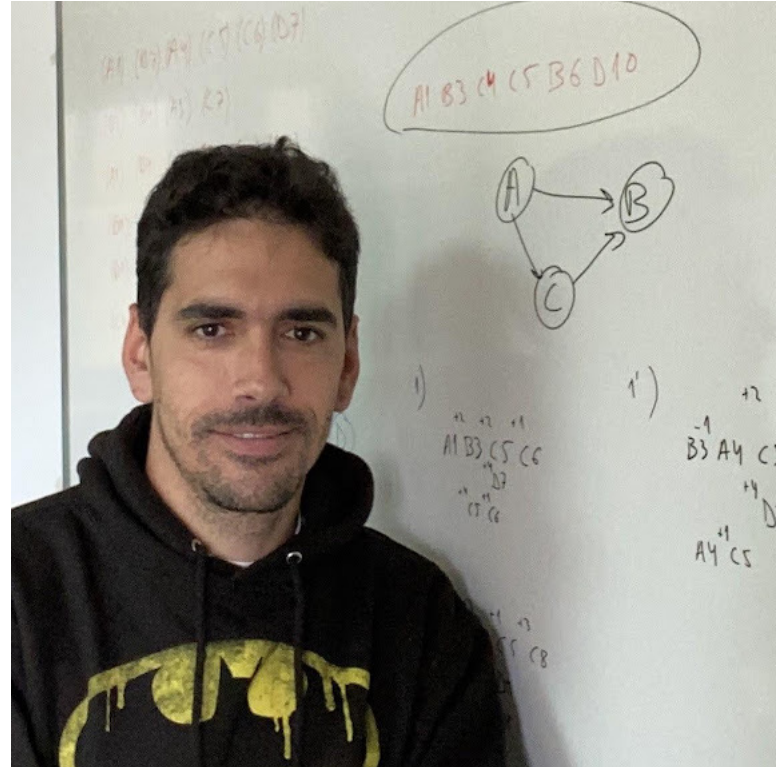
A stylized sunburst graphic in shades of purple and blue, located in the top-left corner of the slide.

# 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 2260 citas en Google Scholar y un índice H de 24. 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.





# Métodos descriptivos supervisados (Parte 2)

*Contrast set mining*

# *Conjuntos de contraste (contrast sets)*

- Introducción

- La minería de conjuntos de contraste es una técnica de minería de datos cuya objetivo es el descubrimiento de patrones discriminativos, cuya frecuencia incrementa de manera significativa de un grupo a otro. Cada grupo puede ser un conjunto de datos diferente
- A diferencia de los patrones emergentes, donde existían dos grupos, los conjuntos de contraste trabajan sobre N grupos, buscando patrones que son muy frecuentes en un grupo y muy poco frecuentes en otro
- Los patrones extraídos se representan en forma de reglas y representan tendencias emergentes entre conjuntos de datos definidos por una propiedad de interés

# Conjuntos de contraste (contrast sets)

- Definición formal

*El problema de minería de conjuntos de contraste se define como la tarea de descubrir patrones cuya frecuencia de aparición se incrementa considerablemente de un conjunto de datos a otro*

$$\max_{\forall \Omega_i, \Omega_j, \Omega_i \cap \Omega_j = \emptyset} |Support(P, \Omega_i) - Support(P, \Omega_j)|$$

# Conjuntos de contraste (contrast sets)

- Ejemplo
  - Economía = normal

Sexo	Edad	Pais	Estudios	Estado civil
Hombre	25	España	Superiores	Soltero
Mujer	45	Alemania	Superiores	Casado
Mujer	62	Italia	Primaria	Casado
Mujer	56	España	Secundaria	Casado

# Conjuntos de contraste (contrast sets)

- Ejemplo
  - Economía = rico

Sexo	Edad	País	Estudios	Estado civil
Mujer	37	Suiza	Doctorado	Soltero
Mujer	55	Alemania	Doctorado	Casado
Hombre	71	Suiza	Primaria	Soltero
Hombre	18	Suiza	Secundaria	Soltero

# Conjuntos de contraste (contrast sets)

- Ejemplo
  - Economía = pobre

Sexo	Edad	País	Estudios	Estado civil
Hombre	18	España	Primaria	Casado
Mujer	21	Alemania	Primaria	Casado
Hombre	31	Italia	Primaria	Divorciado
Hombre	29	Suiza	Secundaria	Soltero



# Conjuntos de contraste (contrast sets)

- Ejemplos

- *Pais = Suiza, Estado civil = Soltero*

- *Conjunto de datos Economía = normal → Soporte 0%*
    - *Conjunto de datos Economía = rico → Soporte 75%*
    - *Conjunto de datos Economía = pobre → Soporte 25%*
    - *Max(0.75, 0.25, 0.50) = 0.75*

- *Edad > 35, Sexo = Mujer*

- *Conjunto de datos Economía = normal → Soporte 75%*
    - *Conjunto de datos Economía = rico → Soporte 50%*
    - *Conjunto de datos Economía = pobre → Soporte 0%*
    - *Max(0.25, 0.75, 0.50) = 0.75*

## *Conjuntos de contraste (contrast sets)*

- Propuesta simple de extracción de conjuntos de contraste
  1. Dividir el conjunto de datos inicial en  $N$  subconjuntos  $D_1, D_2, \dots, D_n$  de acuerdo con alguna propiedad de interés
  2. Aplicar un algoritmo de extracción de patrones sobre cada subconjunto
  3. Evaluar cada patrón resultante en el el resto de subconjuntos
  4. Quedarse con los patrones cuya diferencia máxima entre subconjuntos sea mayor que un umbral

# *Conjuntos de contraste (contrast sets)*

- Propuesta específicas de extracción de conjuntos de contraste
  - STUCCO
  - CIGAR
  - DIFF
  - Gen\_QCsets
  - CSM-SD
  - ECSM
  - DisCOSet
  - Etc

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.