

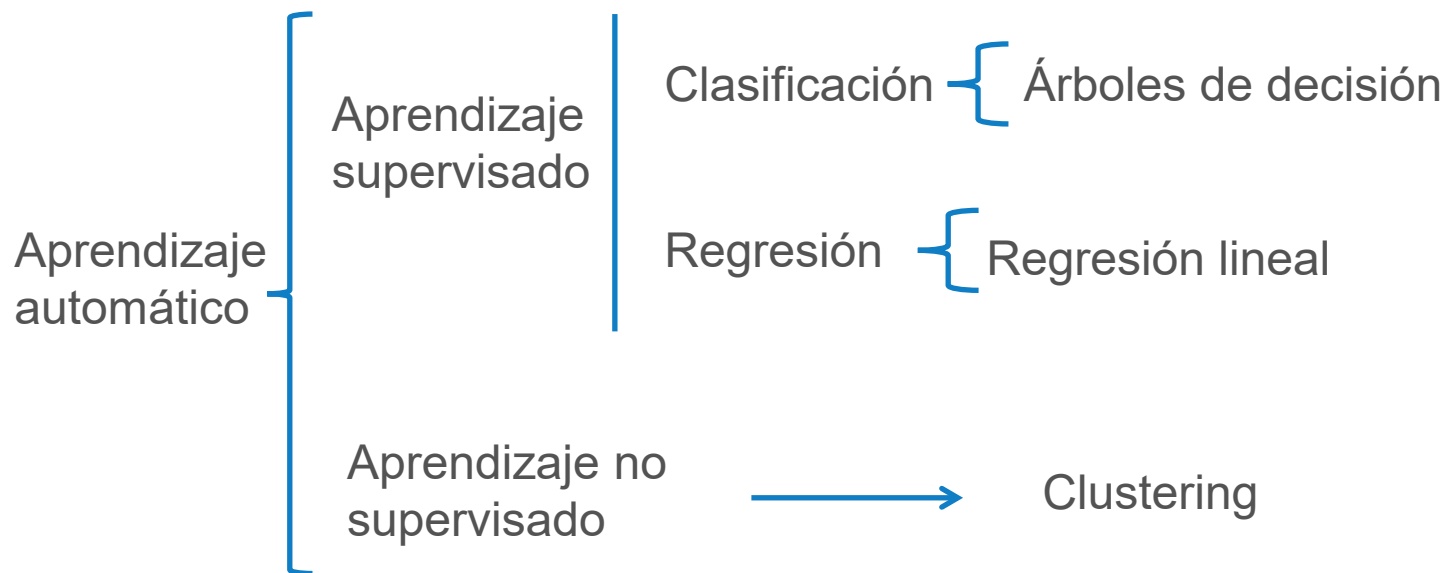
Inteligencia Artificial e Ingeniería del Conocimiento

Elena Verdú Pérez

Aprendizaje Supervisado

¿Cómo estudiar este tema?

SEMANAS	TEMAS	ACTIVIDADES (15.0 PUNTOS)	CLASES EN DIRECTO
Semana7 25-abr-2022 - 29-abr-2022	Tema 6. Aprendizaje supervisado 6.1. ¿Cómo estudiar este tema? 6.2. Introducción al Aprendizaje Automático 6.3. Clasificación		Clase del tema 6
Semana8 02-may-2022 - 06-may-2022	Tema 6. Aprendizaje supervisado (continuación) 6.4. Regresión 6.5. Validación de resultados	Actividad: Laboratorio Aprendizaje Supervisado con Weka (5.0 puntos) Fecha de entrega: 23/05/2022 Test - Tema 06 (0.1 puntos) Fecha de entrega: 03/07/2022	Clase del tema 6 y presentación del laboratorio (2h x 2 turnos)

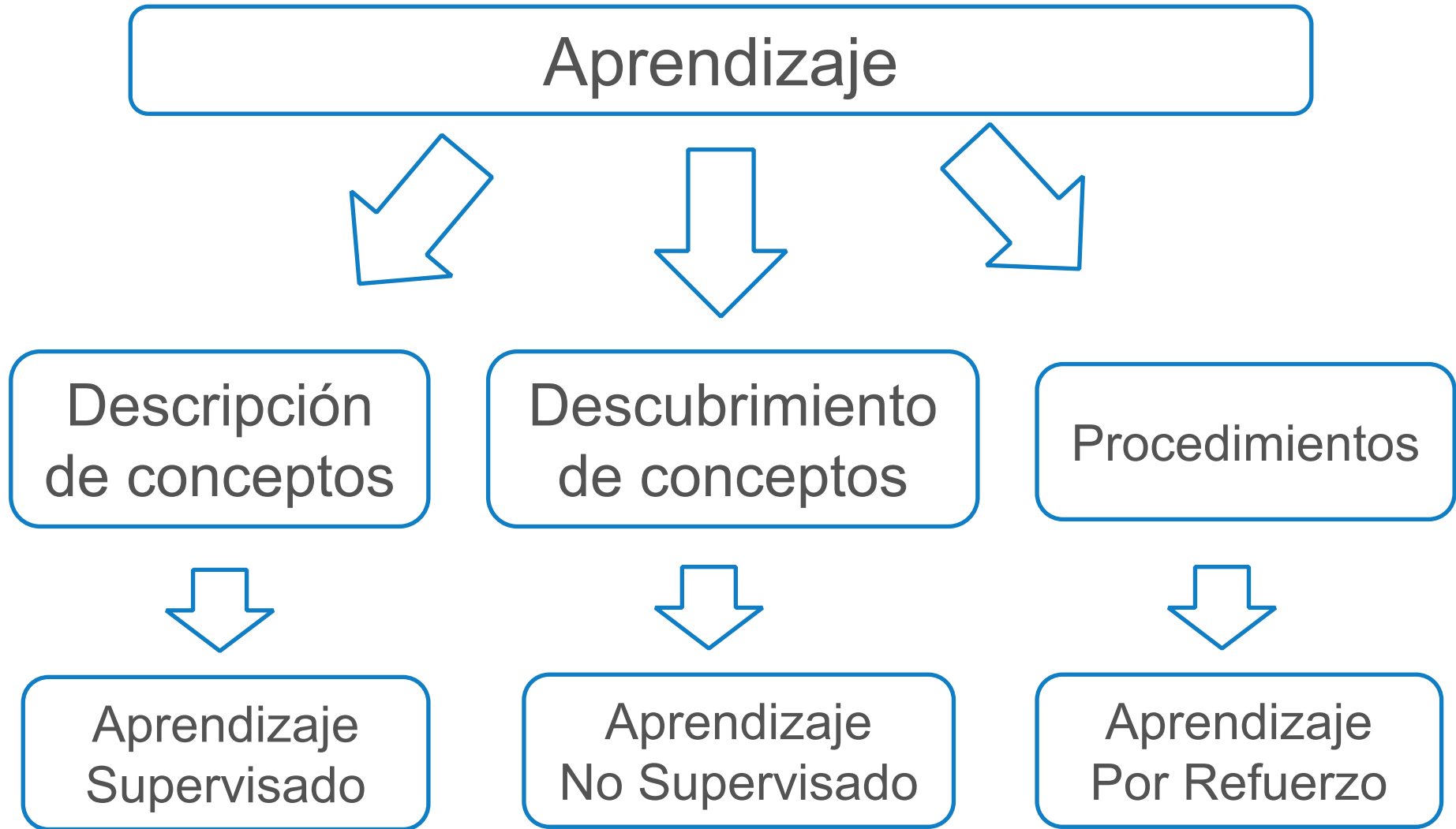


Aprendizaje automático

El aprendizaje automático es la disciplina cuyo foco es la construcción y el desarrollo de algoritmos que aprenden automáticamente a partir de un conjunto de datos.

Una característica del aprendizaje automático es que los algoritmos se diseñan para ser generales (no para un conjunto de datos concreto)

Aprendizaje automático



Aprendizaje supervisado

El aprendizaje supervisado se basa en **algoritmos** que deben encontrar una función que mapee una entrada a su salida esperada.

El algoritmo aprende a partir de una experiencia: un conjunto de ejemplos cuya salida es ya conocida. A este conjunto se le denomina **DATOS DE ENTRENAMIENTO**.



Aprendizaje supervisado

El **aprendizaje supervisado** pretende caracterizar o describir un concepto a partir de instancias o ejemplos del mismo. **La clase de las instancias es conocida.**

The diagram illustrates a supervised learning dataset. A table contains six rows of data. The first row has a blue header with the following attributes: Vivienda, Hijos, Tarjeta, Contrato, and Tipo. The subsequent five rows contain numerical and categorical values. Annotations include: 'Atributos' with arrows pointing to the first four columns; 'Clase' in a blue box with an arrow pointing to the 'Tipo' column; and 'Datos de entrenamiento' with a bracket on the left side of the table.

Vivienda	Hijos	Tarjeta	Contrato	Tipo
Hipoteca	0	Débito	Funcionario	Bueno
Hipoteca	0	Crédito	Asalariado	Malo
Hipoteca	2	Débito	Autónomo	Bueno
Pagada	2	Débito	Asalariado	Bueno
Hipoteca	1	Débito	Asalariado	Malo
Alquiler	2	Débito	Asalariado	Malo

Aprendizaje supervisado

- El modelado se basa en estos puntos/pasos:
 - Modelar los datos disponibles como conjuntos de entrada y salida.
 - Recolectar datos reales (crear conjunto de entrenamiento): datos representativos, evitar ruido.
 - Elección del algoritmo: en función de la estructura de los datos, características del problema.
 - Parametrizar el algoritmo.
 - Evaluar la precisión con datos diferentes a los usados en el entrenamiento (datos de prueba o datos de validación).

Aprendizaje supervisado

❑ El sobreajuste ocurre cuando la función aprendida no generaliza correctamente sino que se ajusta a los datos de entrenamiento únicamente.



Imagen de [pixel2013](#) en [Pixabay](#)

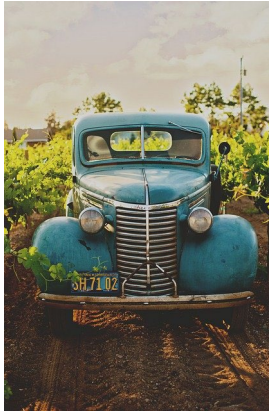


Imagen de [Free-Photos](#) en [Pixabay](#)



Imagen de [Pexels](#) en [Pixabay](#)



Imagen de [Pexels](#) en [Pixabay](#)

?



Imagen de [abdulla binmassam](#) en [Pixabay](#)

Aprendizaje supervisado

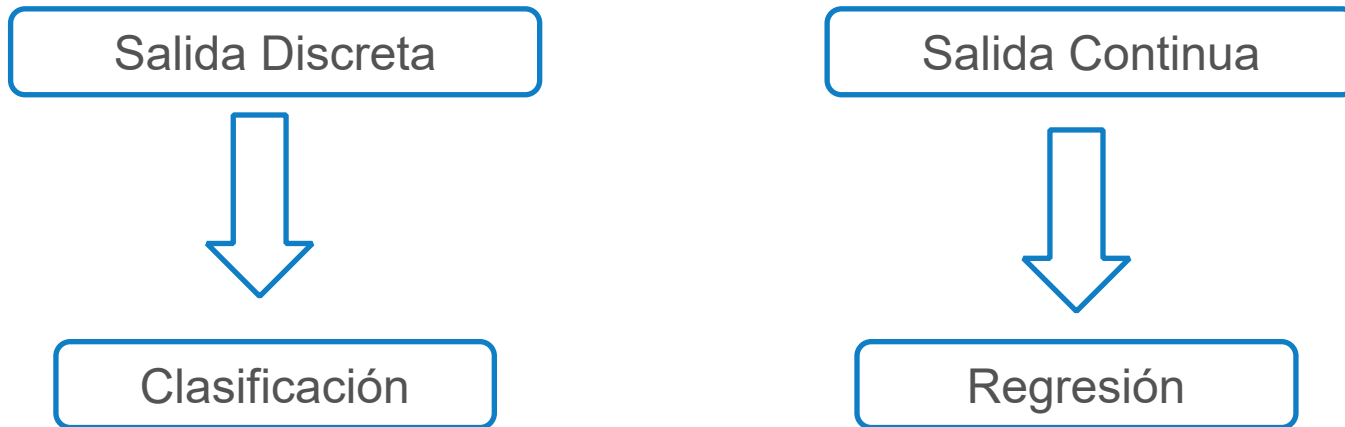
❑ Es importante considerar la legibilidad o **facilidad de interpretar la solución propuesta** por el algoritmo de aprendizaje.

❑ Ejemplo: *asesor de diagnóstico que dados ciertos valores como la temperatura, tensión arterial o hematocrito, determina la enfermedad del paciente tras aprender a partir de una base de datos con pacientes ya diagnosticados.*

Si la función aprendida es una combinación numérica compleja de los valores de entrada, el médico no podrá saber las razones por las que el sistema hace un diagnóstico.

*Si la función aprendida separa los valores/variables en formato regla (**si presión > 13 & pulso < 50 → infarto miocardio**), sí puede explicar el diagnóstico.*

Aprendizaje supervisado

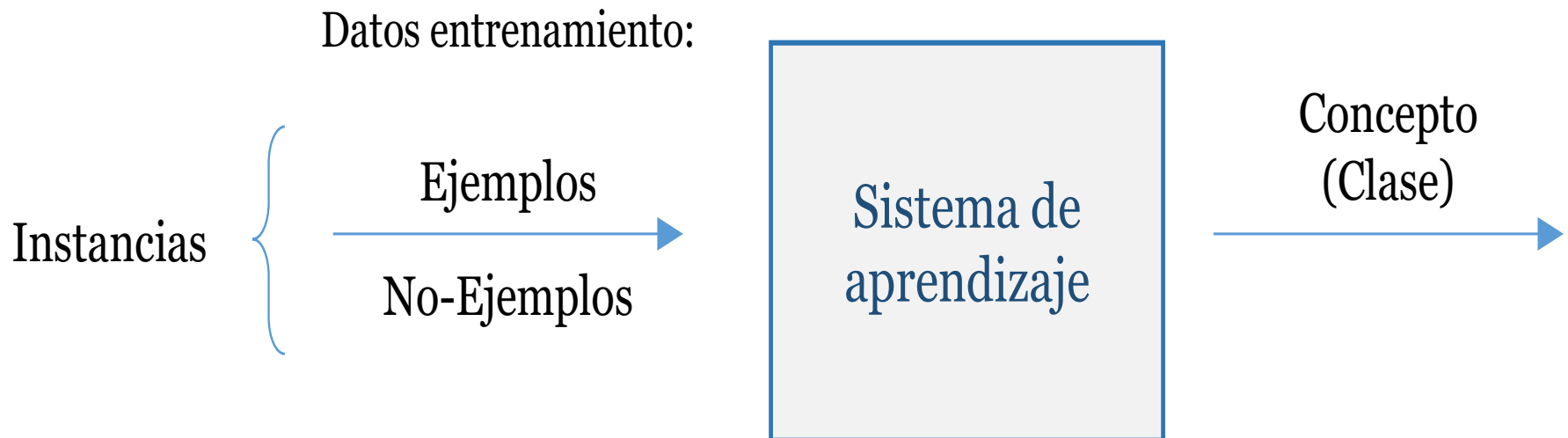


*Ej. Determinar si va a llover o no
(salida es “sí” o “no”)*

*Ej. Predecir la temperatura
máxima del día (cualquier
valor numérico)*

Clasificación

- En aprendizaje supervisado, cuando la salida es discreta, el problema se denomina **problema de clasificación**.
 - Cada posible valor de salida corresponde a una clase.
 - El objetivo es aprender un modelo que determine a qué clase pertenece una instancia futura.



Clasificación

Ejemplo: una entidad bancaria decide clasificar a sus clientes entre buenos y malos en función de retrasos en pagos en el pasado e índice de morosidad. Para facilitar a los agentes de banca el proceso de concesión de créditos, se quiere aprender qué factores influyen en si un cliente será bueno o malo.

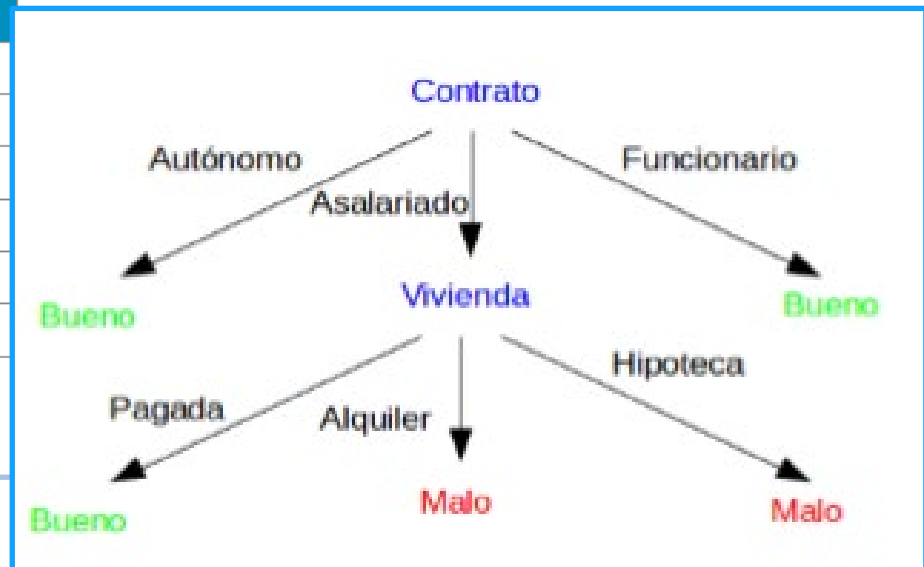
Vivienda	Hijos	Tarjeta	Contrato	Tipo
Hipoteca	0	Débito	Funcionario	Bueno
Hipoteca	0	Crédito	Asalariado	Malo
Hipoteca	2	Débito	Autónomo	Bueno
Pagada	2	Débito	Asalariado	Bueno
Hipoteca	1	Débito	Asalariado	Malo
Alquiler	2	Débito	Asalariado	Malo

Clase

Árboles de decisión

- ❑ Algoritmo de aprendizaje supervisado muy popular.
- ❑ Generan un **modelo clasificador en forma de árbol**:
 - Los nodos hoja se etiquetan con una de las posibles clases.
 - Los nodos internos corresponden a un atributo y cada rama descendiente corresponde a un valor del atributo.

Vivienda	Hijos	Tarjeta	Contrato	Tipo
Hipoteca	0	Débito	Funcionario	Bueno
Hipoteca	0	Crédito	Asalariado	Malo
Hipoteca	2	Débito	Autónomo	Bueno
Pagada	2	Débito	Asalariado	Bueno
Hipoteca	1	Débito	Asalariado	Malo
Alquiler	2	Débito	Asalariado	Malo



Azul: atributos.

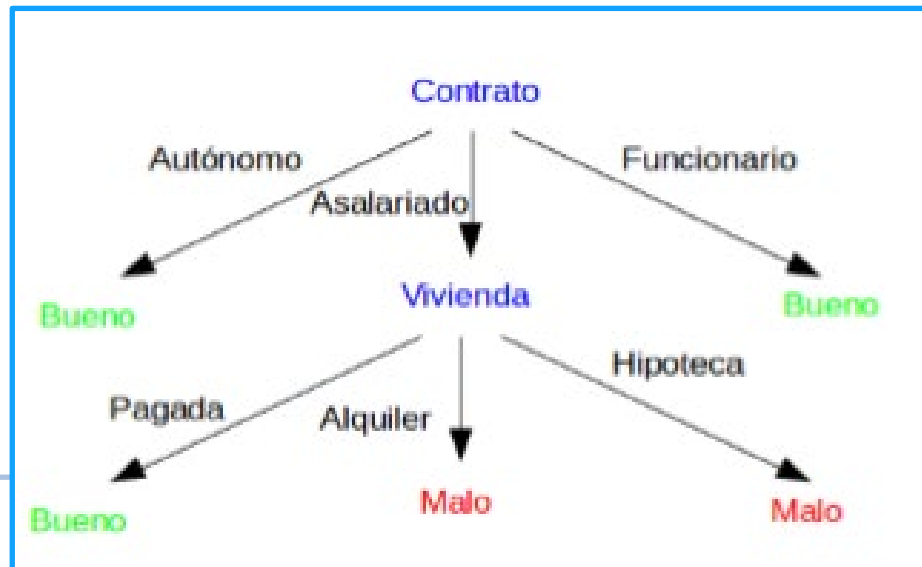
Clases: Verde (bueno), Rojo (malo).

Árboles de decisión

❑ **¿Cómo podemos clasificar una instancia futura?** descendiendo por los nodos siguiendo la rama del valor del atributo correspondiente en cada caso. Cuando se llegue a un nodo hoja, esa es la clase a la que el sistema estimará que la nueva instancia pertenece.

Si se presenta un nuevo cliente que es asalariado y tiene la vivienda pagada ¿qué tipo de cliente es?

Si se presenta un nuevo cliente que es asalariado y tiene una hipoteca ¿qué tipo de cliente es?



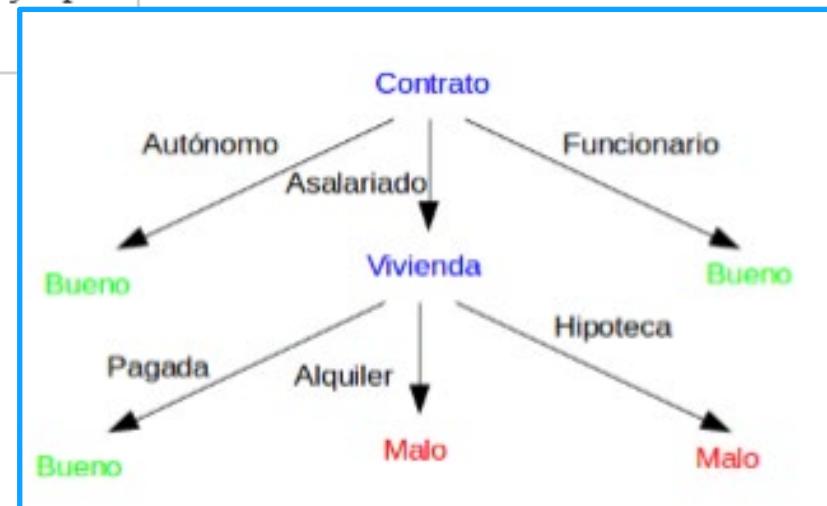
Azul: atributos.

Clases: Verde (bueno), Rojo (malo).

ID3: Algoritmo de Construcción de Árboles de decisión

-PASOS-

- 1 Seleccionar un atributo que maximice la separación de ejemplos.
- 2 Crear un nodo para ese atributo con tantas ramas como valores tenga.
- 3 Propagar los ejemplos de entrenamiento que se correspondan con el valor del atributo (por ejemplo, considerar en un subárbol sólo los asalariados).
- 4 Si todos los ejemplos pertenecen a la misma clase, entonces el nuevo nodo es un nodo hoja etiquetado con la clase correspondiente. Si no, se continúa el proceso eliminando el atributo usado anteriormente y usando los ejemplos propagados por esa rama.



ID3: Algoritmo de Construcción de Árboles de decisión

- ❑ Para seleccionar el atributo que se ha de comprobar en cada nodo **ID3 utiliza la medida de Ganancia de Información.**
- ❑ La Ganancia de Información utiliza la medida de **entropía** que mide la heterogeneidad de un conjunto de datos.

La siguiente fórmula define la entropía H:

$$H = - \sum_{x \in X} p(x) \log_2(p(x))$$

Donde:

- » X es el conjunto de clases del problema.
- » $p(x)$ es la proporción de ejemplos de entrenamiento que pertenecen a la clase x .

ID3: Algoritmo de Construcción de Árboles de decisión

- ❑ La entropía se usa para determinar la ganancia de información cuando se divide el conjunto de datos según un atributo
- ❑ **ID3 selecciona aquel atributo que conlleva mayor Ganancia de Información.**

La fórmula de ganancia de información IG para un atributo a es:

$$IG(a) = H - \sum_{t \in T} p(t)H(t)$$

Donde:

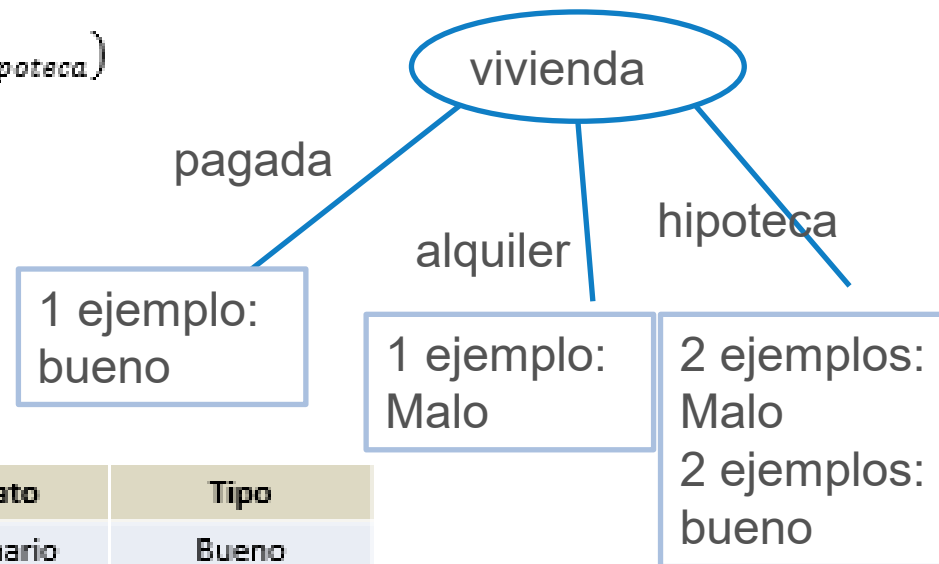
- » T son los subconjuntos de ejemplos de entrenamiento originados de discriminar el set original en base al atributo a .
- » $p(t)$ es la proporción de ejemplos de entrenamiento en t respecto al set original.
- » $H(t)$ es la entropía del subconjunto t .

ID3: Algoritmo de Construcción de Árboles de decisión

❑ Ejemplo: Calculamos entropía discriminando por Vivienda

$$IG(a) = H - \sum_{t \in T} p(t)H(t)$$

$$H(vivienda) = \frac{1}{6}H(V_{pagada}) + \frac{1}{6}H(V_{alquiler}) + \frac{4}{6}H(v_{hipoteca})$$



Vivienda	Hijos	Tarjeta	Contrato	Tipo
Hipoteca	0	Débito	Funcionario	Bueno
Hipoteca	0	Crédito	Asalariado	Malo
Hipoteca	2	Débito	Autónomo	Bueno
Pagada	2	Débito	Asalariado	Bueno
Hipoteca	1	Débito	Asalariado	Malo
Alquiler	2	Débito	Asalariado	Malo

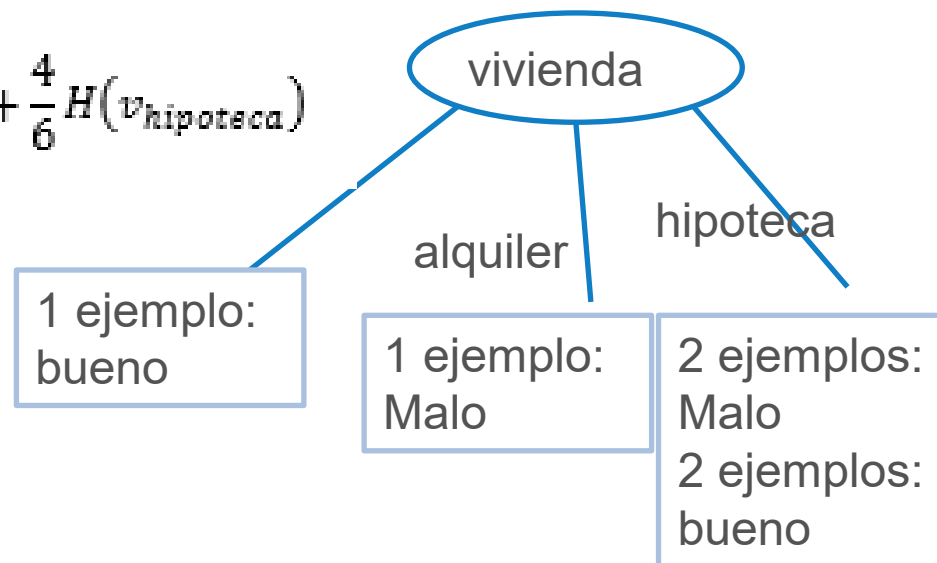
ID3: Algoritmo de Construcción de Árboles de decisión

$$IG(a) = H - \sum_{t \in T} p(t)H(t)$$

$$H(vivienda) = \frac{1}{6}H(V_{pagada}) + \frac{1}{6}H(V_{alquiler}) + \frac{4}{6}H(v_{hipoteca})$$

$$H = - \sum_{x \in X} p(x) \log_2(p(x))$$

$$H(V_{hipoteca}) = -\frac{2}{4} \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \log_2\left(\frac{2}{4}\right) = 1$$



❑ En el caso de $H(V_alquiler)$ y $H(V_pagada)$, todos los ejemplos pertenecen a la misma clase (en ambos casos porque hay un único ejemplo), por lo que, la entropía es 0.

$$H(vivienda) = \frac{1}{6}0 + \frac{1}{6}0 + \frac{4}{6}1 = 0.66$$

ID3: Algoritmo de Construcción de Árboles de decisión

La entropía resultante en los subconjuntos según los distintos atributos que se utilicen para discriminar sería:

❑ $H(\text{vivienda}) = 0.66$

❑ $H(\text{hijos}) = 0.79$

❑ $H(\text{tarjeta}) = 0.81$

❑ $H(\text{contrato}) = 0.54$

ID3 selecciona el atributo que aporta mayor Ganancia de Información ¿Cuál seleccionaría en este ejemplo?

Juego Competencia

Desafío 10: Relativo al tema 6, contesta utilizando como mucho 100 palabras: ¿qué diferencia fundamental existe entre los ejemplos con los que se construye un modelo de aprendizaje supervisado y los ejemplos con los que se construye uno de aprendizaje no supervisado? ¿qué objetivo se busca al construir un modelo de aprendizaje supervisado? ¿qué objetivo se busca al construir un modelo de aprendizaje no supervisado?

Los ejemplos con los que se construye un modelo de aprendizaje supervisado contienen la salida de los datos, mientras que los ejemplos con los que se construye un modelo de aprendizaje no supervisado no contienen la salida de los datos. El objetivo de un modelo de aprendizaje supervisado es encontrar una función que mapee nuevas entradas con su salida esperada. El objetivo de un modelo de aprendizaje no supervisado es encontrar la estructura que define los datos o la relación entre los datos.

Juego Competencia

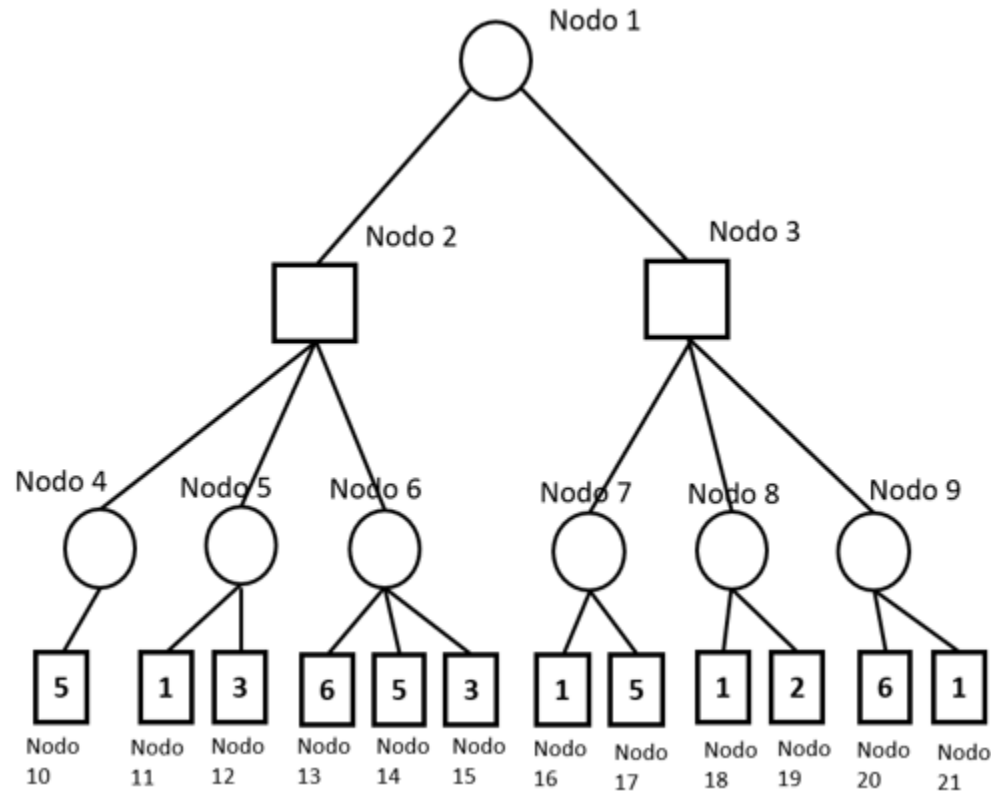
DESAFÍO 11 PROBLEMA MINIMAX Y PODA ALFA BETA

En el siguiente árbol (siguiente diapositiva) se representan todos los posibles estados que se pueden dar en un juego a partir del estado representado por el nodo raíz. El turno de juego inicial corresponde a MAX (el nodo raíz es MAX), y para que MAX gane el juego, la recompensa debe ser superior a 5. Los valores indicados en las hojas del árbol son las recompensas de MAX estimadas de acuerdo a la situación del juego en dicho estado.

1. Aplica el algoritmo Minimax. Indica para cada nodo cuál es el valor de la recompensa propagado. Indica la estrategia de juego más favorecedora para MAX. Además, contesta a la siguiente cuestión: ¿puede MAX ganar?
2. Indica qué subárboles serían podados utilizando la técnica de poda alfa-beta, así como los valores finales de alfa y beta para cada nodo. Supón que se explora el árbol de izquierda a derecha.

En las respuestas no tienes que incluir ninguna imagen, la respuesta se debe dar textualmente tal y como se expone en el ejercicio resuelto que está en el fichero “ejercicio_podaAlfabet4_comoResponderTextualmente.pdf” que se encuentra en la carpeta del Tema 5 en el apartado de Documentación.

Juego Competencia



Plazo: hasta el comienzo de la clase de la semana del 30 de mayo.

Formulario disponible en:

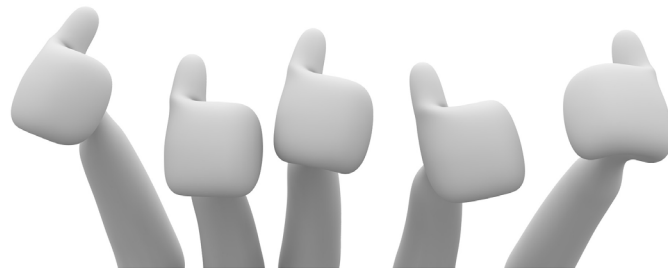
<https://forms.office.com/r/AYTTWwS4n0>

¿Dudas?



¡Muchas gracias por vuestra atención!

¡Feliz y provechosa semana!





www.unir.net