Faculty of Engineering
Cairo University

Machine Learning

Classification and Regression Trees

By:

Shehap Elhadary
Abdelrahman hesham
Omar Walid

Date: 25/11/2025

Table of Contents

1. Introduction

CART (Classification and Regression Trees) is a versatile and interpretable supervised learning algorithm introduced by Breiman et al. in 1984. It can handle both classification and regression tasks, offering transparent decision-making via a flowchart-like tree structure. CART systematically partitions data into homogeneous regions, revealing patterns and feature interactions while producing predictions.

2. Core Concept: Binary Recursive Partitioning

The core of CART is binary recursive partitioning. The algorithm splits the dataset into two groups at each node based on the feature that maximizes improvement in purity or reduction in error. This "greedy" approach continues recursively until a stopping criterion is met, forming a tree structure.

Key characteristics:

- **Binary splits**: Every split divides data into exactly two groups.
- **Greedy approach**: Chooses the locally best split at each step.
- **Leaf predictions**: Classification leaves predict the majority class, regression leaves predict the mean of target values

3. Splitting Criteria

For Classification: Minimizing Impurity:

CART measures node impurity to guide splits.

Gini Impurity (default):

$$\text{Gini} = 1 - \sum_i p_i^2$$

- ***Pi*** = proportion of class ***i*** in the node
- 0 = pure node, maximum impurity depends on class distribution

Entropy (alternative):

$$\text{Entropy} = -\sum_i p_i \log_2 p_i$$

- Measures uncertainty; also 0 for pure nodes, 1 for 50/50 binary split

Information Gain:

$$\text{Gain} = \text{Impurity}(\text{parent}) - \frac{N_{\text{left}}}{N_{\text{parent}}}\text{Impurity}(\text{left}) - \frac{N_{\text{right}}}{N_{\text{parent}}}\text{Impurity}(\text{right})$$

The split with the highest gain is selected.

For Regression: Minimizing Variance

For continuous targets, CART minimizes **variance** (or MSE):

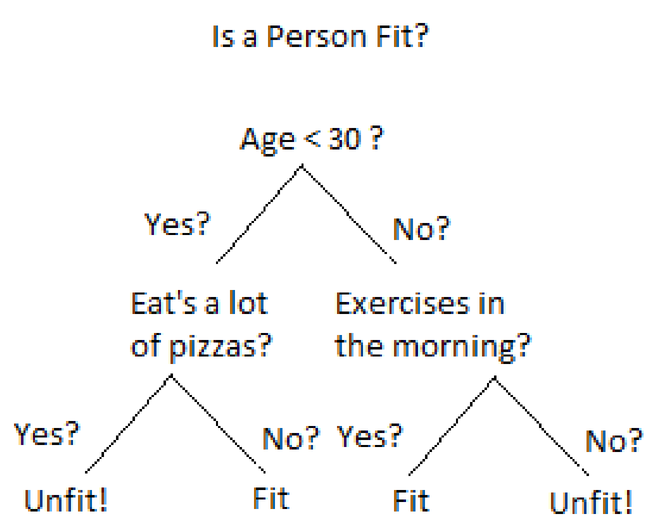
$$\text{Variance Reduction} = \text{Variance(parent)} - \frac{N_{\text{left}}}{N_{\text{parent}}}\,\text{Variance(left)} - \frac{N_{\text{right}}}{N_{\text{parent}}}\,\text{Variance(right)}$$

The goal is to produce child nodes where values are closely clustered around their mean.

4. Building a Tree: Example

Given a small dataset, CART proceeds as follows:

1. **Calculate root impurity** (Gini or variance)
2. **Evaluate all possible splits** for each feature
3. **Select the split with maximum gain**
4. **Repeat recursively** on child nodes until stopping conditions are met



5. Controlling Complexity: Preventing Overfitting

Pre-pruning (early stopping)

- `max_depth`: limits tree levels
- `min_samples_split` / `min_samples_leaf`: minimum samples to split or remain in a leaf
- `min_impurity_decrease`: minimum gain required to split

Post-pruning (Cost-Complexity Pruning)

Grow full tree, then prune weak branches using a complexity parameter α\alphaα:

$$R_\alpha(T) = R(T) + \alpha\,|T|$$

Choose optimal subtree via cross-validation to balance bias and varian

6. Strengths and Limitations

Advantages:

- Highly interpretable ("white-box")
- Handles numerical and categorical data
- Non-parametric, minimal preprocessing required

Disadvantages:

- High variance, sensitive to small changes in data
- Greedy splits may miss global optimum
- Axis-parallel boundaries limit capturing diagonal/non-linear relationships
- Prone to overfitting if unpruned

7. Conclusion

CART remains a cornerstone of machine learning. Its simplicity, interpretability, and recursive splitting logic form the foundation for both standalone models and advanced ensembles. Mastery of CART principles: splits, impurity measures, and pruning is essential for building robust predictive models and understanding modern machine learning algorithms.