

Tema AA–1: Introducción a al Aprendizaje Automático

José A. Alonso Jiménez
Miguel A. Gutiérrez Naranjo

Dpto. de Ciencias de la Computación e Inteligencia Artificial
UNIVERSIDAD DE SEVILLA

¿Qué es el Aprendizaje Automático?

- Any change in a system that allows it to perform better the second time on repetition of the same task or on another task drawn from the same population (*Simon, 1983*).
- El Aprendizaje Automático estudia cómo construir programas que mejoren automáticamente con la experiencia.
- ¿Por qué estudiar Aprendizaje Automático?
 - Recientes avances en la teoría y los algoritmos.
 - Crecimiento desbordante de datos “en línea” (on line).
 - Se dispone de máquinas suficientemente potentes.
 - Interés por parte de la industria.

Algunos ejemplos (I)

- Minería de datos (Data mining): Uso de datos históricos para mejora de decisiones:

Datos médicos → Decisiones médicas

- Conducción autónoma de vehículos: Aprenden a conducirse en autopistas en función de la información visual que reciben.
- Reconocimiento del habla.
- Juegos

Algunos ejemplos (II)

- Programas que se modifican a sí mismos según las costumbres del usuario:
 - Lector de periódicos que aprende los temas de interés del usuario.
 - Gestor de correo electrónico.
 - Casas que aprenden a optimizar el gasto de energía en función de las costumbres y hábitos de sus ocupantes.

Un problema de minería de datos

<i>Patient103</i> time=1	→	<i>Patient103</i> time=2	...	→	<i>Patient103</i> time=n
Age: 23		Age: 23			Age: 23
FirstPregnancy: no		FirstPregnancy: no			FirstPregnancy: no
Anemia: no		Anemia: no			Anemia: no
Diabetes: no		Diabetes: YES			Diabetes: no
PreviousPrematureBirth: no		PreviousPrematureBirth: no			PreviousPrematureBirth: no
Ultrasound: ?		Ultrasound: abnormal			Ultrasound: ?
Elective C-Section: ?		Elective C-Section: no			Elective C-Section: no
Emergency C-Section: ?		Emergency C-Section: ?			Emergency C-Section: Yes
...	

- **Dados:**

- Datos de 9714 pacientes, describiendo datos sobre embarazo y alumbramiento.
- Con 215 características de cada paciente.

- **Objetivo:**

- Aprender a predecir clases de pacientes con alto riesgo de cesárea.

Resultados del estudio

<i>Patient103</i> time=1	→	<i>Patient103</i> time=2	...	→	<i>Patient103</i> time=n
Age: 23		Age: 23			Age: 23
FirstPregnancy: no		FirstPregnancy: no			FirstPregnancy: no
Anemia: no		Anemia: no			Anemia: no
Diabetes: no		Diabetes: YES			Diabetes: no
PreviousPrematureBirth: no		PreviousPrematureBirth: no			PreviousPrematureBirth: no
Ultrasound: ?		Ultrasound: abnormal			Ultrasound: ?
Elective C-Section: ?		Elective C-Section: no			Elective C-Section: no
Emergency C-Section: ?		Emergency C-Section: ?			Emergency C-Section: Yes
...	

- Una de las 18 reglas aprendidas:

- Si la madre es primeriza, los ultrasonidos en el segundo trimestre dan resultados anormales, etc ...
- Entonces será necesaria cesárea

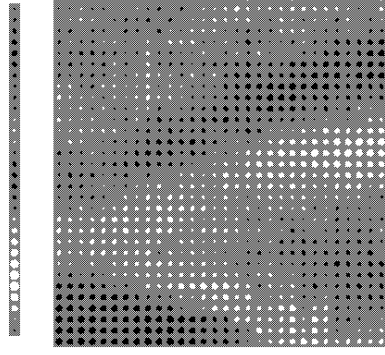
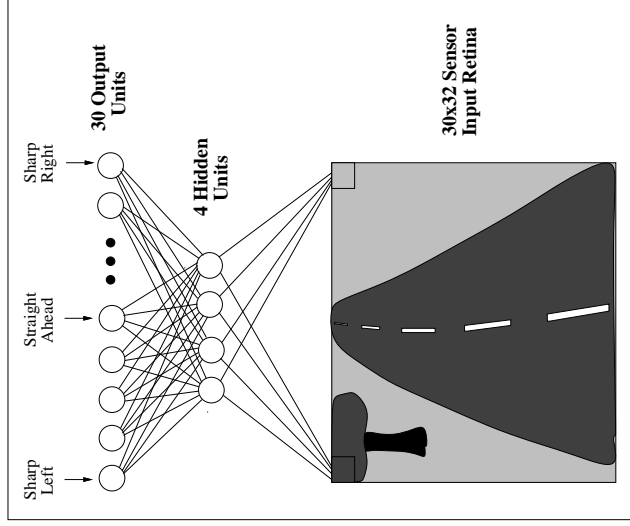
Análisis del riesgo en un crédito

<i>Customer103:</i> (time=t0)	<i>Customer103:</i> (time=t1)	...	<i>Customer103:</i> (time=tn)
Years of credit: 9	Years of credit: 9		Years of credit: 9
Loan balance: \$2,400	Loan balance: \$3,250		Loan balance: \$4,500
Income: \$52k	Income: ?		Income: ?
Own House: Yes	Own House: Yes		Own House: Yes
Other delinquent accts: 2	Other delinquent accts: 2		Other delinquent accts: 3
Max billing cycles late: 3	Max billing cycles late: 4		Max billing cycles late: 6
Profitable customer?: ?	Profitable customer?: ?		Profitable customer?: No
...

- Reglas aprendidas a partir de la síntesis de datos:
- Primera regla:
 - Si $\text{bancos-con-deudas} > 2$ y $\text{Reincidencias} > 1$, entonces $\text{Aceptar} = \text{No}$
- Segunda regla:
 - Si $\text{Bancos-con-deudas} = 0$ y $\text{Ingresos} > 150.000$, entonces $\text{Aceptar} = \text{Sí}$
- ...

Conducción automática

ALVINN [Pomerleau] conduce a 70 mph por autopistas



Hacia dónde vamos (I)

- Hoy: La punta del iceberg
 - Redes neuronales
 - Árboles de decisión
 - Programación lógica inductiva
 - Regresión
 - ...
- Aplicados a bases de datos bien estructuradas
- Interés por parte de la industria

Hacia dónde vamos (II)

- Oportunidades para el futuro:
- Aprendizaje a partir de datos de varias fuentes simultáneas: Bases de datos internas, web,...
- Aprendizaje por experimentación activa
- Aprendizaje de decisiones en lugar de predicciones
- Aprendizaje acumulativo de larga duración
- ¿Lenguajes de programación con aprendizaje incorporado?

Disciplinas relacionadas

- Inteligencia Artificial
- Métodos bayesianos
- Teoría de la complejidad
- Teoría de control
- Teoría de la información
- Filosofía
- Psicología y neurobiología
- Estadística
- ...

Problema de aprendizaje

- Aprendizaje = Mejora de alguna tarea mediante la experiencia
- Tarea (T): Lo que se debe aprender
- Experiencia (E): La que se tiene en relación a lo que se debe aprender
- Rendimiento (R): Medida de la calidad de lo aprendido
- Se dice que un sistema aprende de la experiencia, si el rendimiento R de la tarea T aprendida crece al crecer E .

Problema de aprendizaje

- ¿Qué experiencia?
- ¿Qué debe aprender?
- ¿Cómo representamos el conocimiento?
- ¿Qué algoritmo usaremos para aprenderlo?
- ¿Cómo se mide la mejora?

Ejemplos

- Una forma de aprender a jugar al ajedrez (de mejorar nuestro juego), es jugar contra nosotros mismos. Normalmente la forma de saber si hemos aprendido es jugar contra otros.
- Caracterización de este problema de aprendizaje
 - T : Jugar al ajedrez
 - E : Conjunto de partidas jugadas contra uno mismo
 - R : Porcentaje de partidas ganadas contra otro jugador

Ejemplos

- Una forma de que un sistema aprenda a reconocer palabras en un texto manuscrito, puede ser a partir de una base de datos con imágenes de palabras manuscritas y sus correspondientes transcripciones.
- La forma de saber si el sistema ha aprendido a reconocer palabras, será darle un texto manuscrito y ver cuantas transcripciones correctas hace.
- Caracterización de este problema de aprendizaje
 - *T*: Reconocer palabras manuscritas
 - *E*: Base de datos de palabras con sus transcripciones
 - *R*: Porcentaje de palabras reconocidas

Ejemplos

- Se puede enseñar a un vehículo a conducir automáticamente, guiándose por lo que ve (usando sensores de visión), y suministrándole una base de datos en la que se hayan registrado las imágenes tomadas mientras un conductor humano conducía el vehículo, junto con las correspondientes acciones que hizo.
- Caracterización de este problema de aprendizaje
 - *T*: Conducir un vehículo
 - *E*: Base de datos de imágenes, y las acciones correspondientes, registradas durante conducción por parte de un conductor humano
 - *R*: Distancia recorrida sin cometer ningún error

¿Qué estudiar en A. A.?

- ¿Qué algoritmos pueden aproximar funciones correctamente?
- ¿Cómo influye el número de ejemplos en la exactitud?
- ¿Cómo influye la complejidad de la representación de las hipótesis?
- ¿Cómo influye el ruido?
- ¿Cuáles son los límites teóricos del aprendizaje?
- ¿Cómo puede ayudar el conocimiento a priori?
- ¿Qué esquemas del aprendizaje biológico podemos adoptar?
- ¿Cómo pueden los sistemas alterar su propia representación?

Bibliografía

- MITCHELL, T. M. *Machine Learning* McGraw–Hill, 1997. Capítulo II.
- MARKOV, Z *Machine Learning Course* <http://www-it.fmi.uni-sofia.bg/markov/>
- **Lecturas recomendadas**
 - BRATKO I. *PROLOG Programming for Artificial Intelligence* Addison–Wesley, 1990. Capítulo XVIII