

Introducción al Análisis Inteligente de Datos (IDA)



Maria-Amparo Vila
vila@decsai.ugr.es

Grupo de Investigación en Bases de
Datos y Sistemas de Información
Inteligentes <https://idbis.ugr.es/>
Departamento de Ciencias de la
Computación e Inteligencia Artificial
Universidad de Granada

Introducción al tema

Estructura de la presentación

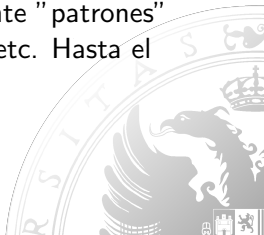
1. Introducción ideas básicas y motivación actual
2. KDD y Análisis Inteligentes de datos.
3. El proceso de KDD y El proceso CRISP-DM
4. Concepto de Minería de Datos
 - 4.1 Minería de datos y Estadística
 - 4.2 Minería de datos y aprendizaje
5. Problemas más importantes en Minería de Datos.
 - 5.1 EDA.Generalización y resumen. DM y DW
 - 5.2 Modelos descriptivos: Agrupamiento
 - 5.3 Modelos descriptivos: Modelización de dependencias
 - 5.4 Modelos predictivos: Clasificación
 - 5.5 Modelos predictivos: Predicción y Análisis de secuencias.
6. Tipos de Minería de datos



IDA, KDD y DM

Introducción al IDA : ideas básicas, motivación histórica

1. Desde el comienzo de la civilización el ser humano ha recopilado datos numéricos: babilonios, egipcios, chinos, griegos romanos, hacían censos, contabilizaban cosechas y recaudaban impuestos etc.
2. También desde el comienzo de la civilización el ser humano ha tratado de describir el mundo que le rodea mediante "patrones" entendidos estos como regularidades: astronomía etc. Hasta el siglo XVIII estas descripciones eran verbales.



IDA, KDD y DM

Introducción al IDA : ideas básicas, motivación histórica

- 3 A partir del siglo XVIII aparece el "método científico" para el estudio de los fenómenos naturales y sociales:



IDA, KDD y DM

Introducción al IDA : ideas básicas, motivación histórica

- 3 A partir del siglo XVIII aparece el "método científico" para el estudio de los fenómenos naturales y sociales:
- La descripción de los fenómenos naturales o sociales, son hipótesis que deben sustentarse mediante la experimentación o la observación



IDA, KDD y DM

Introducción al IDA : ideas básicas, motivación histórica

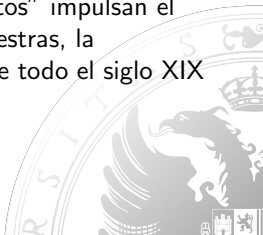
- 3 A partir del siglo XVIII aparece el "método científico" para el estudio de los fenómenos naturales y sociales:
- La descripción de los fenómenos naturales o sociales, son hipótesis que deben sustentarse mediante la experimentación o la observación
 - Es necesario analizar datos y estudiar mediante técnicas numéricas y/o estadísticas las regularidades y relaciones entre los mismos.



IDA, KDD y DM

Introducción al IDA : ideas básicas, motivación histórica

- 3 A partir del siglo XVIII aparece el "método científico" para el estudio de los fenómenos naturales y sociales:
- La descripción de los fenómenos naturales o sociales, son hipótesis que deben sustentarse mediante la experimentación o la observación
 - Es necesario analizar datos y estudiar mediante técnicas numéricas y/o estadísticas las regularidades y relaciones entre los mismos.
 - La falta de "máquinas para calcular y almacenar datos" impulsan el desarrollo de disciplinas tales como la Teoría de Muestras, la Estadística Matemática, el Análisis Numérico durante todo el siglo XIX y parte del XX



IDA, KDD y DM

Introducción al IDA : ideas básicas, motivación histórica

- 4 A partir del uso de los computadores las técnicas de análisis de datos pueden tratar datos de forma masiva y las posibilidades de descripción de fenómenos se amplían:



IDA, KDD y DM

Introducción al IDA : ideas básicas, motivación histórica

- 4 A partir del uso de los computadores las técnicas de análisis de datos pueden tratar datos de forma masiva y las posibilidades de descripción de fenómenos se amplían:
- El tipo y las propiedades de las variables a utilizar se hace menos riguroso
 - Aparecen nuevas técnicas de análisis: EDA (Análisis exploratorio de datos), Taxonomía numérica, Métodos de predicción no probabilísticos etc.



IDA, KDD y DM

Introducción al IDA : ideas básicas, motivación histórica

- 4 A partir del uso de los computadores las técnicas de análisis de datos pueden tratar datos de forma masiva y las posibilidades de descripción de fenómenos se amplían:
 - El tipo y las propiedades de las variables a utilizar se hace menos riguroso
 - Aparecen nuevas técnicas de análisis: EDA (Análisis exploratorio de datos), Taxonomía numérica, Métodos de predicción no probabilísticos etc.
- 5 Dado que parte de estas técnicas se desarrollaron como métodos de aprendizaje dentro de la I.A. se denominaron de **Análisis Inteligente de Datos**



IDA, KDD y DM

Introducción al KDD : ideas básicas, y motivación

- Datos y Bases de datos han crecido de forma vertiginosa
- Un enfoque de consulta a bases de datos clásico no proporciona realmente soluciones para los directivos. Es necesario resumir la información y presentarla de forma inteligible.



IDA, KDD y DM

Introducción al KDD : ideas básicas, y motivación

- Datos y Bases de datos han crecido de forma vertiginosa
- Un enfoque de consulta a bases de datos clásico no proporciona realmente soluciones para los directivos. Es necesario resumir la información y presentarla de forma inteligible.
- A partir de los 80 se pide a un sistema que:
 - Proporcione propiedades de los datos no explícitas.
 - Permita conocer relaciones entre los datos.
 - Proporcione información resumida y/o clasificada.
 - Todas estas facilidades deben estar integradas en un interfaz de usuario amigable e interactivo.



IDA, KDD y DM

Introducción al KDD : ideas básicas, y motivación histórica

Los sistemas de bases de datos clásicos ofrecen muy pocas funcionalidades para llevar a cabo aplicaciones como las que hemos expuesto.



IDA, KDD y DM

Introducción al KDD : ideas básicas, y motivación histórica

Los sistemas de bases de datos clásicos ofrecen muy pocas funcionalidades para llevar a cabo aplicaciones como las que hemos expuesto.

Una nueva forma de aproximación a las bases de datos

Minería de datos (Data Mining, DM)

Extracción de conocimiento (Knowledge Discovery, KDD)



IDA, KDD y DM

Introducción al KDD : ideas básicas, y motivación histórica

Los sistemas de bases de datos clásicos ofrecen muy pocas funcionalidades para llevar a cabo aplicaciones como las que hemos expuesto.

Una nueva forma de aproximación a las bases de datos

Minería de datos (Data Mining, DM)

Extracción de conocimiento (Knowledge Discovery, KDD)

Definición

Minería de Datos (DM o KDD) es un proceso no trivial de identificación de patrones en los datos válidos, novedosos, potencialmente útiles y comprensibles (Frawley et al. 1991)

El término patrón hay que tomarlo en sentido amplio (relaciones, tendencias, agrupamientos, clasificaciones etc..)

IDA, KDD y DM

Introducción al KDD : ideas básicas, y motivación histórica

Definición

Extracción de conocimiento

El proceso de uso de una base de datos para cualquier consulta que se requiera; incluyendo:

- *Preprocesamiento, muestreo y transformaciones,*
- *Aplicación de técnicas de minería de datos para obtener patrones*
- *La evaluación de los resultados de dicha minería para identificar qué patrones se consideran conocimiento*

(Fayyad et al. 1996)

IDA, KDD y DM

Introducción al KDD : ideas básicas, y motivación histórica

Definición

Extracción de conocimiento

El proceso de uso de una base de datos para cualquier consulta que se requiera; incluyendo:

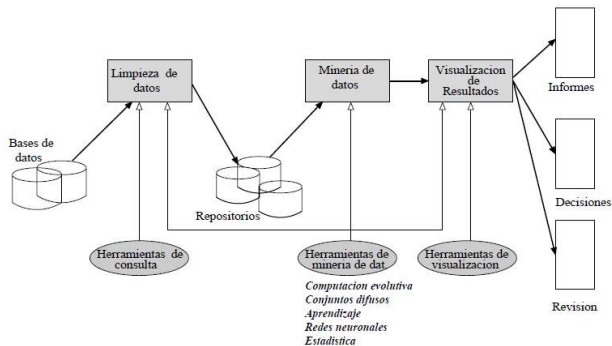
- *Preprocesamiento, muestreo y transformaciones,*
- *Aplicación de técnicas de minería de datos para obtener patrones*
- *La evaluación de los resultados de dicha minería para identificar qué patrones se consideran conocimiento*

(Fayyad et al. 1996)

IDA y KDD se consideran sinónimos. KDD tiene una connotación más empresarial e IDA más científica. DM es una etapa de ambos procesos

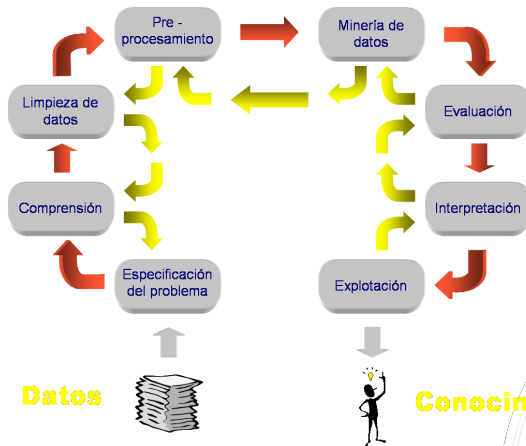
IDA, KDD y DM

Introducción al KDD : etapas en un proceso de KDD



IDA, KDD y DM

El proceso CRISP_DM (CRoss Industry Standard Process for Data Mining)



IDA, KDD y DM

El proceso CRISP_DM :Etapas

Comprensión del proyecto

- ¿Cuál es exactamente el problema?. ¿Qué beneficios se esperan con su solución?
- ¿Qué tipo de solución estamos buscando? ¿Qué respuestas pedimos?
- ¿Qué sabemos acerca del dominio del proyecto?
- ¿Cuál es el riesgo/costo de no resolverlo?



IDA, KDD y DM

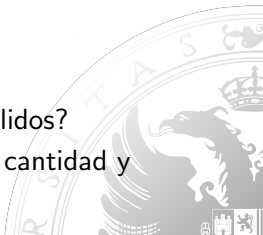
El proceso CRISP_DM :Etapas

Comprension del proyecto

- ¿Cual es exactamente el problema?. ¿Que beneficios se esperan con su solución?
- ¿Qué tipo de solución estamos buscando? ¿Qué respuestas pedimos?
- ¿Que sabemos acerca del dominio del proyecto?
- ¿Cual es el riesgo/costo de no resolverlo?

Comprension de los datos

- ¿De que datos disponemos?
- ¿Son relevantes para el problema?.¿Son fiables, válidos?
- ¿Son los datos suficientes en términos de: calidad, cantidad y temporalidad?



IDA, KDD y DM

El proceso CRISP_DM :Etapas

Preparación de los datos

- ¿Sobre qué datos nos concentramos?
- ¿Como puedo mejorar su calidad?
- ¿Necesitan ser transformados (preprocesados)?



IDA, KDD y DM

El proceso CRISP_DM :Etapas

Preparación de los datos

- ¿Sobre qué datos nos concentramos?
- ¿Como puedo mejorar su calidad?
- ¿Necesitan ser transformados (preprocesados)?

Modelado

- ¿A que tipo(s) de modelo(s)/problema(s) corresponde mi proyecto?
(Selección de modelo)
- ¿Cual es la técnica más adecuada para construir el modelo?(Construcción del modelo)
- ¿Es correcto el modelo desde el punto de vista técnico?.
(Validación del modelo)



IDA, KDD y DM

El proceso CRISP_DM :Etapas

Evaluación

- ¿Satisface el modelo los requerimientos de nuestro proyecto?
- ¿Qué hemos aprendido acerca de nuestro problema a través del modelo?



IDA, KDD y DM

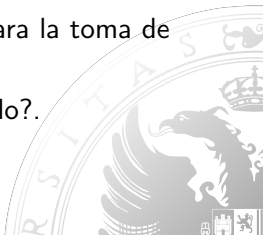
El proceso CRISP_DM :Etapas

Evaluación

- ¿Satisface el modelo los requerimientos de nuestro proyecto?
- ¿Qué hemos aprendido acerca de nuestro problema a través del modelo?

Implantación

- ¿Cómo puede ser útil el conocimiento adquirido para la toma de decisiones?
- ¿Cómo puedo saber si el modelo sigue siendo válido?



Minería de Datos (Data Mining)

Concepto de Minería de Datos

Definition

Entendemos por **Minería de Datos (DM)** el conjunto de técnicas que permiten preparar los datos, construir el modelo y validarlo en un proceso de KDD o IDA



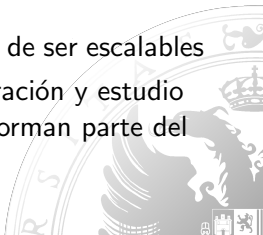
Minería de Datos (Data Mining)

Concepto de Minería de Datos

Definition

Entendemos por **Minería de Datos (DM)** el conjunto de técnicas que permiten preparar los datos, construir el modelo y validarlo en un proceso de KDD o IDA

- Las técnicas de DM son muy variadas así como los problemas que tratan.
- La mayoría de las técnicas de minería de datos han de ser escalables
- Algunos autores opinan que los métodos de preparación y estudio de los datos previos a la selección del modelo no forman parte del DM.



Minería de Datos (Data Mining)

Niveles de Minería de datos



Minería de Datos (Data Mining)

Niveles de Minería de datos

No-dirigida o pura No hay restricciones sobre el sistema, ni tampoco indicaciones acerca de lo que espera el usuario.

Responde a : *Dime algo interesante sobre los datos*

Corresponde al **Análisis Exploratorio de Datos.**



Minería de Datos (Data Mining)

Niveles de Minería de datos

No-dirigida o pura No hay restricciones sobre el sistema, ni tampoco indicaciones acerca de lo que espera el usuario.

Responde a : *Dime algo interesante sobre los datos*

Corresponde al **Análisis Exploratorio de Datos**.

Dirigida El objetivo se especifica algo más específico pero se tiene una actitud **Descriptiva**. Ejemplo: *¿Cómo se agrupan los clientes de un banco'?*

Corresponde a la extracción de patrones



Minería de Datos (Data Mining)

Niveles de Minería de datos

No-dirigida o pura No hay restricciones sobre el sistema, ni tampoco indicaciones acerca de lo que espera el usuario.

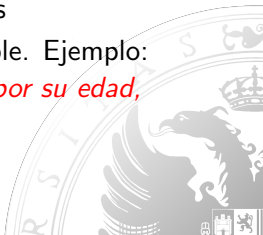
Responde a : *Dime algo interesante sobre los datos*

Corresponde al **Análisis Exploratorio de Datos**.

Dirigida El objetivo se especifica algo más específico pero se tiene una actitud **Descriptiva**. Ejemplo: *¿Cómo se agrupan los clientes de un banco'?*

Corresponde a la extracción de patrones

Predicción Se trata **predecir** el valor de una variable. Ejemplo: *¿Puedo caracterizar un cliente moroso por su edad, tipos de trabajo etc..?*



Minería de Datos (Data Mining)

Niveles de Minería de datos

No-dirigida o pura No hay restricciones sobre el sistema, ni tampoco indicaciones acerca de lo que espera el usuario.

Responde a : *Dime algo interesante sobre los datos*

Corresponde al **Análisis Exploratorio de Datos**.

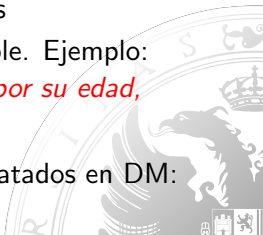
Dirigida El objetivo se especifica algo más específico pero se tiene una actitud **Descriptiva**. Ejemplo: *¿Cómo se agrupan los clientes de un banco'?*

Corresponde a la extracción de patrones

Predicción Se trata **predecir** el valor de una variable. Ejemplo: *¿Puedo caracterizar un cliente moroso por su edad, tipos de trabajo etc..?*

Estos tres niveles se corresponden con los modelos tratados en DM:

Exploratorios, descriptivos y predictivos



Minería de Datos (Data Mining)

Minería de Datos y Estadística



Minería de Datos (Data Mining)

Minería de Datos y Estadística

- Inicialmente la Estadística se desarrolló para resolver problemas de tratamiento de datos sin ayuda de computadores. Posteriormente se enriqueció con el uso de los ordenadores.



Minería de Datos (Data Mining)

Minería de Datos y Estadística

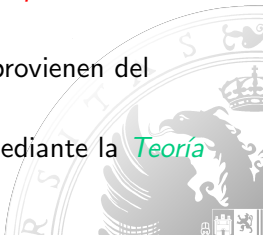
- Inicialmente la Estadística se desarrolló para resolver problemas de tratamiento de datos sin ayuda de computadores. Posteriormente se enriqueció con el uso de los ordenadores.
- La mayoría de las técnicas estadísticas imponen hipótesis muy fuertes a los datos: variables numéricas y continuas, independencia entre variables, igualdad de varianzas etc. *pero pueden ser utilizadas dentro de un problema de DM si dichas hipótesis se cumplen*



Minería de Datos (Data Mining)

Minería de Datos y Estadística

- Inicialmente la Estadística se desarrolló para resolver problemas de tratamiento de datos sin ayuda de computadores. Posteriormente se enriqueció con el uso de los ordenadores.
- La mayoría de las técnicas estadísticas imponen hipótesis muy fuertes a los datos: variables numéricas y continuas, independencia entre variables, igualdad de varianzas etc. *pero pueden ser utilizadas dentro de un problema de DM si dichas hipótesis se cumplen*
- Muchas de las técnicas de validación de modelos provienen del *Diseño de Experimentos*.
- Algunas problemas de escalabilidad se resuelven mediante la *Teoría de muestras*



Minería de Datos (Data Mining)

Minería de datos o Aprendizaje

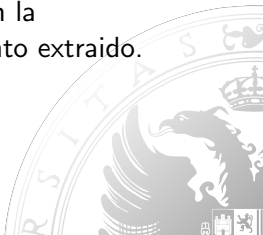
- **Es la Minería de Datos independiente del Aprendizaje?**
- *La diferencia está en la **calidad** de los datos.*



Minería de Datos (Data Mining)

Minería de datos o Aprendizaje

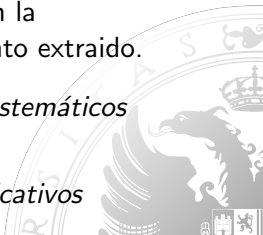
- **Es la Minería de Datos independiente del Aprendizaje?**
- *La diferencia está en la **calidad** de los datos.*
 - *Hay un gran volumen de datos* Se reduce el espacio de búsqueda tanto en filas como en los dominios de las columnas
 - *Los datos varían con el tiempo* El conocimiento adquirido debe poder actualizarse a medida que las bases de datos cambien.
 - *Los datos son incompletos y/o imprecisos* En este punto la teoría de subconjuntos difusos es muy aplicable, tanto en la representación de los datos como en el conocimiento extraído.



Minería de Datos (Data Mining)

Minería de datos o Aprendizaje

- **Es la Minería de Datos independiente del Aprendizaje?**
- *La diferencia está en la **calidad** de los datos.*
 - *Hay un gran volumen de datos* Se reduce el espacio de búsqueda tanto en filas como en los dominios de las columnas
 - *Los datos varían con el tiempo* El conocimiento adquirido debe poder actualizarse a medida que las bases de datos cambien.
 - *Los datos son incompletos y/o imprecisos* En este punto la teoría de subconjuntos difusos es muy aplicable, tanto en la representación de los datos como en el conocimiento extraído.
 - *Los datos pueden ser “ruidosos”, con errores no sistemáticos*
 - *Puede haber gran cantidad de “datos perdidos”*
 - *Algunos datos pueden ser redundantes o no significativos*



Problemas más importantes de DM

Generalización y resumen

- Los datos y los objetos en las bases de datos contienen información muy detallada y a niveles muy primitivos
- La idea básica de la generalización (resumen) es proporcionar descripciones compactas para subconjuntos de datos a un nivel conceptual superior.
- Los datos resumidos se pueden analizar de manera exploratoria y visual. Sugiriendo análisis más sofisticados.



Problemas más importantes de DM

Generalización y resumen

- Los datos y los objetos en las bases de datos contienen información muy detallada y a niveles muy primitivos
- La idea básica de la generalización (resumen) es proporcionar descripciones compactas para subconjuntos de datos a un nivel conceptual superior.
- Los datos resumidos se pueden analizar de manera exploratoria y visual. Sugiriendo análisis más sofisticados.

El enfoque de cubo de datos "data cube"

- La idea básica es usar tablas multidimensionales con datos agregados
- La estructura obtenida se denomina *cubo de datos multidimensional* y se supone almacenada.



Problemas más importantes de DM

Minería de datos y "Data Warehousing"

- **Es el Data Warehousing una parte de la Minería de Datos?**



Problemas más importantes de DM

Minería de datos y "Data Warehousing"

- **Es el Data Warehousing una parte de la Minería de Datos?**
- Aunque los problemas de generalización y resumen se consideran dentro de la minería de datos, realmente corresponden a la primera fase del proceso de KDD.



Problemas más importantes de DM

Minería de datos y "Data Warehousing"

- **Es el Data Warehousing una parte de la Minería de Datos?**
- Aunque los problemas de generalización y resumen se consideran dentro de la minería de datos, realmente corresponden a la primera fase del proceso de KDD.
- La estructura de "cubo de datos" es típica de la representación de información en DW pero fué presentada en "KDD & Data Mining"



Problemas más importantes de DM

Minería de datos y "Data Warehousing"

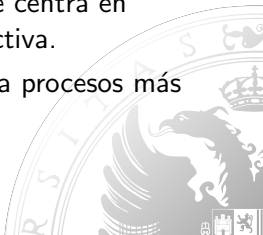
- **Es el Data Warehousing una parte de la Minería de Datos?**
 - Aunque los problemas de generalización y resumen se consideran dentro de la minería de datos, realmente corresponden a la primera fase del proceso de KDD.
 - La estructura de "cubo de datos" es típica de la representación de información en DW pero fué presentada en "KDD & Data Mining"
 - La tecnología de DW es un enfoque comercial y se centra en implementación, desempeño y visualización interactiva.



Problemas más importantes de DM

Minería de datos y "Data Warehousing"

- **Es el Data Warehousing una parte de la Minería de Datos?**
 - Aunque los problemas de generalización y resumen se consideran dentro de la minería de datos, realmente corresponden a la primera fase del proceso de KDD.
 - La estructura de "cubo de datos" es típica de la representación de información en DW pero fué presentada en "KDD & Data Mining"
 - La tecnología de DW es un enfoque comercial y se centra en implementación, desempeño y visualización interactiva.
 - La Minería de Datos es más académica y desarrolla procesos más sofisticados.

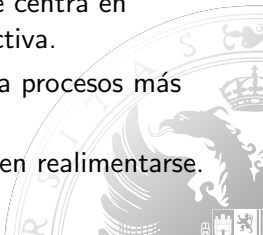


Problemas más importantes de DM

Minería de datos y "Data Warehousing"

• **Es el Data Warehousing una parte de la Minería de Datos?**

- Aunque los problemas de generalización y resumen se consideran dentro de la minería de datos, realmente corresponden a la primera fase del proceso de KDD.
- La estructura de "cubo de datos" es típica de la representación de información en DW pero fué presentada en "KDD & Data Mining"
- La tecnología de DW es un enfoque comercial y se centra en implementación, desempeño y visualización interactiva.
- La Minería de Datos es más académica y desarrolla procesos más sofisticados.
- Realmente son dos etapas de un proceso que pueden realimentarse.



Problemas más importantes de DM

Minería de datos y "Data Warehousing"

Algunas cuestiones "comerciales"

- Las grandes empresas de bases de datos ofertan herramientas de DW.
- Adicionalmente se ofertan herramientas de DM, en muchos casos no propias.
- Las mejores herramientas de DM no son propias de casas de bases de datos
- Existe muy poca ayuda para el usuario



Problemas más importantes de DM

Datos de partida

La estructura de datos más habitual para trabajar con DM es el **Dataset**

items \ variables	V_1	V_2	V_n
i_1	d_{11}	d_{12}	d_{1n}
\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots
i_m	d_{m1}	d_{m2}	d_{mn}



Problemas más importantes de DM

Datos de partida

La estructura de datos más habitual para trabajar con DM es el **Dataset**

items \ variables	V_1	V_2	V_n
i_1	d_{11}	d_{12}	d_{1n}
\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots
i_m	d_{m1}	d_{m2}	d_{mn}

- los items representan los casos
- Las variables pueden ser de muchos tipos
- Puede haber datos perdidos



Problemas más importantes de DM

Modelos descriptivos: Agrupamiento (Clustering)

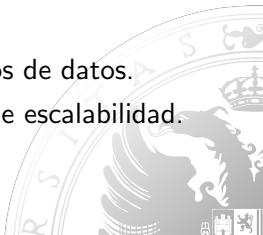
- Consiste en un proceso que agrupa los items de un "dataset" obteniendo un conjunto de "clusters" o clases.
- La mayoría de las técnicas surgen de la Taxonomía Matemática y se basan en la **similitud** entre items.



Problemas más importantes de DM

Modelos descriptivos: Agrupamiento (Clustering)

- Consiste en un proceso que agrupa los items de un "dataset" obteniendo un conjunto de "clusters" o clases.
- La mayoría de las técnicas surgen de la Taxonomía Matemática y se basan en la **similitud** entre items.
- No suelen utilizar conocimiento adicional acerca de la forma de los grupos.
- Existen muy diversas técnicas adaptadas a los tipos de datos.
- Las técnicas más avanzadas resuelven problemas de escalabilidad.



Problemas más importantes de DM

Modelización de dependencias

- *Objetivo:* Describir dependencias significativas entre las variables incluidas en la base de datos



Problemas más importantes de DM

Modelización de dependencias

- *Objetivo:* Describir dependencias significativas entre las variables incluidas en la base de datos Los modelos de dependencias pueden ser:
 - Cualitativas o cuantitativas (dependencias funcionales y análisis de regresión)
 - Dependencias parciales o completas



Problemas más importantes de DM

Modelización de dependencias

● *Objetivo:* Describir dependencias significativas entre las variables incluidas en la base de datos Los modelos de dependencias pueden ser:

- Cualitativas o cuantitativas (dependencias funcionales y análisis de regresión)
- Dependencias parciales o completas

*Cuando se trata de variables cuantitativas, y se espera la existencia de una relación $y = f(x_1, ..x_n)$ tenemos un **modelo predictivo** normalmente de análisis de regresión*



Problemas más importantes de DM

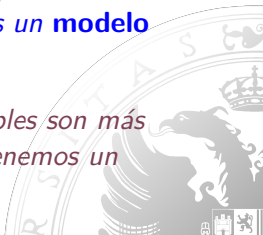
Modelización de dependencias

● **Objetivo:** Describir dependencias significativas entre las variables incluidas en la base de datos. Los modelos de dependencias pueden ser:

- Cualitativas o cuantitativas (dependencias funcionales y análisis de regresión)
- Dependencias parciales o completas

*Cuando se trata de variables cuantitativas, y se espera la existencia de una relación $y = f(x_1, ..x_n)$ tenemos un **modelo predictivo** normalmente de análisis de regresión*

*Cuando no se tiene conocimiento previo, las variables son más generales y se buscan asociaciones entre valores tenemos un **modelo descriptivo***



Problemas más importantes de DM

Problemas descriptivos: reglas de asociación

Descubren asociaciones importantes entre conjuntos de valores de atributos

- Un ejemplo clásico:

Buscar conexiones entre diferentes tipos de productos en una base de datos de ventas. Por ejemplo saber si los clientes que compran leche compran pan



Problemas más importantes de DM

Modelos descriptivos: Clasificación

- Se tiene un conjunto de datos donde una de las variables representa la clase a la que pertenece el ítem.
- Buscamos un *procedimiento de clasificación* que no permita incluir cada ítem nuevo en una clase.
- Las técnicas de clasificación se adaptan de las existentes en Aprendizaje con un especial énfasis en *problemas de escalabilidad*

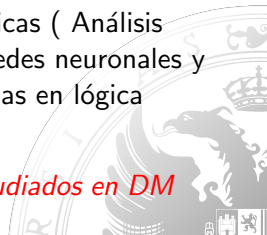


Problemas más importantes de DM

Modelos descriptivos: Clasificación

- Se tiene un conjunto de datos donde una de las variables representa la clase a la que pertenece el ítem.
- Buscamos un *procedimiento de clasificación* que no permita incluir cada ítem nuevo en una clase.
- Las técnicas de clasificación se adaptan de las existentes en Aprendizaje con un especial énfasis en *problemas de escalabilidad*
- Se han adaptado con éxito ID-3, técnicas estadísticas (Análisis Discriminante, Bayesianas), técnicas basadas en redes neuronales y técnicas basadas en "rough sets" y técnicas basadas en lógica difusa.

La clasificación es uno de los problemas mejor estudiados en DM



Problemas más importantes de DM

Series temporales y Análisis de secuencias

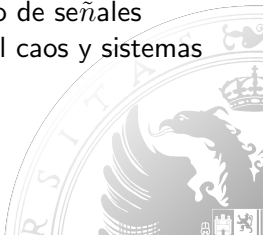
- Este tipo de análisis se aplica a datos que dependen del tiempo y para los cuales se quiere encontrar un patrón temporal.
- Habitualmente estos problemas se han tratado mediante las *Series Temporales* pero estas técnicas de origen estadístico imponen muchas limitaciones a los datos.
- las secuencias suelen ser no estacionarias, no periódicas, irregulares e incluso caóticas.



Problemas más importantes de DM

Series temporales y Análisis de secuencias

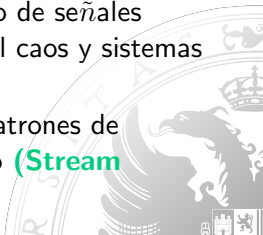
- Este tipo de análisis se aplica a datos que dependen del tiempo y para los cuales se quiere encontrar un patrón temporal.
- Habitualmente estos problemas se han tratado mediante las *Series Temporales* pero estas técnicas de origen estadístico imponen muchas limitaciones a los datos.
- las secuencias suelen ser no estacionarias, no periódicas, irregulares e incluso caóticas.
- Entre las técnicas aplicadas están el procesamiento de señales adaptativo, los algoritmos genéticos, y la teoría del caos y sistemas dinámicos no lineales.



Problemas más importantes de DM

Series temporales y Análisis de secuencias

- Este tipo de análisis se aplica a datos que dependen del tiempo y para los cuales se quiere encontrar un patrón temporal.
- Habitualmente estos problemas se han tratado mediante las *Series Temporales* pero estas técnicas de origen estadístico imponen muchas limitaciones a los datos.
- las secuencias suelen ser no estacionarias, no periódicas, irregulares e incluso caóticas.
- Entre las técnicas aplicadas están el procesamiento de señales adaptativo, los algoritmos genéticos, y la teoría del caos y sistemas dinámicos no lineales.
- Se está dedicando mucho esfuerzo al análisis de patrones de secuencias discretas, dependientes o no del tiempo (**Stream mining**)



Minería de Datos (Data Mining)

Tipos de Minería de datos

Según el ámbito donde se aplique aparecen distintos tipos de DM



Minería de Datos (Data Mining)

Tipos de Minería de datos

Según el ámbito donde se aplique aparecen distintos tipos de DM

Text Mining Cuando se extrae conocimiento de textos.

- Problemas de no estructuración. Forma intermedia
- Problemas de semántica.
- Un problema típico. **Detección de identidades**



Minería de Datos (Data Mining)

Tipos de Minería de datos

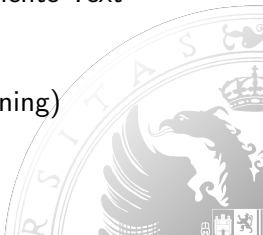
Según el ámbito donde se aplique aparecen distintos tipos de DM

Text Mining Cuando se extrae conocimiento de textos.

- Problemas de no estructuración. Forma intermedia
- Problemas de semántica.
- Un problema típico. **Detección de identidades**

Web Mining Cuando se extrae conocimiento de la Web

- Minería Web de contenidos. Básicamente Text Mining
- Minería Web de uso.
- Minería Web de conexión (Graph Mining)
- Minería sobre redes sociales.



Minería de Datos (Data Mining)

Tipos de Minería de datos

Según el ámbito donde se aplique aparecen distintos tipos de DM

Text Mining Cuando se extrae conocimiento de textos.

- Problemas de no estructuración. Forma intermedia
- Problemas de semántica.
- Un problema típico. **Detección de identidades**

Web Mining Cuando se extrae conocimiento de la Web

- Minería Web de contenidos. Básicamente Text Mining
- Minería Web de uso.
- Minería Web de conexión (Graph Mining)
- Minería sobre redes sociales.

Bioinformática Básicamente Stream Mining y Clustering

