🌲 3. RANDOM FOREST

---

1 Problem Formulation

Same as decision tree but ensemble-based.

Learning objective:

$$f(x) = \frac{1}{T}\sum_{t=1}^{T} f_t(x)$$

---

2 Model Specification

Ensemble of trees:

$$f(x) = \text{majority vote (classification)}$$

$$f(x) = \frac{1}{T}\sum f_t(x) \quad \text{(regression)}$$

---

3 Loss Function

Implicitly minimizes average tree impurity.

---

4 Objective Function

$$\min \frac{1}{T} \sum_{t=1}^{T} R_{emp}(f_t)$$

With:

- Bootstrap sampling
- Feature subsampling

---

5 Optimization Method

- Parallel tree training
- Greedy per tree

Complexity:

$$O(T \cdot nd \log n)$$

---

6 Statistical Interpretation

- Bagging estimator
- Reduces variance
- Approximate Bayesian model averaging

---

7 Regularization & Generalization

Controlled by:

- Number of trees TTT
- Max depth

- Feature subset size

Reduces variance without increasing bias much.

---

8 Theoretical Properties

- Law of Large Numbers → variance reduction
- Consistent under certain assumptions
- Non-convex

---

9 Computational Complexity

Aspect	Complexity
Training	$O(Tnd \log n)$
Inference	$O(T \cdot depth)$
Memory	$O(T \cdot nodes)$

---

10 Limitations

- Less interpretable
- Larger memory usage
- Slower inference

🌲 Random Forest Visualization

```
Data → Tree1 → \
Data → Tree2 →  > Average / Vote → Output
Data → Tree3 → /
```
