DECISION TREE – Theoretical Questions

1. Why are Decision Trees Non-Parametric Models?

A parametric model assumes a fixed functional form with a fixed number of parameters (e.g., linear regression).

Decision trees:

- Do not assume a fixed functional form.
- The number of parameters (nodes, splits) grows with data.
- Model complexity adapts to dataset size.

Thus, trees are non-parametric because their capacity increases with data rather than being fixed.

2. Compare Entropy vs Gini Mathematically

Let class probabilities be $p_1, p_2, \dots, p_K$.

Entropy:

$$H = -\sum_{k=1}^{K} p_k \log\, p_k$$

Gini Impurity:

$$G = 1 - \sum_{k=1}^{K} p_k^2$$

Comparison:

- Both measure node impurity.
- Both minimized when node is pure.
- Entropy grows logarithmically.
- Gini is quadratic.

For binary case (p, 1-p):

$$H(p) = -p\log p - (1-p)\log (1-p)$$
$$G(p) = 2p(1-p)$$

Gini is computationally simpler. Entropy is more sensitive to class imbalance.

3. Why Are Trees Unstable?

Decision trees are unstable because:

- Small changes in data may change the best split.
- Greedy splitting amplifies early decisions.
- Axis-aligned splits are sensitive to noise.

Thus, high variance → unstable model.

4. Computational Complexity of Training

At each node:

- Try p features.
- Sort values: $O(n\log n)$.

Total complexity: $O(pn\log n)$

Worst case (deep tree): $O(pn^2)$

5. Why Greedy Splitting Does Not Guarantee Global Optimum

Tree building solves: $\min_T \sum \text{impurity}(T)$

But greedy splitting:

- Chooses best local split at each node.
- Does not consider future splits.
- Problem is combinatorial and NP-hard.

Thus, greedy optimization ≠ global optimization.

6. How Does Tree Depth Relate to VC Dimension?

For a tree of depth d:

- Maximum number of leaves = 2^d

VC dimension grows roughly as: $VC \approx O(2^d)$

Thus:

- Deeper tree → larger capacity.
- Higher depth → greater overfitting risk.

7. Why Do Trees Overfit Small Datasets?

Trees:

- Keep splitting until pure nodes.
- Small data → random patterns appear meaningful.
- High flexibility → low bias, high variance.

Thus small datasets cause overfitting.

8. Compare CART vs ID3

Property	**CART**	**ID3**
Split type	Binary	Multi-way
Criterion	Gini	Entropy
Handles regression	Yes	No
Pruning	Yes	No (originally)

CART is more general and practical.

9. Derive Pruning as Regularization

Tree objective: $R(T) = R_{emp}(T)$

Pruning adds penalty: $R_\alpha(T) = R_{emp}(T) + \alpha \mid T \mid$

where:

- $|T|$= number of leaves
- α= complexity penalty

This is equivalent to regularization:

Balance fit vs complexity.

10. Why Are Trees Invariant to Monotonic Transformations?

Splits depend on ordering, not absolute values.

If transformation: $x' = f(x)$

where f is monotonic:

Ordering preserved.

Thus, split decisions remain identical.

Hence trees are invariant to monotonic transformations.