



# Introducción a Minería de Datos, Metodologías y Técnicas de Minería de datos

Jose Aguilar  
CEMISID, Escuela de Sistemas  
Facultad de Ingeniería  
Universidad de Los Andes  
Mérida, Venezuela

# Definiciones iniciales



- **Datos:** hechos o medidas que describen características de objetos, eventos o personas, es la materia prima de la que se obtendrá la información.
- **Información:** Datos analizados y presentados en forma adecuada, de interés para un observador en un momento determinado.
- **Conocimiento:** información procesada para emitir juicios que llevan a conclusiones.
- **Meta Conocimiento:** Reglas que permiten obtener conocimiento.

# Principales características de la información



- **Significado (semántica):**

¿Qué quiere decir?

Del significado extraído de una información, cada individuo evalúa las consecuencias posibles y adecúa sus actitudes y acciones de manera acorde a las consecuencias previsibles que se deducen del significado de la información.

- **Importancia (relativa al receptor):**

¿Trata sobre alguna cuestión importante?

La importancia de la información para un receptor, se referirá a en qué grado cambia la actitud o la conducta de los individuos.



# Principales características de la información

- **Vigencia (en la dimensión espacio-tiempo):**

¿Es actual o desfasada?

En la práctica la vigencia de una información es difícil de evaluar, ya que en general acceder a una información no permite conocer de inmediato si dicha información tiene o no vigencia.

.

- **Validez (relativa al emisor):**

¿El emisor es fiable? ¿Puede entregar información no válida?

- **Valor (receptor):**

¿Cuán útil resulta para el destinatario la información?

.

# Situación actual



- La revolución digital ha permitido que la captura de datos sea fácil, y su almacenamiento tenga un costo casi nulo.
- Enormes cantidades de datos son recogidas y almacenadas en BD en la vida diaria.
- **Resultado:** Para analizar estas enormes cantidades de datos, las herramientas tradicionales de gestión de datos y las herramientas estadísticas no son adecuadas.

# Definiciones iniciales



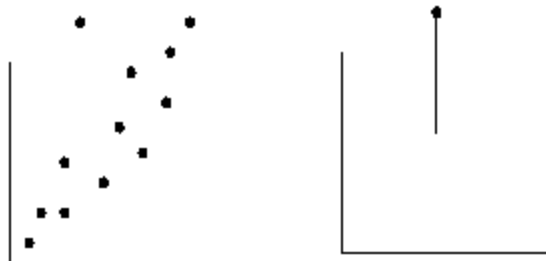
## Conocimiento: Modelo vs. Patrón

Hand, Mannila y Smyth

**Modelo:** Habla de todo el conjunto de datos



**Patrón:** Habla de una región particular de datos.



# Datos y Modelos => Conocimiento

- Los datos se obtienen de:
  - Bases de datos  
(relacionales, espaciales, temporales, documentales, multimedia, etc)
  - World Wide Web
- Modelos descriptivos: identifican patrones que explican o resumen los datos:
  - Reglas de asociación
  - Clustering:
- Modelos predictivos: estiman valores de variables de interés (a predecir) a partir de valores de otras variables
  - Regresión
  - Clasificación

# Situación actual



- Los datos por sí solos no producen beneficio directo.

**Su verdadero valor consiste en poder extraer información útil para la toma de decisiones**

- Tradicionalmente se analizaban datos con la ayuda de técnicas estadísticas (resumiendo y generando informes) o validando modelos sugeridos manualmente por los expertos.



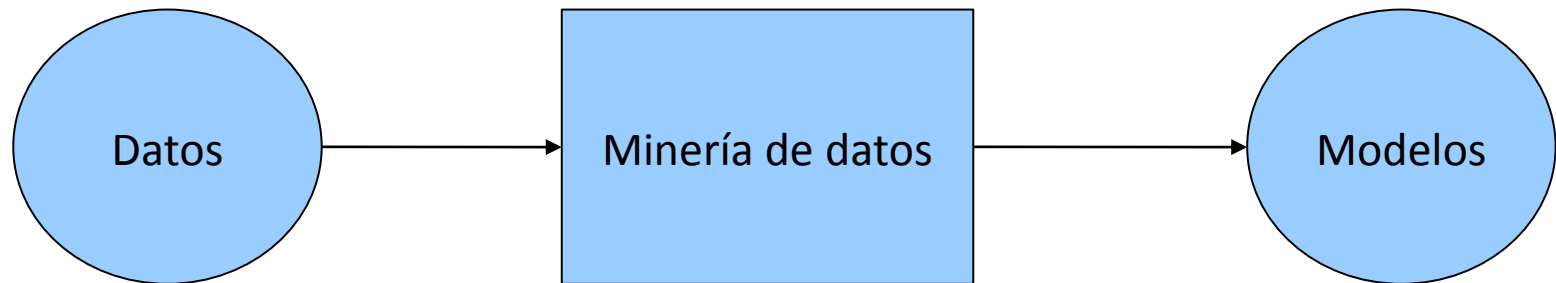
# Situación actual



- Estos procesos son irrealizables a medida que aumenta el tamaño de los datos.
- Bases de datos con un nº de registros del orden de  $10^9$  y  $10^3$  de dimensión, son fenómenos relativamente comunes.
- La tecnología informática puede automatizar este proceso.

**Minería de datos**

# Una visión simplificada de la minería de datos



- Los “modelos” son el producto de la minería de datos...
- ...y dan soporte a las estrategias de decisión que se tomen



# ¿Qué es la Minería de Datos?

La minería de datos tiene como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten hacia la toma de decisión.

Minería de datos es la exploración y análisis de grandes cantidades de datos con el objeto de encontrar patrones y reglas significativas (conocimiento)



# ¿Qué es la Minería de Datos?

- Es un mecanismo de explotación que consiste en la búsqueda de información valiosa en grandes volúmenes de datos.
- Ligada a las bodegas de datos (información histórica) con la cual los algoritmos de minería de datos obtienen información necesaria para la toma de decisiones.



# ¿Qué es la Minería de Datos?

Análisis de grandes volúmenes de datos para encontrar relaciones no triviales, y para resumirlos de manera que sean entendibles y útiles.

**Hand, Mannila y Smyth**

Extracción de patrones y modelos interesantes, potencialmente útiles y datos en base de datos de gran tamaño.

**Hand**



# ¿Qué es la Minería de Datos?

- Data Mining (Minería de datos)
  - Proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos (Witten and Frank, 2000)
- Knowledge Discovery in Databases – KDD (Descubrimiento de Conocimiento en Bases de Datos)
  - Proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles, a partir de los datos (Fayyad y col. 1996)



# Minería de datos

- Proceso de utilizar datos “crudos” para inferir importantes relaciones entre ellos
- Colección de técnicas poderosas para analizar grandes volúmenes de datos
- No existe un solo enfoque para minería de datos sino un conjunto de técnicas que se pueden utilizar de manera independiente o en combinación

# Lo que no es la minería de datos



Búsqueda de un número de teléfono en el directorio telefónico

Usar un motor de búsqueda en Internet para obtener información acerca de "Amazon"

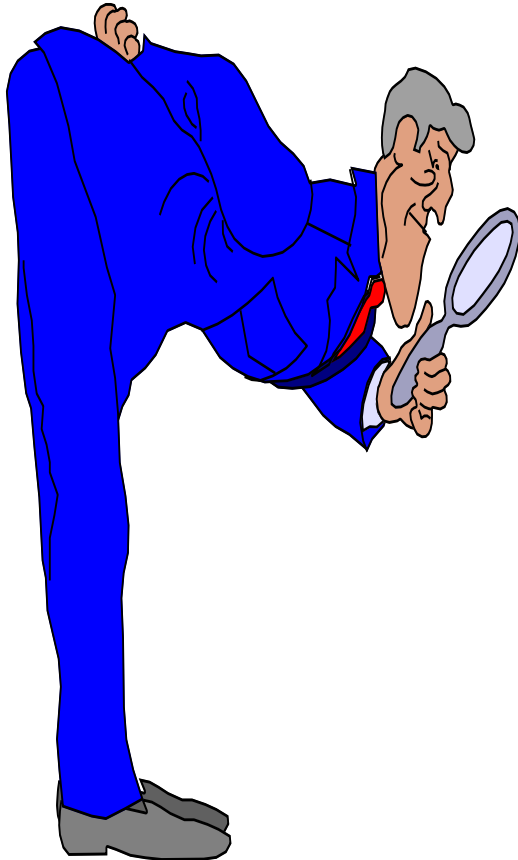
La fuerza bruta crujido de datos a granel

La presentación de los datos de diferentes maneras

Una tarea intensiva de base de datos

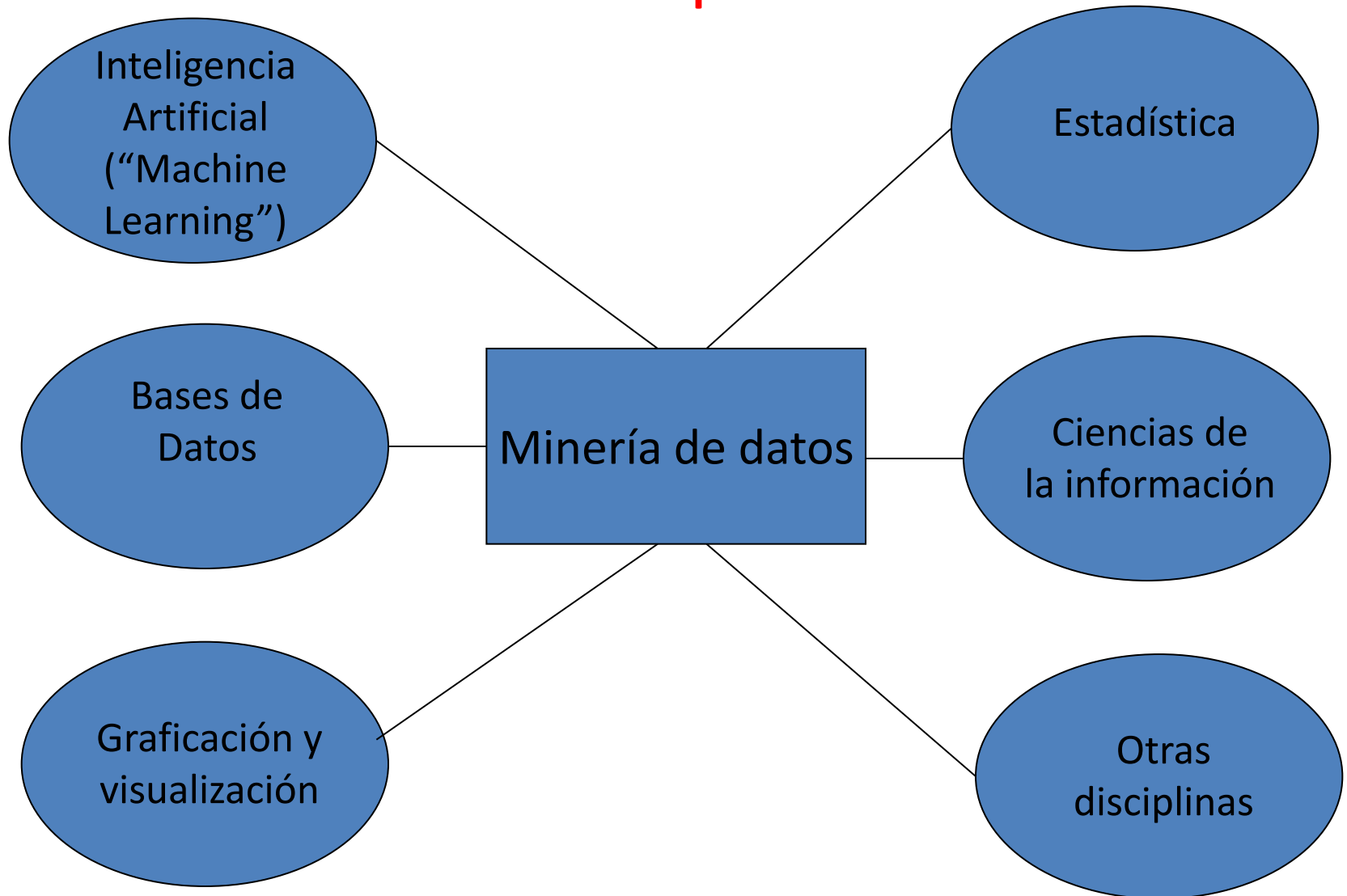


# Qué si es Data Mining?



- Una palabra de moda para una clase de técnicas que encuentran patrones en los datos
- Un conjunto de técnicas que se encuentran las relaciones que anteriormente no han sido descubiertos

# La minería de datos es un campo multidisciplinario



# Estadística vs Minería de datos

	<b>Estadística</b>	<b>Minería de datos</b>
Construcción de modelos	Ceñido a premisas y teoremas	Mayor libertad en la construcción, interpretable
Búsqueda	Test de la razón de la verosimilitud	Metaheurísticos
Transparencia	Más complicados de interpretar	Más claros y sencillos
Validación	No	Sí

# Minería de Datos vs. Estadística

Además, en minería de datos tenemos:

- 1.- Mayor volumen de datos (diferencia fundamental)
- 2.- Mayor complejidad de los datos (miles de atributos o dimensiones).
- 4.- Énfasis está en buscar modelos más que en verificar un modelo fijo.



# OLAP vs Minería de datos

OLAP	Minería de datos
¿Cuál es la proporción media de accidentes entre fumadores y no fumadores?	¿Cuál es la mejor predicción para accidentes?
¿Cuál es la factura telefónica media de mis clientes y de los que han dejado la compañía?	¿Dejara X la compañía? ¿Qué factores afectan a los abandonados?
¿Cuánto es la compra media diaria de tarjetas robadas y legítimas?	¿Cuáles son los patrones de compra asociados con el fraude de tarjetas?

**OLAP proporciona una muy buena vista de lo que está sucediendo, pero no se puede predecir lo que va a suceder en el futuro o por qué está sucediendo**

# ¿Qué es la Minería de Datos?

- **Métodos Descriptivos**

Encontrar patrones interpretable que describen los datos.

- **Métodos de Predicción**

Utilizar algunas variables para predecir los valores desconocidos o futuros de otras variables.

# ¿Qué es la Minería de Datos?

La minería de datos se puede dividir en:

- **Minería de datos predictiva (mdp):** usa primordialmente técnicas estadísticas.
- **Minería de datos para descubrimiento de conocimiento (mddc):** usa principalmente técnicas de inteligencia artificial.

# Tipos de aplicaciones de la minería de datos



- Clasificación
- Pronóstico (Predicción)
- Asociación
- Agrupación o segmentación (Clustering)



# Tipos de aplicaciones de la minería de datos

- Clasificación [predictivo]
- Clustering [descriptivo]
- Descubrimiento Regla Asociación [descriptivo]
  - Análisis de dependencia de datos
  - correlación y causalidad
- Descubrimiento Patrones Secuenciales [descriptivo]
  - Análisis de series de tiempo, asociaciones secuenciales
- Regresión [predictivo]
- Tendencia y Desviación de detección [predictivo]
- Filtros Colaborativos [predictivo]
- Resumir
- Descripción de Conceptos-
  - Descripción de características
  - descripción discriminante



# Clasificación

- Examinar las características de un nuevo objeto y asignarle una clase o categoría de acuerdo a un conjunto de tales objetos previamente definido
- Ejemplos:
  - Clasificar aplicaciones a crédito como bajo, medio y alto riesgo
  - Detectar reclamos fraudulentos de seguros



# Pronóstico

- Predecir un valor futuro con base a valores pasados
- Ejemplos:
  - Predecir cuánto efectivo requerirá un cajero automático en un fin de semana



# Asociación

- Determinar cosas u objetos que van juntos
- Ejemplo:
  - Determinar que productos se adquieren conjuntamente en un supermercado



# Agrupación o segmentación

- Dividir una población en un número de grupos más homogéneos
- No depende de clases pre-definidas a diferencia de clasificación
- Ejemplo:
  - Dividir la base de clientes de acuerdo con los hábitos de consumo

# ALGORITMOS DE APRENDIZAJE

**1. SUPERVISADOS:** predicen el valor de un atributo de un conjunto de datos conocidos otros atributos.

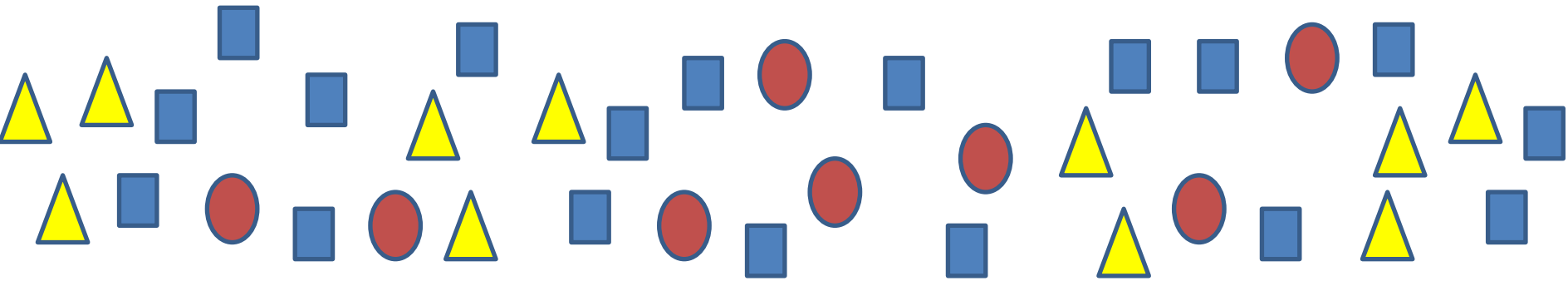
- Clasificación, Predicción
- Ejemplos: Algoritmos genéticos: Redes neuronales, Árboles de decisión, Regresión

**2. NO SUPERVISADOS:** descubren patrones y tendencias en los datos sin tener ningún tipo de conocimiento previo acerca de cuales son los patrones buscados

- Clustering, Análisis de enlace, Análisis de frecuencia

# Aprendizaje supervisado

El proceso de modelado se realiza sobre un conjunto de ejemplos formado por entradas al sistema y la respuesta que debería dar para cada entrada.



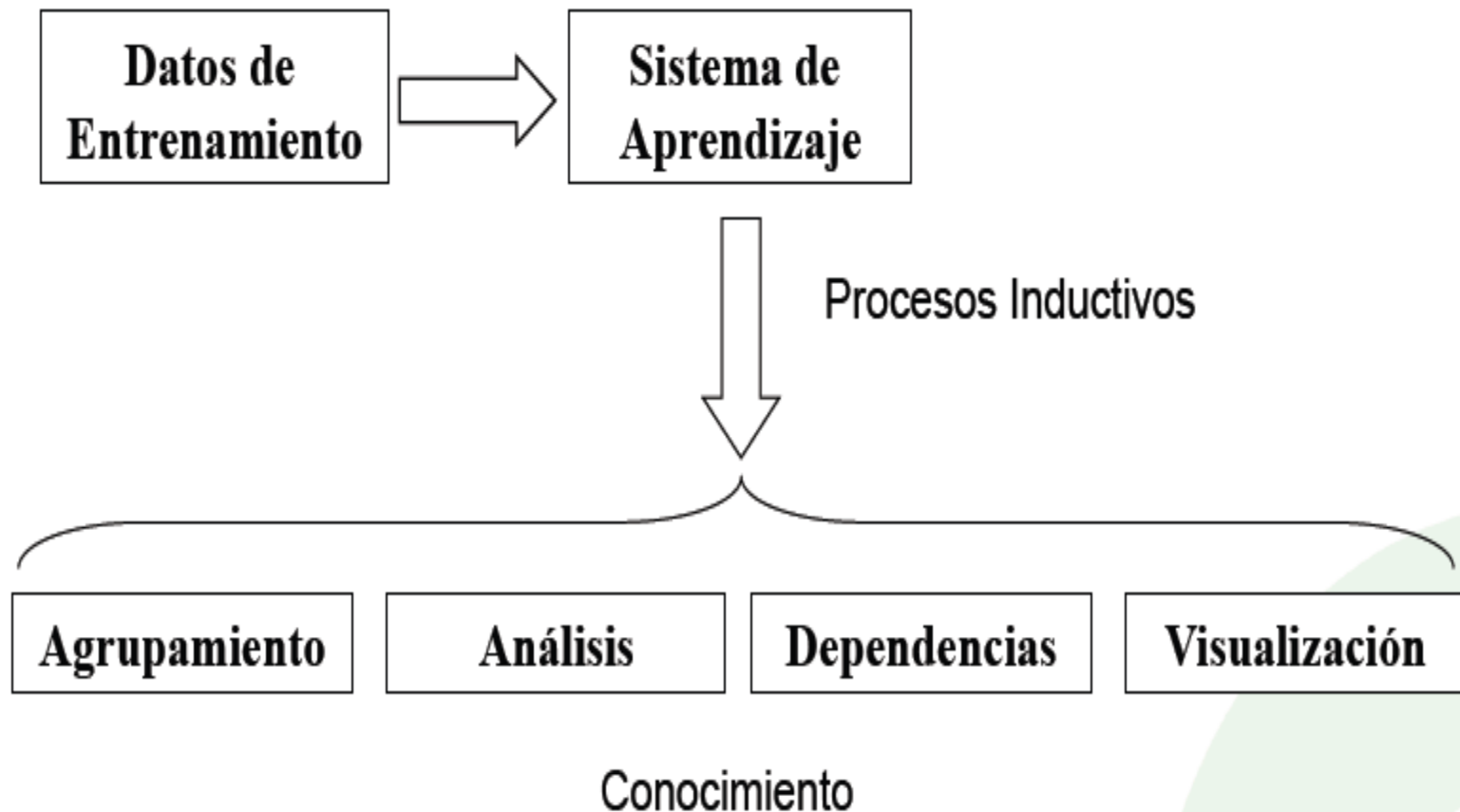
# Aprendizaje no supervisado

Todo el proceso de modelado se lleva a cabo sobre un conjunto de ejemplos formado tan sólo por entradas al sistema.

- No se tiene información sobre las categorías de esos ejemplos.
- Por lo tanto, en este caso, el sistema tiene que ser capaz de reconocer patrones para poder etiquetar las nuevas entradas.

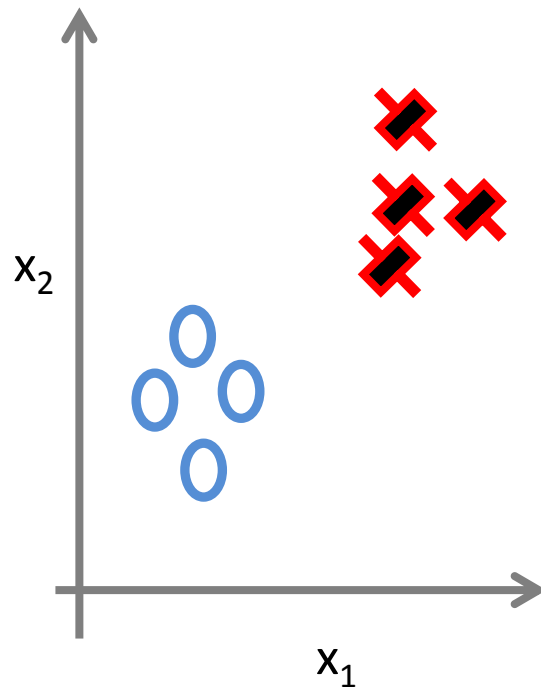


# Aprendizaje no supervisado

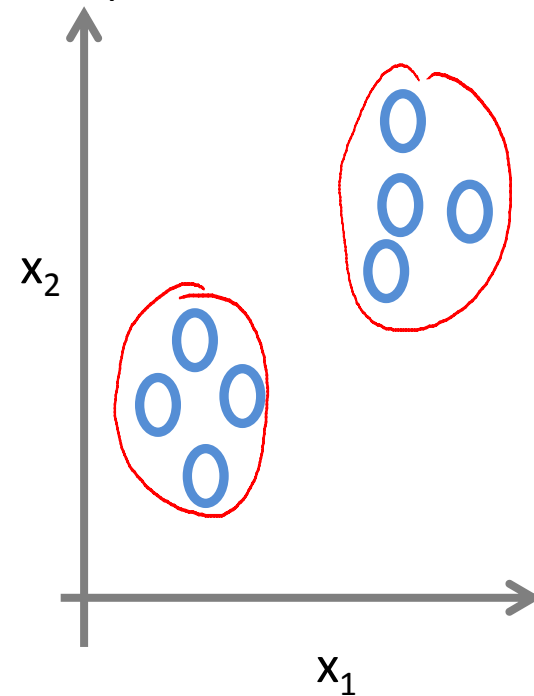


# Aprendizaje no supervisado

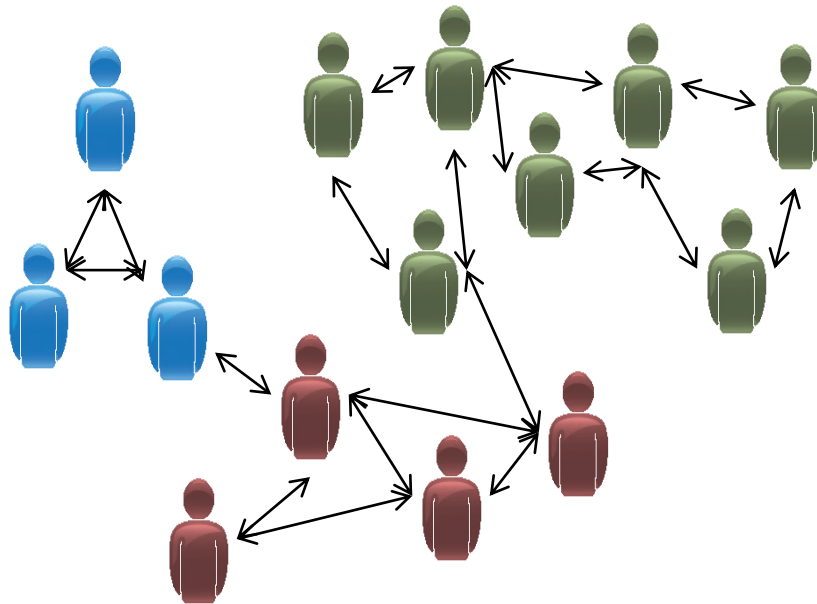
Aprendizaje supervisado



Aprendizaje no supervisado



# Aprendizaje no supervisado



Social network analysis

# Minería de Secuencia de Datos

- Buscar Similitud en serie temporal de datos
- Regresión y Análisis de Tendencias en series temporales de datos
- Minería Patrón Secuencial en secuencias simbólicas
- Clasificación de Secuencia
- Alineación de secuencias biológicas

# Minería de Gráfos y Redes

- Minería de Patrón de Gráfo
- Modelado estadístico de Redes
- Agrupación y clasificación de grafos y redes homogéneas
- Agrupación, clasificación de las Redes heterogéneos
- Descubrimiento, clases, y Predicción de Enlace en Redes de Información
- Búsqueda de Similitud en Redes de Información:
- Evolución de las redes de información social

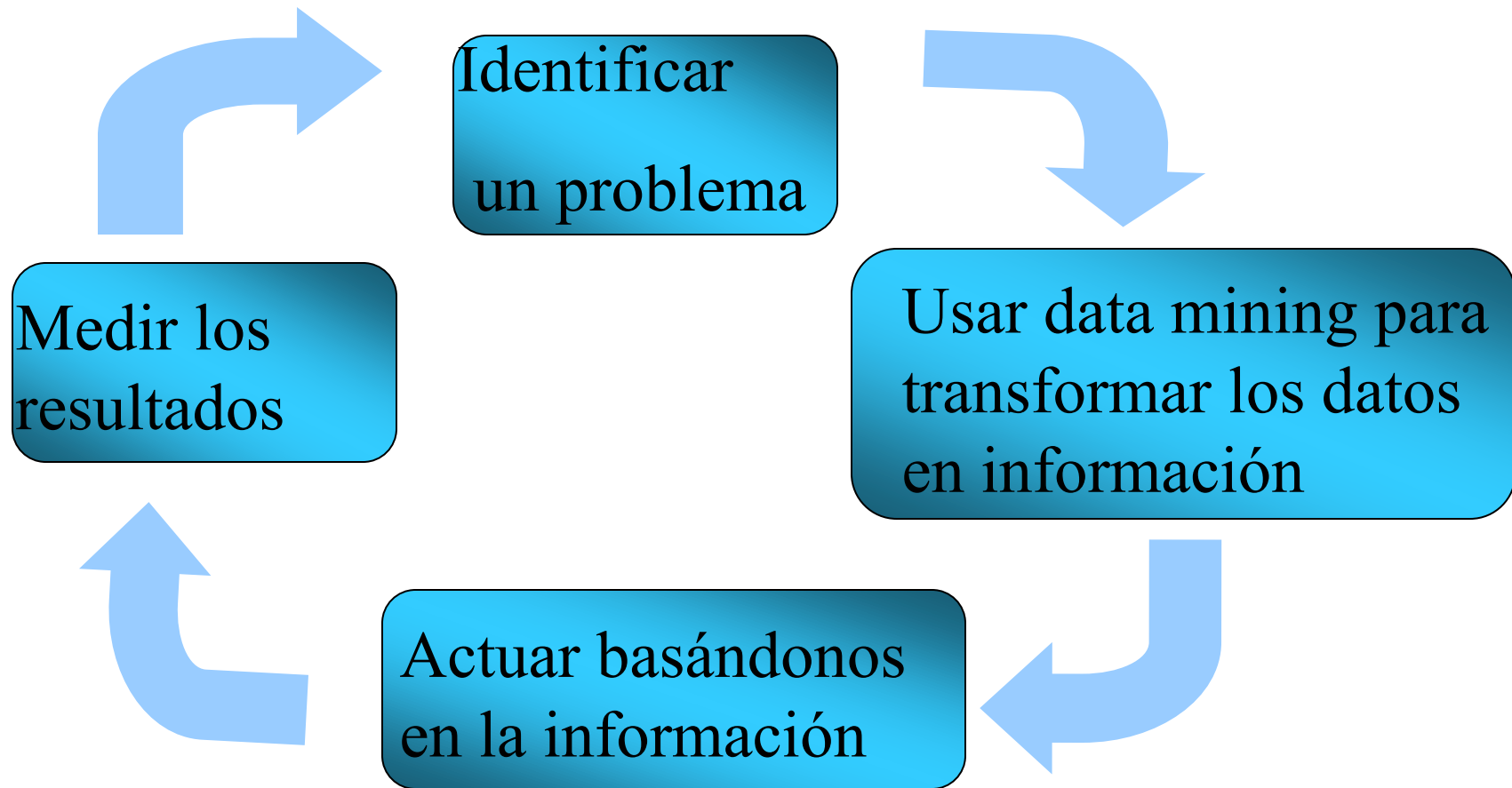
# Minería en otras clases de Datos

- Minería de Datos Espaciales
- Minería espacio-temporal y Object ivos en movimiento
- Minería Cyber-físico de datos del sistema: salud, control de tráfico aéreo, simulación de inundaciones
- Minería de datos multimedia
- Minería de datos de texto
- Minería de datos Web
- Minería de datos Streams

# Importante

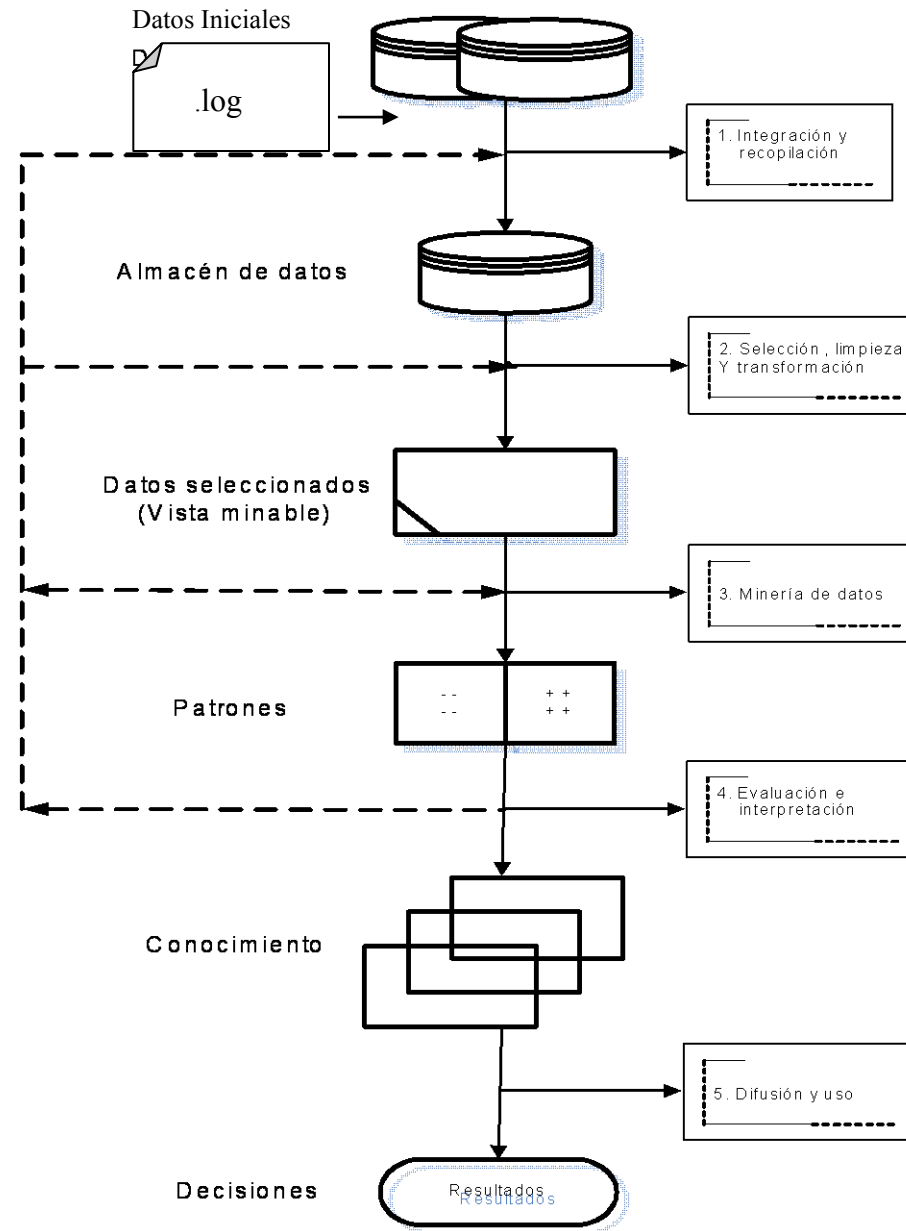
- La promesa de Data Mining es encontrar los patrones
- Simplemente el hallazgo de los patrones no es suficiente
- Debemos ser capaces de entender los patrones, responder a ellos, actuar sobre ellos,

# El ciclo de data mining





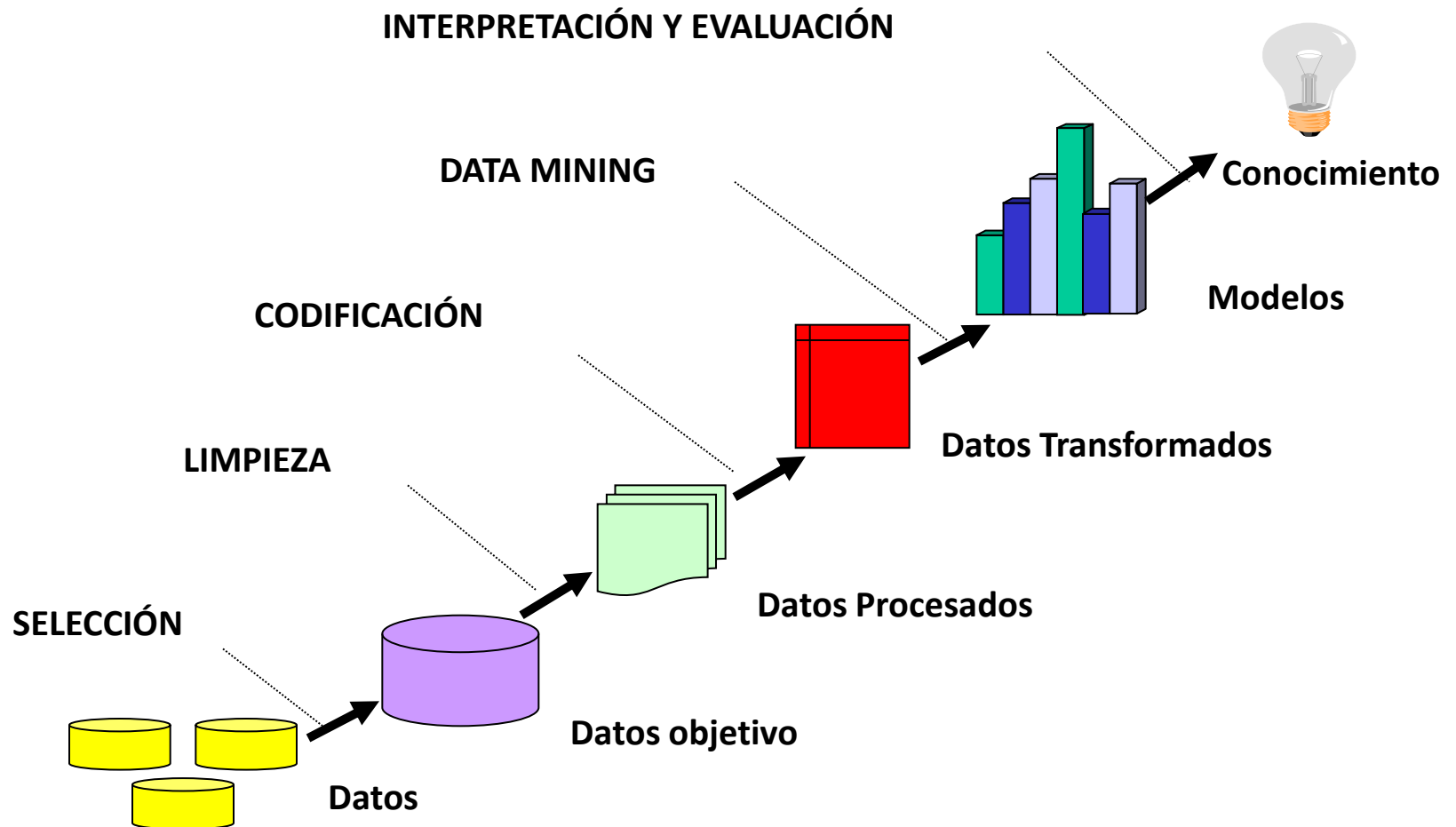
# Esquema básico: Minería de datos



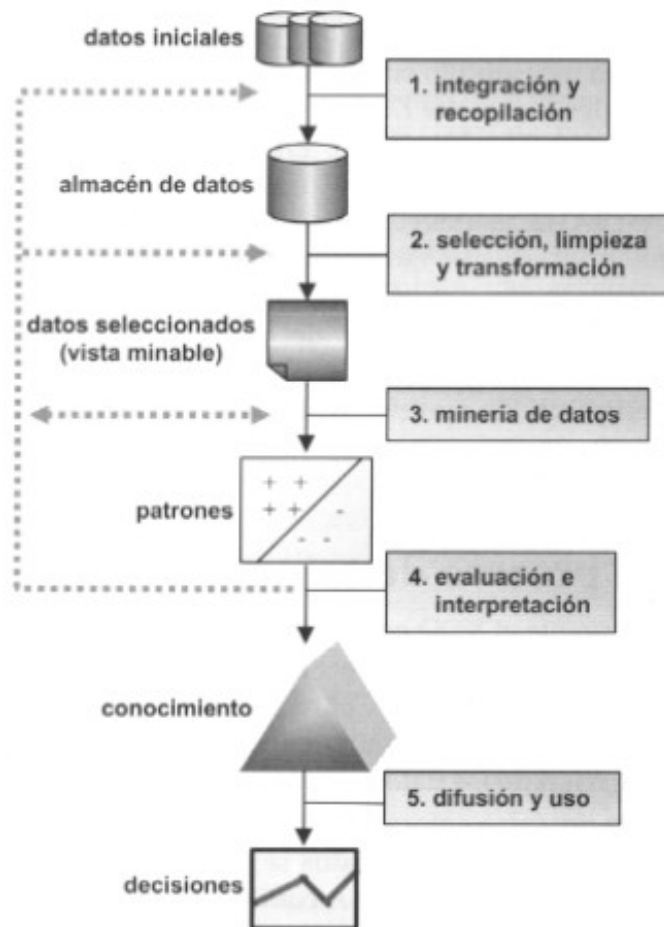
# KDD – Knowledge Discovery in Databases

- La minería de datos se encuadra dentro de un proceso mucho mayor conocido como KDD (Knowledge Discovery from Databases)
- Fases del proceso iterativo e interactivo
  1. Integración y recopilación de datos
  2. Selección, limpieza y transformación
  3. Minería de datos
  4. Evaluación e interpretación
  5. Difusión y uso

# El Proceso de KDD



# KDD



# Fase 1

- Integración y recopilación de datos
  - Decisiones estratégicas basadas en el análisis, la planificación y la predicción: datos en varios departamentos
  - Cada fuente de datos distintos formatos de registro, diferentes grados de agregación, diferentes claves primarias, ....
  - Integración de múltiples bases de datos: almacenes de datos (data warehousing)
  - Almacén de datos aconsejable cuando el volumen de información es grande. No estrictamente necesario (archivos de texto, hojas de cálculo, ...)

# Fase 2

- Selección, limpieza y transformación
  - Calidad del conocimiento descubierto depende (además del algoritmo de minería) de la calidad de los datos analizados
  - Presencia de datos que no se ajustan al comportamiento general de los datos (*outliers*)
  - Presencia de datos perdidos (*missing values*)
  - Selección de variables relevantes (*feature subset selection*)
  - Selección de casos aleatoria en bases de datos de tamaño ingente
  - Construcción de nuevas variables que faciliten el proceso de minería de datos
  - Discretización de variables continuas

# Fase 3

- Minería de datos
  - Modelos descriptivos
    - Reglas de asociación
    - Clustering
  - Modelos predictivos
    - Regresión: regresión lineal, regresión logística
    - Clasificación supervisada: clasificadores Bayesianos, redes neuronales, árboles de clasificación, inducción de reglas, K-NN, combinación de clasificadores

# Fase 4

- Evaluación e interpretación
  - Técnicas de evaluación: validación simple (training + test), validación cruzada con *k-fold*, bootstrap
  - Reglas de asociación: cobertura (soporte), confianza
  - Clustering: variabilidad intra y entre
  - Regresión: error cuadrático medio
  - Clasificación supervisada: porcentaje de bien clasificados, matriz de confusión, análisis ROC
  - Modelos precisos, comprensibles (inteligibles) e interesantes (útiles y novedosos)



# Fase 5

- Difusión y uso
  - Difusión: necesario distribuir, comunicar a los posibles usuarios, integrarlo en el *know-how* de la organización
  - Medir la evolución del modelo a lo largo del tiempo (patrones tipo pueden cambiar)
  - Modelo debe cada cierto tiempo de ser:
    - Reevaluado
    - Reentrenado
    - Reconstruido

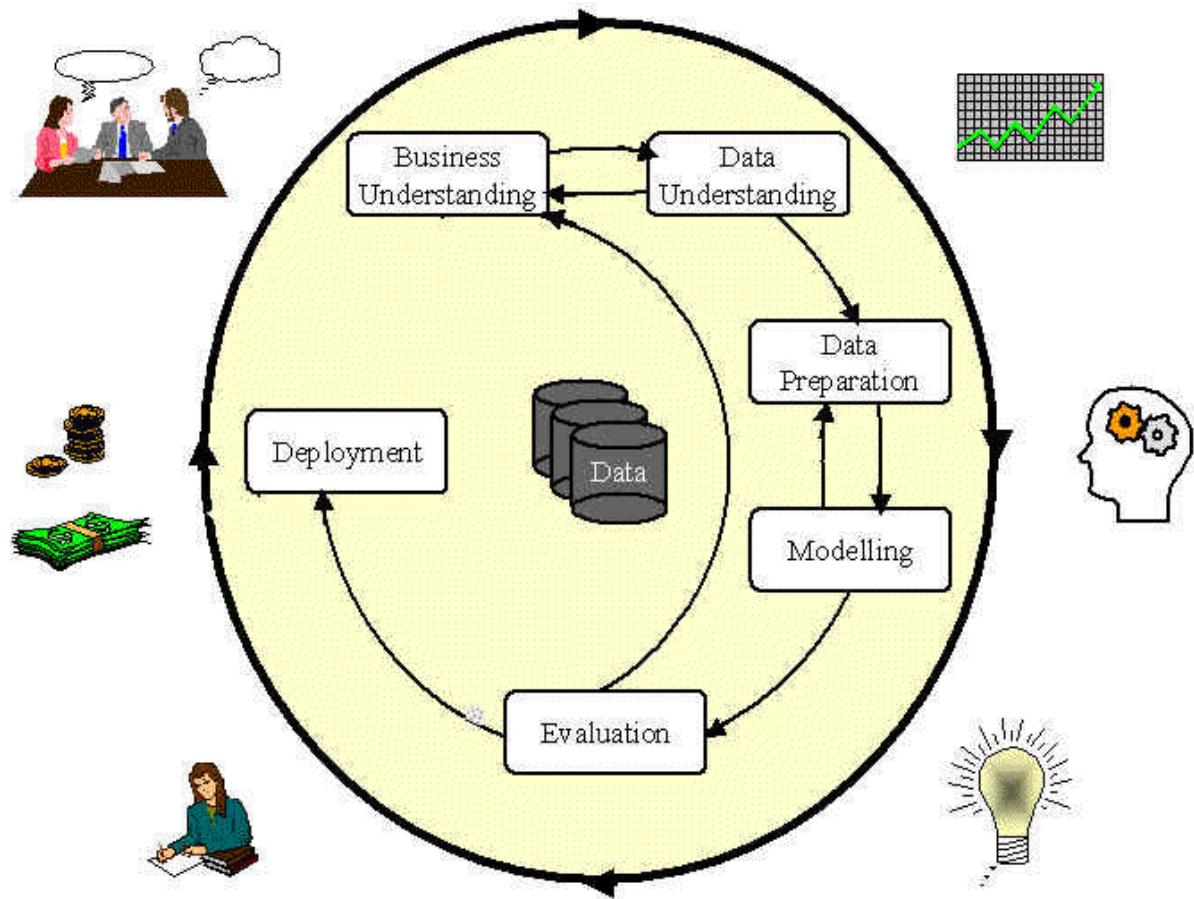
# CRISP-DM

El proceso de extracción de datos debe ser fiable y repetible por personas con pocos conocimientos de minería de datos

CRISP-DM proporciona un marco uniforme para  
directores  
documentación de la experiencia

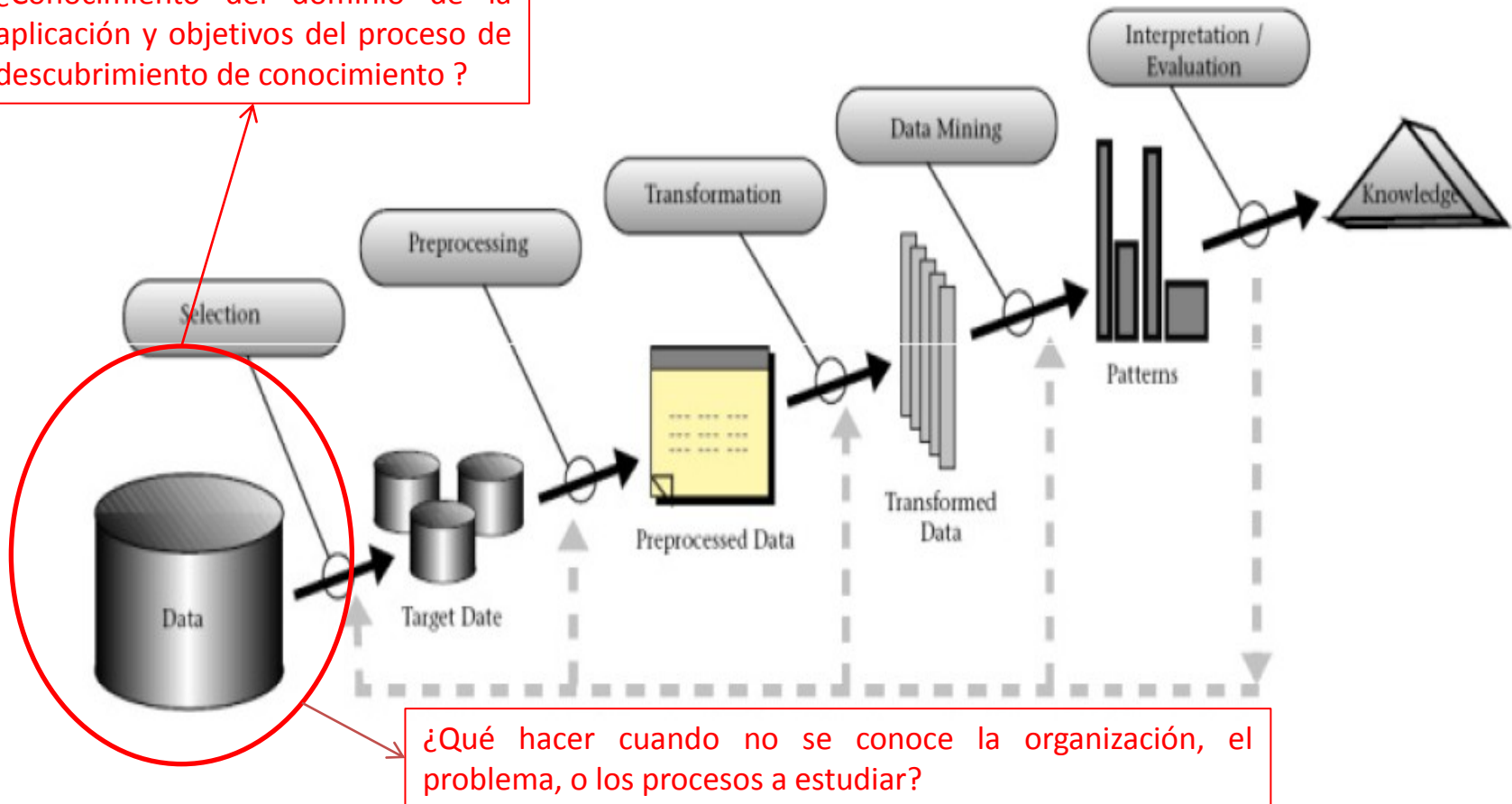
CRISP-DM es flexible para tener en cuenta las diferencias  
Diferentes problemas de negocio  
Datos diferentes

# Fases en CRISP-DM



# Problemas

¿Conocimiento del dominio de la aplicación y objetivos del proceso de descubrimiento de conocimiento ?



¿Qué hacer cuando no se conoce la organización, el problema, o los procesos a estudiar?

# Metodología ULA

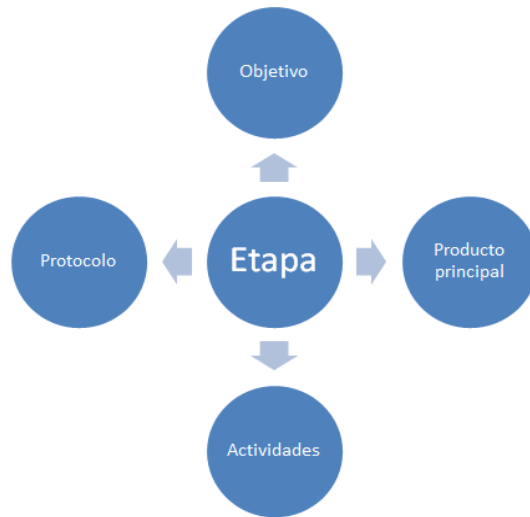
- Esta metodología surge de la necesidad de conducir de manera adecuada un proyecto de MD, donde además, se integra un guía detallada de lo que se debe hacer en cada etapa del proceso de KDD, para llenar algunos vacíos que presentan otras metodologías a la hora de enfrentarse con organizaciones, que no conocen los problemas que se presentan, y que podría aportar la MD para la mejora de la misma.
- Esta metodología permite identificar y conceptualizar la solución de un problema desde la perspectiva del desarrollo de aplicaciones basadas en MD, para un proceso de cualquier institución/empresa

# Metodología ULA

- Todas las fases y actividades propuestas pretenden abarcar todo el dominio de conocimiento que puede encontrarse en una organización.
- Sin embargo, en función de los expertos disponibles y del tiempo previsto de desarrollo, no todas las actividades propuestas en cada fase podrían desarrollarse en su totalidad.

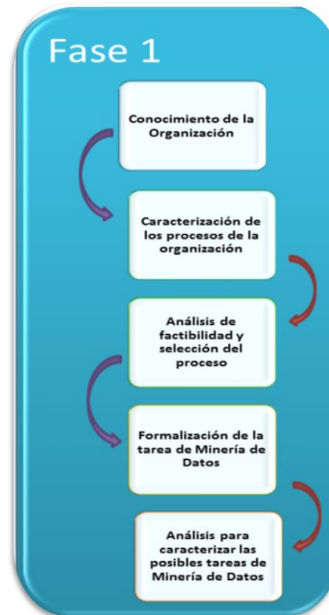


# Metodología ULA



Elementos que conforman cada etapa de las fases de la metodología.

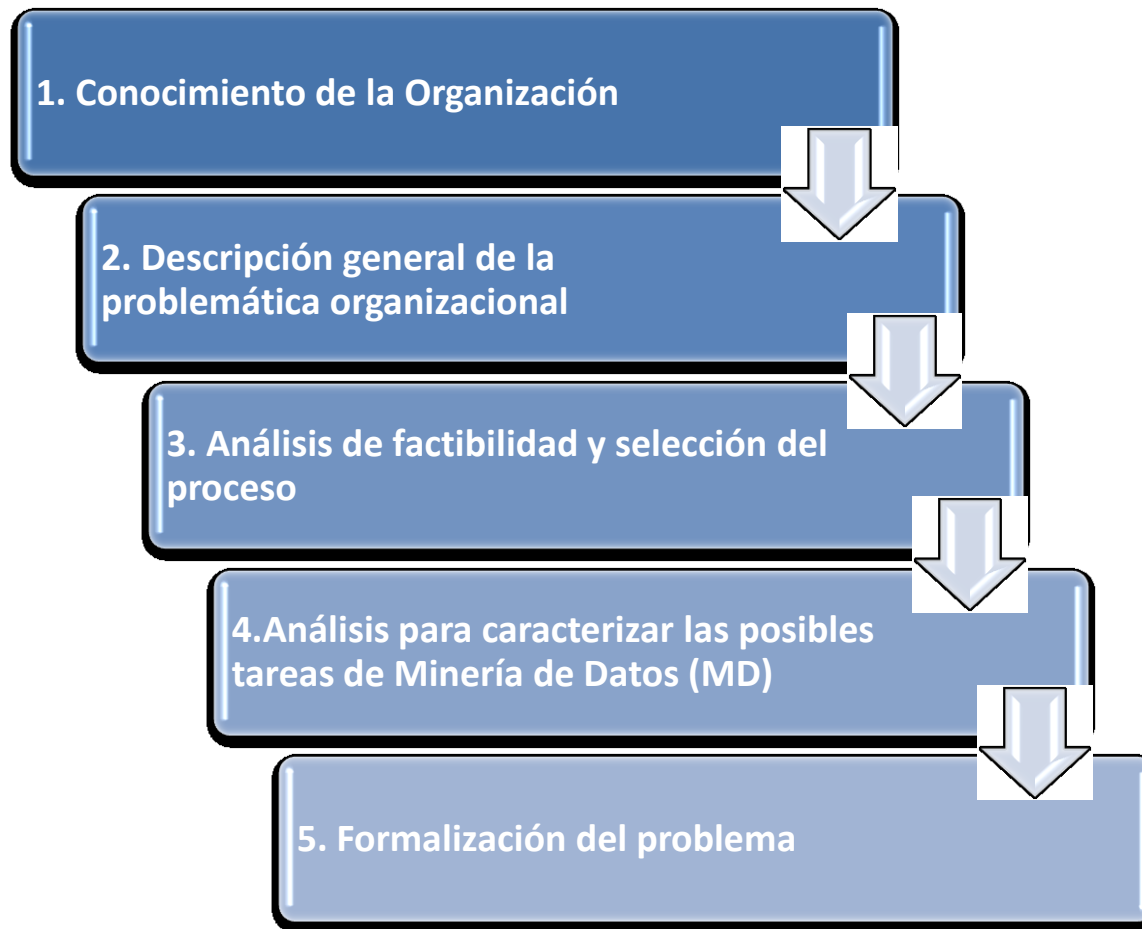
# ***Fase 1: Conocimiento de la Organización***



Esta fase tiene como finalidad realizar un proceso de ingeniería de conocimiento, orientado a organizaciones/empresas, de las cuales no se conoce o se tiene poca información del (de los) problema(s), o los procesos a estudiar. Se enfoca en identificar y conceptualizar la solución de un problema, desde la perspectiva del desarrollo de aplicaciones basadas en MD.



# ***Fase 1: Conocimiento de la Organización***



# Etapa 1: Conocimiento de la Organización

1. Objetivo {
- Conocer la organización/empresa, sus objetivos, procesos, objetos y actores

## 2. Protocolo de la Fase:

- Descripción de los elementos de la institución/empresa y sus características. Objetivos, Procesos , Objetos y Actores.
- Descripción de las relaciones entre estos elementos.
- Organización de estos elementos.


# Etapa 2: Descripción general del problema

1. Objetivo {
- Conocer los procesos sobre los cuales se puede enfocar el proyecto de minería de datos.

## 2. Protocolo de la Fase:

- Familiarización con los procesos sobre los cuales se puede realizar la ingeniería de conocimiento
- Identificación de la fuente de conocimiento
- Familiarización con los ambientes computacionales donde se encuentran los datos a ser utilizados en cada proceso.

# Etapa 3: Análisis de factibilidad y selección del proceso

1. Objetivo
- 
- Analizar los procesos con la información proporcionada/recogida.

## 2. Protocolo de la Fase:

- Revisión de los procesos propuestos por los expertos
- Disponibilidad del experto o grupo de expertos
- Análisis de las fuentes de información sobre los procesos

# Etapa 4: Análisis para caracterizar las posibles tareas de Minería de Datos

## 1. Objetivo

- Caracterizar las posibles tareas de minería de datos a realizar en el(los) proceso(s) seleccionado(s) en la fase anterior (objetivos, requerimientos, factibilidad, etc.), con la finalidad de escoger las tareas de MD de interés a desarrollar.

## 2. Protocolo de la Fase:

- Selección y descripción de los actores.
- Descripción de los escenarios actuales y posibles escenarios futuros de la institución/empresa.
- Especificación de los requerimientos para los posibles escenarios futuros (donde se puedan aplicar tarea(s) de MD)
- Elaboración de los casos de uso para los requerimientos funcionales

# Etapa 5: Formalización del Problema

## 1. Objetivo

- Definir el(los) problema(s) formales de MD.

## 2. Protocolo de la fase

- Desarrollo de un informe, con la conceptualización del proceso a estudiar, la caracterización de sus problemáticas operacionales y del uso de la MD en dicho proceso.

# Caso de Estudio: Empresa Petrolera

## ***Fase 1: Conocimiento de la organización:***

Se trata de una empresa que se encarga de la exploración, extracción, producción, mejoramiento y comercialización de crudo extrapesado.

## ***Fase 2: Caracterización de los procesos de la organización***

La cadena de valor de la empresa se muestra en la siguiente figura, donde el proceso principal objeto de estudio se concentra en la tercera etapa de la cadena de valor.



# Caso de Estudio: Empresa Petrolera

## ***Fase 3: Selección del Proceso***

Se estudió cada uno de los subproceso (objetivos, actividades, productos, etc.), y se obtuvo la interacción entre ellos.

En la tabla se ilustra este proceso de priorización y selección, considerando sólo los dos procesos que resultaron mejor ponderados en este caso de estudio.

Criterios	CDU	DCU
Importancia para la organización	5	5
Propósito de la MD	5	5
Interacciones entre procesos	2	4
Procesos dependientes	5	3
Importancia de la calidad del producto	4	4
Seguridad Industrial	4	5
Replicabilidad de la herramienta desarrollada	5	4
Cantidad de Expertos	5	5
Fuentes de información	5	5
Confidencialidad de la información	3	3
¿Qué información se recoge del proceso para ser almacenada?	5	5
Con que frecuencia se recoge la información almacenada	4	4
¿Qué herramientas se cuentan, para recolectar y manipular la información?	4	4



# Caso de Estudio: Empresa Petrolera

## *Fase 4: Análisis para caracterizar las posibles tareas de Minería de Datos (MD)*

- Descripción del escenario actual

Resultados que se obtienen	Actor(es) asociado(s)	Variables Asociadas	Actividades que se realizan
<b>Gasoil directo (SRGO), nafta pesada y residuo atmosférica.</b>	<ul style="list-style-type: none"> <li>• Expertos asociados al proceso</li> <li>• Ingenieros de Procesos</li> <li>• Operadores</li> <li>• Unidad de destilación atmosférica</li> </ul>	<ul style="list-style-type: none"> <li>• Tren de precalentamiento: temperatura de la carga.</li> <li>• Desaladores: tiempo para el asentamiento y separación del agua del petróleo, presión.</li> <li>• Hornos de crudo: temperatura</li> <li>• Columna de crudo: presión, temperatura, rata de vapor de despojamiento.</li> </ul>	<ul style="list-style-type: none"> <li>• Carga del crudo.</li> <li>• Precalentamiento del crudo diluido.</li> <li>• Desalado.</li> <li>• Precalentamiento del crudo desalado.</li> <li>• Generación de cortes de crudo en la columna.</li> </ul>

# Caso de Estudio: Empresa Petrolera

- Descripción del escenario futuro

Resultados que se desean obtener	Actor(es) asociado(s)	Variables Asociadas	Actividades de MD que se realizarían	Funcionalidades nuevas
<b>Predicción de la calidad del producto, para optimizar el proceso</b>	<ul style="list-style-type: none"><li>• Expertos asociados al proceso</li><li>• Operadores</li><li>• Columna de crudo</li></ul>	Presión, temperatura de tope y rata de vapor de despojamiento de la columna de crudo.	Predicción	<ul style="list-style-type: none"><li>• Predicción de las características del producto, según las condiciones de funcionamiento de la torre de crudo.</li><li>• Ayudar a optimizar el proceso de producción, generando información para orientar a los actores en la toma de decisiones con la predicción (es) resultante(s).</li></ul>

## ***Fase 2: Preparacion de los Datos***

- Para aplicar MD sobre un problema en específico, es necesario contar con un historial de datos asociado al problema en estudio.
- Esto conlleva realizar distintas operaciones con los datos, con la finalidad de acondicionarlos para desarrollar un modelo de MD. Por lo tanto, en esta fase se plantea realizar la preparación de los datos, que serán utilizados para el desarrollo de la herramienta de MD.
- Para realizar tareas de MD es necesario tener los datos integrados en una sola vista, que denominaremos vista minable la cual está compuesta por una tabla con todas las variables del proceso y los datos a considerar en el estudio de MD.

# ***Fase 2: Preparacion de los Datos***

En particular, para realizar este proceso nosotros definimos dos tipos de vista minable:

- Vista Minable conceptual (VMC): describe en detalle cada una de las variables a ser tomadas en cuenta para la tarea de MD, para cada escenario futuro seleccionado. La misma está compuesta por todas las variables de interés, y algunos campos adicionales de importancia para realizar el proceso de tratamiento de datos.
- Vista Minable operativa (VMO): Es el resultado de cargar los datos del historial y de realizar la etapa de tratamiento de datos, con la información de la VMC.

Tanto en la VMC, como en la VMO, se identifican ciertas variables llamadas “*variables objetivo*”. Estas variables se detectan una vez planteado el escenario futuro y la tarea de MD a realizar. Dichas variables son las que permitirán la consecución de los objetivos de MD, ya que las mismas son las que se desean predecir, clasificar, calcular, inferir, en otras palabras, es la que deseamos obtener con la tarea de MD.

# ***Fase 2: Preparacion de los Datos***

Etapas que conforman la fase 2.



# ***Caracterización de los datos del Dominio de la aplicación***

## ***Objetivos***

- Ubicar y comprender los datos asociados a el(los) escenario(s) futuro(s)
- Construcción de una VMC que tenga las variables de interés para el caso de estudio
- Construcción de una VMO
- Definición de la(s) variable(s) objetivo(s) en la vista minable operativa

## ***Productos principales***

- Características de los repositorios donde se encuentran los datos
- VMC
- VMO
- Descripción de la(s) variable(s) objetivo(s)
-

# ***Caracterización de los datos del Dominio de la aplicación***

## ***Protocolo de la etapa***

### ***Comprensión de la fuente de datos de entrada:***

- Comprensión de los datos asociados a las variables
- Determinación de los repositorios de datos

### ***Construcción de la VMC:***

- Realizar un primer filtrado..
- Establecer las relaciones entre las variables seleccionadas (dependencia entre variables, redundancia, variables que son producto de fórmulas).
- Extender la VMC

# ***Caracterización de los datos del Dominio de la aplicación***

## ***Protocolo de la etapa***

- ***Construcción de la VMO:***
  - Si se encuentran en diferentes repositorios, ubicarlos
  - Observar la organización en la que están dispuestos los datos en cada repositorio, y como se almacenan
  - Definir una estrategia para unificar los datos en un solo repositorio (escoger una herramienta, puede ser un lenguaje de procesamiento matemático).
  - Integrar los formatos.
  - Crear la VMO resultante de la integración de los datos asociados a las variables escogidas en la VMC (fusión de tablas, integración de bases de datos, entre otros).
- ***Definir las variables objetivo:***
  - Identificar a qué conclusiones puede llegar el experto humano teniendo en cuenta las entradas
  - Observar el objetivo en el escenario futuro seleccionado e identificar ¿Cuál de las variables llevan a dicho objetivo?
  - Escoger la(s) variable(s) objetivo(s)



# ***Tratamiento de datos***

## ***Objetivos***

- Esta etapa se centra en generar datos de calidad, es decir datos sin anomalías, sin inconsistencias de formato, sin capturas erróneas, sin campos vacíos; aplicando métodos de limpieza, transformación y reducción sobre la vista minable operativa.

## ***Productos principales***

- VMO

# ***Tratamiento de datos***

## ***Limpieza:***

- Estudiar la representación de cada una de las variables.
- Buscar anomalías de representación.
- Definir alguna estrategia de limpieza para erradicar dichas anomalías y obtener data consistente.
- Escoger las herramientas tecnológicas para realizar las operaciones de limpieza de datos definidas.

# *Tratamiento de datos*

## ***Transformación:***

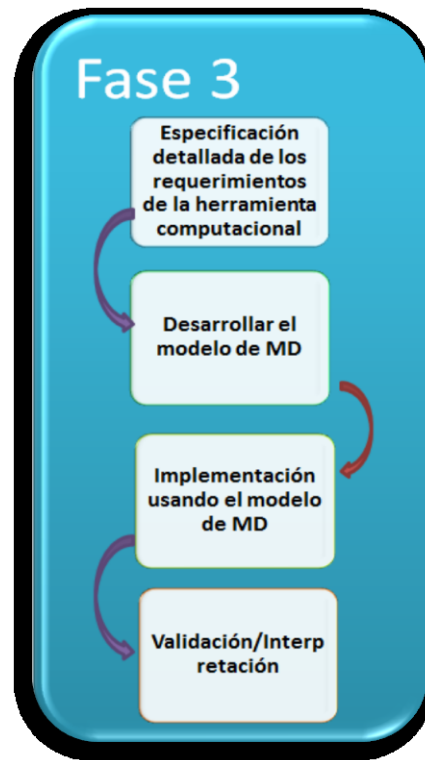
- Estudiar las representaciones de cada una de las variables
- Identificar las representaciones que se puedan transformar en otra representación más conveniente o fácil de utilizar a la hora de aplicar la tarea de MD, como por ejemplo, cambiar las unidades de las variables para trabajar con un único sistema de medición. Otro ejemplo puede ser cambiar las unidades de medición de las variables.
- Ordenar dichas transformaciones que se desean aplicar en una tabla, para observar las equivalencias.
- Aplicar la transformación con un software seleccionado
- Identificar las variables que potencialmente se pueden normalizar, un ejemplo de ello es normalizar los valores de las variables en un rango que no sea muy grande (por ejemplo un número de 0 a 1).
- Definir la función(es) de normalización para cada una de las variables seleccionadas en el paso anterior y ordenarla en tablas.
- Aplicar la función(es) de normalización en las variables seleccionada
- Describir en tablas cada una de las transformaciones y normalizaciones realizadas.

# *Tratamiento de datos*

- ***Reducción:***
- Realizar análisis estadísticos para reducir variables que posean una alta relación lineal, como por ejemplo un análisis de correlación.
- De ser necesario, combinar variables por un método seleccionado, tal como el PCA (del inglés *Principal Component Analysis*) que es considerado también un método para reducción de variables.
- Identificar las posibles variables que se pueden reducir.
- Justificar la reducción de las mismas
- Construir la nueva vista minable con las nuevas variables reducidas

# ***Fase 3: Desarrollo de herramientas de MD***

Esta fase busca generar una herramienta de *software* que permita utilizar el modelo de MD, en los escenarios futuros donde se identificaron su potencial uso. El producto principal de esta fase es la aplicación de MD como tal, y el modelo de MD.



# Técnicas de Minería de datos:

- Árboles de decisión,
- Reglas de asociación,
- Redes Neuronales Artificiales,
- Las tablas de decisión

# Arboles de decisión

Los árboles de decisión son unos de los algoritmos clasificadores más conocidos y usados en las tareas de Data Mining, ya que son una forma de representación sencilla para clasificar instancias.

# ÁRBOLES DE DECISION

Árboles de decisión son particiones secuenciales de un conjunto de datos

**Objetivo:** Segmentar la población para encontrar grupos homogéneos según una cierta variable de respuesta. Esta técnica permite representar de forma gráfica una serie de reglas sobre la decisión que se debe tomar en la asignación de un determinado elemento a una clase o valor de salida.



# Tablas de decisión

Es la forma más simple y más rudimentaria para representar la salida de la máquina de aprendizaje.

Outlook	Temperature	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

# Ejemplos

## Criterios

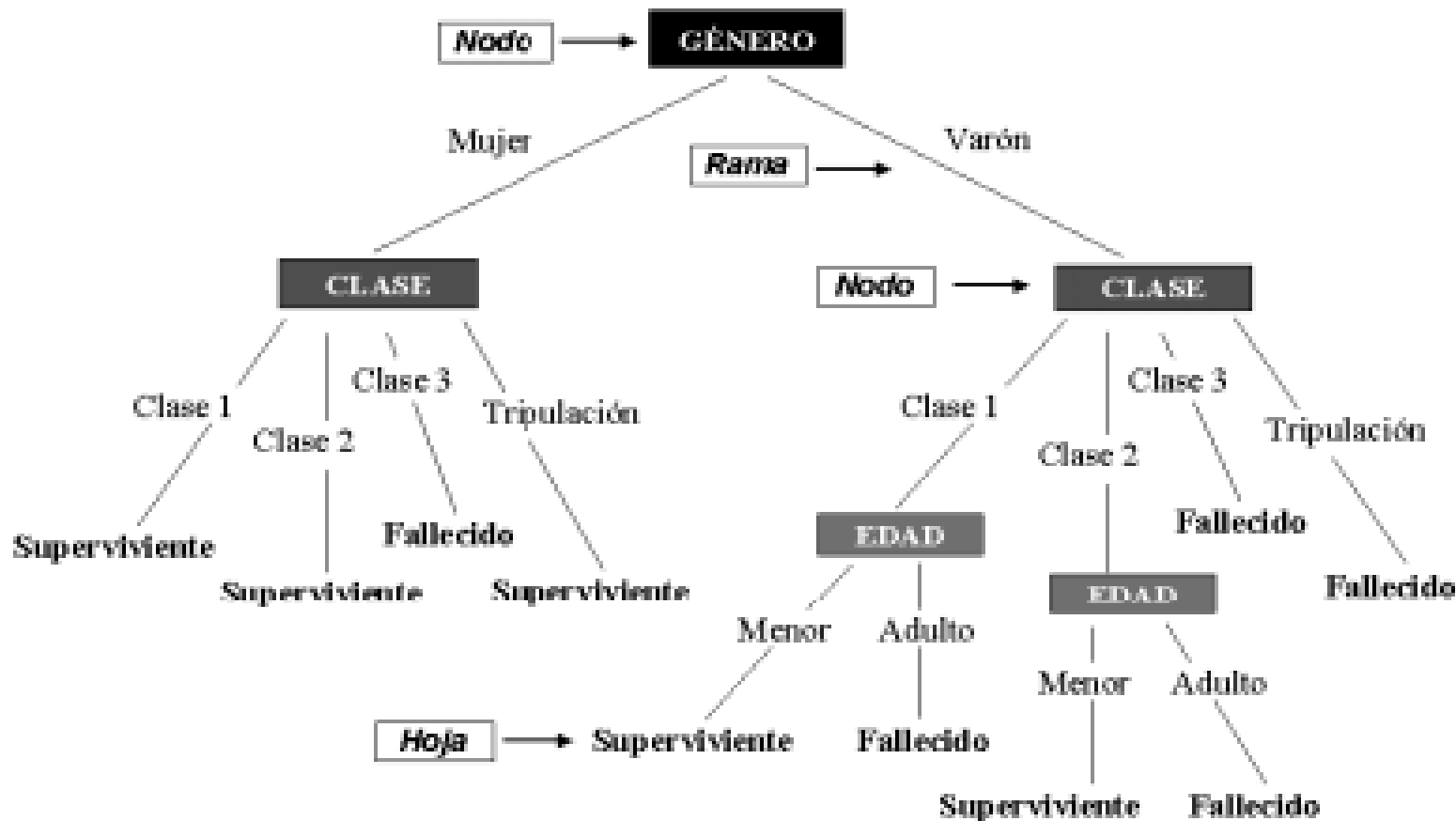
Qué aprendo?

Ejemplos

Ej	Alt	Bar	Dia	EdM	Patr	Prec	EdD	Tipo	RES	T --->	Espera
X1	S	N	N	S	Alg	\$\$\$	N	Franc	S	0-10	S
X2	S	N	N	S	Ilen	\$	N	Jap	S	10-15	N
X3	N	S	N	N	Alg	\$	N	Hamb	N	0	S
...											
X12	S	S	S	S	Ilen	\$	N	Hamb	N	10	S

# ÁRBOLES DE DECISION

Los árboles de decisión suelen ser empleados en tareas de clasificación, y también, aunque en menor medida, en tareas de predicción

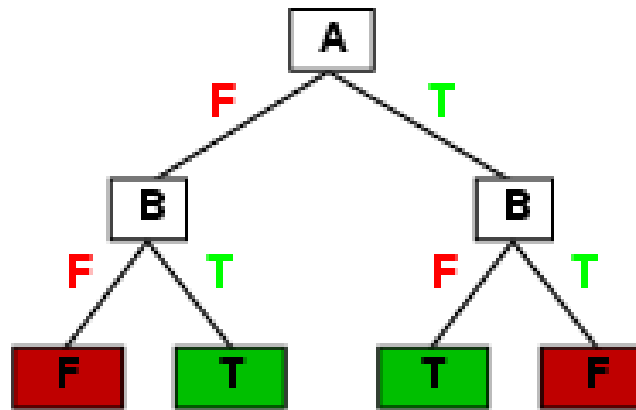


Ej. Acontecimientos relativos al hundimiento del Titanic

# Árbol de Decisión

- Puede expresar cualquier función a partir de sus atributos de entrada.
- Un árbol de decisión es consistente para cualquier conjunto de entrenamiento, cuando hay un **camino a una hoja para uno o varios ejemplos**
- Basado en la idea de **tablas de la verdad**:

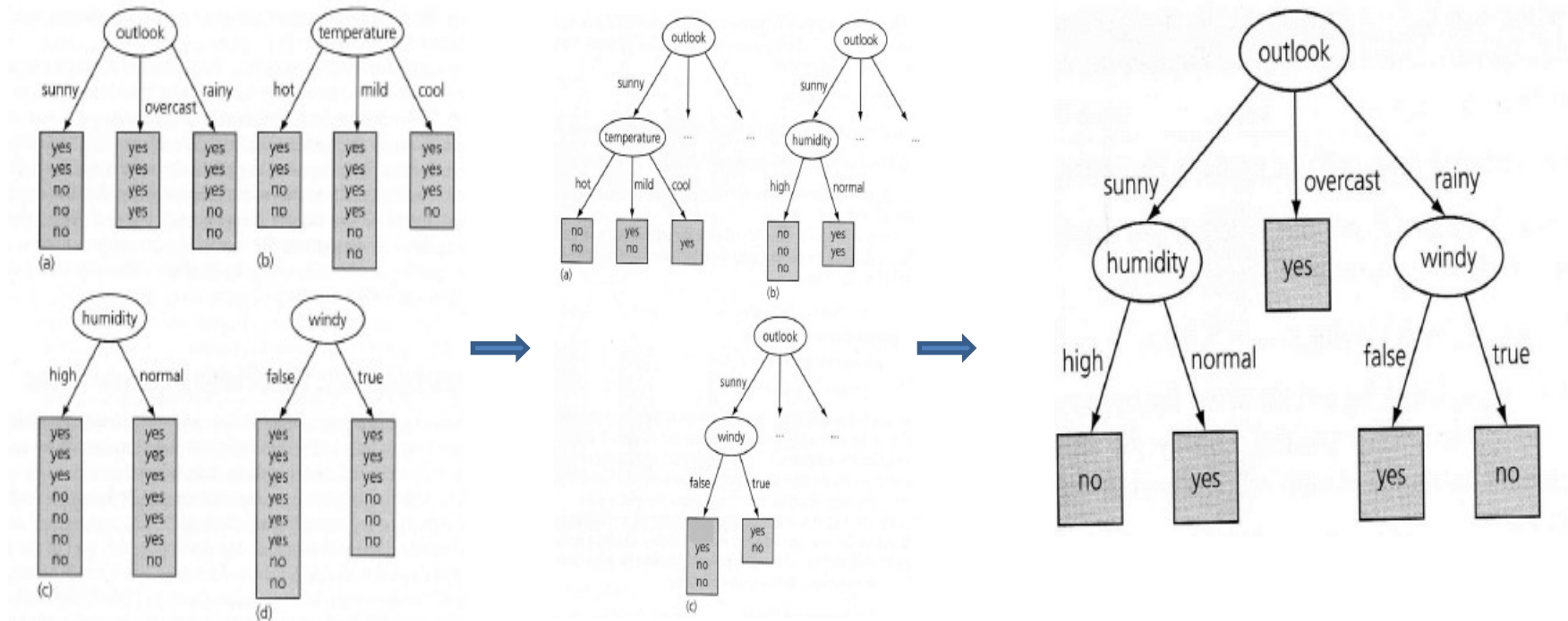
A	B	A xor B
F	F	F
F	T	T
T	F	T
T	T	F



Es una estrategia de aprendizaje inductivo

# Arboles de decisión

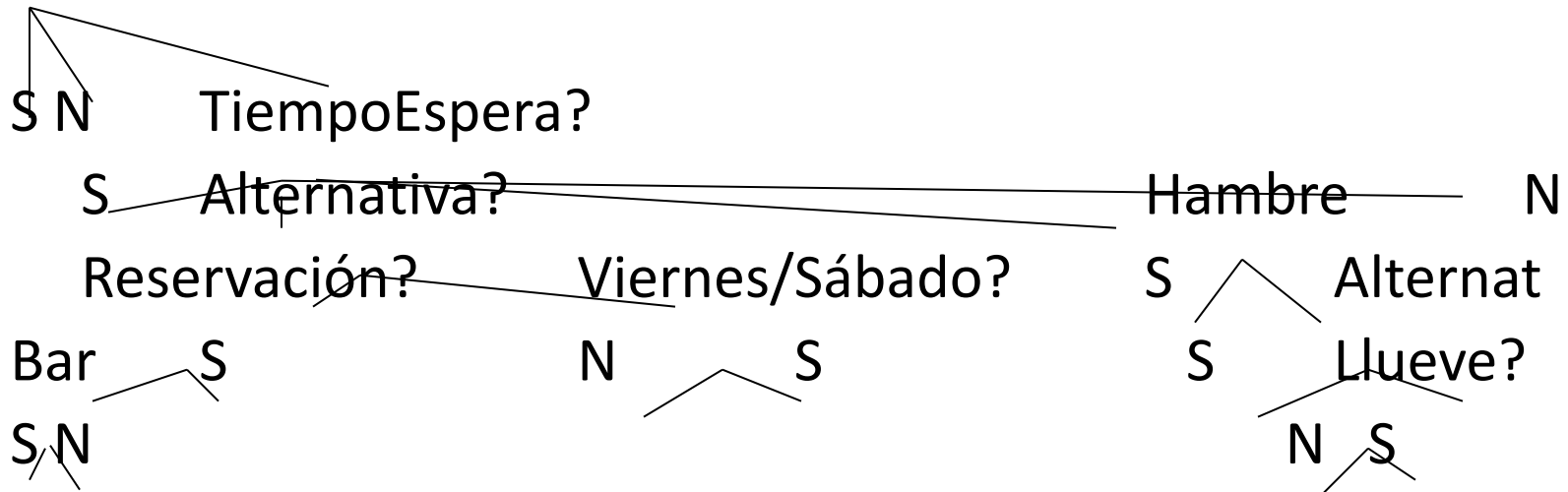
## transformaciones



# Árbol de Decisión

- Para nuestro ejemplo inicial:

Patrón?



# Árbol de Decisión

- Idea: escoger atributo "más significativo" como raíz del (sub)-árbol

## ¿Cómo?

- Si hay + y - ejemplos escoger atributo que mejor los divida (mayor discriminante)
- Si hay particiones con + y -, buscar un 2do atributo para seguir partiendo

## Macroalgoritmo AD(ejemplos, atributos)

Si ejemplos no vacíos entonces

    Si ejemplos clasificados entonces

        regresar (clasificación)

    de lo contrario

        mejor: escoger\_atributo(atributos, ejemplos)

        arbol: un nuevo árbol de decisión con *mejor* como raíz

        por cada valor  $V_i$  de mejor

            Subejemplos: ejemplos con  $mejor = V_i$

            Subarbol: AD(Subejemplos, atributos)

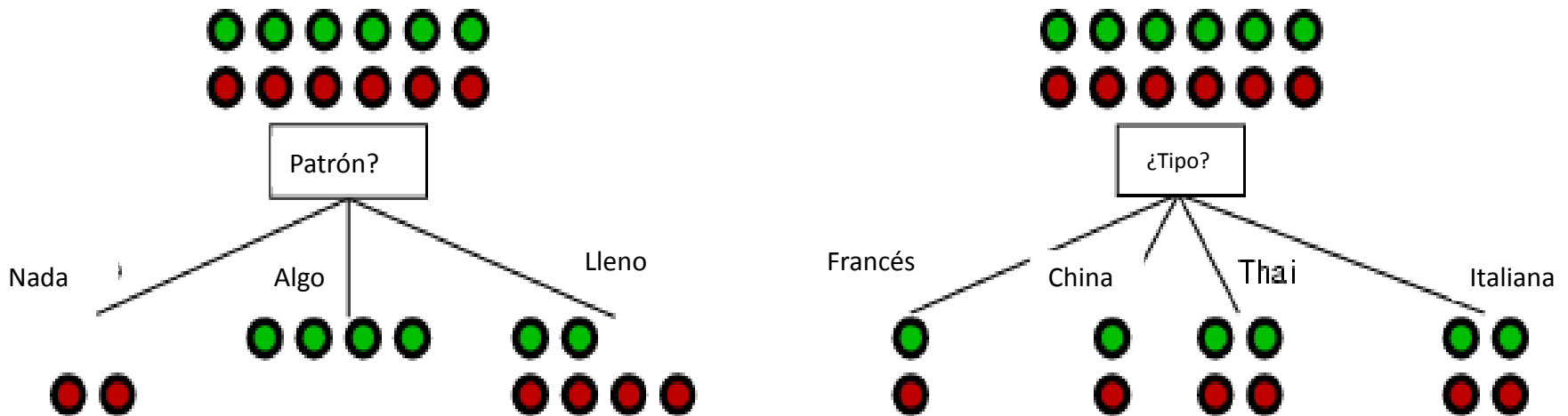
            Arbol: actualizar(nueva rama con etiqueta  $V_i$  y Subarbol)

Regresa(arbol)

# Escoger un atributo

aprender reglas (clases)

- ¿*Patrón* es una mejor escogencia que *Tipo*?



Basado en concepto de *contenido de información*

Parte de 
$$Info(p, n) = -p \log_2(p) - n \log_2(n)$$

Es una medida de la entropía (grado de desorden) de los ejemplos

n: numero de ejemplos -

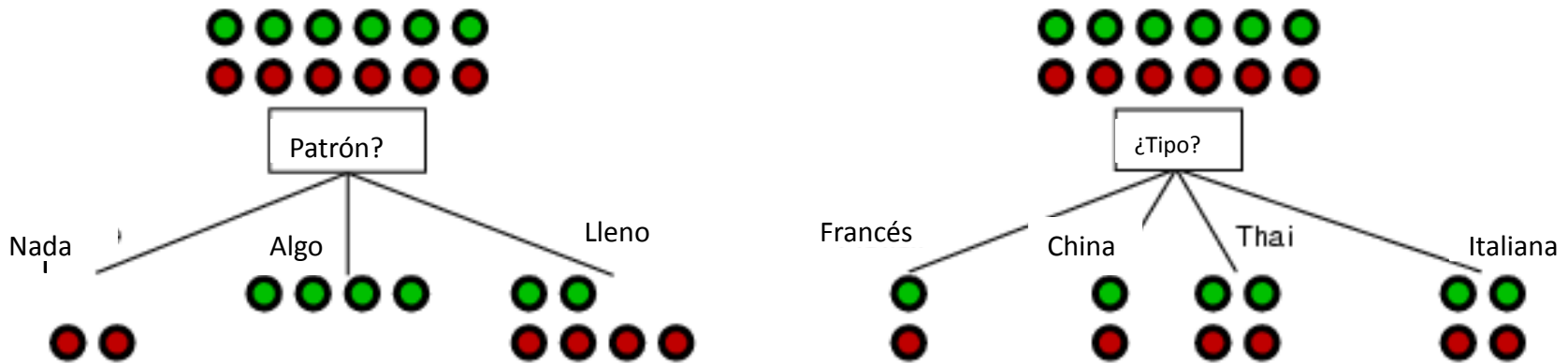
p: numero de ejemplos +



# Escoger un atributo

aprender reglas (clases)

¿*Patrón* es una mejor escogencia que *Tipo*?



Donde:

$I$  es entropía de los ejemplos:

y

$$IG(A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - \text{resto}(A)$$

v. posibles valores de A  
 $p_i$  y  $n_i$ ? ver siguiente lamina

$$I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$$

$$\text{resto}(A) = \sum_{i=1}^v \left| \frac{p_i - n_i}{p+n} \right| I\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right)$$

# Arbol de Decisión y Lógica de Predicado

¿Quién es  $p_i$ ?  $p_i$  puede ser 
$$p_i = \frac{|E_i^+|}{|E_i^+| + |E_i^-|}$$

Donde  $E_i^+$  es el porcentaje de ejemplos clasificados como + por el valor  $v$  del atributo A

## Una Formula general para escoger a los atributos:

Como hay que elegir el atributo con mayor información (menor entropía), otra posibilidad es calcular una función de merito (FM)

$$FM(A) = \sum_{i=1}^v r_i \inf o(p_i, n_i)$$

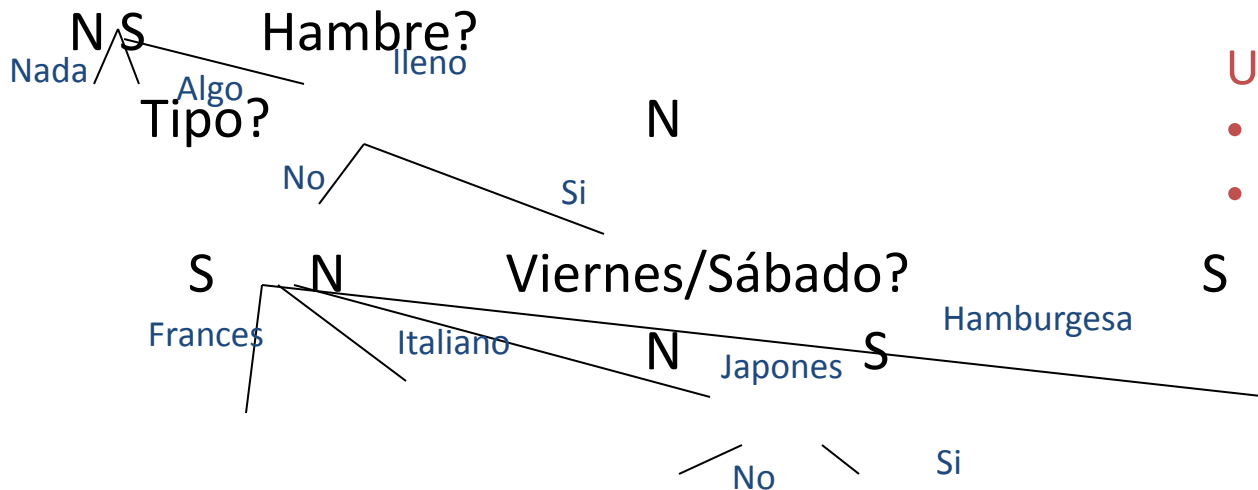
$p_i$  = % ejemplos clasificados como + en la rama  $i$

$$r_i = \left| \frac{p_i - n_i}{p + n} \right|$$

# Arbol de Decisión y Lógica de Predicado

$$\forall r \text{ espera}(r) \Rightarrow \text{Patrón}(r, \text{ algo}) \vee (\text{Patrón}(r, \text{ full}) \wedge \text{NoHambre}(r) \wedge \text{tipo}(r, \text{ francés})) \vee (\text{Patrón}(r, \text{ full}) \wedge \text{NoHambre}(r) \wedge \text{tipo}(r, \text{ hamburguesa})) \vee (\text{Patrón}(r, \text{ full}) \wedge \text{NoHambre}(r) \wedge \text{tipo}(r, \text{ Japones})) \wedge \text{viernes/Sabado}(r)$$

# Patrón?



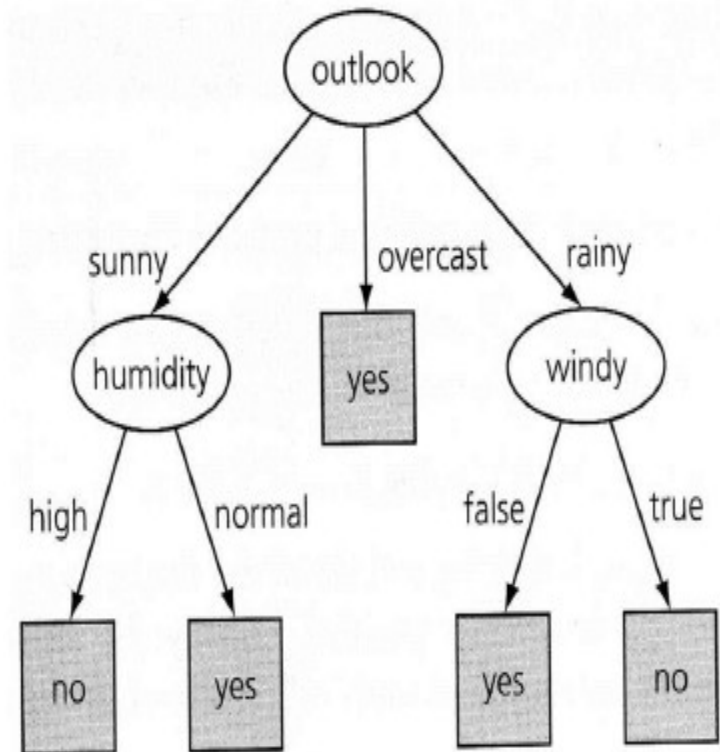
## Uso de operadores:

- Para unir ramas O
- Para seguir una rama Y

# Construcción de árboles de decisión

Se completa el árbol completando cada rama hasta cumplir un cierto compromiso:

- Número mínimo de elementos de una hoja.
- **Cobertura:** Mínimo número (o porcentaje) de casos posibles cubiertos correctamente de la BD.
- **Precisión:** Error de clasificación menor de un umbral puesto. Por ejemplo: precisión del 80%. Significa, que pararemos en esa hoja cuando el número de clases clasificadas correctamente sea mayor o igual al 80%.



# Podado de un Árbol

¿Cómo decidir si desea  
reemplazar un nodo interno con  
una hoja?

Imaginemos que la verdadera probabilidad de error en el nodo es  $q$ , y que las  $N$  instancias son generados por un proceso de Bernoulli con parámetro  $q$ , de la que  $E$  son los errores. El intervalo de confianza viene dado por:

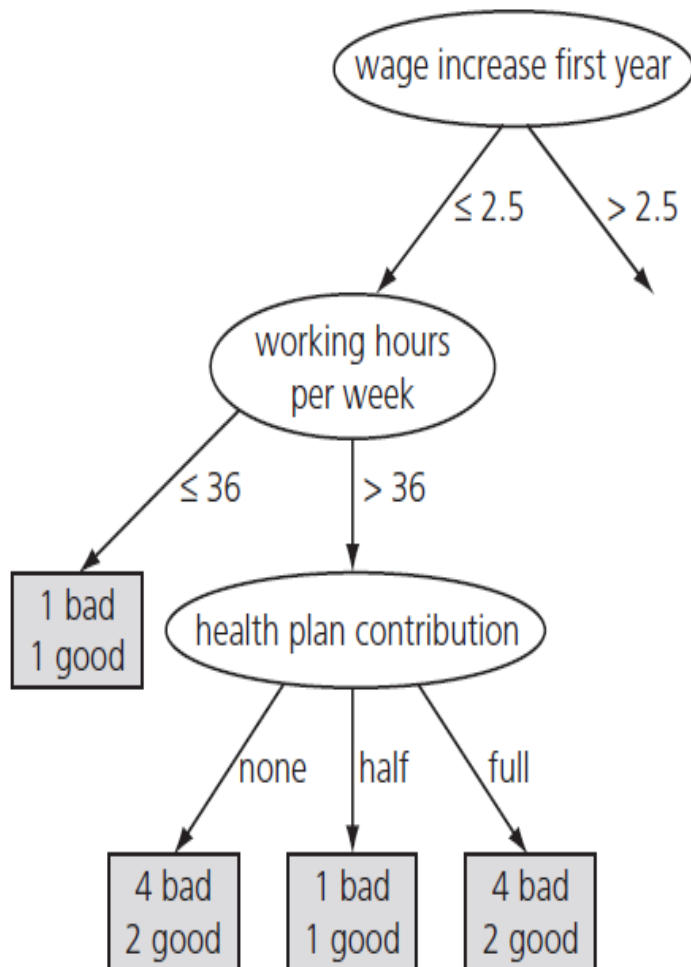
$$\Pr\left[\frac{f - q}{\sqrt{q(1-q)/N}} > z\right] = c,$$

donde  $N$  es el número de muestras,  $f = E / N$  es el porcentaje de error observado, y  $q$  es la tasa de error. Al igual que antes, esto conduce a un límite superior de confianza para  $q$ .

Ahora usamos ese límite superior de confianza como una estimación (pesimista) para la tasa de error  $e$  en el nodo:

$$e = \frac{f + \frac{z^2}{2N} + z \sqrt{\frac{f}{N} - \frac{f^2}{N} + \frac{z^2}{4N^2}}}{1 + \frac{z^2}{N}}.$$

# Podado de un Árbol



None:  $E = 2$ ,  $N = 6$ , y por lo que  $f = 0,33'$ .  $e = 0,47$ . tasa de error de formación es del 33%, se utilizará la estimación pesimista del 47%.

Half:  $E = 1$ ,  $N = 2$ ,  $e = 0.72$ .

Full: Tiene el mismo valor de  $e$  como el primero.

El siguiente paso es combinar las estimaciones de error para estos tres hojas en la relación entre el número de ejemplos que se refieren, 6: 2: 6, lo que conduce a una estimación de error combinado de 0,51.

Health plan contribution:  $f = 5/14$ .  $e = 0.46$ . Debido a que este es menor que el error de estimación combinada de los tres niños, ellos se podan.

Working hours per week: La estimación de error para la primera, con  $E = 1$  y  $N = 2$ , es  $e = 0,72$ , y para el segundo es  $e = 0,46$ . La combinación de estos, 2 : 14, conduce a un valor que es mayor que la estimación del error para el nodo de horas de trabajo, por lo que el subárbol se poda y se sustituye por un nodo hoja.

# Dedución de reglas rudimentarias

Relation: weather.symbolic					
No.	1: outlook Nominal	2: temperature Nominal	3: humidity Nominal	4: windy Nominal	5: play Nominal
3	overcast	hot	high	FALSE	yes
7	overcast	cool	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
10	rainy	mild	normal	FALSE	yes
14	rainy	mild	high	TRUE	no
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes

## Evaluando los atributos de los datos

	Attribute	Rules	Errors	Total errors
1	outlook	sunny → no overcast → yes	2/5 0/4	4/14
2	temperature	rainy → yes hot → no* mild → yes cool → yes	2/5 2/4 2/6 1/4	5/14
3	humidity	high → no normal → yes	3/7 1/7	4/14
4	windy	false → yes true → no*	2/8 3/6	5/14

outlook: sunny → no  
           overcast → yes  
           rainy → yes

# Modelización estadística

## Datos de tiempo

Outlook	Temperature	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

## probabilidades

Outlook			Temperature			Humidity			Windy		Play		
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	2	3	hot	2	2	high	3	4	false	6	2	9	5
overcast	4	0	mild	4	2	normal	6	1	true	3	3		
rainy	3	2	cool	3	1								
sunny	2/9	3/5	hot	2/9	2/5	high	3/9	4/5	false	6/9	2/5	9/14	5/14
overcast	4/9	0/5	mild	4/9	2/5	normal	6/9	1/5	true	3/9	3/5		
rainy	3/9	2/5	cool	3/9	1/5								



# Modelización estadística

para un día dado

Outlook	Temperature	Humidity	Windy	Play
sunny	cool	high	true	?

Probabilidad de que sea si:  $2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = 0.0053$

Probabilidad de que sea no:  $3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14 = 0.0206$

Normalización

$$\text{Probability of yes} = \frac{0.0053}{0.0053 + 0.0206} = 20.5\%$$

$$\text{Probability of no} = \frac{0.0206}{0.0053 + 0.0206} = 79.5\%$$

# REGLAS DE ASOCIACION

Nos permite predecir patrones de comportamientos futuros sobre ocurrencias simultaneas de valores de variables. Técnica no supervisada

Una asociación entre dos atributos ocurre cuando la frecuencia con la que se dan dos o más valores determinados de cada uno conjuntamente es relativamente alta.

Las reglas de asociación intentan descubrir asociaciones o conexiones entre objetos.

***Consecuencia  $\Leftarrow$  Antecedente 1 Antecedente 2 ... Antecedente m.***

Ejemplo, en un supermercado se analiza si los pañales y las compotas se compran conjuntamente.

# REGLAS DE ASOCIACION: ejemplo

## Gestión estante del supermercado.

- Objetivo: Identificar los elementos que se compran juntos por bastante muchos clientes.
- Enfoque: Procesar los datos de punto de venta recogidos con escáneres de código de barras para encontrar dependencias entre elementos.
- Una regla clásica
  - Si un cliente compra pañales y leche, entonces es muy probable que compre compotas.

# Reglas de Asociación

- Pueden predecir cualquier atributo, o solo la clase, o predecir combinaciones de atributos.
- La **cobertura** de una regla de asociación es el número de instancias para las cuales ella predice correctamente (*soporte*).
- La **precisión** (*confianza*) es el número de instancias que predice correctamente, expresado como una proporción de todas las instancias a las que se aplica.

Outlook	Temperature	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

# Reglas de Asociación

se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos

Por ejemplo, en la tabla anterior con la regla:

If temperature = cool then humidity = normal

Otra regla es:

If windy = false and play = no then outlook = sunny and humidity = high

# Reglas de Asociación

## Items con cobertura mayor o igual a 2

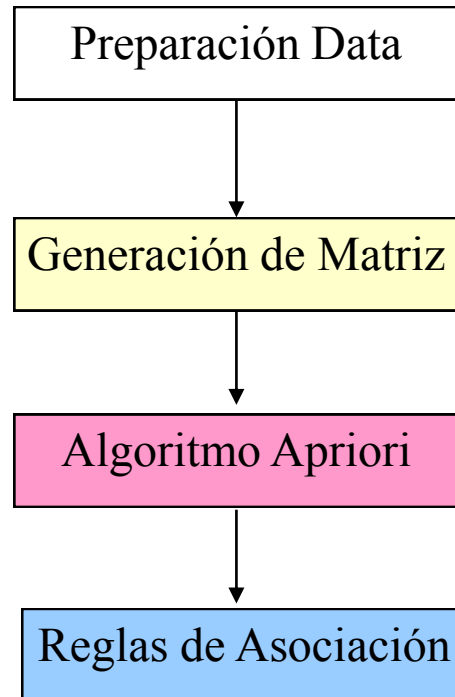
	One-item sets	Two-item sets	Three-item sets	Four-item sets		One-item sets	Two-item sets	Three-item sets	Four-item sets
1	outlook = sunny (5)	outlook = sunny temperature = mild (2)	outlook = sunny temperature = hot humidity = high (2)	outlook = sunny temperature = hot humidity = high play = no (2)	...	...	...	...	...
2	outlook = overcast (4)	outlook = sunny temperature = hot (2)	outlook = sunny temperature = hot play = no (2)	outlook = sunny humidity = high windy = false play = no (2)	38	humidity = normal windy = false (4)	humidity = normal windy = false play = yes (4)	humidity = high windy = false play = no (2)	
3	outlook = rainy (5)	outlook = sunny humidity = normal (2)	outlook = sunny humidity = normal play = yes (2)	outlook = overcast temperature = hot windy = false play = yes (2)	39	humidity = normal play = yes (6)			
4	temperature = cool (4)	outlook = sunny humidity = high (3)	outlook = sunny humidity = high windy = false (2)	outlook = rainy temperature = mild windy = false play = yes (2)	40	humidity = high windy = true (3)			
5	temperature = mild (6)	outlook = sunny windy = true (2)	outlook = sunny humidity = high play = no (3)	outlook = rainy humidity = normal windy = false play = yes (2)	47	...	...		
...	...	...	...	...		...	...		

# Reglas de Asociación

- Las reglas se obtienen a partir de los itemsets  
humidity = normal, windy = false, play = yes
- Esto nos lleva a las 7 reglas potenciales:
  - If humidity = normal and windy = false → play = yes 4/4**
  - If humidity = normal and play = yes → windy = false 4/6
  - If windy = false and play = yes → humidity = normal 4/7
  - If humidity = normal → windy = false and play = yes 4/6
  - If windy = false → humidity = normal and play = yes 4/8
  - If play = yes → humidity = normal and windy = false 4/9
  - If → humidity=normal and windy=false and play=yes 4/12

# Ejemplo de Reglas de Asociación

Encontrar las asociaciones que se producen entre los diferentes sitios de la página Web cuando los usuarios acceden a ésta.





# Reglas de Asociación

$$X \rightarrow Y$$

[/public/about.jsp ]---->/public/team.jsp

Soporte:

Soporte ( $X \rightarrow Y$ ) = Probabilidad ( $X \cup Y$ )

Confianza:

Confianza ( $X \rightarrow Y$ ) = Probabilidad ( $X / Y$ )

# Reglas de Asociación



## MineroWeb

**Ingreso** | **Procesamiento** | **Salidas**

[Estadísticas de Uso](#) | [K-Medias](#) | [Patrones Secuenciales](#) | [Reglas de Asociación](#)

Reglas Generadas: 14

Para mejor entendimiento de la regla, siga el esquema:  
\*El (Confianza)% de usuarios que visitaron  
(antecedente), visitarán (consecuente)\*

Soporte:  %    Confianza:  %       

N°	Regla	Soporte	Confianza
1	[/public/about.jsp ]---->/public/team.jsp	15,21%	57,14%
2	[/public/findUsers.jsp ]---->/public/portalDocument.jsp	13,04%	83,33%
3	[/public/findUsers.jsp ]---->/public/team.jsp	13,04%	83,33%
4	[/public/portalDocument.jsp ]---->/index.jsp	15,21%	57,14%
5	[/public/portalDocument.jsp ]---->/public/team.jsp	15,21%	71,42%
6	[/index.jsp ]---->/public/team.jsp	17,39%	62,5%
7	[/loginError.jsp ]---->/private/mycourses/index.jsp	10,86%	80%

# Reglas de Asociación

## Preparación de Data

### Registro\_Log

id	Id_se...	id_user	ip	Solicitud	fecha	bytes
8	2	11	200.110.86.82	/loginError.jsp	2006-02-26 00:03:00	3641
13	2	11	200.110.86.82	/private/mycourses/	2006-02-26 00:04:00	4785
16	2	11	200.110.86.82	/private/mycourses/website/...	2006-02-26 00:05:00	5717
19	2	11	200.110.86.82	/private/download/1048/3676...	2006-02-26 00:09:00	50688
24	2	11	200.110.86.82	/private/mycourses/website/...	2006-02-26 00:10:00	0
25	2	11	200.110.86.82	/private/mycourses/website/...	2006-02-26 00:10:00	4100
44	2	11	200.110.86.82	/private/mycourses/website/...	2006-02-26 00:19:00	5717
53	2	11	200.110.86.82	/js/tiny_mce/plugins/previe...	2006-02-26 00:21:00	0
110	4	11	200.110.86.82	/js/util.js	2006-02-26 01:03:00	0
176	4	11	200.110.86.82	/private/mycourses/website/...	2006-02-26 01:08:00	8778

### Registro\_Paginas\_Site

### Registro\_Sesion

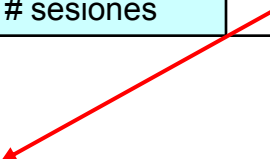
id_session	id_user	ip	hora_Inicio	hora_fin	num_pag..
3	31	201.2...	2006-02-26 00:54:00	2006-02-26 01:24:00	10
14	30	201.2...	2006-02-26 11:20:00	2006-02-26 11:27:00	9
30	23	200.6...	2006-02-26 16:41:00	2006-02-26 16:43:00	2
38	17	200.2...	2006-02-26 18:46:00	2006-02-26 18:46:00	0
1	6	200.1...	2006-02-26 00:01:00	2006-02-26 00:02:00	1
2	11	200.1...	2006-02-26 00:01:00	2006-02-26 00:29:00	42
7	11	200.1...	2006-02-26 01:36:00	2006-02-26 01:44:00	14
10	3	200.1...	2006-02-26 10:17:00	2006-02-26 10:23:00	3
11	32	201.2...	2006-02-26 10:33:00	2006-02-26 10:33:00	2
13	1	200.1...	2006-02-26 11:14:00	2006-02-26 11:15:00	4

id_pagina	url
2	/index.jsp
7	/private/mybriefc...
16	/private/mycourse...
20	/private/mycourse...
22	/private/mycourse...
26	/private/mycourse...
30	/private/mycourse...
32	/private/myprofil...
35	/public/findUsers...
36	/public/portalDoc...

# Reglas de Asociación

## Generación Matriz

Sesión / Página	1	2	3	4	5	.....	# páginas
1	0	1	0	1	0	.....	0
2	1	0	1	1	0	.....	0
3	1	1	0	1	0	.....	0
4	0	1	1	1	0	.....	0
5	1	0	0	0	0	.....	0
6	0	1	0	0	1	.....	0
:	:	:	:	:	:	.....	0
:	:	:	:	:	:	.....	0
# sesiones	0	1	0	1	0	.....	0

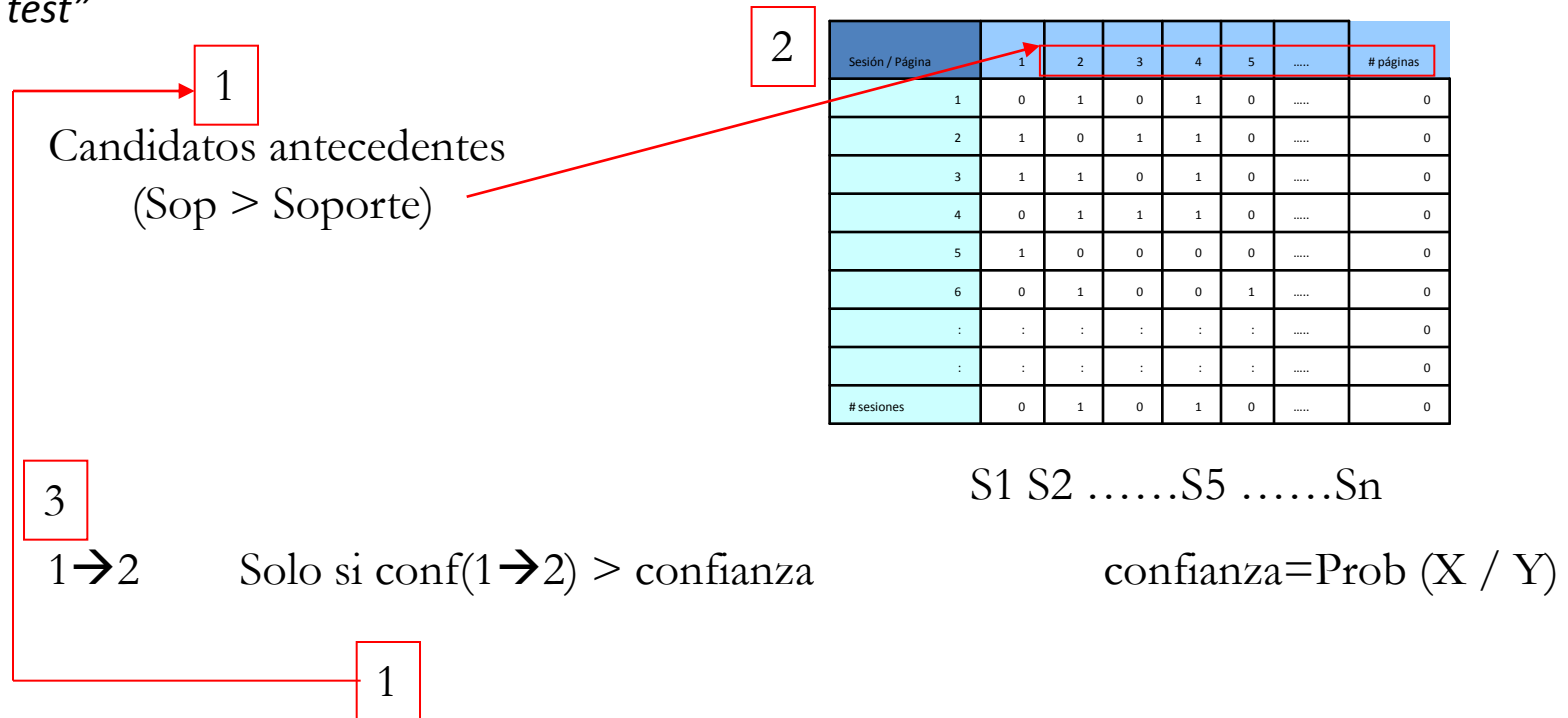

$$S1 = (0 + 1 + 1 + 0 + 1 + 0 + \dots + 0) / \# \text{ sesiones}$$

# Reglas de Asociación

## Algoritmo Apriori (matriz , soporte, confianza)

Usa *conocimiento a priori* de las propiedades de los ítems (páginas) frecuentes que ya se han encontrado.

*“Si un conjunto no puede pasar un test, todos sus súper conjuntos también fallarán el mismo test”*

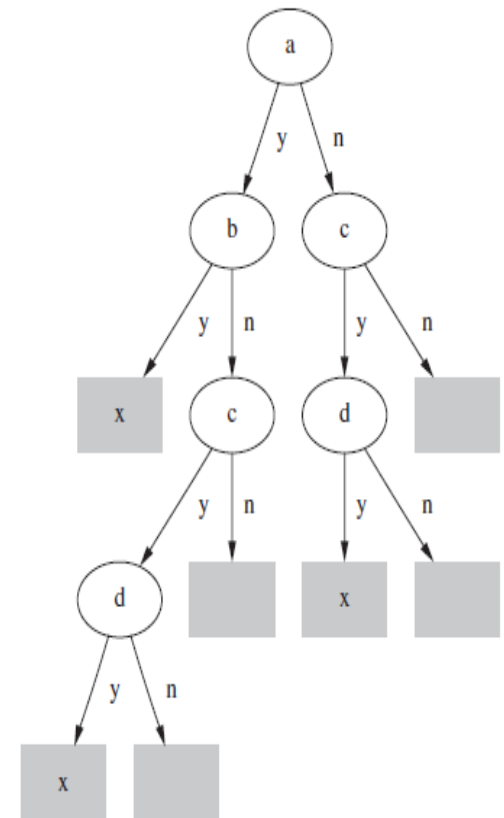
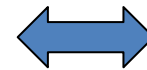


# Reglas de clasificación

Las reglas de clasificación son una alternativa popular a los árboles de decisión,

Por ejemplo:

```
If outlook = sunny and humidity = high then play = no
If outlook = rainy and windy = true    then play = no
If outlook = overcast                  then play = yes
If humidity = normal                   then play = yes
If none of the above                   then play = yes
```



# Utilidad de una categoría

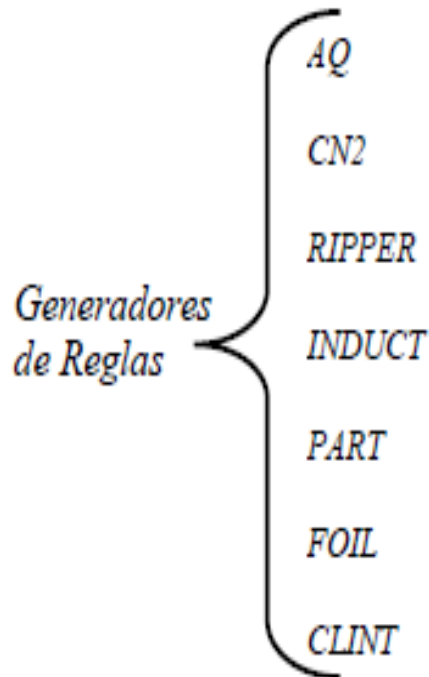
- Mide la calidad general de una partición

$$CU(C_1, C_2, \dots, C_k) = \frac{\sum_{\ell} \Pr[C_{\ell}] \sum_i \sum_j \left( \Pr[a_i = v_{ij} | C_{\ell}]^2 - \Pr[a_i = v_{ij}]^2 \right)}{k}$$

$\Pr[a_i = v_{ij} | C_{\ell}]$  es una estimación de la probabilidad de que el atributo  $a_i$  tiene un valor  $v_{ij}$ , en el grupo  $C_{\ell}$

donde  $C_1, C_2, \dots, C_k$  son los  $k$  grupos; la suma exterior es de estos grupos; las siguientes sumas interiores de los atributos  $a_i$ , y sus posibles valores  $v_{i1}, v_{i2}, \dots$ .

# Generación de reglas



- Algunas reglas inducidas pueden derivar de la construcción de un árbol de decisión, siendo primero generado el árbol de decisión y después trasladado a un conjunto de reglas
- Otros algoritmos se basan en el uso de técnicas de aprendizaje con lógica de predicados (ILP, Inductive Logic Programming). (FOIL, FFOIL, CLINT, etc.)



# RNA y Aprendizaje

# ***INTRODUCCIÓN A LAS RNAs***

¿CÓMO LA RED NEURONAL HUMANA ESTA  
DISEÑADA?

¿CÓMO EL CEREBRO PROCESA LA INFORMACIÓN?

¿CON QUÉ ALGORITMOS Y ARITMÉTICA EL CEREBRO  
CALCULA?

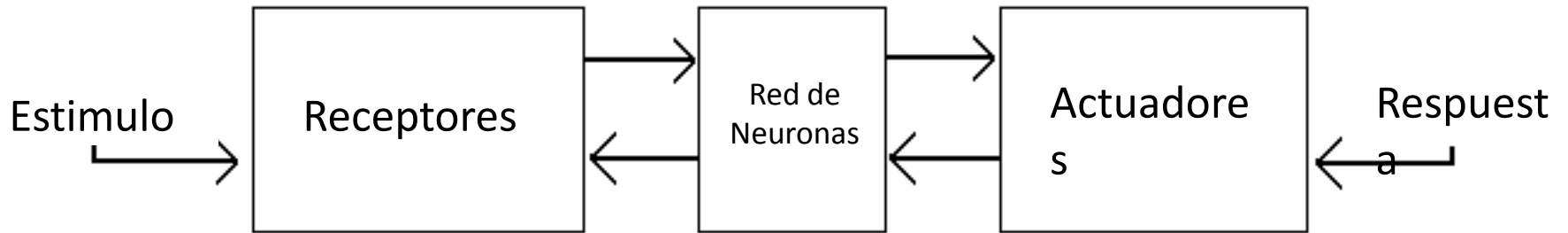
¿CÓMO PUEDE EL CEREBRO IMAGINAR?

¿CÓMO PUEDE EL CEREBRO INVENTAR?

¿QUÉ ES PENSAR?

¿QUÉ ES SENTIR?

# SISTEMA NERVIOSO



# ***MODELO BIOLOGICO***

SISTEMA NEURONAL



CONTROL CENTRALIZADO DE LAS  
FUNCIONES BIOLOGICAS

- **CEREBRO ~ 100 MIL MILLONES DE NEURONAS  
Y 10000 CONEXIONES POR NEURONA**

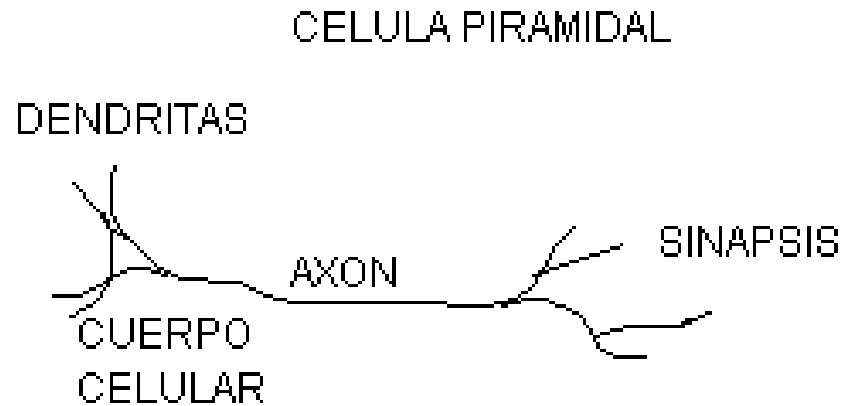
# ***MODELO BIOLOGICO***

- **NEURONAS: CELULAS VIVAS**
- **CARACTERISTICAS:**
  - ELEMENTOS SIMPLES INTERCONECTADOS
  - FUNCIONAMIENTO EN PARALELO, ASINCRÓNICA Y NO ALGORÍTMICAMENTE
  - INTERACCIONES COMPLEJAS

# ***NEURONA***

- UNIDAD FUNDAMENTAL DEL SISTEMA NERVIOSO ESPECIALIZADAS EN CIERTAS TAREAS
- PROCESADOR DE SEÑALES ELÉCTRICAS (DESCARGAS EN EL CUERPO CELULAR) Y BIOQUÍMICAS (NEUROTRANSMISORES)
- RECIBE Y COMBINA SEÑALES DESDE MUCHAS NEURONAS

# NEURONA

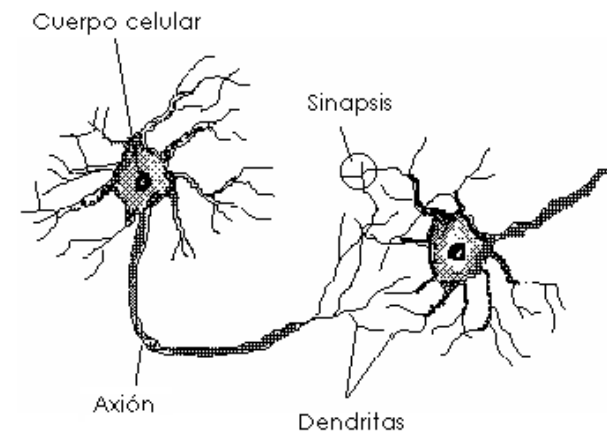


- **AXÓN:** LINEA DE TRANSMISIÓN
- **DENDRITAS:** ZONAS RECEPTORAS
- **SINAPSIS:** EXCITADORAS E INHIBIDORAS
- SEÑALES ELECTRICAS Y QUIMICAS

# SINAPSIS

UNIDAD FUNCIONAL QUE  
INTERRELACIONA LAS NEURONAS

- **NEUROTRANSMISOR:** GENERA POLARIZACIÓN PARA LA MEMBRANA POSTSINÁPTICA
- **POTENCIAL POSTSINÁPTICO:** PUEDE SER POSITIVO (EXCITACIÓN) O NEGATIVO (INHIBICIÓN)

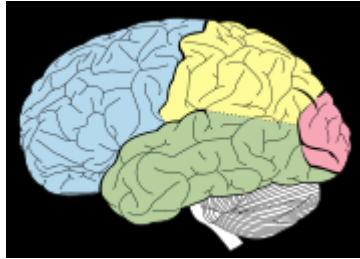




# ***REDES NEURONALES***

- MUCHAS **CONEXIONES PARALELAS** ENTRE NEURONAS
- MUCHAS CONEXIONES PROVEEN **MECANISMOS DE RETROALIMENTACIÓN** PARA LAS NEURONAS
- ALGUNAS NEURONAS PUEDEN **EXCITAR** UNAS NEURONAS MIENTRAS **INHIBEN** A OTRAS

# ***REDES NEURONALES***



- EJECUTAN UN PROGRAMA QUE ES **DISTRIBUIDO**
- TIENEN PARTES **PRE-HECHAS** Y OTRAS QUE **EVOLUCIONAN**

# ***CAPACIDADES RED NEURONAL***

- Procesamiento paralelo
- Adaptativa
- Asociativa
- Auto-organización
- Generalización, clasificación, extracción y optimización

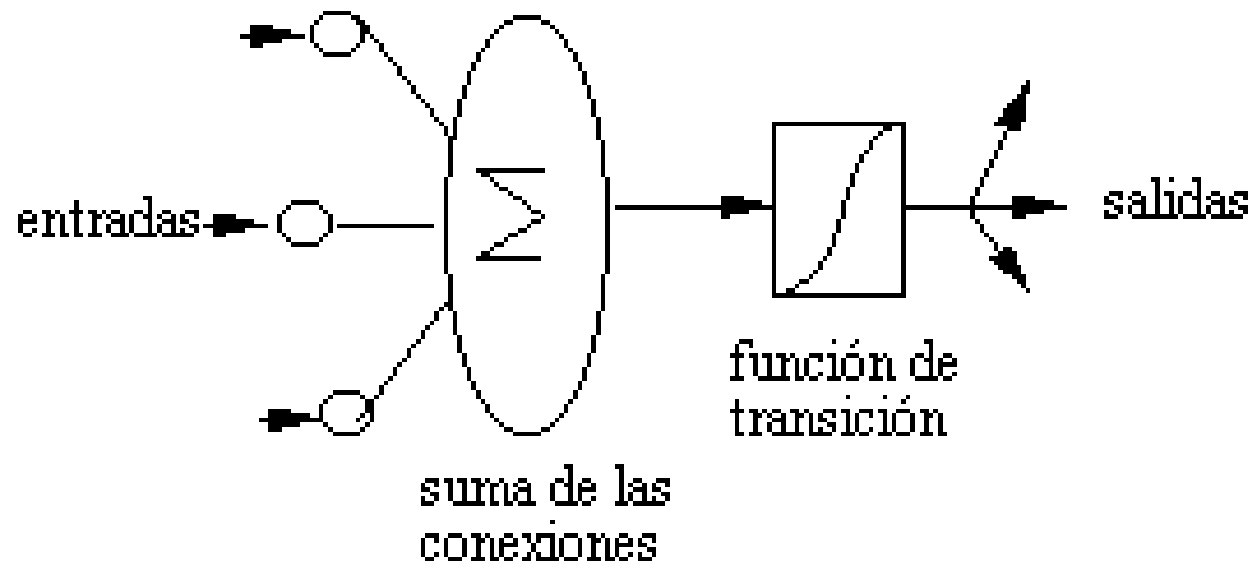
# COMPARACION RED NEURONAL

Neurona Biológica	Neurona Artificial
Señales que llegan a la sinapsis	Entradas a la neurona
Carácter excitador o inhibidor de la sinapsis de entrada	Pesos de entrada
Estimulo total de la neurona	Sumatoria de pesos por entradas
Activación o no de la neurona	Función de activación
Respuesta de la neurona	Función de salida

# COMPARACION RED NEURONAL

Aspectos	Computador	Cerebro Humano
Unidades de Cálculo	CPUs	$10^{11}$ neuronas
Unidades de Almacenamiento	RAM y disco duro	$10^{11}$ neuronas Y $10^{14}$ sinapsis
Ciclos	Mherz	$10^{-3}$ segundos
Banda Ancha	Capacidad de transmisión	$10^{14}$ conex. (bits)/segundo
Actualización/seg.	Capacidad de procesamiento paralelo	$10^{14}$

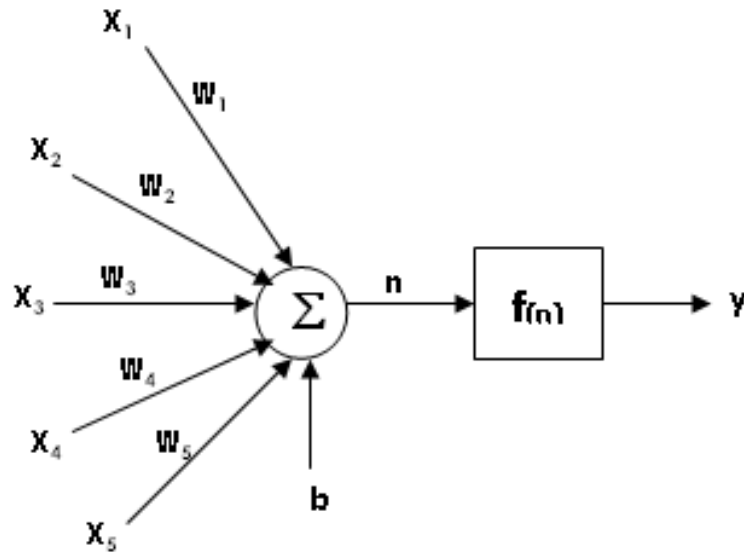
# ***COMO TRABAJA UNA NEURONA ARTIFICIAL***



# COMO TRABAJA UNA NEURONA ARTIFICIAL

$X_1, X_2, \dots, X_n$  son las **señales de entrada** y cada una pasa a través de un peso  $W$ , llamado **peso sináptico** de la conexión, cuya función es análoga a la de la **función sináptica de la neurona biológica**

El **nodo sumatorio** acumula todas las señales de entrada multiplicadas por los pesos y las pasa a la salida a través de una **función de activación** o transferencia  $f(n)$ , ( $b$  es el sesgo).



# ***COMO TRABAJA UNA RED NEURONAL***

1. El conjunto de unidades de procesamiento (neuronas formales).
2. El estado interno o de activación de las neuronas.
3. Las conexiones entre las neuronas.
4. Las conexiones con el ambiente.



# ***COMO TRABAJA UNA NEURONA***

**5. La regla de propagación  $h_i(t) = g(w_{ij}, x_j(t))$**

Ej. 
$$h_i(t) = \sum_j w_{ij} x_j(t)$$

**6. La función de activación**

$$a_i(t) = f_i(a_i(t-1), h_i(t))$$

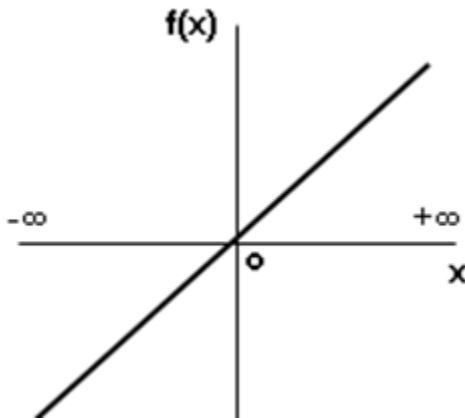
**7. La función de transición o de salida**

$$y_i(t) = F_i(a_i(t))$$

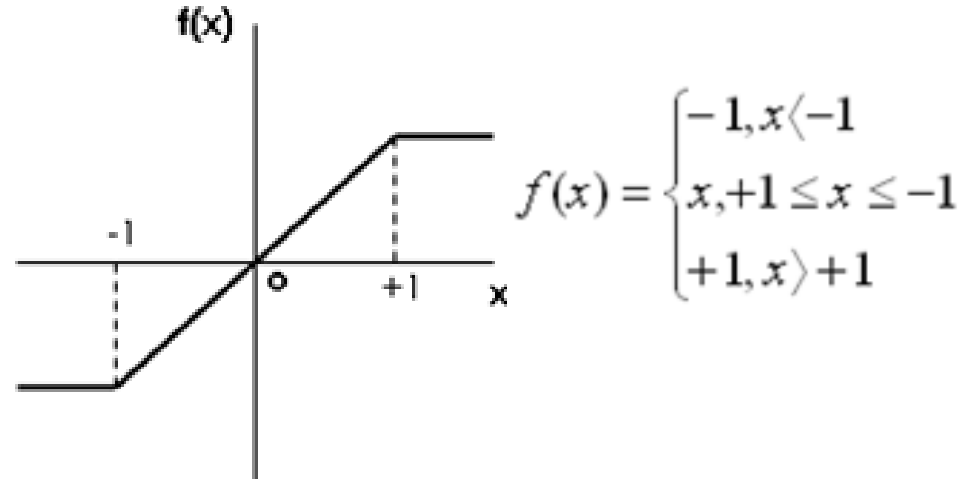
# Función de activación

Función identidad o función lineal:

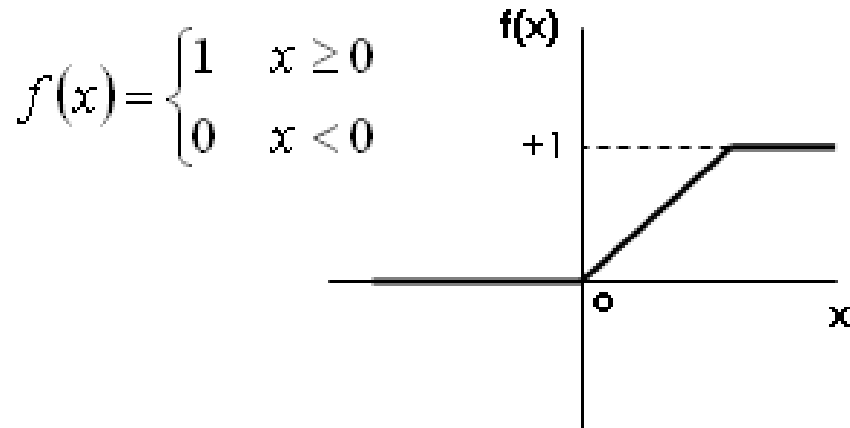
$$f(x) = x$$



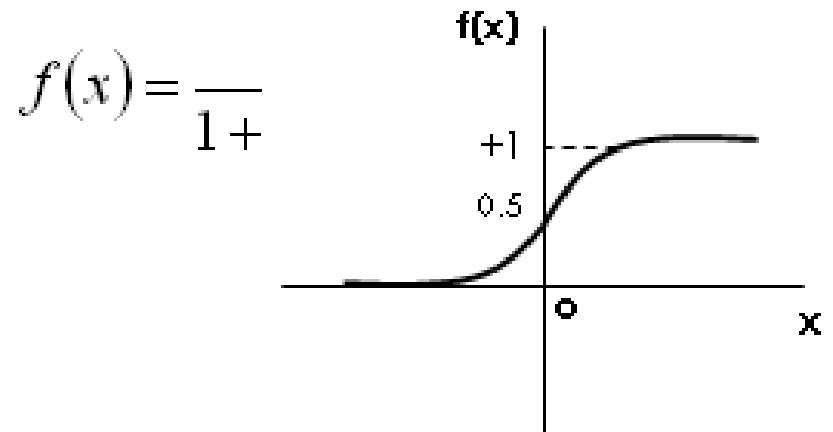
Función lineal por tramos



Función escalón

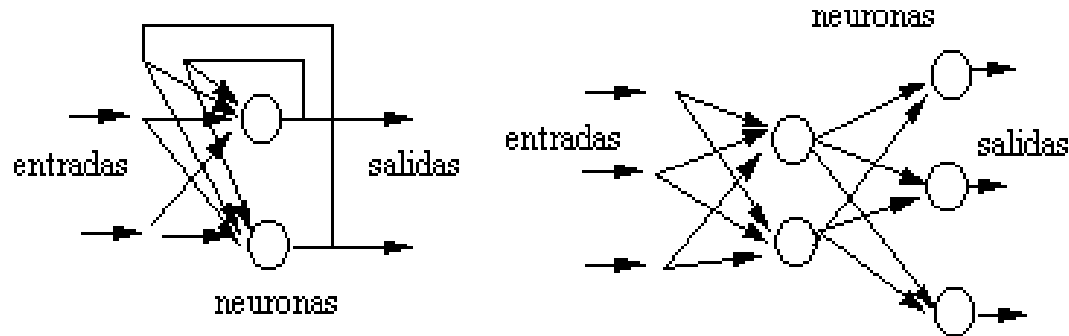


Función sigmoideal



# COMO TRABAJA UNA RED DE NEURONAS

## 8. La topología o arquitectura de la red



- **conexión total** (todas las neuronas interconectadas) o **conexión parcial** (por ejemplo, las redes de capas).
- **Realimentada o unidireccional**

# *Topologías de las RNA*

## **Redes monocapa:**

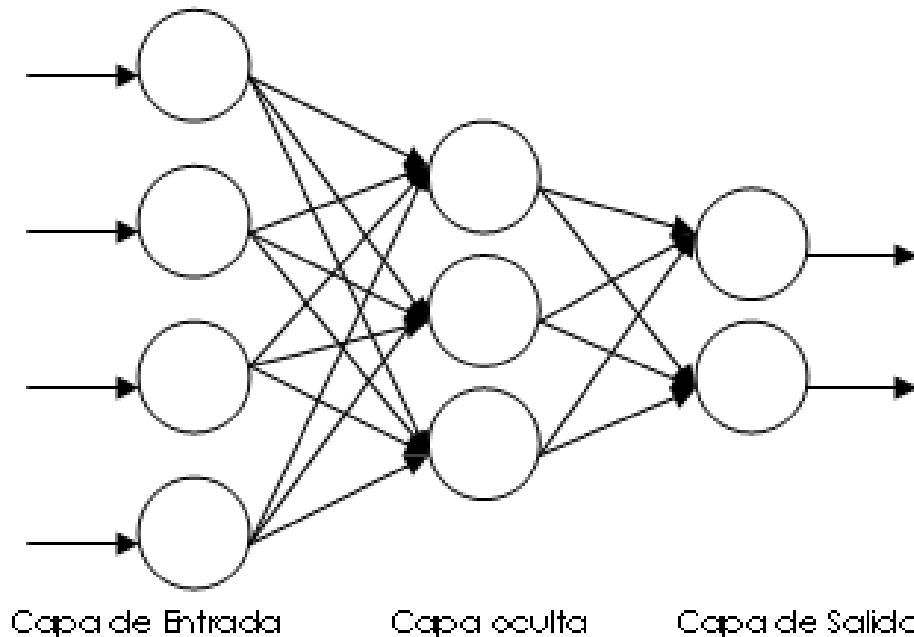
- Redes con una sola capa.
- Para unirse las neuronas crean conexiones laterales para conectar con otras neuronas de la única capa.

## **Redes multicapas:**

- Generalización de las anteriores donde existe un conjunto de capas intermedias entre la entrada y la salida llamadas *capas ocultas*.
- Pueden ser:
  - Propagación hacia adelante
  - Propagación hacia atrás
  - Redes recurrentes
  - Redes de alimentación lateral

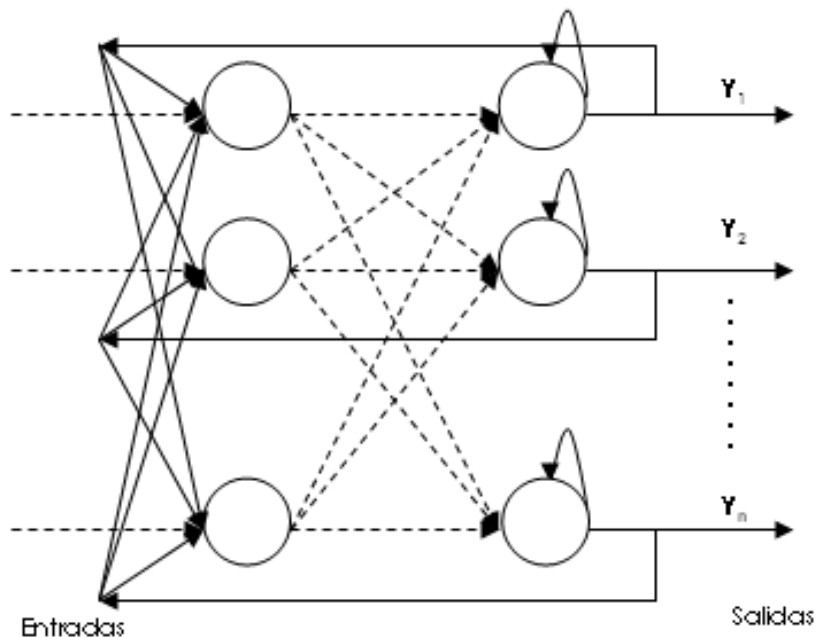
# Redes Multicapas

- **Capa de Entrada:** está constituida por los nodos de entrada, que reciben directamente la información de las fuentes externas a la red.
- **Capas Ocultas:** no tienen contacto con el exterior ya que se encuentran ubicadas entre la capa de entrada y la capa de salida. La cantidad de capas ocultas dependerá del problema en estudio y deben especificarse en la arquitectura.
- **Capa de Salida:** está constituida por los nodos que transfieren la información a la salida de la red y de acuerdo al tipo de problema en estudio se determinará el número de neuronas de salida.

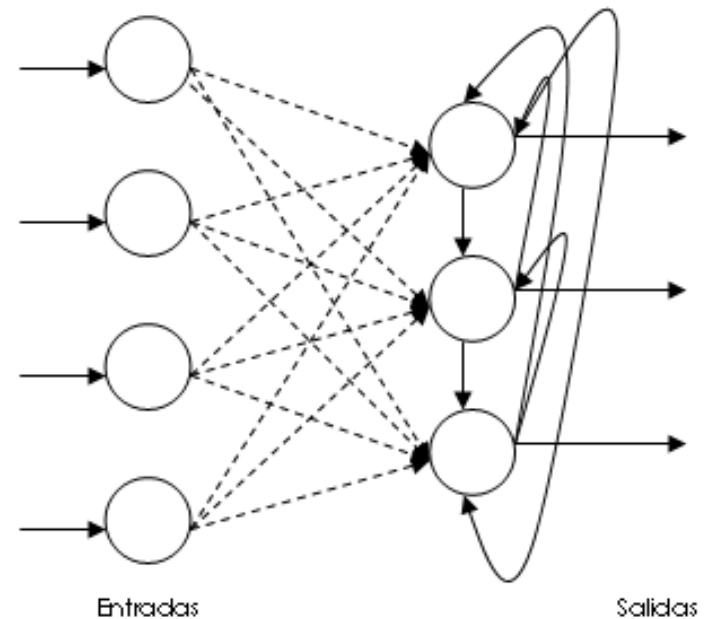


# Redes Multicapas

## Redes recurrentes



## Redes de alimentación lateral



# **Aprendizaje en las RNs**

# APRENDIZAJE

- El aprendizaje de una RNA se basa en un proceso que **permite que la red aprenda a comportarse según unos objetivos específicos**.
- El aprendizaje le da la capacidad a la RNA de **cambiar su comportamiento**, es decir su proceso de entrada-salida, como resultado de los cambios en el medio.
- En particular, las reglas de aprendizaje son procedimientos que se siguen para **ajustar los parámetros de la red** a partir de un proceso de estimulación por el entorno de la red
- La mayoría de las veces consiste en **determinar un conjunto de pesos**
- El aprendizaje es esencial para la mayoría de las arquitecturas de RNA, por lo que la **elección de un algoritmo de aprendizaje** es algo de gran importancia en el diseño de una red.



# APRENDIZAJE

- Al finalizar la fase de entrenamiento/aprendizaje de una RNA, se espera que la red haya aprendido lo suficiente para **resolver otro problema similar** satisfactoriamente.
- No existe en la literatura una metodología que indique la **manera de escoger el tipo o forma de aprendizaje** de la red para obtener resultados óptimos.
- Tipo de aprendizaje viene determinado por la **forma en que los parámetros** se deben adaptar

# ***MEMORIAS ASOCIATIVAS***

- RN ALMACENAN INFORMACIÓN APRENDIDA REFLEJADA EN SUS PESOS
- AL APLICARLE UNA ENTRADA LA RNA RESPONDE CON UNA SALIDA ASOCIADA A DICHA INFORMACIÓN DE ENTRADA



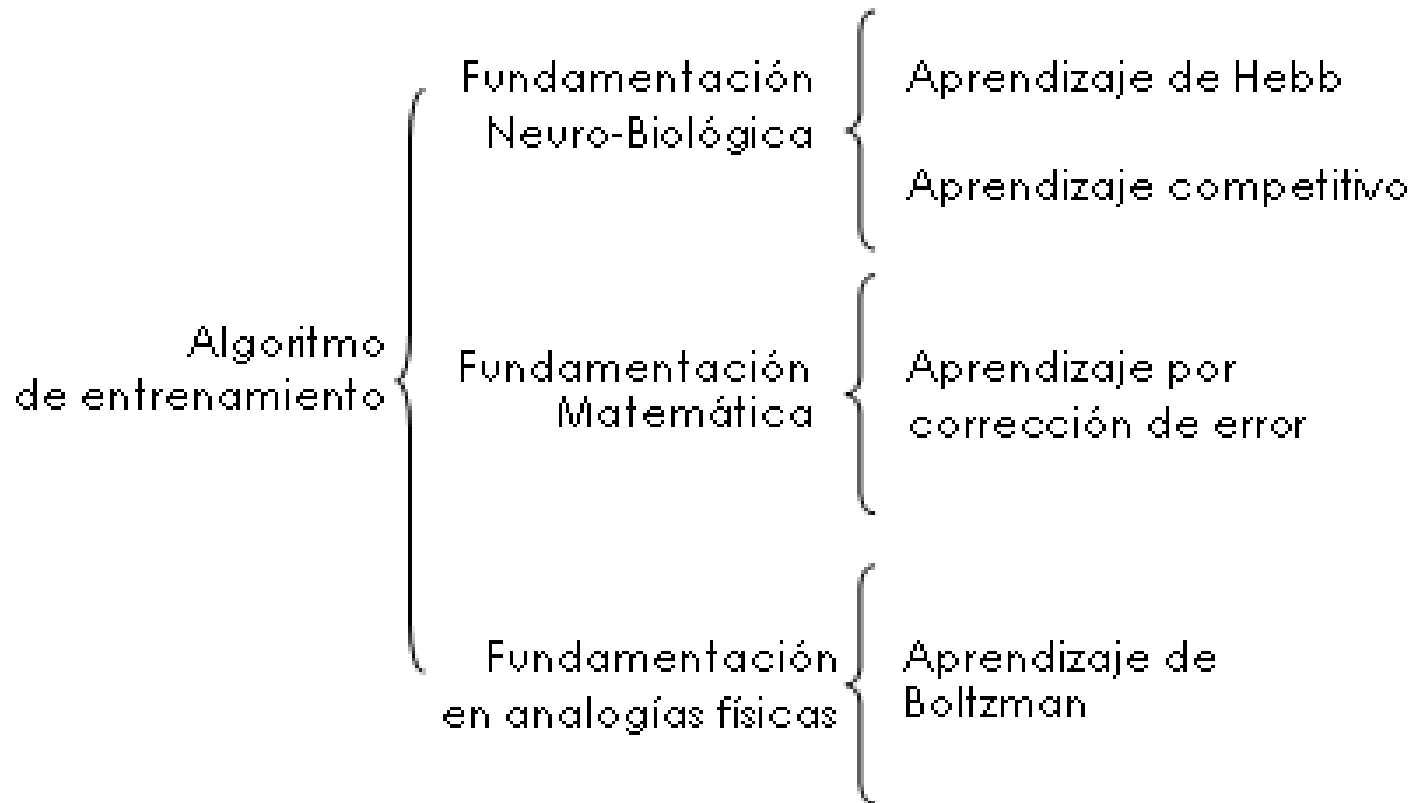
**ASOCIACIÓN ENTRADA/SALIDA**

# ***APRENDIZAJE***

MODIFICAR PESOS DE LAS  
CONEXIONES DE LAS NEURONAS  
(CREAR, DESTRUIR, MODIFICAR)

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t)$$

# APRENDIZAJE



**Clasificación de los Algoritmos de Aprendizaje basados en su fundamentación conceptual**

# APRENDIZAJE

**A. PARADIGMAS DE APRENDIZAJE:** *Define como se relaciona con su entorno. Se distinguen por el tipo de retroalimentación que se le ofrece al alumno.*

- **supervisado:** el crítico proporciona la salida correcta.
- **no supervisado,** no se proporciona retroalimentación en absoluto.
- **Basado en recompensa:** la crítica proporciona una evaluación de la calidad (el "premio") de lo hecho por el alumno.

# ***APRENDIZAJE***

- En los agentes se pueden usar todas
- En el caso de múltiples agentes, los métodos supervisados no son fáciles de aplicar

**Mas usado los métodos de recompensa.**

- Aprendizaje basado en recompensas puede ser dividido en dos subconjuntos:
  - **Métodos de aprendizaje por refuerzo:** estiman funciones de valor
  - **Métodos estocásticos**, tales como la computación evolutiva, recocido simulado.

# ***APRENDIZAJE***

## ***B. ALGORÍTMOS DE APRENDIZAJE: DEFINE REGLAS DE APRENDIZAJE (MODIFICACIÓN DE LOS PESOS)***

*CORRECCIÓN DE ERROR  
BOLTZMAN  
HEBBIANO  
COMPETITIVO  
EVOLUTIVO*

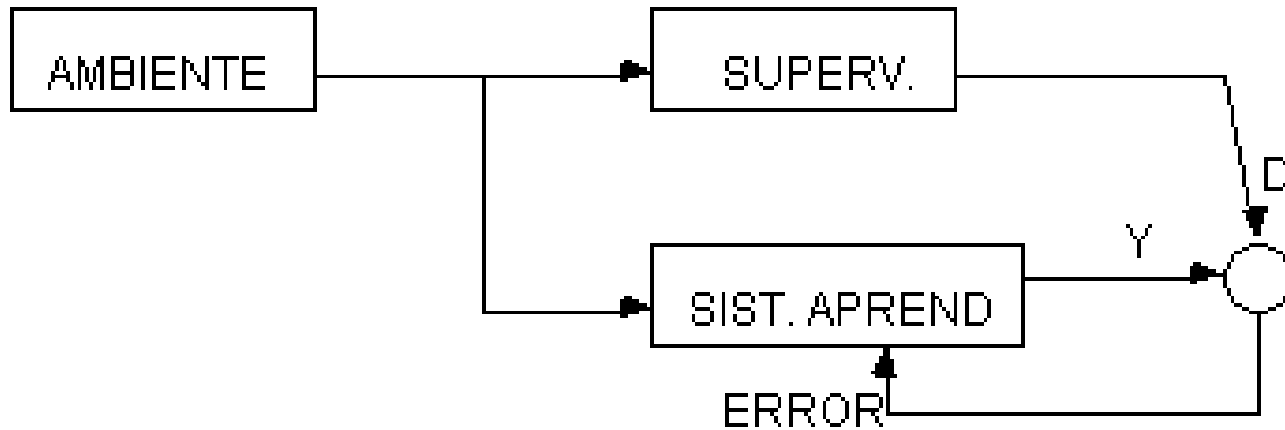
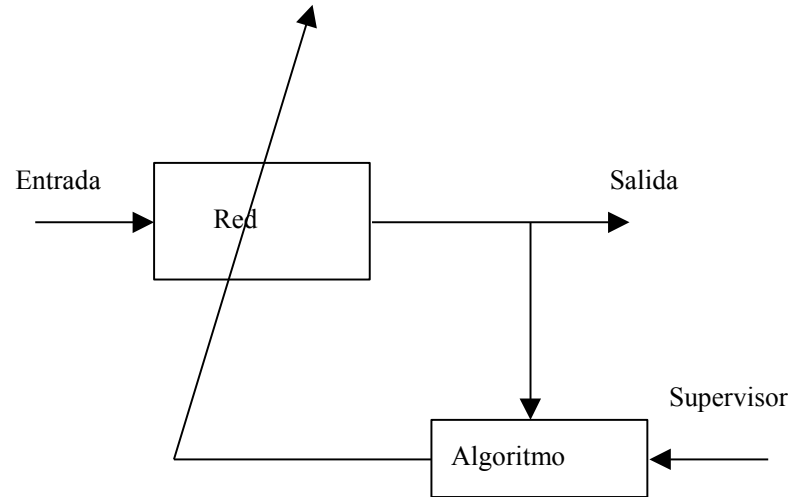
# ***SUPERVISADO***

Respuesta correcta para cada ejemplo dada

- SE DAN DATOS DE ENTRADA Y SALIDA OBJETIVO
- SALIDA RED DEBE CONCORDAR CON LA DESEADA



# ***SUPERVISADO***



# ***CORRECCIÓN DE ERROR***

CONOCIDO TAMBIEN COMO **DESCENSO DE GRADIENTE**

$$E_k(t) = D_k(t) - Y_k(t)$$

$D_k$ : respuesta deseada

$Y_k$ : respuesta neurona k

$$Y_k = F(X_k)$$

$X_k$ : entrada neurona k

$$\Delta W_{ij}(t) = \alpha E_i(t) X_j(t)$$

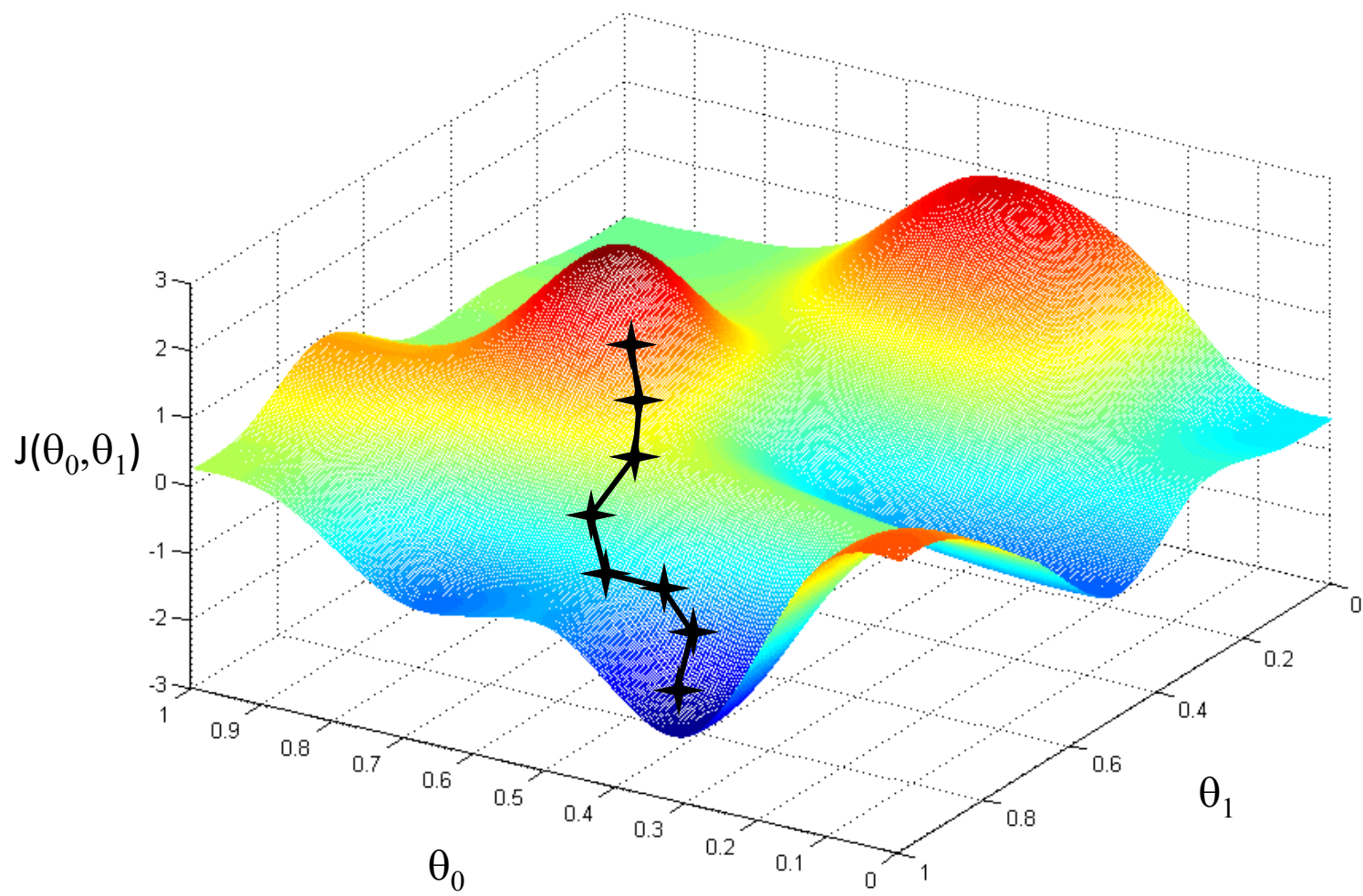
$\alpha$ : tasa de aprendizaje

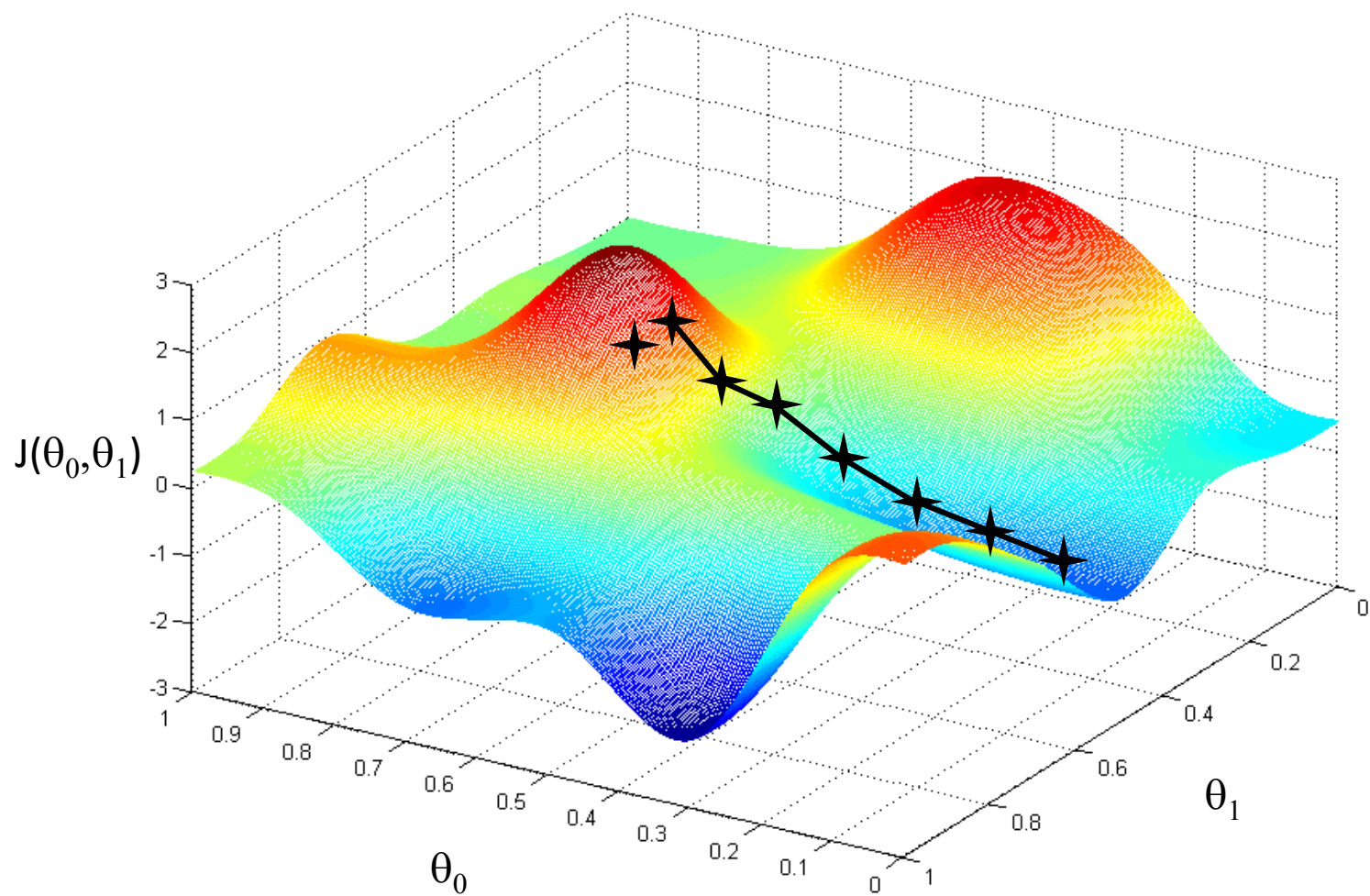
# ***CORRECCIÓN DE ERROR***

## **ALGORITMO**

1. CALCULAR EDO. DE LA RED ( $Y_i$ )
2. CALCULAR ERROR ( $E_i$ )
3. AJUSTAR PESOS

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t)$$





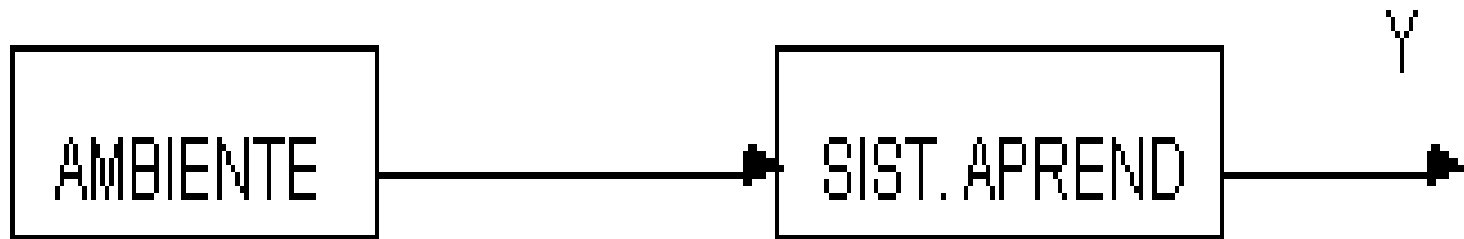
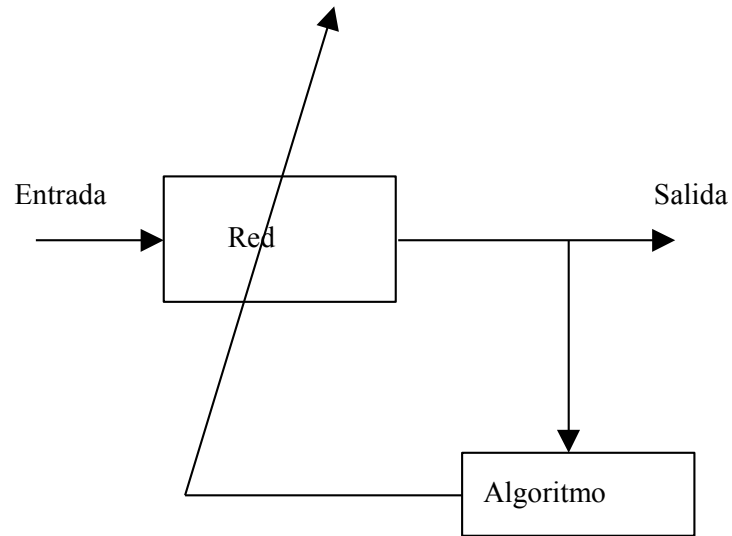
# *Algoritmo de un RNA*

1. Inicialización de los pesos y umbral
2. Fase de **entrenamiento**
  1. Presentación de las entradas y salida deseada
  2. Adaptación de los pesos
3. Fase de **Reconocimiento**
  1. Presentación de una entrada dada
  2. Salida reconocida

# ***NO SUPERVISADO (AUTOORGANIZADO)***

- NO RECIBE INFORMACIÓN DE SU ENTORNO (Se reciben patrones sin la respuesta deseada)
- CON LOS DATOS SE BUSCAN CORRELACIONES O REGULARIDADES EN EL CONJUNTO DE ENTRADAS:
  - EXTRAER RASGOS
  - AGRUPAR PATRONES SEGÚN SU SIMILITUD
- MAPAS AUTOORGANIZADOS

# ***NO SUPERVISADO (AUTOORGANIZADO)***





# ***HEBBIANO***

- MÁS VIEJO
- DOS O MAS NEURONAS ACTIVADAS  
SIMULTANEAMENTE

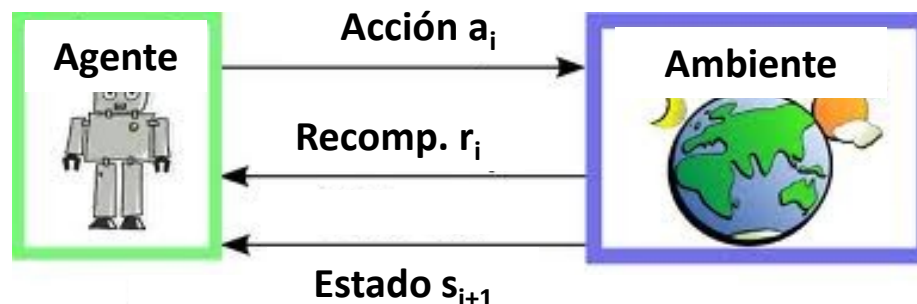
=> REFORZAR LA CONEXIÓN ENTRE ELLAS

$$\Delta W_{ij} = \alpha Y_i Y_j$$

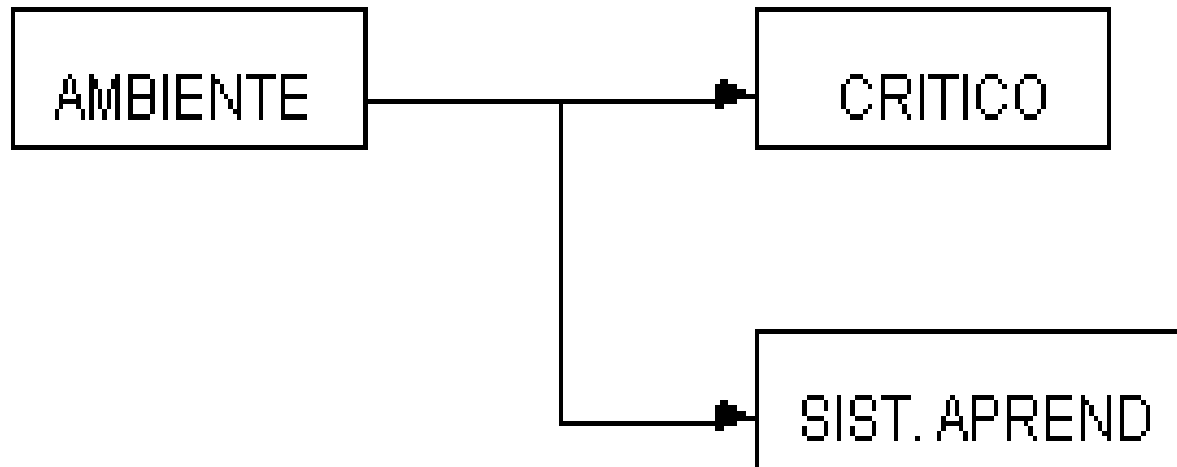
# REFORZADO

## Recompensa ocasional

- SUPERVISOR INDICA SI SALIDA SE AJUSTA A LO DESEADO O NO (que bien o mal se esta haciendo, no si es la salida deseada!!)
- SUPERVISOR HACE PAPEL DE CRÍTICO MÁS QUE DE MAESTRO (premio-castigo)



# ***REFORZADO***



# ***REFORZADO***

- Particularmente útiles en los ámbitos en los que exista información de reforzamiento (expresado como penalizaciones o recompensas) proporcionada después de una secuencia de acciones realizadas en el ambiente.
- Métodos comunes: Q-Learning y diferencia temporal- (TD )
  - **Q-Learning:** aprende la utilidad de llevar a cabo acciones que me lleven a ciertos estados,
  - **TD** aprender la utilidad de estar en ciertos estados.

# REFORZADO

- Todos los métodos de aprendizaje por refuerzo **están inspirados** en
  - fórmulas de actualización de la utilidades esperadas
  - exploración del espacio de estados.
- **La actualización** es a menudo una suma ponderada de:
  - valor actual utilidad,
  - refuerzo obtenido al realizar una acción y
  - utilidad esperada por el siguiente estado alcanzado, después se realiza la acción.

# ***Algoritmo de un RNA***

1. Presentación de las entradas
2. Calculo de la salida actual
3. Adaptación de los pesos

# Tareas de Aprendizaje

- Aproximación
- Asociación
  - Autoasociativa
  - Heteroasociativa
- Clasificación
- Predicción
- Control                      planta:  $u(t), y(t)$       modelo:  $r(t), d(t)$        $\lim |d(t) - y(t)| = 0$
- Filtraje

# Modelos Neuronales

Clasificación por tipo de aprendizaje y arquitectura

Híbridos: RBF (RADIAL BASIC FUNCTION)

Supervisados

Realimentados : feed-propagation

Unidireccionales PERCEPTRON, M RN, BOLTZMAN, backpropagation

No supervisados

Realimentados: ART, HOPFIELD

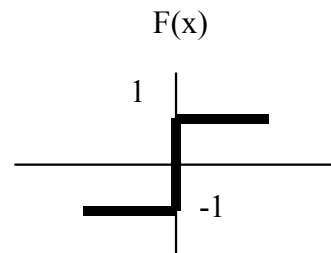
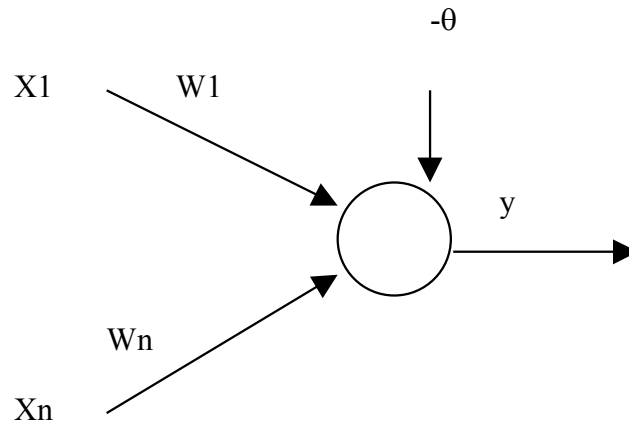
Unidireccionales: KOHONEN

Reforzados



# PERCEPTRÓN

- 1<sup>ER</sup> MODELO DE RED DE NEURONAS ARTIFICIALES (ROSEMBLATT 1958)
- APRENDE PATRONES SENCILLOS (2 CLASES)
- 1 NEURONA



$$Y = F(\sum W_i X_i - \theta)$$

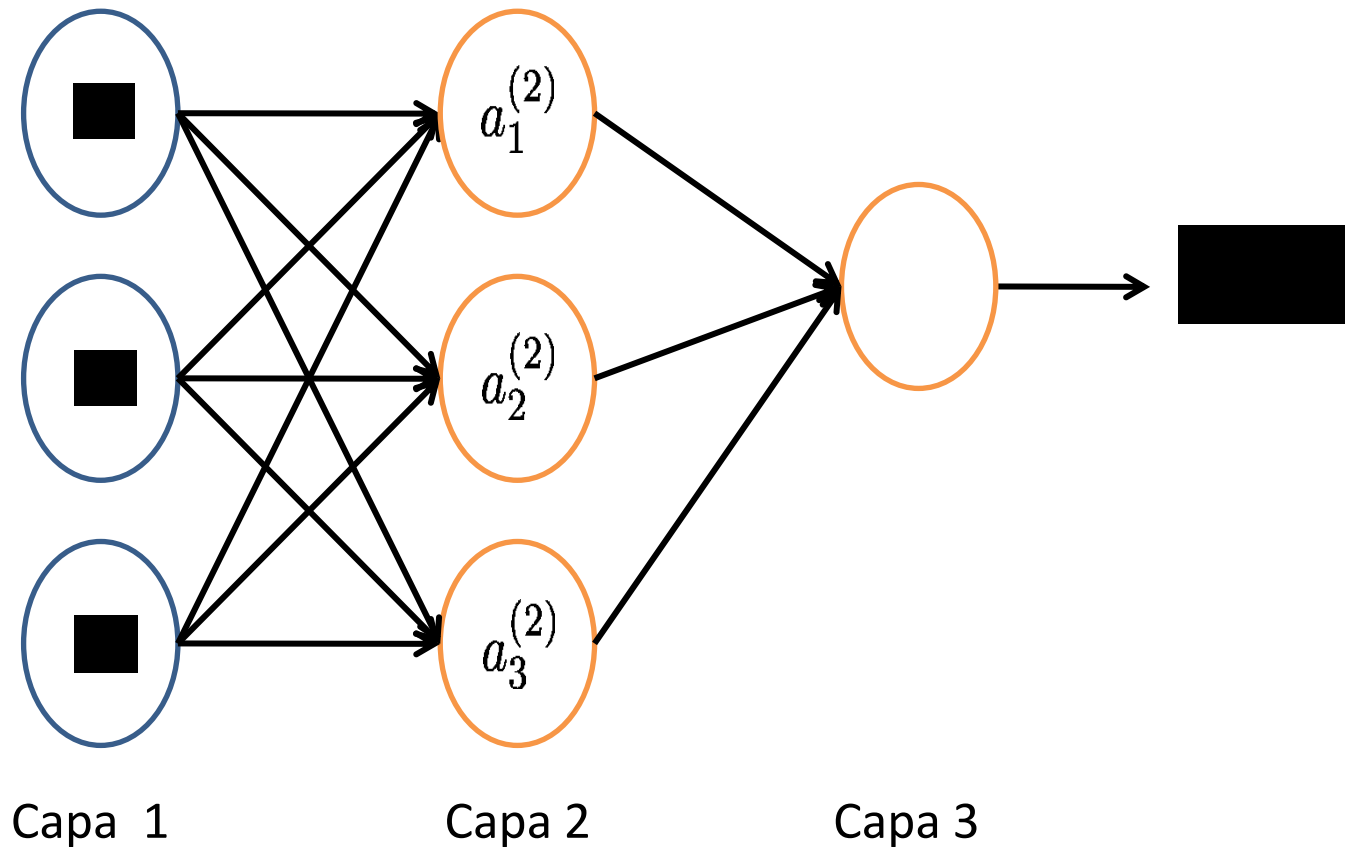
# PERCEPTRÓN

- REGIONES QUE INDICA A QUE PATRÓN PERTENECE CADA CLASE SEPARADAS POR UN HIPERPLANO
  - => PATRONES SEPARABLES GEOMÉTRICAMENTE
  - => DOS ENTRADAS LINEA RECTA  $x_2 = w_1 x_1 / w_2 + \theta / w_2$
  - => TRES ENTRADAS PLANO
- NO RESUELVE OR-EXCLUSIVO

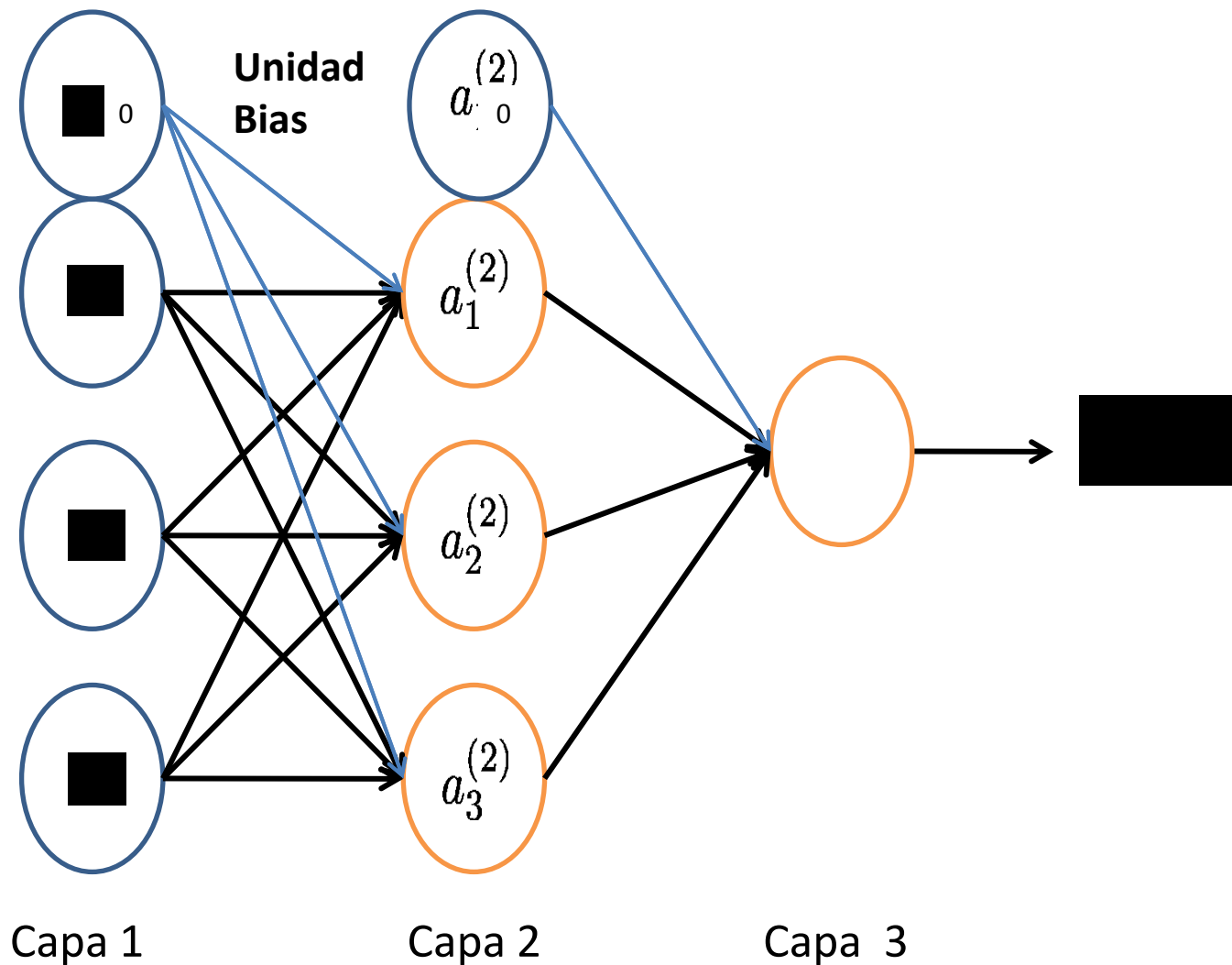
# PERCEPTRÓN

- APRENDIZAJE: SUPERVISADO
- ALGORÍTMO:
  1. INICIAR PESO Y UMBRAL
  2. PRESENTAR PAR ENTRADA-SALIDA
  3. CALCULAR SALIDA ACTUAL  
 $Y(t)$
  4. ADAPTAR LOS PESOS  
 $W_i(t) = W_i(t) + \alpha[d(t) - Y(t)]X_i(t)$   
HASTA QUE  $d(t) - y(t)^2$  valor pequeño
  5. REGRESAR AL PASO 2

# Modelo de Redes Neuronales



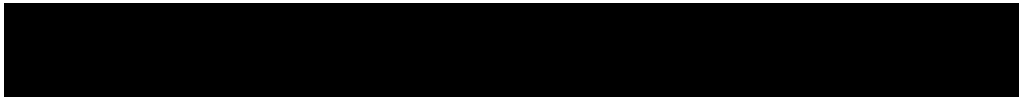
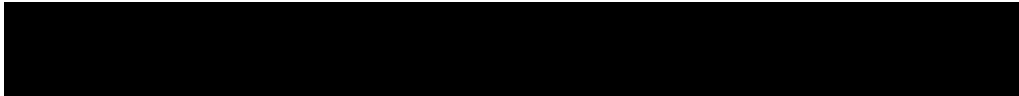
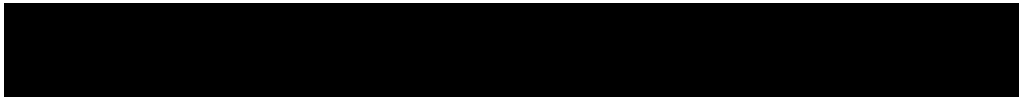
# Modelo de Redes Neuronales

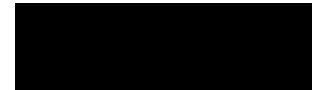
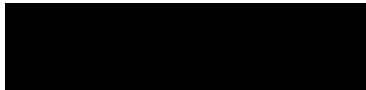


# Modelo de Redes Neuronales

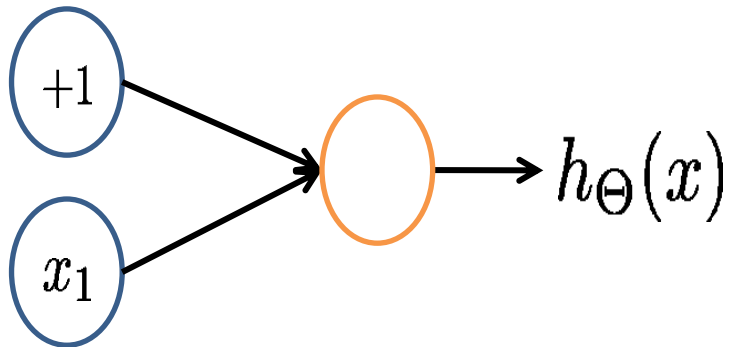
$a_i^{(j)}$  = “activación” unidad  $i$  en capa  $j$


$\Theta^{(j)}$  = Matriz de pesos entre capa  $j$  y capa  $j + 1$





**Negation:**

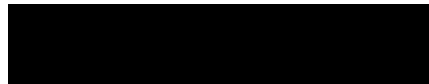
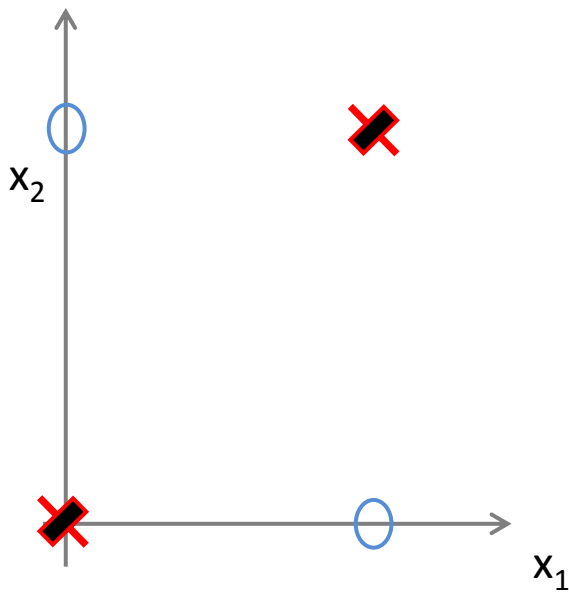


	$h_{\Theta}(x)$
0	1
1	0

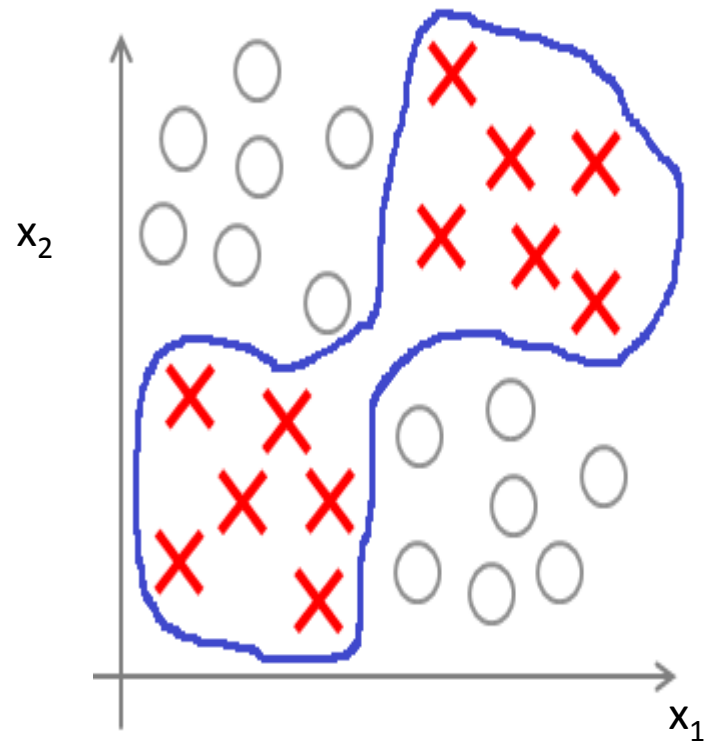


# Ejemplo de clasificación no lineal: XOR/XNOR

$x_1, x_2$  are binary (0 or 1).



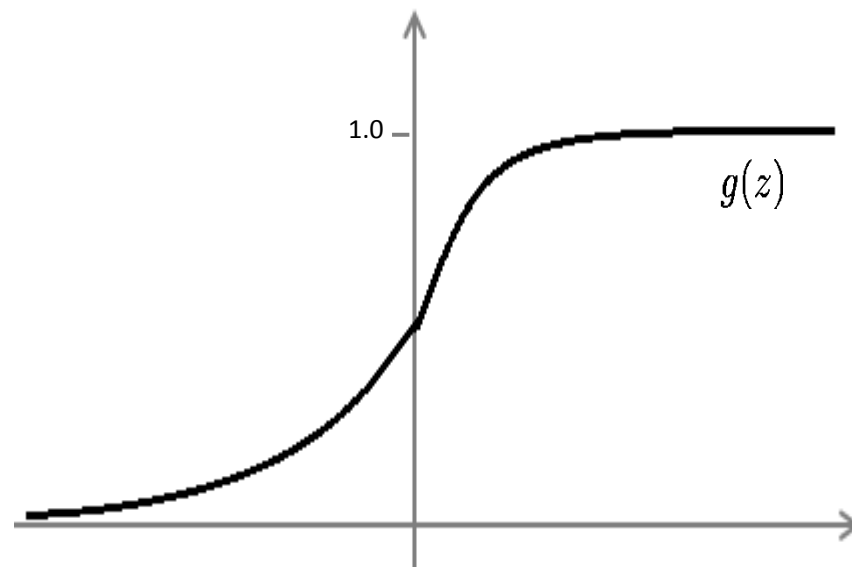
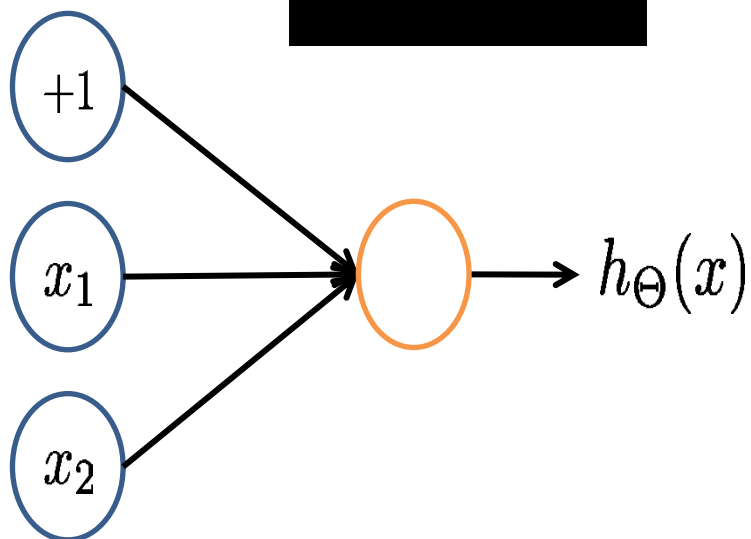
$\text{NOT } (x_1 \text{ XOR } x_2)$







AND

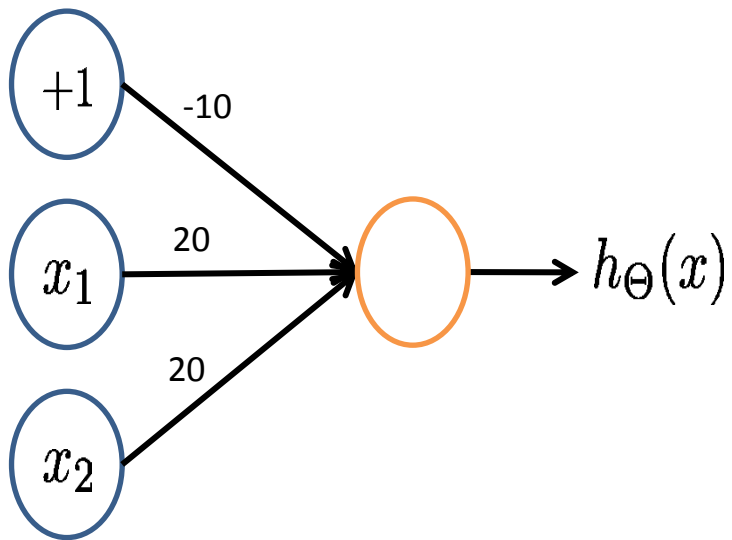
$$x_1, x_2 \in \{0, 1\}$$





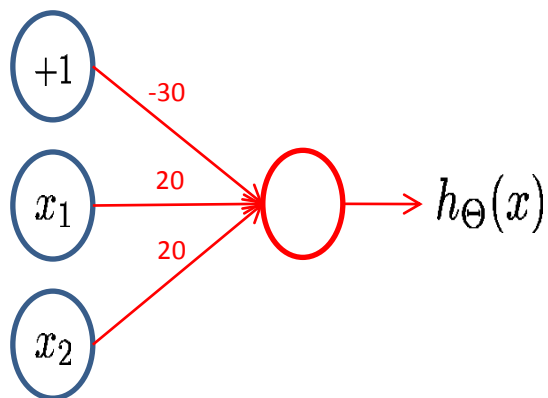
		$h_{\Theta}(x)$
		
0	0	0
0	1	0
1	0	0
1	1	1

$$h_{\Theta}(x) = g(-30 + 20x_1 + 20x_2)$$

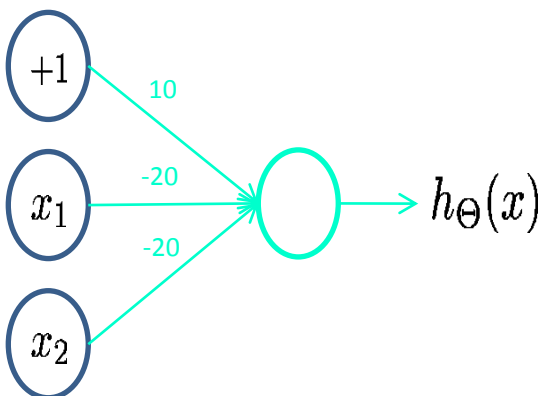
OR



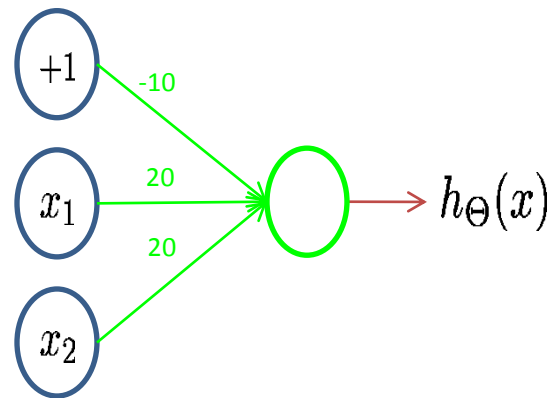
		$h_{\Theta}(x)$
0	0	0
0	1	1
1	0	1
1	1	1



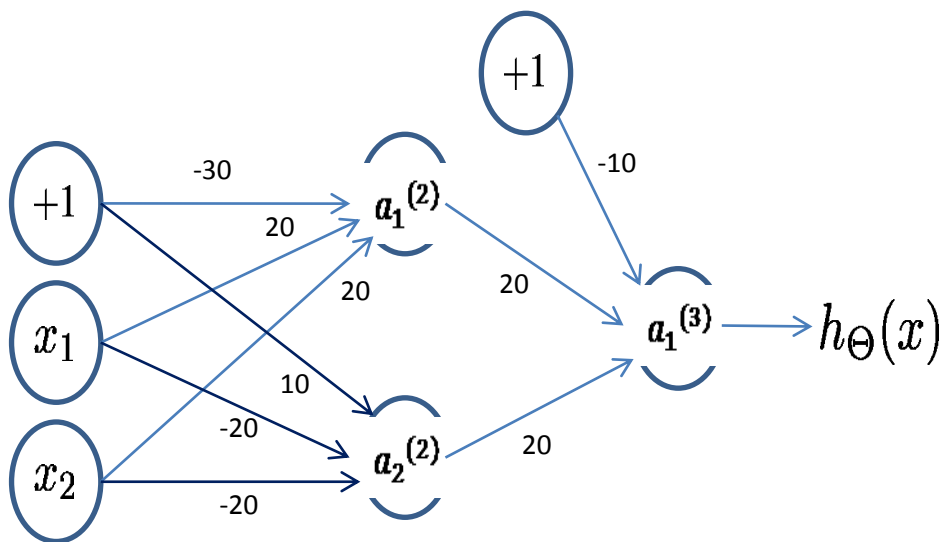
$x_1$  AND  $x_2$



(NOT  $x_1$ ) AND (NOT  $x_2$ )



$x_1$  OR  $x_2$



$x_1$	$x_2$	$a_1^{(2)}$	$a_2^{(2)}$	$h_{\Theta}(x)$
0	0	0	1	1
0	1	0	0	0
1	0	0	0	0
1	1	1	0	1

# Clasificación multi-clase



Peatón



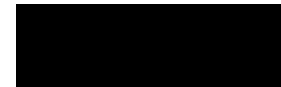
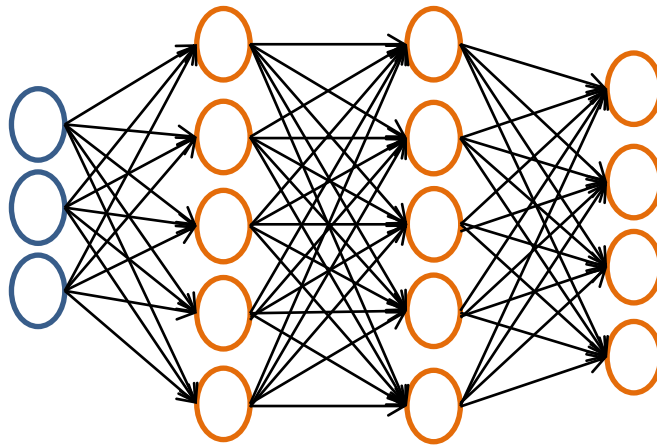
Carro



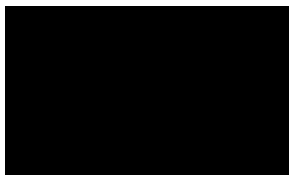
Motocicleta



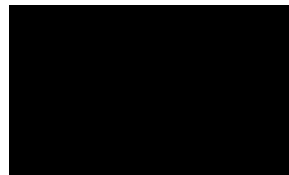
Camión



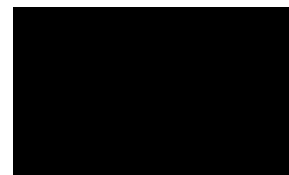
Será



**peatón**



**carro**



**motocicleta**

etc.

# REDES BAYESIANAS

# Redes Bayesianas

Las redes bayesianas son grafos dirigidos acíclico cuyos nodos representan variables aleatorias en el sentido de Bayes

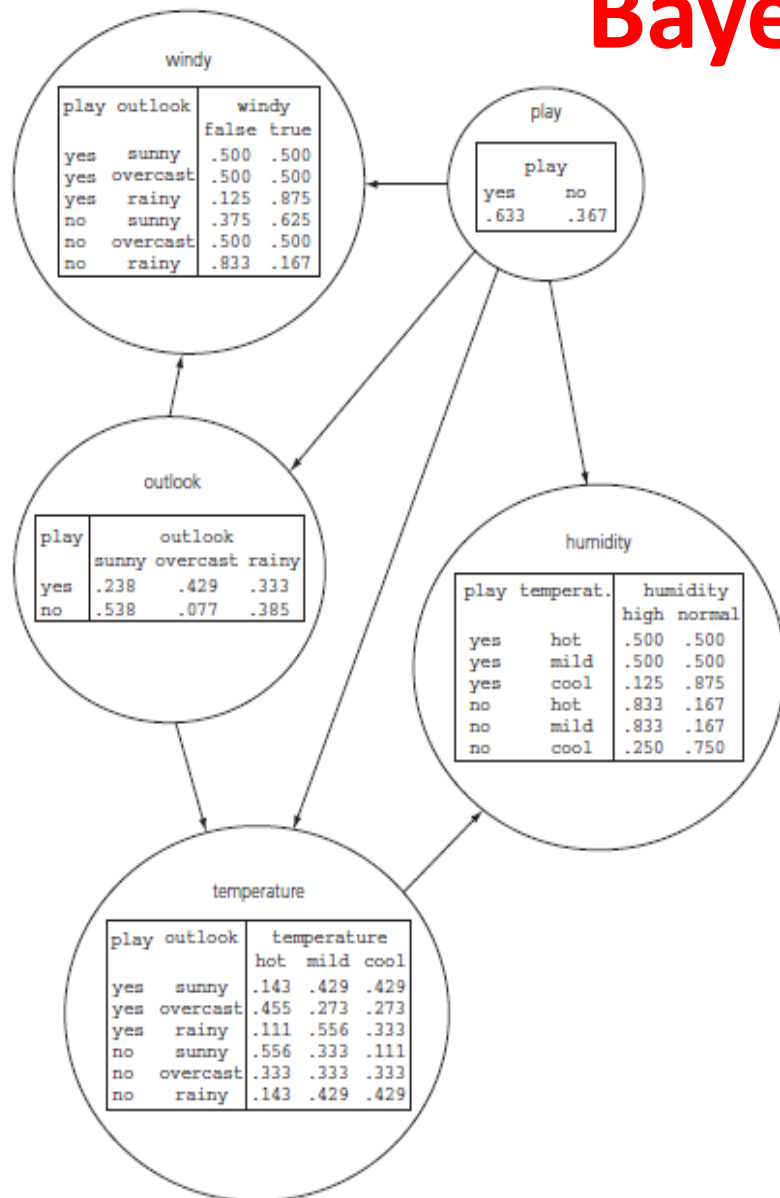
En el teorema de Bayes se expresa la probabilidad condicional de un evento aleatorio A dado B en términos de la distribución de probabilidad condicional del evento B dado A y la distribución de probabilidad marginal de sólo A. Pueden ser cantidades observables, variables latentes, parámetros desconocidos o hipótesis.

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

# Redes Bayesianas

- ❑ Las aristas representan dependencias condicionales
- ❑ Los nodos que no se encuentran conectados representan variables las cuales son condicionalmente independientes de las otras.
- ❑ Cada nodo tiene asociado una función de probabilidad que toma como entrada un conjunto particular de valores de las variables padres del nodo y devuelve la probabilidad de la variable representada por el nodo.

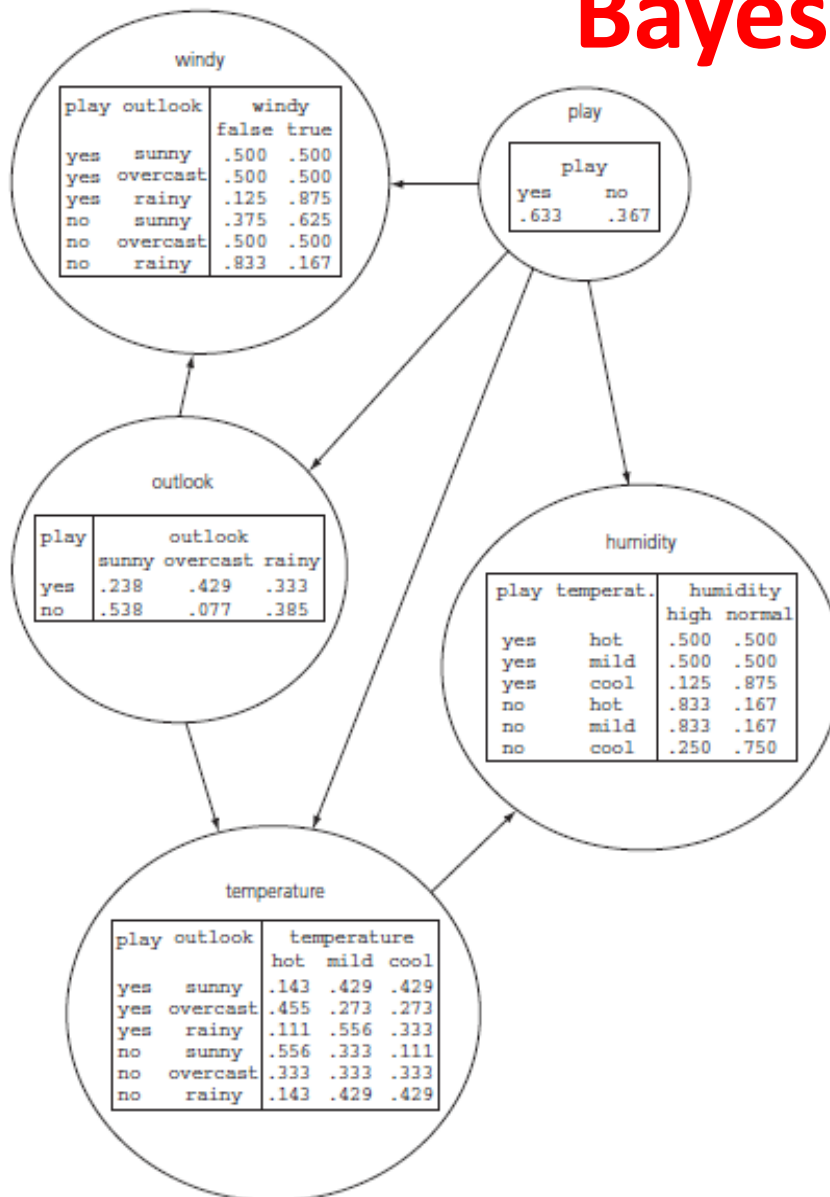
# Haciendo predicciones con Redes Bayesianas



Por ejemplo, considerar la posibilidad de una instancia con valores **perspectivas = lluvias, temperatura = frío, humedad = alto, y con viento = true.**



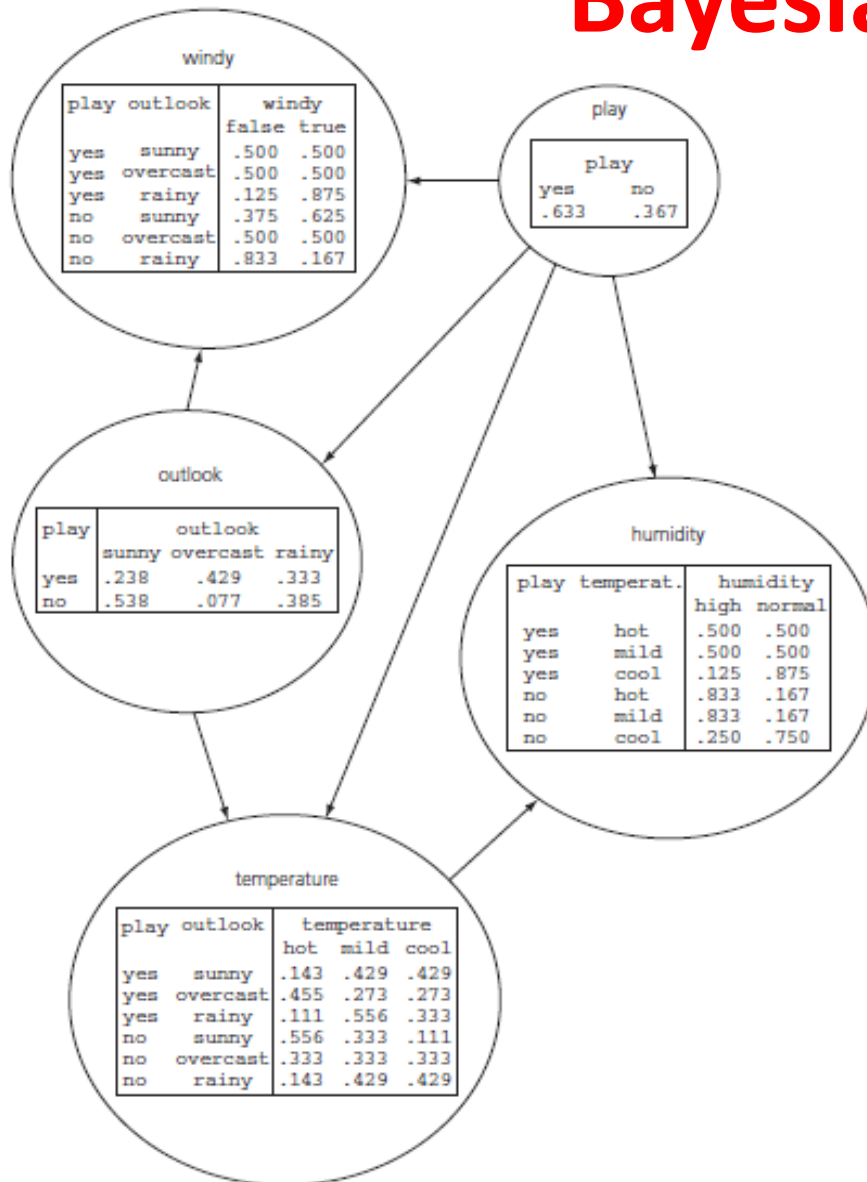
# Haciendo predicciones con Redes Bayesianas



Para calcular la **probabilidad para jugar = no**, en la red da probabilidad:

- 0.367 desde el nodo Play,
- 0.385 desde outlook,
- desde temperature 0.429,
- 0.250 de humidity, y
- 0.167 de windy

# Haciendo predicciones con Redes Bayesianas



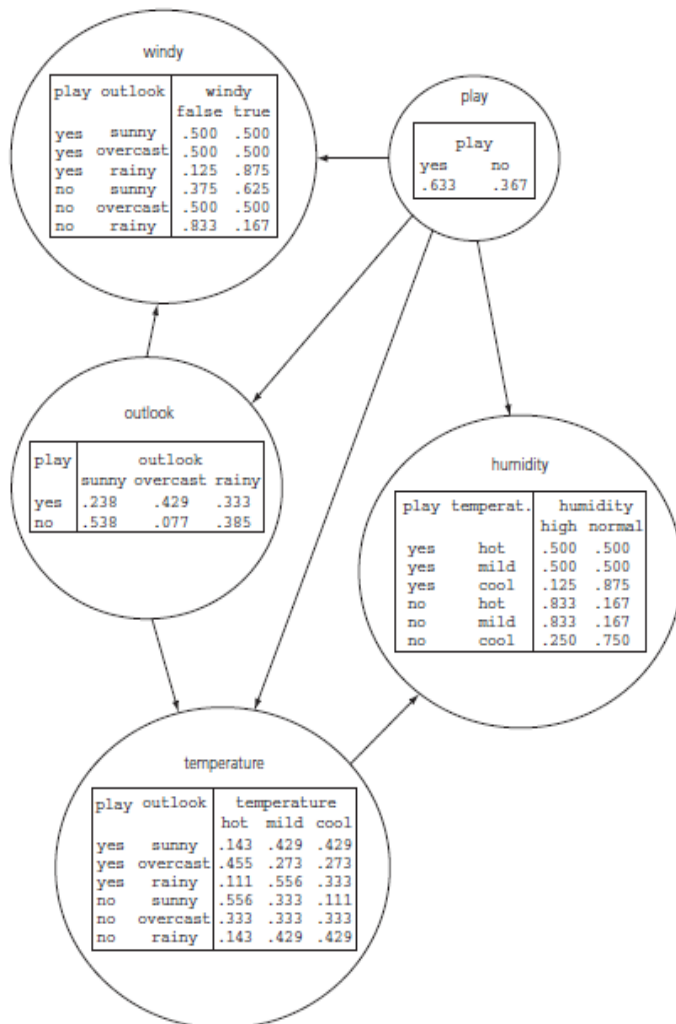
El producto es 0,0025. El mismo cálculo para el juego = yes es 0.0077.

Sin embargo, estos no son la respuesta final:

las probabilidades finales deben sumar 1,

# Haciendo predicciones con Redes Bayesianas

En realidad, son las probabilidades conjuntas  $\Pr [\text{play} = \text{no}, E]$  y  $\Pr [\text{play} = \text{yes}, E]$  donde  $E$  representada los valores de los atributos de la instancia que llegan a play.



Para obtener las probabilidades condicionales  $\Pr [\text{play} = \text{no} | E]$  y  $\Pr [\text{play} = \text{yes} | E]$ , normalizar las probabilidades conjuntas dividiéndolas por su suma.

Esto da probabilidad 0,245 para jugar = no y 0.755 para jugar = yes

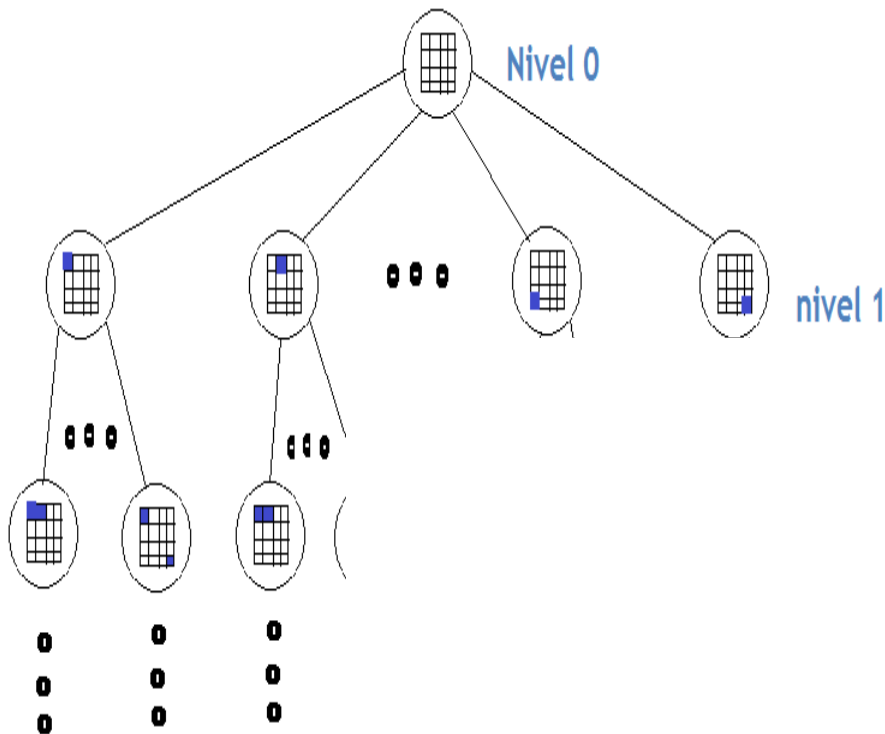
# Aprendizaje de Redes Bayesianas

El aprendizaje, en general, de redes bayesianas consiste en inducir un modelo, estructura y parámetros asociados, a partir de datos.

Este puede dividirse naturalmente en dos partes:

- **Aprendizaje estructural.** Obtener la estructura o topología de la red.
- **Aprendizaje paramétrico.** Dada la estructura, obtener las probabilidades asociadas.

# Manejo de Incertidumbre



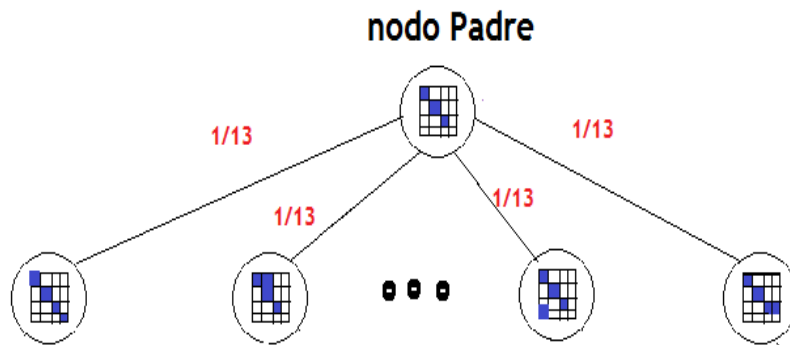
Red bayesiana para el  
manejo de incertidumbre

$$MUE = \max(U(a_i) * \sum_{i=0}^n P_i(a_i/P_i))$$

**Caso juego:** Según la función MUE la mejor acción será aquella en la cual la razón dada entre la utilidad y la probabilidad de que el oponente obtenga una mala jugada sea máxima.

# Modelo Matemático de Aprendizaje

Se tiene el siguiente Árbol con 13 nodos



Red bayesiana en su estado de máxima confusión

**Según acción del adversario sea buena o no, la rama debe ser premiada (o penalizada) y las del resto de hermanos inversamente modificadas (aprendizaje reforzado)**

Para actualizar las ramas se pueden usar los siguientes valores:

- $p_{obj} = 6/10$  se suma (resta) a la rama evaluada para premiar (castigar)
- $p_{resto} = 2/10$  se resta (suma) al resto de ramas para penalizar (premiar)

# Aprendizaje Paramétrico

**Estimador de máxima verosimilitud :** Se estiman las probabilidades en base a la frecuencia de los datos.

**Nodos raices** Se estima la probabilidad marginal. Por ejemplo:  $P(A_i) \sim N_{A_i}/N$ , donde  $N_{A_i}$  es el numero de ocurrencias del valor  $i$  de la variable  $A_i$  y  $N$  es el numero total de casos o registros.

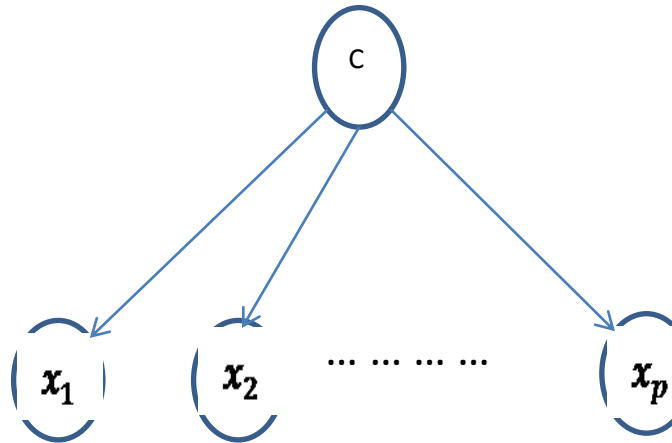
**Resto Nodos** Se estima la probabilidad condicional de la variable dados sus padres. Por ejemplo:  $P(B_i | A_j, C_k) \sim N_{B_i A_j C_k} / N_{A_j C_k}$ , donde  $N_{B_i A_j C_k}$  es el numero de casos en que  $B = B_i$ ,  $A = A_j$  y  $C = C_k$  y  $N_{A_j C_k}$  es el numero de casos en que  $A = A_j$  y  $C = C_k$ .

# Aprendizaje Estructural

- Aprendizaje de Arboles
- Aprendizaje NaiveBayes



# Aprendizaje NaiveBayes



- Las variables predictoras son condicionalmente independientes dada la variable clase

- <sup>+</sup>

$$P(C_i | E) = P(C_i)P(E_1 | C_i)P(E_2 | C_i)...P(E_n | C_i) | C_i) / P(E)$$