



Inteligencia en Redes de Comunicaciones

Tema 6
Aprendizaje

Julio Villena Román, Raquel M. Crespo García, José Jesús García Rueda
{jvillena, rcrespo, rueda}@it.uc3m.es



Universidad
Carlos III de Madrid

El objetivo de este tema es realizar una introducción a los conceptos, fundamentos y técnicas básicas de aprendizaje computacional. En primer lugar se formalizará el concepto de aprendizaje y se describirán las diferencias entre aprendizaje inductivo y deductivo, y luego inductivo supervisado y no supervisado. En la segunda parte del tema se estudiarán las técnicas de aprendizaje más ampliamente utilizadas, en concreto, de aprendizaje inductivo supervisado se estudiará el algoritmo de los K-vecinos más cercanos, el algoritmo C4.5 de árboles de decisión y las redes neuronales, y respecto a aprendizaje inductivo no supervisado se presentan las técnicas de clustering y de análisis de dependencias y asociaciones (algoritmo *a-priori*).

Objetivo

El objetivo es dotar
a los sistemas de inteligencia artificial
con la cualidad humana del **aprendizaje**,
esto es,
ser capaz de adquirir nuevos conocimientos
a partir de las experiencias
(relación con el entorno)

El objetivo del aprendizaje computacional...

Índice

- ▶ Formalización y tipos de aprendizaje
- ▶ Aprendizaje inductivo
 - ▶ Inducción supervisada
 - ▶ Inducción no supervisada
- ▶ Aprendizaje deductivo
- ▶ Otros tipos
 - ▶ Aprendizaje abductivo
 - ▶ Algoritmos genéticos
 - ▶ Aprendizaje por analogía
 - ▶ ...

Índice del tema.

Aprendizaje o adquisición de conocimiento

▶ **Tres aspectos:**

- ▶ Incrementar el conocimiento adquirido
- ▶ Realizar tareas con mayor eficiencia o exactitud
- ▶ Llevar a cabo nuevas tareas

“Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the task or tasks drawn from the same population more efficiently and more effectively the next time” – Herb Simon

Fundamentos

- ▶ Basado en una actividad de diálogo de un sistema inteligente (agente) con el medio y/u otros agentes
- ▶ Siempre hay relación con el entorno, incluso en los sistemas que aprenden por introspección
- ▶ Es imprescindible una capacidad de memorizar
- ▶ Consiste en una abstracción:
 - ▶ **nivel simbólico:** procesos de manipulación de símbolos
 - ▶ **nivel de algoritmo:** secuencia de decisiones y acciones
 - ▶ **nivel del conocimiento:** indica el objetivo, las reglas del juego y demás elementos de información

Formalización

▶ Esquema genérico de inferencia (razonamiento):

$$CB \cup P \Rightarrow C$$

- ▶ CB = conocimiento base, todo lo conocido dependiente o no del dominio
 - ▶ P = premisa (regla o hecho del dominio de aplicación)
 - ▶ C = consecuente
- ▶ Tarea de aprendizaje:
identificar correctamente un elemento desconocido D, a partir de una información I sobre él

Formalización del aprendizaje como ecuación de esquema de inferencia.

Tipos de inferencia

- ▶ **inferencia deductiva**: particularizar la estructura que refleja el conocimiento disponible para justificar la información surgida

$$CB \cup P \Rightarrow i?$$

- ▶ **inferencia inductiva**: establecer una hipótesis P que permita determinar el papel de la información C respecto a la estructura lógica contenida en CB

$$CB \cup i? \Rightarrow C$$

Grandes tipos de aprendizaje: inductivo y deductivo.

Ejemplos

Deducción

Cuando llueve y sales sin paraguas, te mojas

Yo no tengo paraguas

∪

Hoy llueve

⇒

Hoy me voy a mojar

Inducción

“El maestro de esgrima” me gusta

“El club Dumas” me gusta

“La Reina del Sur” me gusta

“La tabla de Flandes” me gusta

“La carta esférica” me gusta

∪

(todos son de A. Pérez Reverte)

⇒

Los libros de A. Pérez Reverte me gustan

El caso de la izquierda es un ejemplo típico de aplicación de inferencia deductiva. A partir de un conjunto de reglas genéricas (p.ej. “Cuando llueve y sales sin paraguas, te mojas”), unido a un hecho concreto (“Hoy llueve”), se puede deducir una conclusión (en este caso, “Hoy me voy a mojar”).

En cambio, en el caso de la izquierda se parte de hechos concretos sobre el dominio. En este caso, hay una serie de libros cada uno de los cuales me gusta individualmente. A ello se suma el conocimiento base de que todos ellos comparten como característica común que su autor es A. Pérez Reverte. A partir de esta información, se puede formular la hipótesis de que “Los libros de A. Pérez Reverte me gustan”.

Tipos de aprendizaje según el proceso

- ▶ Aprendizaje **inductivo**
 - ▶ inducción supervisada
 - ▶ vecino más cercano (*nearest neighbour*)
 - ▶ árboles de decisión (*decision trees*)
 - ▶ redes neuronales (*neural networks*)
 - ▶ inducción no supervisada
 - ▶ agrupamiento (*clustering*)
 - ▶ análisis de dependencias
 - ▶ redes neuronales (*neural networks*)
- ▶ Aprendizaje **deductivo**
- ▶ Aprendizaje de **naturaleza mixta**
 - ▶ Aprendizaje abductivo
 - ▶ Algoritmos genéticos
 - ▶ Causa-efecto
 - ▶ aprendizaje por analogía
 - ▶ razonamiento basado en casos

El aprendizaje puede clasificarse, en función del proceso, en tres categorías:

-Aprendizaje inductivo: consiste en inferir una regla general o premisa a partir de información sobre hechos concretos (tal como se ha visto en los ejemplos anteriores).

-Aprendizaje deductivo: consiste en aplicar inferencia deductiva para particularizar la estructura que refleja el conocimiento disponible dado un hecho o premisa concreto

-Aprendizaje de naturaleza mixta: en este grupo se incluyen otros tipos de procesos de aprendizaje no basados ni en inferencia inductiva ni deductiva

Tipos de inducción según los datos

▶ Aprendizaje supervisado

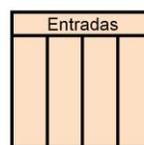
- ▶ Dado un conjunto de ejemplos en la forma de pares (*entradas, salidas*), encontrar las reglas que mejor modelan la relación entre las entradas y las salidas



$$S_i = f(E_1, E_2, \dots)$$

▶ Aprendizaje no supervisado

- ▶ Extraer conclusiones a partir de un conjunto de ejemplos sin etiquetar (*sólo entradas*)



- Agrupaciones
- Relaciones entre variables

El aprendizaje inductivo puede a su vez clasificarse en dos categorías en función de los datos disponibles para dicho aprendizaje: Aprendizaje supervisado y aprendizaje no supervisado

El aprendizaje supervisado consiste en inferir una relación (reglas, función matemática, modelo...) entre una(s) entrada y una(s) salida(s). En este caso, los datos necesarios para el proceso de aprendizaje son un conjunto de pares entrada-salida. Es decir, un conjunto de casos en los que conocemos los valores de salida correspondientes dado un determinado conjunto de valores de entrada.

El objetivo último de este tipo de aprendizaje (supervisado) es conocer la relación entre los parámetros de entrada y las variables de salida de modo que, para un caso nuevo en el que sólo se conozcan los valores de entrada, se pueda deducir la salida correspondiente. Es decir, en el conjunto de datos que se utilizan como base para el proceso de aprendizaje se dispone de más información que en futuras aplicaciones del modelo inferido. Así, durante la etapa de aprendizaje los datos disponibles incluyen tanto los parámetros de entrada como la salida correspondiente. Mientras que una vez completado el proceso de aprendizaje, el modelo generado servirá para deducir la salida correspondiente (desconocida a priori) dados unos determinados parámetros de entrada.

Un ejemplo probablemente ayudará a clarificar este concepto. Uno muy típico es una entidad financiera que desea crear un sistema inteligente para ayudar a decidir sobre la concesión de créditos a los clientes. En este caso, los datos de entrada son las características del cliente en el momento de solicitar el préstamo: saldo, ingresos, garantías, etc. y la salida es la variable que se desea predecir: si el cliente devolvió o no el préstamo. En una primera etapa, se utilizarán los datos históricos de préstamos previos para que el sistema “aprenda” a decidir sobre la solvencia del cliente. En esta etapa de aprendizaje no sólo se dispone de las características del cliente sino también de si dicho cliente efectivamente devolvió el préstamo. Con ello se intenta establecer una relación para intentar predecir el resultado a partir sólo de la información de entrada y se puede evaluar la precisión de dichas predicciones. Una vez se ha generado el modelo, podrá utilizarse con nuevas solicitudes. Pero en estos nuevos casos, obviamente, no dispondremos de información sobre la futura devolución o no del préstamo (¡el cliente está solicitándolo!).

Por el contrario, en el caso del aprendizaje no supervisado simplemente se trata de inferir la potencial estructura subyacente en el conjunto de datos. En este caso, no hay variables de entrada y de salida, todas las variables se utilizan como entradas. La información disponible sobre cada caso de ejemplo para el aprendizaje es la misma que la que se recibirá en el futuro para nuevos casos. ¿Cómo podemos aprender algo incluso aunque no tengamos información sobre una ‘salida esperada’? Incluso así se puede aprender, sólo que otro tipo de información. Un ejemplo bien conocido de este tipo de aprendizaje es deducir a partir de los datos históricos que la mayoría de las llamadas telefónicas que hace un determinado cliente se dirigen a un número reducido de números.

Terminología

Entrada(s) Salida(s)

Temp	Tiempo	Día	Ropa	Elección
26°C	Sol	S	Informal	Andar
2°C	Nieve	L	Informal	Conducir
17°C	Nublado	M	Informal	Andar

Ejemplo
(o individuo)

Variable, atributo o dato

- Variable nominal o categórica: valores discretos
- Variable numérica: valores no discretos
- Variable de salida → también suele llamarse *clase*
- Clase true/false → ejemplo *positivo/negativo*
- Clase discreta/numérica → problema de *clasificación / regresión*

 Universidad Carlos III de Madrid

IRC 2011/2012 - 11

Antes de explicar más en detalle cada uno de los tipos de aprendizaje, así como técnicas disponibles para cada uno de ellos se necesita definir la terminología en que nos vamos a basar para referirnos a los datos.

En función del conjunto de valores que puedan tomar, las variables se clasifican en:

-Nominales o categóricas: toman valores discretos, sin un orden o relación de distancia predefinido entre ellos (por ejemplo: verdadero/falso; la variable color con valores rojo, amarillo y verde; días de la semana –en este caso hay una secuencia, pero no podemos afirmar que el martes es mayor que el lunes-, etc.)

-Numéricas: variables cuyo rango de valores tiene una determinada relación de orden entre ellos. Pueden ser a su vez discretas (por ejemplo, la edad) o continuas (por ejemplo, el peso).

Formalización del aprendizaje supervisado

- ▶ Dado el conjunto de datos:

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$$

$\{x_1^1, x_1^2, \dots, x_1^n\}$
entradas

salidas

- ▶ Encontrar una hipótesis (función) que aproxime lo mejor posible el valor de y a partir de los valores de x



Formalmente, el aprendizaje supervisado consiste en, dado un conjunto de datos formado por un conjunto de pares (entrada, salida), encontrar un hipótesis (función) que aproxime lo mejor posible el valor de la salida en función de los valores de la entrada. Tanto las entradas como las salidas pueden ser a su vez n-tuplas compuestas de n valores.

Aspectos del aprendizaje supervisado

- ▶ Basado en cierta manera en la memorización
 - ▶ Posible limitación computacional por los recursos disponibles

Temp	Tiempo	Día	Ropa	Elección
26°C	Sol	S	Informal	Andar
2°C	Nieve	L	Informal	Conducir
17°C	Nublado	M	Informal	Andar
17°C	Nublado	M	Informal	

En estas diapositivas se presentan ciertas consideraciones sobre el aprendizaje inductivo supervisado. En este caso, se hace referencia a que todos los algoritmos de aprendizaje supervisado se basan en cierta forma en la memorización. En el ejemplo presentado, la predicción sería “Andar” porque en el pasado la elección en circunstancias análogas fue “Andar”.

Aspectos del aprendizaje supervisado (2)

- ▶ Tarea incremental e incierta: las reglas se asumen como correctas mientras no aparezcan casos que lo desmientan

Temp	Tiempo	Día	Ropa	Elección
17°C	Sol	L	Informal	Conducir
17°C	Sol	L	Informal	Conducir
17°C	Sol	L	Informal	Andar
17°C	Sol	L	Informal	Andar
17°C	Sol	L	Informal	Andar
17°C	Sol	L	Informal	Andar
17°C	Sol	L	Informal	

El aprendizaje supervisado es una tarea incremental, puesto que evoluciona según se reciben nuevos datos de entrada. E incierta, puesto que reglas que inicialmente parecen correctas pueden descartarse posteriormente con la llegada de nuevos datos que las desmienten.

Por ejemplo, basándonos en los datos de la tabla, si tuviésemos información únicamente sobre los dos primeros casos llegaríamos a la conclusión que siempre que la temperatura son 17°, hace sol, es lunes y llevamos ropa informal cogeremos el coche. Sin embargo, si posteriormente llegan los casos de las filas 3 a 6, veremos que la mayoría de días que se cumplen las condiciones anteriores se opta por ir andando (y no en coche, como erróneamente habíamos concluido al basarnos sólo en los primeros datos).

Aspectos del aprendizaje supervisado (3)

- ▶ Posible existencia de ruido en los datos
- ▶ El nº de reglas inducidas depende inversamente del nº de ejemplos presentados

Temp	Tiempo	Día	Ropa	Elección
23°C	Sol	V	Formal	Conducir
13°C	Sol	D	Informal	Andar
21°C	Lluvia	X	Informal	Andar
28°C	Sol	L	Informal	Conducir
12°C	Sol	S	Formal	Conducir
26°C	Sol	S	Informal	Andar
-5°C	Nieve	L	Informal	Conducir
22°C	Sol	M	Informal	Andar
17°C	Lluvia	L	Informal	

El número de reglas inducidas depende inversamente del número de ejemplos presentados, puesto que cuanto más información tengamos más restrictiva será. Si sólo tenemos unos cuantos datos, podemos hacer muchas hipótesis sobre la predicción (por ejemplo, que con ropa informal, la elección es “Andar”, o también que los lunes la elección es “Conducir”), pero a medida que tenemos más datos de entrenamiento, se van invalidando algunas de esas hipótesis.

Finalmente, es importante destacar que a menudo existe ruido en los datos de entrenamiento.

Aprendizaje inductivo supervisado

- ▶ **Aprendizaje inductivo**
 - ▶ inducción supervisada
 - ▶ vecino más cercano (*nearest neighbour*)
 - ▶ árboles de decisión (*decision trees*)
 - ▶ redes neuronales (*neural networks*)

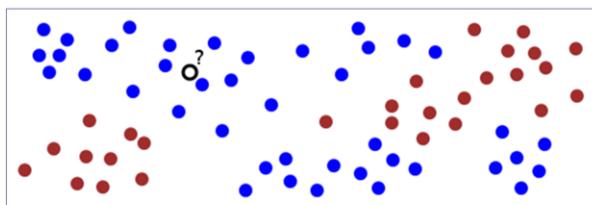
- ▶ **Evaluación**
- ▶ **Sobreentrenamiento**

Existen numerosos algoritmos y modelos basados en aprendizaje supervisado. La elección de una técnica y otra dependerá de factores como el número de datos disponibles, el tipo de las variables de entrada y salida (numérico/categorico), si existe o no ruido en los datos o casos contradictorios, etc.

Vecino más cercano (*Nearest Neighbour*)

- ▶ Almacenar todos los datos
- ▶ Dado un ejemplo, busca el ejemplo almacenado más parecido y devuelve su salida como respuesta
 - ▶ Suele usarse la distancia Euclídea:

$$d(x^i, x^k) = \sqrt{\sum_j (x_j^i - x_j^k)^2}$$



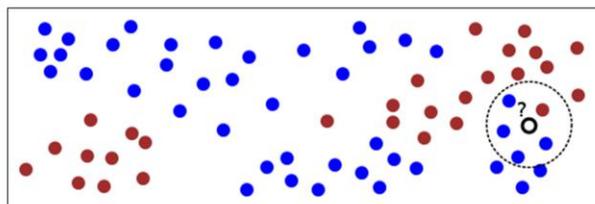
El modelo del vecino más cercano (Nearest Neighbour) es la técnica más sencilla de aprendizaje supervisado. Está basado exclusivamente en memorización, sin aplicar ningún tipo de procesamiento para ‘resumir’ la información de los casos de muestra.

Consiste en almacenar todos los ejemplos de muestra conocidos. Para cada caso nuevo que llega, busca el ejemplo almacenado más ‘parecido’ –es decir, aquél cuya distancia al objetivo es mínima- y devuelve su salida como respuesta. Como métrica de distancia para identificar el caso más próximo suele usarse la distancia Euclídea, pero puede usarse en su lugar otra métrica si resulta más adecuada para el problema en cuestión.

Por ejemplo, suponiendo que los puntos de la imagen se correspondan con coordenadas en el mar (azul) o tierra (rojo), el punto ‘?’ se clasificaría como mar, puesto que el más próximo es de dicha clase.

K-NN

- ▶ Se buscan los **K ejemplos más parecidos**, y como valor de salida se da:
 - ▶ Variables **numéricas**: el valor medio
 - ▶ Variables **categorías**: el valor más frecuente

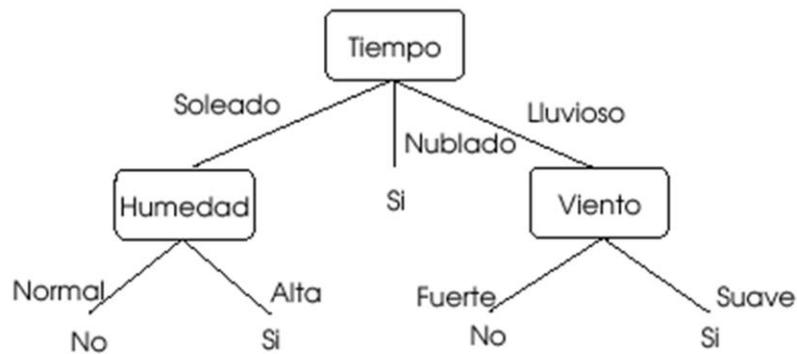


El modelo k-NN es una optimización del vecino más cercano (Nearest Neighbour), que intenta minimizar el riesgo de errores debidos a casos próximos a la frontera o espúreos.

En lugar de limitarse a buscar el ejemplo más próximo, busca los N casos más próximos. Para variables de salida discretas, el algoritmo k-NN devuelve como resultado la moda (el valor más frecuente). Para variables continuas, se devuelve como resultado la media.

Continuando con el ejemplo anterior, si se intenta clasificar el punto situado en la interrogación (“?”) el algoritmo del vecino más cercano daría como resultado que pertenece a la clase tierra (rojo). Sin embargo, al estar rodeado de puntos azules (mar) es probable que éste sea un valor más correcto para esas coordenadas. El algoritmo k-NN (con $k=5$) evaluaría los 5 casos más próximos y daría como resultado mar (azul), puesto que la mayoría pertenecen a esa categoría.

Árboles de decisión



**IF (tiempo=Soleado AND Humedad=Alta) OR
(tiempo=Lluvioso AND Viento=Suave) OR
tiempo=Nublado
THEN Coger_Paraguas=Si**

Ejemplo del algoritmo de árboles de decisión con el criterio para decidir si coger o no el paraguas al salir de casa según las condiciones meteorológicas.

Árboles de decisión (2)

- ▶ Algoritmo de aprendizaje por inducción supervisada que pretende modelar los datos de ejemplo mediante un árbol, llamado **árbol de decisión**
- ▶ En este árbol, los nodos intermedios son los **atributos de entrada** de los ejemplos presentados, las ramas representan **valores de dichos atributos** y los nodos finales son los valores de la **clase**
- ▶ Para elegir qué atributos y en qué orden aparecen en el árbol, se utiliza una función de evaluación: **ganancia de información**
- ▶ Pueden reescribirse en forma de reglas IF-THEN
- ▶ Algoritmo base: **ID3** (inicial, sólo datos nominales), **C4.5** (también datos numéricos)

Base teórica

1) La **entropía** es la cantidad de bits, en promedio, que harían falta para codificar los ejemplos

≈ Medida de la “cantidad de información” representada

$$Et(S) = -p \log_2(p) - n \log_2(n) \text{ (caso binario)}$$

(p y n = proporciones de ejemplos positivos/negativos)

$$Et(s) = \sum_i (-t_i \log_2(t_i)) \text{ (en general)}$$

Base teórica de los árboles de decisión: entropía.

Base teórica (2)

2) La **ganancia de información** sería la reducción de entropía del conjunto al clasificar usando el atributo

$$G(S,A) = Et(S) - \sum_i (t_i/T \cdot Et(S_{t_i}))$$

para todos los valores diferentes del atributo A:

t_i =número de ejemplos de S etiquetados con ese valor

T=número total de ejemplos

$Et(S_{t_i})$ =entropía de los ejemplos etiquetados con ese valor

- ▶ Se selecciona el atributo con mayor ganancia de información en cada iteración

Base teórica de los árboles de decisión: ganancia de información.

Ejemplo:
Modelo psicológico de la felicidad de un estudiante

Estudia	Vida social	Aprueba	Feliz
Poco	Poco	Poco	NO
Poco	Poco	Mucho	SI
Poco	Algo	Algo	NO
Poco	Algo	Mucho	SI
Poco	Mucho	Poco	SI
Algo	Poco	Poco	NO
Algo	Poco	Mucho	NO
Algo	Mucho	Poco	SI
Mucho	Poco	Poco	NO
Mucho	Poco	Algo	NO
Mucho	Algo	Mucho	NO
Mucho	Mucho	Algo	SI

Ejemplo de cálculo de árbol de decisión: modelo psicológico de la felicidad de un estudiante.

Atributo de la primera división

$$Et(S) = -5/12 \log_2(5/12) - 7/12 \log_2(7/12) = 0,979$$

$$Et(\text{Estudia=Poco}) = -3/5 \log_2(3/5) - 2/5 \log_2(2/5) = 0,970$$

$$Et(\text{Estudia=Algo}) = -1/3 \log_2(1/3) - 2/3 \log_2(2/3) = 0,918$$

$$Et(\text{Estudia=Mucho}) = -1/4 \log_2(1/4) - 3/4 \log_2(3/4) = 0,811$$

$$\mathbf{G(S,Estudia)} = Et(S) - 5/12 \cdot 0,970 - 3/12 \cdot 0,918 - 4/12 \cdot 0,811 = \mathbf{0,075}$$

$$Et(\text{VidaSocial=Poco}) = 0,650$$

$$Et(\text{VidaSocial=Algo}) = 0,918$$

$$Et(\text{VidaSocial=Mucho}) = -3/3 \log_2(3/3) - 0/3 \log_2(0/3) = 0$$

$$\mathbf{G(S,VidaSocial)} = Et(S) - 6/12 \cdot 0,650 - 3/12 \cdot 0,918 - 3/12 \cdot 0 = \mathbf{0,425}$$

$$Et(\text{Aprueba=Poco}) = 0,970$$

$$Et(\text{Aprueba=Algo}) = 0,918$$

$$Et(\text{Aprueba=Mucho}) = 1,0$$

$$\mathbf{G(S,Aprueba)} = Et(S) - 5/12 \cdot 0,970 - 3/12 \cdot 0,918 - 4/12 \cdot 1,0 = \mathbf{0,012}$$

Elección del atributo de la primera división (raíz del árbol).

Primera división (Vida social)



Atributo de la segunda división

$$Et(S) = -1/6 \log_2(1/6) - 5/6 \log_2(5/6) = 0,650$$

$$Et(\text{Estudia=Poco}) = 1,0$$

$$Et(\text{Estudia=Algo}) = 0$$

$$Et(\text{Estudia=Mucho}) = 0$$

$$\mathbf{G(S,Estudia)} = Et(S) - 2/6 \cdot 1,0 - 2/6 \cdot 0 - 2/6 \cdot 0 = \mathbf{0,316}$$

$$Et(\text{Aprueba=Poco})=0$$

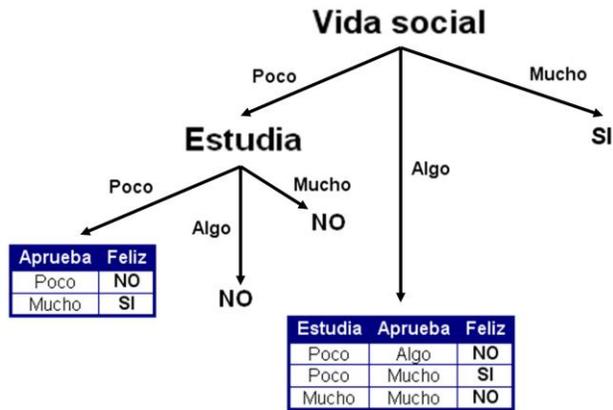
$$Et(\text{Aprueba=Algo})=0$$

$$Et(\text{Aprueba=Mucho})=1,0$$

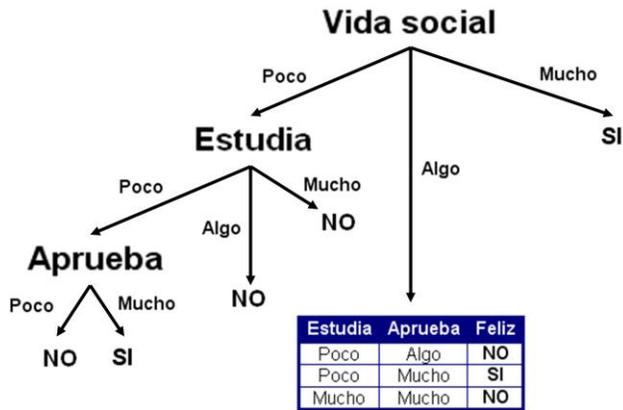
$$\mathbf{G(S,Aprueba)} = Et(S) - 3/6 \cdot 0 - 1/6 \cdot 0 - 2/6 \cdot 1,0 = \mathbf{0,316}$$

Elección del atributo de la segunda división, rama de la izquierda.

Segunda división (Estudia)

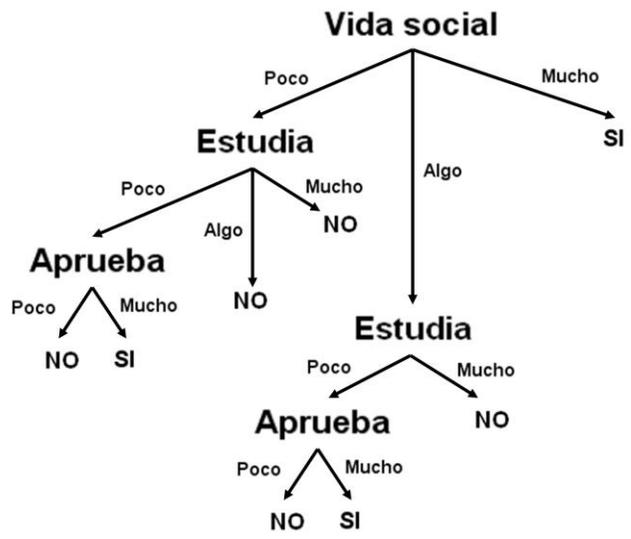


Tercera división (Aprueba)



Elección del atributo de la tercera división, rama de la izquierda.

Árbol completo



Reglas

Si VS=M entonces F=SI

Si VS=P y E=P y AP=P entonces F=NO

Si VS=P y E=P y AP=M entonces F=SI

Si VS=P y E=A entonces F=NO

Si VS=P y E=M entonces F=NO

Si VS=A y E=P y AP=A entonces F=NO

Si VS=A y E=P y AP=M entonces F=SI

Si VS=A y E=M entonces F=NO

Si VS=M entonces F=SI

En otro caso F=NO (Regla por defecto)

Los árboles de decisión se pueden convertir en reglas muy fácilmente, que son muy sencillas de implementar en cualquier lenguaje de programación.

Reglas simplificadas

Si $VS=M$ entonces $F=SI$

Si $VS=P$ y $E=P$ y $AP=P$ entonces $F=NO$

Si $VS=P$ y $E=A$ entonces $F=NO$

Si $VS=P$ y $E=M$ entonces $F=NO$

Si $VS=A$ y $E=P$ y $AP=A$ entonces $F=NO$

Si $VS=A$ y $E=M$ entonces $F=NO$

Si $VS=M$ entonces $F=SI$

Si $E=P$ y $AP=M$ entonces $F=SI$

En otro caso $F=NO$ (Regla por defecto)

Podado de árboles (*pruning*)

- ▶ Para reducir el **tamaño** de los árboles de decisión y aumentar su **grado de generalización**
- ▶ Elimina las ramas del árbol que corresponden a un número de ejemplos menor que un umbral dado
- ▶ Enlaza fragmentos “parecidos” de árbol para reducir el número de nodos y así generalizar las reglas, simplificándolas

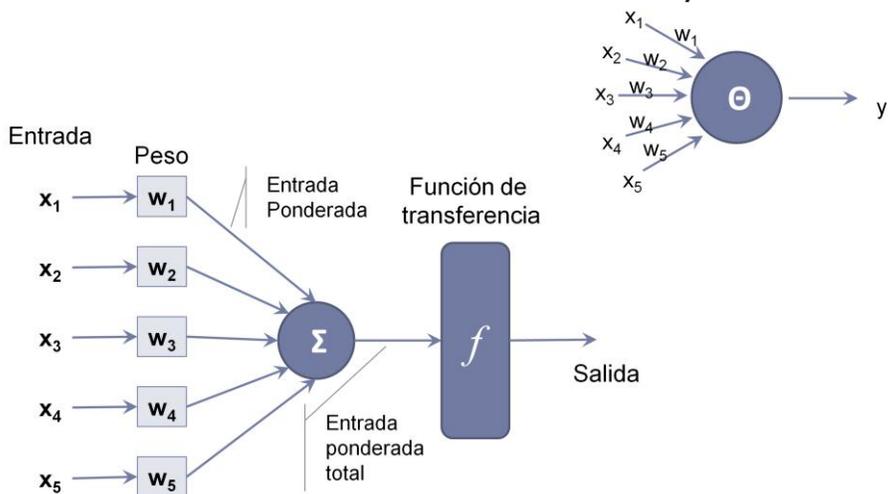
Redes neuronales

- ▶ **Tres cualidades básicas del cerebro humano:**
 - ▶ El conocimiento está distribuido en un (gran) número de neuronas en el cerebro
 - ▶ Las neuronas se comunican unas con otras
 - ▶ El cerebro es adaptable (puede aprender)
- ▶ Como su nombre indica, las redes neuronales están inspiradas en el modelo biológico y son una simplificación matemática basada en el estímulo/respuesta de las neuronas
- ▶ Son capaces de “aprender” (modelar) la relación existente entre sus entradas y salidas

Fundamentos biológicos de las redes neuronales.

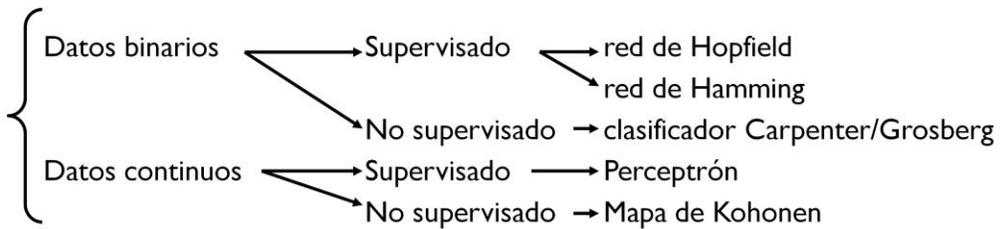
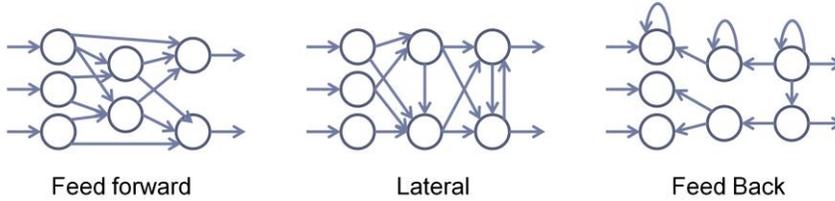
Modelo matemático de una neurona

- ▶ Modelan la relación existente entre entradas y salidas



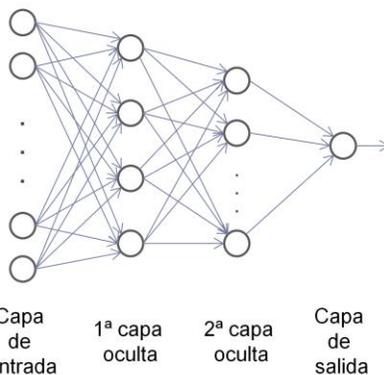
Modelos de redes neuronales

- ▶ Las topologías de conexión de las neuronas dan lugar a los distintos tipos de redes neuronales



Perceptrón multicapa

- ▶ La estructura de red neuronal más sencilla
- ▶ Organización en capas:
 - ▶ Capa de entrada
 - ▶ Capa(s) oculta(s)
 - ▶ Capa de salida
- ▶ Entrenamiento: algoritmo de retropropagación (*backpropagation*)



Capacidades del perceptrón

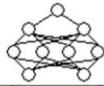
Estructura	Problema X-OR	Regiones cóncavas	Regiones
2 capas 			
3 capas 			
4 capas 			

Imagen original de: DataEngine (manual de producto), MIT GmbH

Evaluación del aprendizaje

- ▶ Para evaluar el aprendizaje, el conjunto inicial de datos se divide en dos (con diferentes posibles muestreos):
 - ▶ Conjunto de **entrenamiento** (*training*)
 - ▶ El que se usa para modelar (entrenar el modelo)
 - ▶ En torno al 80% de los datos totales
 - ▶ Conjunto de **evaluación** (*test*)
 - ▶ El que se usa para evaluar
 - ▶ El 20% de los datos restantes
- ▶ Una vez entrenado el modelo, se ejecuta sobre el conjunto de test, y se calcula el **error**:

$$\left. \begin{aligned} \text{Error}_{\text{absoluto}} &= |\text{Valor}_{\text{real}} - \text{Valor}_{\text{predicho}}| \\ \text{Error}_{\text{relativo}} &= \frac{|\text{Valor}_{\text{real}} - \text{Valor}_{\text{predicho}}|}{\text{Valor}_{\text{real}}} \end{aligned} \right\} \sum \text{ todos los ejemplos}$$

Validación cruzada

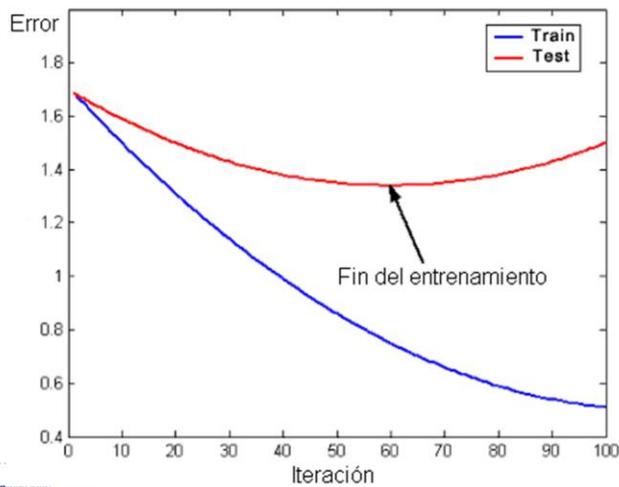
- ▶ El método anterior puede sufrir de error de muestreo (incorrecta selección de los conjuntos de entrenamiento y evaluación)
- ▶ Para resolverlo, se utiliza la **validación cruzada** (*cross validation*)
 - ▶ Se divide el conjunto inicial de datos en N partes
 - ▶ Se entrena/evalúa N veces, cada vez con un conjunto de test diferente
 - ▶ El error final es la media de las N medidas de error

Sobreentrenamiento (*overfitting*)

- ▶ Cuando el algoritmo de aprendizaje se ajusta tanto a los datos de entrada que **pierde su capacidad de generalizar**
 - ▶ El error de cálculo de los ejemplos futuros será alto
- ▶ Es necesario detener el proceso de entrenamiento en un determinado punto

Sobreentrenamiento (*overfitting*)

- ▶ El sobreentrenamiento se detecta si en una nueva iteración de aprendizaje **aumenta el error de test**



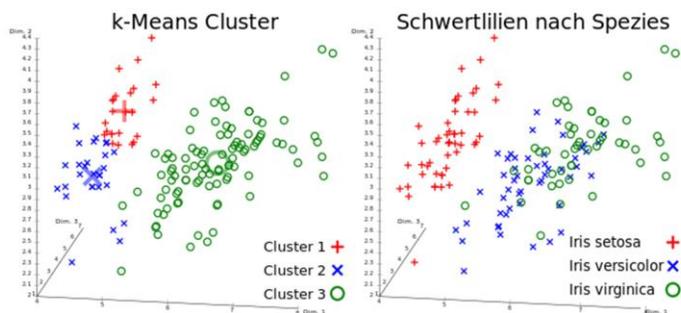
Aprendizaje inductivo no supervisado

- ▶ **Aprendizaje inductivo**
 - ▶ inducción no supervisada
 - ▶ agrupamiento (*clustering*)
 - ▶ análisis de dependencias
 - ▶ redes neuronales (*neural networks*)



Agrupamiento (*clustering*)

- ▶ Inducción no supervisada
- ▶ El objetivo del sistema es detectar agrupamientos o estructuras intrínsecas en el conjunto de datos, identificando grupos homogéneos de individuos parecidos



Conjunto de datos de flores de iris: clustering usando k means (izda.) vs. especies reales (derecha).
Las marcas más grandes (en el gráfico de la izda.) representan los centroides.
Visualización creada usando ELKI. Autor: Chire (<http://commons.wikimedia.org/wiki/User:Chire>)
Fuente: http://commons.wikimedia.org/wiki/File:Iris_Flowers_Clustering_kMeans_de.svg



Universidad
Carlos III de Madrid

IRC 2011/2012 - 43

Iris setosa

Iris versicolor

Iris virginica

Idea básica

- ▶ Definir cada individuo por sus propiedades
- ▶ Definir una medida de comparación entre individuos
 - ▶ Distancia Euclídea → algoritmos C-Medias (C-Means)

$$d(x^i, x^k) = \sqrt{\sum_j (x_j^i - x_j^k)^2}$$

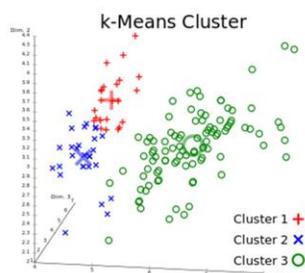
- ▶ Crear grupos, asociando a un mismo grupo a todos los individuos muy parecidos entre sí y muy diferentes a los demás

(algoritmos clásicos)

Técnicas

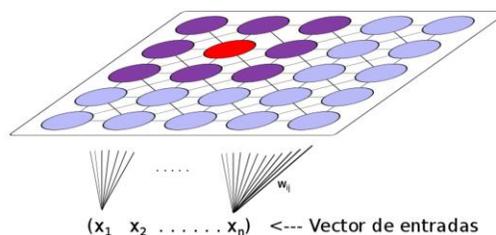
▶ Algoritmos clásicos

- ▶ Iterativo
- ▶ Aglomerativo
- ▶ Por división



▶ Algoritmos neuronales

- ▶ Mapas autoorganizativos de Kohonen



Iterativo

Inicialmente, asignación aleatoria de los elementos a los grupos

Iterativamente se van recalculando los centros de los grupos como punto medio de los objetos en ellos contenidos (algoritmo C-Means)

Aglomerativo

Inicialmente, tantos grupos como elementos

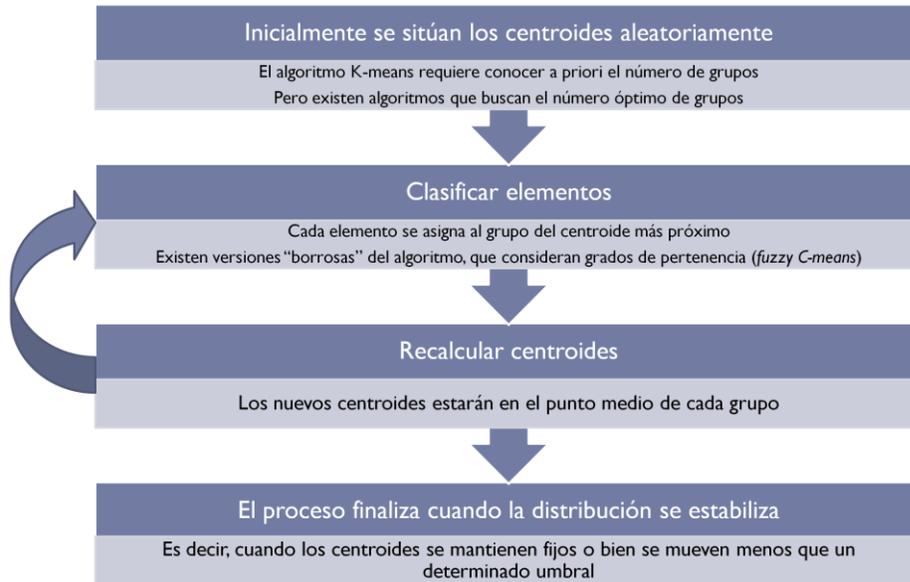
Sucesivamente se van uniendo los grupos cuyo grado de parecido supere un umbral

Por división

Inicialmente un único grupo con todos los elementos

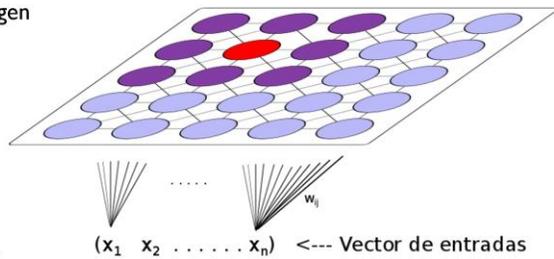
Sucesivamente se va dividiendo el grupo en grupos más pequeños en los que los elementos tienen un mayor grado de parecido

Clustering iterativo (K-means)



Mapas de Kohonen (mapas autoorganizativos)

- ▶ Cada vez que llega un nuevo elemento:
 - ▶ Se activa la neurona más próxima
 - ▶ La más parecida al elemento
 - ▶ Neurona roja en la imagen
 - ▶ Se ajustan los pesos
 - ▶ Para acercar la neurona al elemento de entrada
 - ▶ Se propaga el ajuste a las neuronas más cercanas a la activada
 - ▶ Por niveles
 - ▶ El efecto de ajuste va disminuyendo con la distancia
 - ▶ Neuronas moradas en la imagen



Análisis de dependencias

- ▶ **Objetivo:** Descubrir relaciones significativas entre los datos
- ▶ **Concepto de “transacción”**
(secuencia de sucesos u observaciones, relacionadas entre sí)
- ▶ Si interviene el tiempo, las secuencias se denominan **series temporales**



Técnicas

▶ **Análisis de asociaciones**

- ▶ Detecta elementos en una transacción que implican la presencia de otros elementos
- ▶ Reglas de asociación $A, B, C \dots \Rightarrow Z$
 - ▶ **sopORTE**: % de transacciones totales donde aparecen los elementos A, B, C...
 - ▶ **confianza**: % de las transacciones en las que aparecen A, B, C... y también aparece Z
- ▶ También llamado 'análisis de la cesta de la compra'
 - ▶ En este caso, el objetivo es detectar qué productos se compran juntos

▶ **Análisis de secuencias**

- ▶ Lo mismo, salvo que el orden de los sucesos es relevante en las transacciones

Aprendizaje deductivo

- ▶ Aprende de los conceptos conocidos, por ejemplo, de las reglas de juego y el objetivo
- ▶ El proceso es iterativo, añadiendo lo aprendido al conocimiento inicial para aprender nuevas cosas
- ▶ **Aprendizaje analítico**: utilizando conocimiento del dominio, razonar por qué un objeto o evento es una instancia positiva de un concepto
- ▶ **Aprendizaje basado en explicación (EBL)**: a partir de por qué un objeto satisface una definición funcional, formular una definición estructural

Ejemplo

- ▶ Quiero reconocer el objeto “copa”
 - ▶ Se parte de:
liquidoen(x) y estable(x) y puedobeberen(x) \rightarrow copa(x)
(definición funcional)
 - ▶ Con el resto de conocimiento, busco otra manera
(generalmente más directa) de reconocer el objeto
“copa”, llegando a:
vidrio(x) y cóncavo(x) y pequeño(x) \rightarrow copa(x)
(definición estructural)
- (ejemplo: diagnóstico de enfermedades mediante pruebas y no síntomas)

Aprendizaje de naturaleza mixta

- ▶ **Aprendizaje de naturaleza mixta**
 - ▶ Aprendizaje abductivo
 - ▶ Algoritmos genéticos
 - ▶ Causa-efecto
 - ▶ aprendizaje por analogía
 - ▶ razonamiento basado en casos

Aprendizaje abductivo

- ▶ Ni inductivo ni deductivo
- ▶ En este aprendizaje, el objetivo es completar los procesos originados por “teorías incompletas” mediante el establecimiento de **suposiciones válidas**
- ▶ Es decir, dada una conclusión conocida, proponer “hipótesis que la expliquen”
Datos CAUSA(A,C) y C
Inferir A
- ▶ Se centra en construir una **explicación** sobre un evento introduciendo suposiciones

Ejemplo de abducción

- ▶ **Ejemplo:**
para todo X , $\text{llueve_sobre}(x)$ implica $\text{mojado}(x)$
- ▶ **Por tanto,**
si algo está mojado, se puede hipotetizar que ha llovido

- ▶ **¡Contrario a la lógica clásica!**

Algoritmos genéticos

- ▶ Ni inductivo ni deductivo
- ▶ El conjunto de reglas se considera como una población de “pseudo-organismos”
- ▶ Estos organismos, al igual que los biológicos, pueden:
 - ▶ cruzarse unos con otros (operadores de cruce sexuales)
 - ▶ mutar (aleatoriamente) (operadores de mutación)
- ▶ ...con lo que producen nuevos individuos de la población
- ▶ y se produce la selección natural: sólo “sobreviven” los mejores individuos (operadores de selección)
- ▶ Hay una función de evaluación de la población, de tal forma que el proceso se “congela” si se alcanza el objetivo

Aprendizaje por analogía

- ▶ Ni inductivo ni deductivo
- ▶ Intenta emular la capacidad humana de aplicar el mismo razonamiento sobre problemas “parecidos” anteriores

- ▶ **Analogía transformacional**: se recupera de la memoria los problemas “parecidos”, la solución recordada se toma como “estado inicial” y se aplican transformaciones
- ▶ **Analogía derivacional**: no sólo se transfiere entre ambos problemas la solución, sino las decisiones adoptadas (alternativas desechadas, razones...)

Ejemplo de analogía

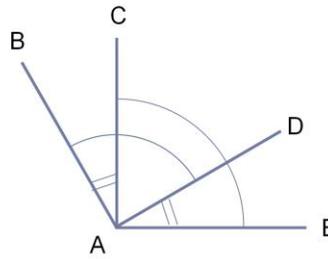
Solución previa



- ▶ $WX=YZ$
- ▶ $XY=XY$
- ▶ $WX+XY = XY+YZ$
- ▶ $WY=XZ$

Nueva solución

- ▶ $BAC = DAE$
- ▶ $CAD = CAD$
- ▶ $BAC+CAD = CAD+DAE$
- ▶ $BAD=CAE$



Razonamiento basado en casos (*case-based reasoning* - CBR)

- ▶ Es una clase particular de razonamiento por analogía
- ▶ Consiste en resolver problemas a partir de **problemas** “parecidos” anteriores (casos):
 - ▶ adaptando soluciones antiguas para resolver problemas nuevos, o
 - ▶ recuperando casos anteriores para iluminar aspectos de la situación actual
- ▶ Necesita:
 - ▶ métodos para extraer el conocimiento relevante de los casos
 - ▶ integrar el caso en la estructura del conocimiento
 - ▶ almacenar el problema para ser seleccionado en casos similares.

Y más...

- ▶ Linear/Polynomic regression
- ▶ Support Vector Machines (SVM)
- ▶ Radial-Base Functions (RBF)
- ▶ Naïve-Bayes
- ▶ Principal Component Analysis (PCA)
- ▶ Feature Selection
- ▶ ...

