

L U M B R E



Árboles de decisión

Grado: 5

Materia: Aprendizaje Supervisado · Machine Learning

11/10/2026

Antes de empezar

Un árbol de decisión es la metáfora más humana del aprendizaje automático: preguntas de sí o no encadenadas hasta llegar a una conclusión. Divide y vencerás sobre el espacio de características, buscando en cada corte dejar los grupos lo más puros posible.

QUÉ APRENDERÁS

- Explicar cómo un árbol construye un clasificador dividiendo recursivamente los datos y midiendo la pureza de cada partición con gini o entropía.

Árboles de decisión

Un árbol es una cascada de preguntas

Un árbol no aprende una fórmula: aprende una serie de preguntas. Cada nodo es una condición sobre una característica, cada rama una respuesta, cada hoja una decisión. Esa estructura lo hace legible para un humano como pocos modelos.

VIDEO

Entiende los árboles de decisión en 3 minutos

Vídeo en español que resume cómo un árbol parte los datos con preguntas sucesivas hasta llegar a una decisión.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=hBssDI6LNKQ>

Dividir para reinar

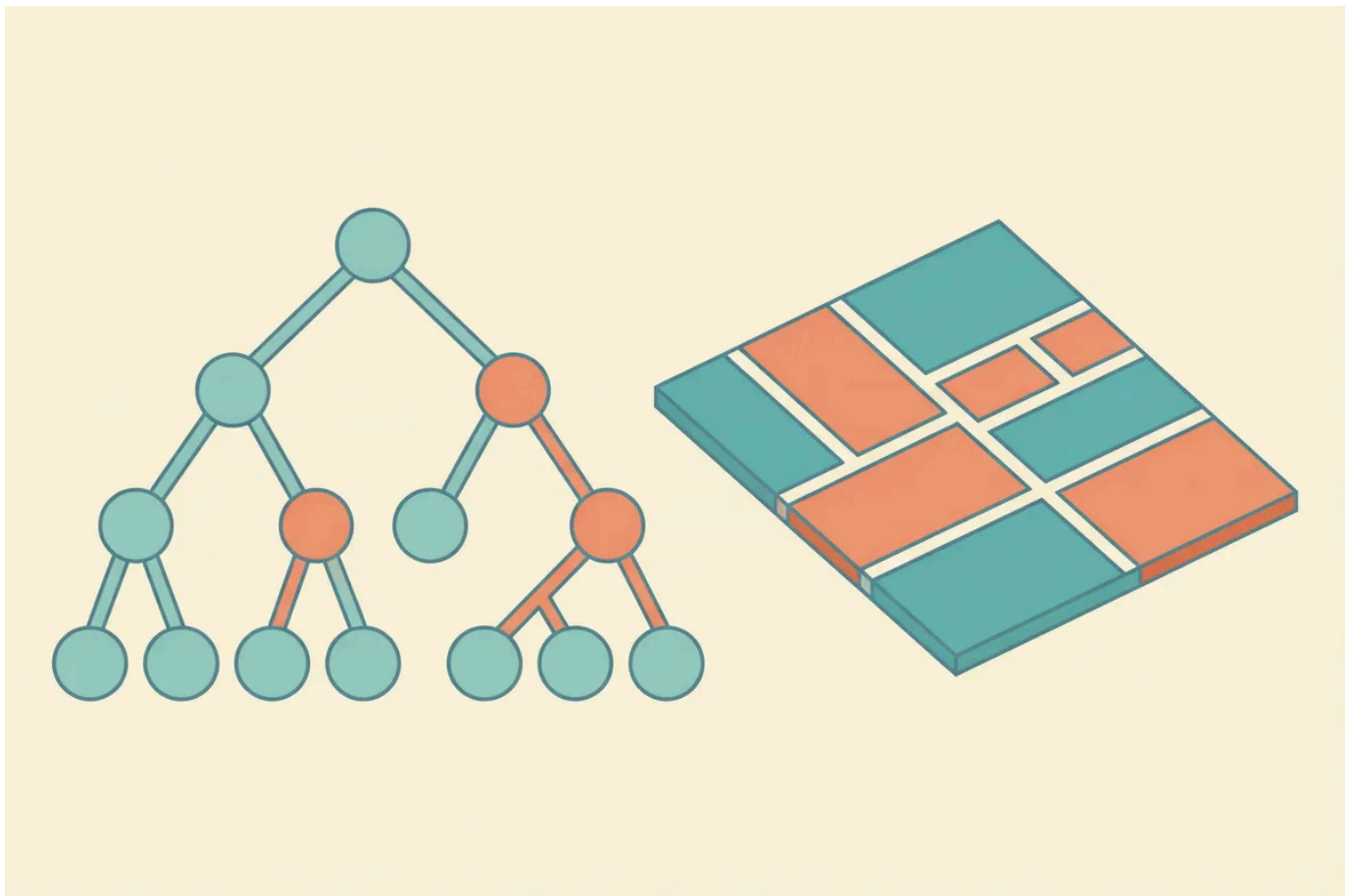
El árbol parte los datos con un corte sobre una característica —«¿edad mayor que treinta?»— y luego subdivide cada mitad con otro corte. Repite recursivamente hasta que los grupos son puros o se agota el presupuesto de profundidad. Cada división reduce el espacio a trozos más manejables.

La pureza como criterio

¿Qué corte elegir entre los infinitos posibles? El que deje los hijos más puros: con casos de una sola clase. La pureza se mide con funciones de impureza, y el algoritmo busca en cada nodo el corte que más la reduce.

Gini frente a entropía

El índice de gini castiga la mezcla de clases de forma cuadrática; la entropía, heredada de la teoría de la información, la mide en bits. En la práctica dan árboles casi idénticos: la entropía penaliza un poco más el desorden, pero la diferencia rara vez importa.



Las dos caras del mismo modelo: la cascada de preguntas y el espacio que dibuja, partido en rectángulos alineados con los ejes, uno por hoja.

La ganancia de información

Lo que guía el reparto es la reducción de impureza que logra un corte, ponderada por el tamaño de cada hijo. El árbol glotón elige en cada paso el corte de mayor ganancia, una estrategia voraz que no garantiza el óptimo global pero funciona sorprendentemente bien.

Cuándo parar de dividir

Si el árbol creciera sin freno, seguiría particionando hasta dejar un solo ejemplo por hoja: memorizaría el entrenamiento. Por eso hay criterios de parada: profundidad máxima, mínimo de muestras por hoja o ganancia mínima exigida.

Leer un árbol

De la raíz a una hoja se recorre una regla: «si edad mayor que treinta y ingresos menores que cuarenta, entonces sí». Cada camino es una conjunción de condiciones, y el conjunto de hojas, una región rectangular del espacio. Esa legibilidad es su gran activo.

El precio: sobreajuste

Los árboles profundos son de las máquinas de sobreajustar más eficaces que existen: recortan el ruido como si fuera señal. Un único árbol suele ser malo en generalización; su fortaleza real aparece cuando se combina con otros en bosques.

Inestabilidad

Cambiar unos pocos datos o el orden de las filas puede producir un árbol radicalmente distinto. Los cortes voraces de la raíz se propagan de forma caótica. Esa inestabilidad, no la falta de potencia, es lo que motivó el bagging.

Errores frecuentes

- Dejar el árbol crecer sin límite y celebrar el cero error en entrenamiento.
- Confiar en un único árbol para generalizar, ignorando su inestabilidad.
- Asumir que, por interpretable, el árbol refleja el mecanismo real del fenómeno.

Ejemplo resuelto: decidir quién cancela el servicio

Un operador quiere un árbol legible que explique las bajas. El proceso:

1. Elegir el criterio: gini por defecto, rápido y casi idéntico a entropía en resultados.
2. Dejar crecer: construir el árbol completo para ver qué cortes ganan más en cada nodo.
3. Podar: exigir mínimo de muestras por hoja y profundidad máxima para no memorizar.
4. Leer los caminos: traducir las ramas principales en reglas comprensibles para negocio.

5. Comparar la base: verificar que el árbol le gana a predecir siempre la clase mayoritaria.

¿Para qué sirve en la realidad?

Donde hay que explicar la decisión —denegaciones, triaje, segmentación—, el árbol es lenguaje directo: cada rama es una pregunta que un humano puede revisar y discutir. Y como pieza de los bosques, es el motor de casi todo modelo tabular moderno.

1. ¿Qué mide una función de impureza como gini o entropía en un nodo del árbol?

- ☐ La profundidad alcanzada por ese nodo
- ☐ El grado de mezcla de clases entre los ejemplos del nodo
- ☐ El número total de características
- ☐ La tasa de aprendizaje del reparto

2. Un árbol de decisión muy profundo tiende a sobreajustar los datos de entrenamiento.

- ☐ Verdadero
- ☐ Falso

No linealidad gratis

A diferencia de la logística, el árbol no exige que las relaciones sean rectas: los cortes paralelos a los ejes aproximan fronteras de cualquier forma. Paga por esa flexibilidad con fragmentación del espacio y bordes escalonados.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Árboles



4. Ordena el proceso de construcción de un árbol:

Elementos:

- evaluar el nodo actual y su impureza
- probar cortes posibles sobre las características
- quedarse con el de mayor ganancia de información
- repetir en los hijos o cerrar la región en una hoja

Nodo	. . .
Corte	. . .
Ganancia	. . .
Hoja	. . .

MLU Explain: árboles de decisión <https://mlu-explain.github.io/decision-tree/>
Explicación visual interactiva del reparto recursivo y la pureza.

5. Une cada término del árbol con su definición.

5.	castigo cuadrático de la mezcla de clases	-----
5.	desorden medido en bits	-----
5.	reducción de impureza que logra un corte	-----
5.	región final del espacio asociada a una decisión	-----

Opciones: gini · entropía · ganancia de información · hoja

Dividir y preguntar

Árboles de decisión

Construcción

cortes voraces

mayor ganancia

Pureza

gini

entropía

Frenos

profundidad

mínimo por hoja

poda

Riesgos

sobreajuste

inestabilidad

El mapa mental interactivo es interactivo en la versión online.



¿Qué es un árbol de decisión? <https://www.ibm.com/es-es/think/topics/decision-trees>
Explicación de referencia en español del método y sus aplicaciones.

6. Tu árbol clasifica perfecto el entrenamiento pero irregular el futuro. ¿Qué dos o tres parámetros de parada ajustarías primero y por qué?

Referencia curricular

Grado:	5
Materia:	Aprendizaje Supervisado · Machine Learning

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-supa-003-arboles-de-decision>

Material educativo generado desde Lumbre · lumbre.lat