

# Introducción al Tratamiento Inteligente de la Información

Tratamiento Inteligente de la Información y Aplicaciones

Juan A. Botía, [juanbot@um.es](mailto:juanbot@um.es)

Departamento de Ingeniería de la Información y las Comunicaciones, Universidad de Murcia

October 6, 2009

- 1 Datos, información y conocimiento
- 2 Conceptos
- 3 Valores nulos e incorrectos
- 4 Análisis Inteligente de datos
  - La minería de datos
  - El aprendizaje automático
- 5 Selección de modelos

## Los Datos (numéricos, de texto, imágenes)

En texto los símbolos básicos son las palabras (elevada complejidad del espacio)

- Retos al manejar texto: la búsqueda y el *matching*
- los objetos del análisis cuando trabajamos con texto son los propios textos

## Los Datos (numéricos, de texto, imágenes) (y II)

Numéricos se generan como mapeo del mundo que se está estudiando a una representación conveniente para el análisis

- el objeto del análisis es un determinado sistema a partir del cual se han obtenido caracterizaciones numéricas
- La manipulación directa del mundo objeto de estudio no es conveniente.
  - ▶ para determinar cual de dos grupos de hombres pesa más, podríamos poner a cada grupo en un lado de una báscula gigante y ver para que lado se inclina, o bien sumar los pesos de cada grupo y compararlos

## Los datos y las técnicas de análisis

- El conjunto de técnicas y algoritmos que se presentan en esta asignatura descansan sobre una visión idealizada del análisis de los datos
- Sin embargo son potencialmente confusos, y podemos encontrar en ellos las siguientes características no deseables:
  - ▶ Datos ausentes
  - ▶ Datos no registrados
  - ▶ Datos de población diferente a la que es objetivo de estudio

## Datos ausentes

- Un estudio puede generar un conjunto de datos en los cuales no se encontraran todos los necesarios
- Ejemplo: un estudio en el cual se trata de estudiar las migrañas, favoreciéndolas de alguna manera.
  - ▶ el tipo de sujeto más interesante sería aquel más propenso a la jaqueca
  - ▶ se podría dar el caso de tener que expulsar del estudio a ese tipo de individuo por su propio bien si presenta un cuadro de dolores excesivo

## Datos no registrados

- Debidos a
  - ▶ errores humanos
  - ▶ hechos casuales inesperados
- Ejemplo: [1], pag. 9 menciona el haberse perdido de una tabla una columna numérica importante por el hecho de haberse quedado fuera del tamaño del papel en donde se estaba imprimiendo para conservarse.

## Datos con target incorrecto

- En ensayos clínicos, los pacientes deberían obtenerse con un muestreo aleatorio
- Lo que ocurre en realidad es bien distinto
  - ▶ se obtienen de entre los que han acudido a un hospital determinado y además han pasado un criterio de admisión al estudio estricto
- En estas condiciones pueden generarse medidas totalmente anómalas (i.e. *outliers*) fuera de rango por
  - 1 no constituyen casos genuinos de la población
  - 2 derivados de errores humanos (e.g. al teclear el valor)
  - 3 derivados de errores en los elementos de medida electrónicos.

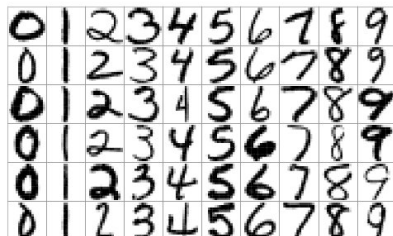


## Consecuencias de irregularidades en los datos

- Inmediata: el análisis de los datos es bastante arduo: si lo que se buscan son pequeñas regularidades entre los datos, el efecto de estas anomalías puede ser tan significativo estadísticamente como las regularidades encontradas.
- antes de cualquier tentativa de análisis de unos datos determinados, es necesario examinar previamente los mismos

## Datos: ejemplares y atributos

- Conjunto de datos grande: ¿en ejemplares o atributos?
- Grande en ejemplares: transacciones anuales de un banco
- Grande en atributos: caracteres escritos a mano alzada y representados gráficamente



## Conceptos

- Los datos están compuestos por conceptos, ejemplares y atributos [5]
- Las regularidades que buscamos en los datos representan conceptos, que son lo que buscamos en realidad
- Un concepto debe ser interpretable por el humano para que pueda descifrarse, entenderse, validarse en definitiva.
- un ejemplar es un ejemplo de un elemento que responde al concepto que se está buscando
- tratamos de obtener regularidades apartir de situaciones puntuales que han tenido lugar en el sistema a modelar
- Problema: existen problemas en los cuales los conceptos a aprender están estructurados de tal forma que no es posible aprenderlos apartir de diferentes situaciones
  - ▶ Secuencias temporales no divisibles en ejemplares

# Búsqueda de conceptos

- Cuatro tipos de conceptos típicos
  - 1 Clasificadores
  - 2 Asociaciones entre atributos
  - 3 Agrupadores de ejemplares
  - 4 Predictores numéricos

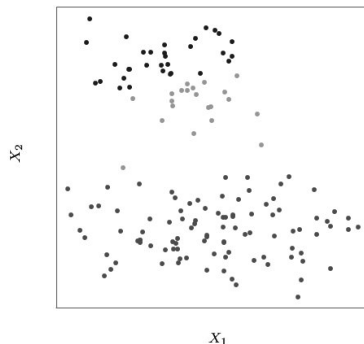
## Asociaciones entre atributos

- Esas asociaciones nos aportan información sobre los datos
- La minería de uso web es la aplicación de la minería de datos al descubrimiento de patrones de uso [4]
  - ▶ Patrones son asociaciones entre URLs
  - ▶ sean tres URLs  $\{r_1, r_2, r_3\}$  de un Web.
  - ▶ Una regla de asociación entre esas tres URLs sería

$$r_1 \Rightarrow r_2, r_3$$

## Agrupaciones entre ejemplares

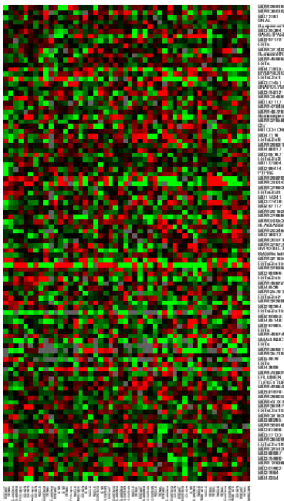
- lo que se obtiene es el mismo conjunto de datos dispuestos en diferentes grupos
- cada grupo representa un nuevo concepto que se ha descubierto en los datos



## Predictores numéricos (regresión)

- se busca aproximar el comportamiento de un sistema mediante una función, lineal o no lineal, en  $\Re$
- Ejemplo de microarreglos de ADN
  - ▶ Usado para determinar la presencia de un gen concreto en una cadena de ADN
  - ▶ Ejemplar: secuencia de nucleótidos de varios miles de genes
  - ▶ Mediante fluoroscopia cada uno de los genes refleja su presencia mediante la intensidad del color rojo (presente) y verde (ausente), gris no está disponible
  - ▶ Cada gen una fila y cada ejemplar (humano con cancer de un tipo determinado) una columna

# Predictores numéricos (regresión) (II)





# Atributos

- valor de un atributo para un ejemplar particular es una medida de la propiedad a la cual se refiere el atributo
- tipos de valores pueden dividirse en dos grandes grupos: los valores numéricos y los valores nominales
  - 1 Numéricos: miden magnitudes con números, ya sean enteros o reales
  - 2 Nominales: se refieren a valores con distintos símbolos (etiquetas que distinguen unos de otros)
    - ★ suma, multiplicación o comparación de tamaño no tienen sentido, solo la de igualdad.
    - ★ Como nominales un poco especiales podemos encontrar los ordinales

## Valores nulos

- El significado de los valores nulos debe ser interpretado cuidadosamente
- Primeramente se debe determinar la causa de su existencia
  - ▶ funcionamiento erróneo del equipo de medida
  - ▶ cambios en la parte de recogida de datos
  - ▶ en el diseño del experimento
  - ▶ integración de diferentes conjuntos de datos casi idénticos
  - ▶ en encuestas: negación a responder
- Debemos responder a la pregunta: ¿Es significativa la presencia de valores nulos para la extracción de conclusiones o no?

## Valores nulos (II)

- La mayoría de veces se asume implícitamente que no existe significancia particular en la existencia de un nulo
- una interpretación más cuidadosa puede llevar a aportar algún significado adicional a la existencia de un valor ausente (e.g. “no testeado”)

*imagínese una base de datos de casos médicos en la que, en determinados atributos aparecen los resultados de determinados tests médicos. Un médico determina si a un paciente debe, o no, realizarse cada uno de los tests disponibles. El hecho de no haber realizado un determinado test a un paciente (i.e. un valor nulo en la columna correspondiente del ejemplar en cuestión) es importante ya que simplemente mirando los tests que no se han hecho a un paciente se podría realizar un diagnóstico certero.*

## Datos incorrectos

- Datos incorrectos pueden llevar a conclusiones incorrectas.
- Su existencia se justifica ya que, a menudo, los datos sobre los que se va a realizar el análisis no se han obtenido precisamente pensando en ello (no se han obtenido cuidadosamente, no se han corregido)
- En datos nominales
  - ▶ errores tipográficos
  - ▶ *Pepsi* o *Pepsi cola*
- En datos numéricos
  - ▶ se deben detectar visualizando el valor de la variable correspondiente para determinar valores fuera de rango
- tuplas duplicadas, datos obsoletos o intencionalmente erróneos.

## Minería de Datos (*Data Mining*)

- Trata del desarrollo de métodos y técnicas para obtener algún significado de los datos en bases de datos
- El problema a tratar es el de transformar los datos *en crudo*, demasiado voluminosos para ser entendidos fácilmente, en otras formas que puedan ser
  - 1 más compactas (un informe)
  - 2 más abstractas (una aproximación descriptiva)
  - 3 más útiles (un modelo predictivo para estimar valores en casos futuros)

# Motivaciones para la minería de datos

- Transformación de datos en conocimiento
  - ▶ Procesamiento manual, tradicionalmente
    - 1 Empresas dedicadas a la salud analizan tendencias y cambios en los consumidores cada cuatro meses para planificar inversiones y actuaciones futuras
    - 2 Geólogos planetarios analizan imágenes de planetas lejanos para catalogar, a mano, los diferentes accidentes geológicos que se encuentran
    - 3 Procesamiento lento, caro y subjetivo
  - ▶ Al crecer dramáticamente el volumen de los datos, este tipo de análisis de datos manual llega a ser impracticable en muchos dominios

## Motivaciones para la minería de datos (y II)

- El crecimiento del tamaño en bases de datos se está produciendo en dos formas
  - 1 El número de registros  $N$  en la base de datos
  - 2 El número de  $d$  atributos en cada registro
- Las bases de datos con un contenido de un billón de registros es cada vez más común (i.e. ciencias astronómicas)
- Las bases de datos con atributos entre 100 y 1000 en cada registro también (i.e. aplicaciones de diagnóstico médico)

## Aplicaciones en el mundo real de la minería de datos

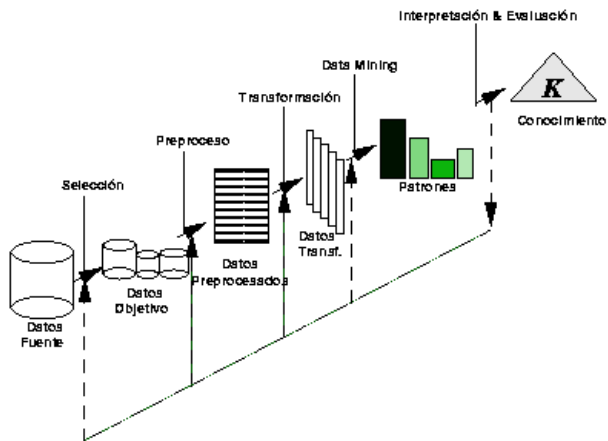
- Marketing: análisis de bases de datos de clientes para identificar grupos y predecir su comportamiento
  - ▶ American Express incrementó sus clientes en un ratio de entre el 10% y el 15%
  - ▶ Análisis de la cesta de la compra
- Inversiones: muchas compañías usan minería de datos para inversiones
  - ▶ LBS usa reglas, redes neuronales y algoritmos genéticos para gestión de carteras por un total de 600 millones de dólares
- Detección de fraude: como por ejemplo, de tarjetas de crédito
- Fabricación: procesos de producción de grandes equipos
  - ▶ el sistema CASSIOPEE, fabricado entre GE y SNECMA (*Aero Space Propulsion and Equipment Group*) lo aplicaron tres compañías europeas para el diagnóstico y predicción de problemas en el Boeing 737.
- Telecomunicaciones, agricultura, diseño web y un largo etcétera



# KDD, definición

- KDD es el proceso no trivial de identificar patrones válidos, nuevos, potencialmente útiles e interpretables en datos
- Datos: casos en una base de datos
- Patrón: expresión en un determinado lenguaje de un modelo acerca de los datos
- No trivial: implica el uso de un proceso de búsqueda e inferencia

# El ciclo de KDD



# Aprendizaje Automático: justificación

¿Puede resultar útil que las máquinas aprendan?

- Argumentos

- ① Coste en desarrollo de S.I. complejos

- ★ Problemas dinámicos
    - ★ Aplicaciones informáticas poco flexibles
    - ★ Se necesitan soluciones adaptativas

- ② Aprendizaje como cualidad básica de la I.A.

”No hay inteligencia en una aplicación que siempre ejecuta las mismas acciones, cometiendo sistemáticamente los mismos errores”

## Aprendizaje Automático: concepto

- El término “Aprendizaje Automático” está íntimamente ligado al aprendizaje humano.
- En el DRAE encontramos

### Definición

*Aprender es adquirir el conocimiento de alguna cosa por medio del estudio o de la experiencia.*

- En ([3], página 2) encontramos

### Definición

*Se dice de un programa de ordenador que aprende a partir de la experiencia  $E$ , con respecto a alguna clase de tareas  $T$  y una medida de rendimiento  $P$ , si su rendimiento en las tareas  $T$ , medido mediante  $P$ , mejora con la experiencia  $E$ .*

## Taxonomías del Aprendizaje Automático

- Taxonomía según [2], basada en tres dimensiones:
  - 1 Estrategia de aprendizaje usada como base
  - 2 Representación del conocimiento a adquirir (o aprender).
  - 3 Dominio de aplicación

Sin Inferencia



Aprendizaje por memorización

Aprendizaje por instrucción

Aprendizaje por analogía

Aprendizaje a partir de ejemplares

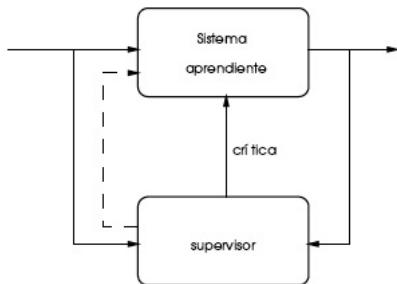
Aprendizaje a partir de observaciones  
y descubrimiento

Aprendizaje por refuerzo

## Enfoques históricos: conductista

- Enfoque Conductista

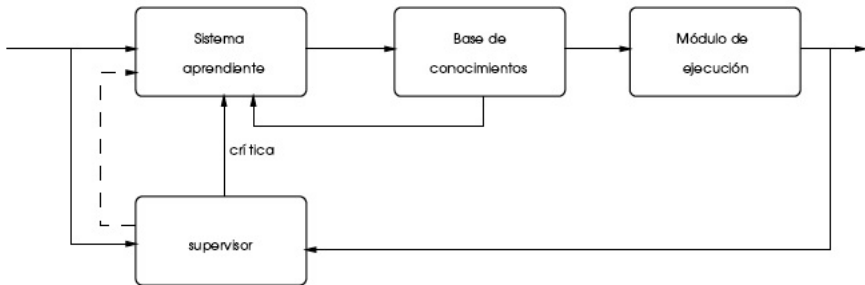
- ▶ Inspirado en el concepto de *refuerzo* de la psicología conductista
- ▶ Aprendizaje implica modificaciones estructurales o paramétricas
- ▶ Modificaciones guiadas por un supervisor



## Enfoques históricos: cognoscitivo

- Enfoque Cognoscitivo

- ▶ Lo más importante es la formación de conocimiento inteligible, su representación y mejora
- ▶ La base de conocimiento evoluciona durante el aprendizaje
- ▶ El módulo de ejecución se encarga de transformar ese conocimiento en actuaciones que reflejan el aprendizaje



## Diseñando un sistema de aprendizaje

Vamos a diseñar un sistema que aprenda a jugar a las damas





## Diseñando un sistema de aprendizaje (y II)

Seguiremos el siguiente orden de tareas (ver [3], página 7)

- 1 Elegir el tipo de experiencia para el entrenamiento
- 2 Elegir una función objetivo
- 3 Elegir una representación para la función objetivo
- 4 Elegir un algoritmo para aproximar la función objetivo

## Selección de $E$

Hay tres decisiones de diseño a tomar para obtener  $E$

- 1 Realimentación directa o indirecta en cada elemento de  $E$ 
  - ▶ Problema de asignación de crédito
- 2 Nivel de control sobre la  $E$  en el entrenamiento
  - ▶ El instructor podría seleccionar situaciones de partida en el tablero y mostrar para cada una el movimiento correcto.
  - ▶ Además, el sistema aprendedor podría preguntar al instructor por estados de tablero que el considera confusos y este último podría mostrarle qué movimientos sería los adecuados para él.
  - ▶ Otra opción podría ser aquella en la que el sistema aprendedor tiene control total sobre los estados del tablero y las partidas, como se haría al jugar contra sí mismo.
- 3 Calidad en la experiencia  $E$ .

## Selección de $E$

- Decidiremos que nuestro sistema va a aprender jugando contra sí mismo.
- Ahora
  - ▶  $T$ : jugar a las damas,
  - ▶  $P$ : porcentaje de partidas ganadas y
  - ▶  $E$ : juegos jugados contra sí mismo.

## Elección de la función objetivo

- ¿cómo evaluar los movimientos del juego, para decidir cuál efectuar en cada momento?
- Podríamos definir una función

$$\textit{EligeMovimiento} : B \longrightarrow M$$

- Vamos a elegir una función muy similar

$$V : B \longrightarrow \mathbb{R}$$

## Elección de la función objetivo (y II)

- Podríamos definir inicialmente  $V$ , para un  $b \in B$  mediante:
  - ① si  $b$  es un estado final del tablero en el que ganamos,  $V(b) = 100$
  - ② si  $b$  es un estado final del tablero en el que perdemos,  $V(b) = -100$
  - ③ si  $b$  es un estado final del tablero en el que empatamos,  $V(b) = 0$
  - ④ si  $b$  no es un estado final,  $V(b) = V(b')$ , en donde  $b'$  es el mejor estado final que se puede conseguir, comenzando en  $b$  y jugando de forma optimal hasta el final del juego.
- Cuando  $b$  cumple el caso 4 tenemos una estupenda explosión combinatoria. Esa definición de  $V$  no es operacional.
- Necesitamos encontrar una aproximación de  $V$  que denotamos con  $\hat{V}$ .

## Representación de $\hat{V}$

Para la definición de  $\hat{V}$  podríamos escoger

- 1 una tabla con una entrada para cada posición distinta para cada estado posible en el tablero de juego,
- 2 una red neuronal artificial,
- 3 una función cuadrática, o
- 4 la que vamos a usar: una combinación lineal de los siguientes factores
  - ▶  $x_1$ : número de piezas negras en el tablero
  - ▶  $x_2$ : número de piezas blancas en el tablero
  - ▶  $x_3$ : número de reinas negras en el tablero
  - ▶  $x_4$ : número de reinas blancas en el tablero
  - ▶  $x_5$ : número de piezas negras amenazadas por las blancas
  - ▶  $x_6$ : número de piezas blancas amenazadas por las negras

$$\hat{V}(b) = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6$$

## Representación de $\hat{V}$

La tarea de aprendizaje se va a reducir a encontrar los  $w_i$ . Ahora, redefiniendo nuestro sistema aprendedor, respecto a la definición original de aprendizaje

- $T$ : jugar a las damas
- $P$ : porcentaje de juegos ganados en el torneo mundial
- $E$ : juegos jugados contra sí mismo
- Función objetivo:  $V : B \rightarrow \mathfrak{R}$
- Representación de  $V$

$$\hat{V}(b) = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6$$

## Evaluando los ejemplos de entrenamiento

- La única información que el sistema aprendedor tiene, para cada ejemplo, es si la partida se ganó o no a partir de una posición  $b$ .
- Necesitamos ejemplos de entrenamiento con valores específicos de bondad para valores determinados de estados en el tablero → situación ambigua
- Mitchell sugiere una función muy sencilla, y efectiva

$$V_{train}(b) \leftarrow \hat{V}(sucesor(b))$$

- Justificación intuitiva diciendo que  $\hat{V}$  será más precisa conforme nos vamos acercando a las posiciones finales de cada partida.



## Ajustando los $w_i$

- Necesitamos encontrar una  $\hat{V}$  que minimice el error

$$E = \sum_{\langle b, V_{train}(b) \rangle \in \text{ejemplos entrenamiento}} (V_{train}(b) - \hat{V}(b))^2$$

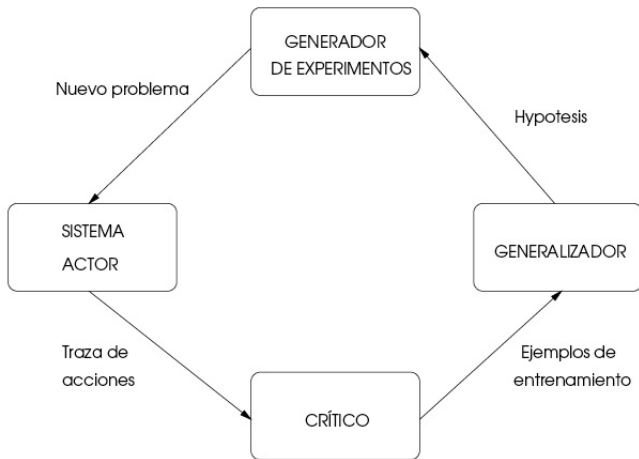
- Sugerimos el algoritmo LMS (*least mean squares*),

*Regla de actualización de pesos LMS*

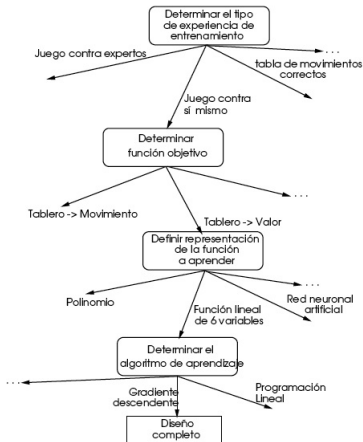
- ▶ Usar los  $w_i$  actuales para el cálculo de  $\hat{V}(b)$
- ▶ Para cada  $w_i$ , actualizarlo según

$$w_i \leftarrow w_i + \mu(V_{train}(b) - \hat{V}(b))x_i$$

## Diagrama estructural del sistema final



# Resumen de la metodología



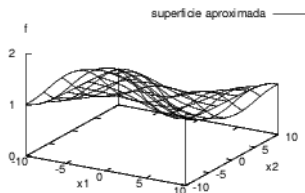
## La dualidad del aprendizaje computacional

- El **sesgo de representación** determina cómo se representa el espacio de hipótesis en donde se busca un modelo de inducción
- El **sesgo de búsqueda** define cómo el algoritmo evoluciona a través del espacio de hipótesis para encontrar un modelo de inducción aceptable

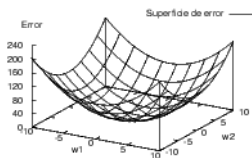
*Vamos a estudiar los paradigmas básicos de aprendizaje supervisado, desde esos dos ángulos*

## Sesgos en redes neuronales

- Sesgo de representación  $\hat{f}_r = x_1 w_1 + x_2 w_2 + y$

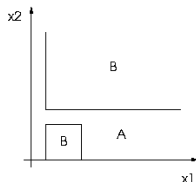


- Sesgo de búsqueda  $E = \frac{1}{2} \sum_{x \in D} (c_i - o_i)^2$



## Sesgos en inducción de árboles/reglas de decisión

- Sesgo de representación



- Sesgo de búsqueda

En la familia ID3

- Entropía
- Atributos con muchos valores
- Distribución de clases poco balanceada

En algoritmos AQ, CN2

- Cubrimiento secuencial
- Búsqueda beam (mantiene varios complejos)

## Sesgos en Algoritmos Evolutivos

- Sesgo de representación → independiente del algoritmo



- Sesgo de búsqueda
  - ▶ Búsqueda *beam* (en varios haces)
  - ▶ Aleatoria
  - ▶ Saltos bruscos en el espacio de búsqueda

## Sesgos en Bayes *naïve*

- Sesgo de representación  $\rightarrow$  valores de probabilidad para  $p(C_i)$  y  $p(x|C_i)$
  
  
  
  
  
  
  
  
  
  
- Sesgo de búsqueda  $\rightarrow$  no hay. Simplemente contamos repeticiones de valores en el conjunto de entrenamiento.

$$p(C_i|x) = p(C_i|x_1, x_2, \dots, x_n) = \frac{p(C_i)p(x|C_i)}{\sum_j p(C_j)p(x|C_j)}$$



## Conclusión inicial

Por su heterogeneidad, no existe una técnica de aprendizaje supervisado que se comporte mejor que el resto para todo tipo de problema y por lo tanto ...

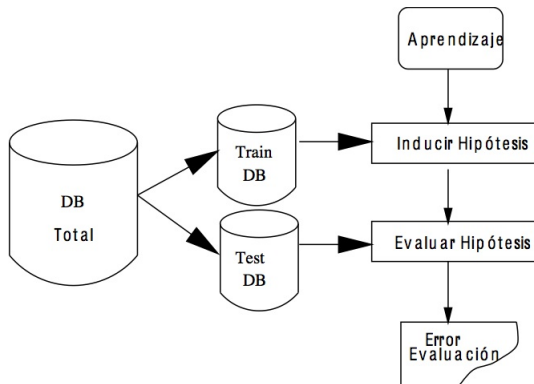
Necesitamos conocer todas las técnicas para poder enfrentarnos con garantías de éxito a problemas en los que poder aplicar el Aprendizaje Automático

# Estimación de la calidad de un modelo inductivo

- ¿Para qué sirve?
  - ▶ Para medir su capacidad de generalización
  - ▶ Por propia necesidad de los algoritmos (proceso de búsqueda)
  - ▶ Para seleccionar un modelo entre un conjunto
  - ▶ Para combinar modelos
- Técnicas
  - ▶ Holdout
  - ▶ Validación cruzada de  $k$  pliegues
  - ▶ Proceso bootstrap

# Holdout

- Enfoque tradicional (más simplista)
- Se divide el conjunto  $D$  en  $D_l$  y  $D_t$ , tal que  $\frac{|D_l|}{|D|} = 0.75(\text{approx.})$
- Aprendizaje sobre  $D_l$ , estimación sobre  $D_t$



## Validación cruzada $k$ pliegues

- Se divide el conjunto  $D$  en  $k$  partes  $\{D_1, D_2, \dots, D_k\}$  iguales, y disjuntas
- Se realizan  $k$  procesos de aprendizaje, usando en el proceso  $i$  el conjunto  $D_i$  como test, y el resto para el aprendizaje
- Estimación del error según

$$acc_{cv} = \frac{1}{n} \sum_{j=1}^n \sum_{\langle v_i, y_i \rangle \in D_j} \delta(\mathcal{I}(D - D_j, v_i), y_i),$$

en donde

- ▶  $|D| = n$ ,
- ▶  $\mathcal{I}(A, v)$  la salida para el valor  $v$  del modelo inducido por  $\mathcal{I}$  en el conjunto  $A$  y
- ▶  $\delta(x, y)$  es la diferencia entre las predicciones  $x$  e  $y$

# Bootstrap

- Sea  $O$  de tamaño  $n$ .
- Una muestra bootstrap se hace tomando  $n$  muestras, del conjunto con remplazamiento.
- La probabilidad de que una instancia cualquiera no se haya escogido es de

$$(1 - 1/n)^n \approx e^{-1} \approx 0.368$$

- Con esas  $n$  muestras se compone el conjunto de entrenamiento,  $D_t$ .
- El resto van al conjunto de test,  $D_t$ .
- Con  $n$  razonablemente grande, el conjunto de test contendrá un 36.8% de las instancias
- El conjunto de aprendizaje contendrá un 63.2%.
- El estimador bootstrap se obtiene mediante la expresión

$$acc_{boot} = \frac{1}{b} \sum_{i=1}^b (0.632e_{i_i} + 0.368acc_D)$$

- ▶ Siendo  $b$  el número de clasificadores obtenidos,
- ▶  $e_{i_i}$  el error de evaluación para el clasificador  $i$
- ▶  $acc_s$  es el estimador de rendimiento en el conjunto  $D$ .

## Evaluación de Hipótesis de Regresión - Errores absolutos

- Sea  $F$  la hipótesis de regresión
- Sea  $\{v_i^1, v_i^2, \dots, v_i^n\}$  el conjunto de instancias
- Error cuadrático medio

$$MSE = \frac{1}{n} \sum_{j=1}^n (F(v_i^j) - v_o^j)^2$$

- ▶ El más común
- ▶ A veces se usa el error *rooted*, para expresar el error en las mismas unidades que las magnitudes que se aproximan.
- Error absoluto medio

$$MAE = \frac{1}{n} \sum_{j=1}^n |F(v_i^j) - v_o^j|$$

- ▶ Se comporta mejor que el anterior cuando determinados valores de

## Evaluación de Hipótesis de Regresión - Errores relativos

- Cuando es más importante saber la proporción del error
- Error cuadrático relativo

$$RSE = \frac{\sum_{j=1}^n (F(v_i^j) - v_o^j)^2}{\sum_{j=1}^n v_o^j - \bar{v}_o},$$

en donde  $\bar{v}_o = \frac{1}{n} \sum_{j=1}^n v_o^j$ .

- Error absoluto relativo

$$RAE = \frac{\sum_{j=1}^n |F(v_i^j) - v_o^j|}{\sum_{j=1}^n |v_o^j - \bar{v}_o|},$$

# Bibliografía



Michael Berthold and David J. Hand.  
*Intelligent Data Analysis. An Introduction.*  
Springer, 2003.  
Second edition.



Ryszard S. Michalski.  
A theory and methodology of inductive learning.  
In R.S. Michalski, J.G. Carbonell, and T.M. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach*, volume 1, pages 83–129. Springer, 1983.



Tom M. Mitchell.  
*Machine Learning.*  
McGraw-Hill, 1997.



Jaideep Srivastava, Robert Cooley, Mukund Deshpande, and Pang-Ning Tan.  
Web usage mining: Discovery and applications of usage patterns from web data.  
*SIGKDD Explorations*, 1(2):12–23, 2000.



Ian H. Witten and Eibe Frank.  
*Data Mining. Practical Machine Learning Tools and Techniques with JAVA Implementations.*  
Morgan Kauffman, 2000.