🌳 2. DECISION TREE

---

1 Problem Formulation

Input space:

$$\mathcal{X} \subseteq \mathbb{R}^d$$

Output space:

- Classification: finite labels
- Regression: R\mathbb{R}R

Learning objective:
Minimize expected impurity.

---

2 Model Specification

Hypothesis Function

Piecewise constant function:

$$f(x) = c_m \quad \text{if } x \in R_m$$

Parameter Space

- Split feature index
- Split threshold
- Tree structure

Structural Assumptions

- Axis-aligned splits
- Recursive partitioning

---

3 Loss Function

Classification

Gini impurity:

$$G = 1 - \sum_k p_k^2$$

Entropy:

$$H = -\sum_k p_k \log p_k$$

Regression

$$\ell = (y - \hat{y})^2$$

---

4 Objective Function

Minimize impurity after split:

$$\min_{j,t} \left[\frac{N_L}{N} I_L + \frac{N_R}{N} I_R \right]$$

---

5 Optimization Method

- Greedy recursive splitting
- No global optimization
- Complexity: $O(nd\log n)$

---

6 Statistical Interpretation

- Non-parametric estimator
- No explicit probability model
- Class probability = proportion in leaf

---

7 Regularization & Generalization

Controlled by:

- Max depth
- Min samples per leaf
- Pruning

High variance model.

---

8 Theoretical Properties

- Non-convex
- No global optimality guarantee
- Consistent under infinite data

---

9 Computational Complexity

Aspect	Complexity
Training	$O(nd \log n)$
Inference	$O(\text{depth})$
Memory	O(nodes)

🔟 Limitations

- Overfitting
- Unstable to small changes
- Biased toward features with many splits