

Árboles de Decisión - CART

CART (Classification and Regression Trees)

Definición

CART es un **algoritmo de aprendizaje supervisado** utilizado para construir **árboles de decisión** en tareas de **clasificación** y **regresión**.

Características principales

- ✓ **División recursiva:** Divide el conjunto de datos en subconjuntos más pequeños utilizando una métrica de pureza como el **Gini** (para clasificación) o el **Error Cuadrático Medio (MSE)** (para regresión).
- ✓ **Estructura tipo árbol binario:** Cada nodo interno representa una decisión basada en una característica, y cada hoja representa un resultado.
- ✓ **Podado del árbol:** CART permite **pruning** (poda) para evitar el sobreajuste y mejorar la generalización.
- ✓ **Compatibilidad con variables numéricas y categóricas:** Puede trabajar con ambos tipos de datos.
- ✓ **Método determinista:** Para los mismos datos y parámetros, genera el mismo árbol.

CART (Classification and Regression Trees)

Ventajas

- ✓ Fácil de interpretar y visualizar.
- ✓ No requiere preprocesamiento exhaustivo.
- ✓ Puede manejar datos con ruido y valores faltantes.

Desventajas

- ✓ Sensible a cambios en los datos (alta varianza).
- ✓ Puede sobreajustar si el árbol es muy profundo.
- ✓ No es el mejor en precisión comparado con métodos como Random Forest o Boosting.