

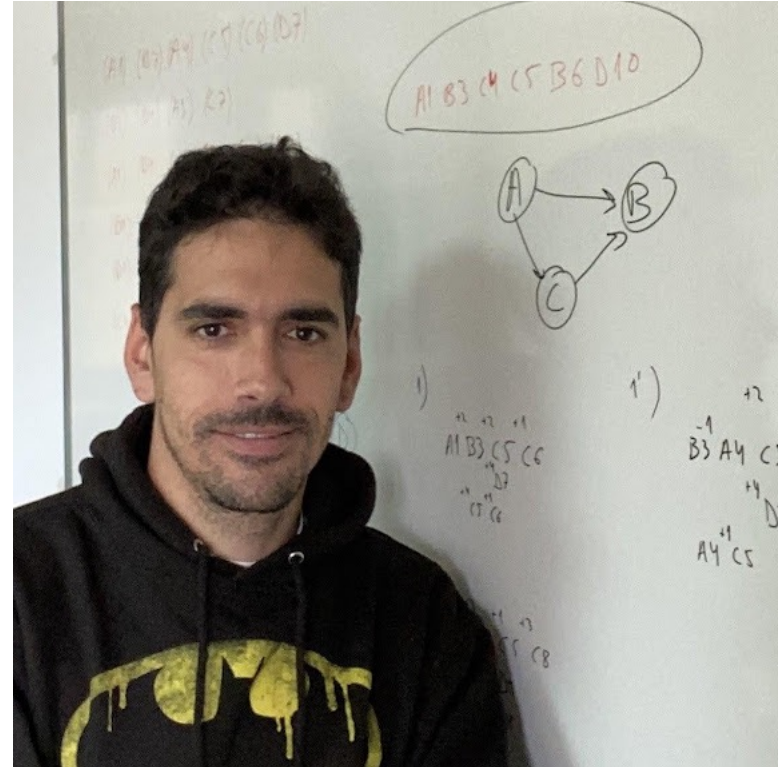
A stylized sunburst graphic in shades of purple and blue, located in the top-left corner of the slide.

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)

Clasificación asociativa

Clasificación asociativa

- **Introducción**

- En tareas predictivas (predecir una o varias variables de salida) a partir de nuevos datos, existen diferentes metodologías que permiten construir modelos exactos:
 - Sistemas basados en reglas
 - Árboles de decisión
 - Maquinas de vector soporte (SVM)
 - Redes neuronales
 - Etc
- Los clasificadores basados en reglas permiten un alto grado de interpretabilidad, permitiendo comprender y explicar los resultados

Clasificación asociativa

- **Introducción**

- Los sistemas basados en reglas pueden dividirse en dos grandes grupos:
 - Inducción de reglas
 - Clasificación basada en reglas de asociación, también conocida como clasificación asociativa
- La clasificación asociativa incluye, generalmente, dos procesos:
 - Extracción de reglas descriptivas mediante algoritmos de reglas de asociación
 - Construcción de un clasificador a partir de las reglas obtenidas en el paso anterior
- Ventajas de la clasificación asociativa:
 - Extraen todas las reglas posibles, algunas de las cuales pueden ser omitidas por otras técnicas de clasificación
 - Capacidad de actualizar reglas sin necesidad de modificar el conjunto de reglas total

Clasificación asociativa

- Algoritmo simple
 - La metodología más simple consiste en un sistema de votación:
 1. Extraer todas las reglas de asociación que satisfagan unos determinados umbrales de soporte y confianza
 2. Se comprueba que reglas satisfacen una nueva entrada para ver qué salida se produce
 3. Se establecen criterios de desempate si fuese necesario
 - Prevalecen reglas mas pequeñas
 - Prevalecen reglas con mejor confianza
 - etc
 - Modificaciones del sistema de votación:
 - Número de reglas
 - Peso de las reglas en base a valores de confianza
 - Peso de las reglas en base a conocimiento del experto

Clasificación asociativa

- Algoritmo simple

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

Clasificación asociativa

- Algoritmo simple
 - Extraer todas las reglas de asociación que satisfagan unos determinados umbrales de soporte y confianza
 - Soporte mínimo 30%
 - Confianza mínima 60%

Universitarios -> Si (Soporte = 0.33, Confianza = 1.00)

Casado -> Si (Soporte = 0.50 , Confianza = 1.00)

Hombre -> Si (Soporte = 0.33, Confianza = 0.67)

Mujer -> Si (Soporte = 0.33, Confianza = 0.67)

Primaria -> No (Soporte = 0.33 , Confianza = 0.67)

Universitarios, Casado -> Si (Soporte = 0.33, Confianza = 1.00)

Clasificación asociativa

- Algoritmo simple
 - Se comprueba que reglas satisfacen una nueva entrada para ver qué salida se produce

Hombre, soltero y con estudios universitarios

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

Clasificación asociativa

- Algoritmo simple
 - Se comprueba que reglas satisfacen una nueva entrada para ver qué salida se produce

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

Universitarios -> Si (Soporte = 0.33, Confianza = 1.00)

~~Casado -> Si (Soporte = 0.50, Confianza = 1.00)~~

Hombre -> Si (Soporte = 0.33, Confianza = 0.67)

~~Mujer -> Si (Soporte = 0.33, Confianza = 0.67)~~

~~Primaria -> No (Soporte = 0.33, Confianza = 0.67)~~

~~Universitarios, Casado -> Si (Soporte = 0.33, Confianza = 1.00)~~

Salida = Si

Clasificación asociativa

- Algoritmos
 - Algoritmo CBA
 - Primer algoritmo de clasificación asociativa, propuesto en 1998
 - Uso de algoritmo Apriori para obtener todas las reglas de asociación por encima de un umbral de soporte y confianza
 - Ordena reglas en base a criterios de predecencia:
 - Orden decreciente de valores de confianza
 - A igualdad de confianza, orden decreciente de valores de soporte
 - A igualdad de soporte, orden de generación de las reglas según Apriori
 - Usando las reglas ordenadas, se evalúa sobre los datos de entrenamiento que registros son clasificados correctamente e incorrectamente por la regla.
 - Se seleccionan las reglas que mejor clasifican

Clasificación asociativa

- Algoritmos
 - CBA2
 - CMAR
 - CPAR
 - MAC
 - ACCF
 - L3
 - ACN
 - ADT
 - ACAC
 -

LAC: Library for Associative Classification

<https://github.com/kdis-lab/lac/>

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.