

Sesión 6: Árboles de Decisión (CART), Poda, ROC y AUC

Oscar Leonardo Rincón León

Objetivo

Aplicar árboles de decisión e interpretar reglas. Evaluar rendimiento con métricas visuales como la curva ROC y el AUC.

¿Qué es un árbol de decisión?

- Modelo predictivo basado en reglas jerárquicas.
- Representa decisiones con condiciones anidadas.
- CART: algoritmo clásico que construye árboles binarios.

Características de CART

- Árboles binarios: cada nodo tiene dos ramas.
- Clasificación: impureza de Gini o entropía.
- Regresión: minimiza el error cuadrático medio (MSE).
- Salida: reglas tipo “si... entonces...”.

- Nodo raíz: nodo inicial con todos los datos.
- Nodo interno: evalúa una condición de división.
- Hoja: nodo terminal con una clase o valor.
- Profundidad: número de niveles del árbol.

Sobreajuste y profundidad

- Árboles profundos capturan ruido del entrenamiento.
- Pérdida de generalización: sobreajuste.
- Árboles poco profundos: subajuste.
- Balance necesario mediante validación.

¿Qué es la poda?

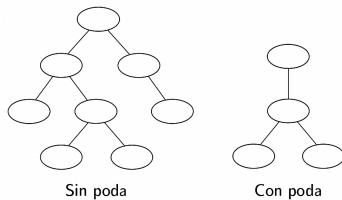
- Reduce complejidad del árbol para evitar sobreajuste.
- Mejora la generalización.

Tipos de poda:

- Poda previa: limitar profundidad o tamaño mínimo.
- Poda posterior: eliminar nodos irrelevantes tras el entrenamiento.

Ejemplo visual de poda

Ejemplo visual de poda



- Matriz de confusión.
- Precisión, Recall, F1.
- Curva ROC y AUC.

- ROC: compara sensibilidad (TPR) vs tasa de falsos positivos (FPR).
- Muestra el rendimiento según distintos umbrales.
- Ideal: curva cercana al vértice superior izquierdo.

AUC: Área bajo la curva

- Mide capacidad de discriminación.
- $AUC = 1$: modelo perfecto. $AUC = 0.5$: aleatorio.
- Permite comparar modelos sin fijar umbral.

- Clasificación de hogares vulnerables.
- Segmentación por riesgo financiero.
- Justificación de reglas ante actores sociales.

- Entrenar árbol con y sin poda.
- Visualizar árbol y extraer reglas.
- Comparar desempeño con matriz y AUC.

- ¿Qué variables fueron más relevantes?
- Interpretabilidad vs precisión.
- ¿Es útil un árbol para justificar decisiones?