

Árboles de decisión y sistemas de reglas

Latex

Oscar Painen Briones

2024



Universidad
de **Aysén**



Contenidos

1 Introducción

- ▶ **Introducción**
- ▶ Clasificación: Árboles de decisión
- ▶ Aprendizaje de Reglas por Cobertura
- ▶ Poda y Reestructuración
- ▶ Conclusión

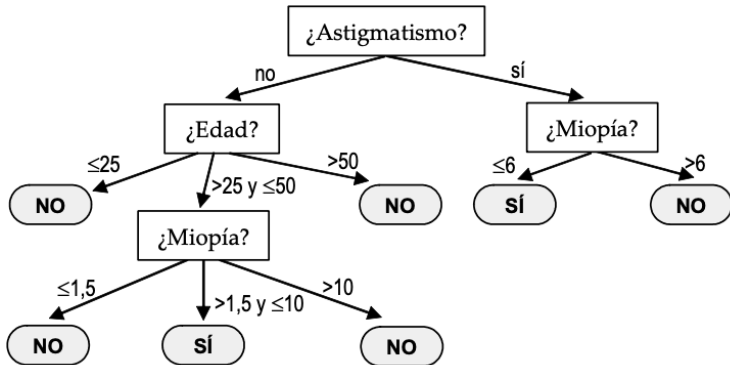
Introducción

1 Introducción

- **Árboles de decisión y sistemas de reglas.**
- Ambos son métodos fundamentales en la **minería de datos** utilizados en tareas como la clasificación y regresión.
- Los árboles de decisión son populares por su capacidad de manejar datos categóricos y numéricos, generando reglas claras y fácilmente interpretables.
- Los sistemas de reglas utilizan una estructura “Si-Entonces”, proporcionando mayor comprensibilidad.

Ejemplo de Árbol de Decisión

1 Introducción



Árbol de decisión para determinar una cirugía ocular.

Ejemplo de Árbol de Decisión

1 Introducción

¿Operación?

SI Astig.=No Y $25 < \text{Edad} \leq 50$ Y $1.5 < \text{Miopía} \leq 10$ ENTONCES SÍ

SI Astig.=Sí Y $\text{Miopía} \leq 6$ ENTONCES SÍ

EN OTRO CASO NO

Reglas correspondientes al árbol.



Contenidos

2 Clasificación: Árboles de decisión

► Introducción

► **Clasificación: Árboles de decisión**

► Aprendizaje de Reglas por Cobertura

► Poda y Reestructuración

► Conclusión

Árboles de decisión

2 Clasificación: Árboles de decisión

Clasificación con Árboles de Decisión

- Los árboles de decisión funcionan dividiendo el espacio de datos en subconjuntos más pequeños.
- Cada nodo representa una decisión basada en un **atributo**, mientras que las hojas contienen las **clases** predichas.
- Algoritmos como **ID3**, **C4.5** y **CART** son utilizados para construir los árboles según la partición.
- Se utilizan criterios como la **entropía** y la **ganancia de información** para seleccionar los atributos.

Diagrama del Proceso de Clasificación

2 Clasificación: Árboles de decisión

ALGORITMO Partición(N :nodo, E :conjunto de ejemplos)

SI todos los ejemplos E son de la misma clase c **ENTONCES**

Asignar la clase c al nodo N .

SALIR; // Esta rama es pura, ya no hay que seguir partiendo. N es hoja.

SI NO:

Particiones := generar posibles particiones.

MejorPartición:= seleccionar la mejor partición según el criterio de partición.

PARA CADA condición i de la partición elegida.

Añadir un nodo hijo i a N y asignar los ejemplos consistentes a cada hijo (E_i).

Partición(i , E_i). // Realizar el mismo procedimiento global con cada hijo.

FIN PARA

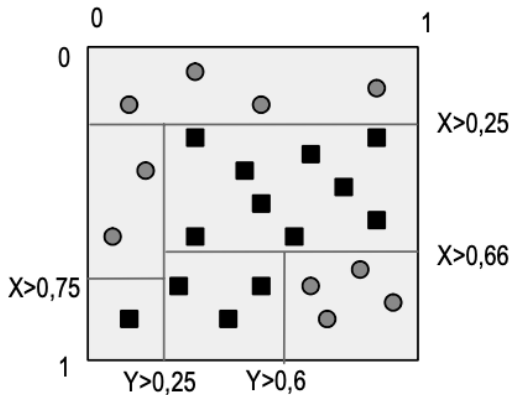
FIN SI

FIN ALGORITMO

Para clasificar un conjunto de ejemplos E , se invoca con la llamada Partición(R, E), donde R es un nodo raíz de un árbol por empezar.

Particiones posibles

2 Clasificación: Árboles de decisión



Selección de particiones

2 Clasificación: Árboles de decisión

Criterio	$f(p^1, p^2, \dots, p^c)$
Error Esperado	$\min(p^1, p^2, \dots, p^c)$
GINI (CART)	$1 - \sum (p_i)^2$
Entropía (gain)	$\sum p_i \cdot \log(p_i)$
DKM	$2(\prod p_i)^{1/2}$



Contenidos

3 Aprendizaje de Reglas por Cobertura

- ▶ Introducción
- ▶ Clasificación: Árboles de decisión
- ▶ **Aprendizaje de Reglas por Cobertura**
- ▶ Poda y Reestructuración
- ▶ Conclusión

Aprendizaje de Reglas por Cobertura

3 Aprendizaje de Reglas por Cobertura

- En lugar de construir un árbol, se genera un **conjunto de reglas** para cubrir el espacio de datos.
- Las reglas siguen la estructura “**Si-Entonces**” para relacionar los atributos con las clases.
- El proceso es secuencial, cubriendo porciones del conjunto de datos en cada paso.
- Las reglas resultantes son más fáciles de interpretar comparadas con otros algoritmos como redes neuronales.

Ejemplo de Reglas “Si-Entonces”

3 Aprendizaje de Reglas por Cobertura

¿Operación?

1. **SI** Astig.=Sí **Y** Miopía>6 **ENTONCES** NO
2. **SI** $25 < \text{Edad} \leq 50$ **Y** $\text{Miopía} \leq 6$ **ENTONCES** SÍ
3. **SI** Edad>50 **ENTONCES** NO
4. **SI** Edad ≤ 25 **ENTONCES** NO
5. **SI** Miopía>10 **ENTONCES** NO
6. **EN OTRO CASO** Operación=Sí



Contenidos

4 Poda y Reestructuración

- ▶ Introducción
- ▶ Clasificación: Árboles de decisión
- ▶ Aprendizaje de Reglas por Cobertura
- ▶ **Poda y Reestructuración**
- ▶ Conclusión

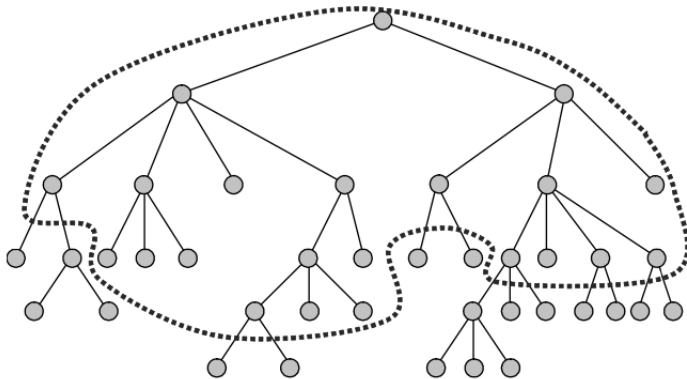
Poda en Árboles de Decisión

4 Poda y Reestructuración

- La **poda** es una técnica utilizada para evitar el sobreajuste, eliminando ramas innecesarias.
- Se emplean métodos como la **poda de complejidad de costo** y la **poda posterior al árbol**.
- La poda mejora la generalización y rendimiento en datos nuevos.
- La **reestructuración** del árbol busca mejorar su rendimiento, ajustando nodos y combinando reglas.

Proceso de Poda

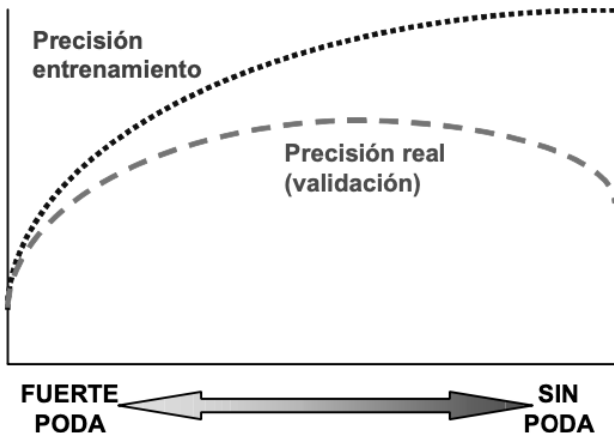
4 Poda y Reestructuración



Ejemplo de poda.

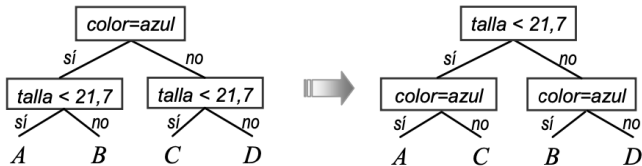
Nivel de Poda Óptimo

4 Poda y Reestructuración



Otros operadores

4 Poda y Reestructuración



Ejemplo de poda.



Contenidos

5 Conclusión

- ▶ Introducción
- ▶ Clasificación: Árboles de decisión
- ▶ Aprendizaje de Reglas por Cobertura
- ▶ Poda y Reestructuración
- ▶ Conclusión

Conclusión

5 Conclusión

- Los árboles de decisión y los sistemas de reglas son poderosas herramientas para la clasificación y análisis de datos.
- Ambos enfoques ofrecen **interpretabilidad** y son eficientes en la minería de datos.
- Las técnicas de **poda** y **reestructuración** son esenciales para mejorar el rendimiento y evitar el sobreajuste.

Árboles de decisión y sistemas de reglas

Thank you for listening!