RANDOM FOREST – Theoretical Questions

1. Why Does Bagging Reduce Variance but Not Bias?

Bagging averages multiple trees:

$$\hat{f}_{bag} = \frac{1}{B}\sum_{b=1}^{B} f_b\,(x)$$

Bias: $E[\hat{f}_{bag}] = E[f(x)]$

Bias unchanged.

Variance: $Var(\hat{f}_{bag}) = \frac{1}{B^2}\sum Var(f_b) + \frac{1}{B^2}\sum Cov(f_i, f_j)$

Variance reduces when trees not perfectly correlated.

2. Prove That Averaging Reduces Variance

If: $Var(f_i) = \sigma^2$

and correlation ρ,

Then: $Var(\bar{f}) = \frac{\sigma^2}{B} + \rho\sigma^2\frac{B-1}{B}$

If $\rho < 1$,

Variance decreases as B increases.

3. Derive Out-of-Bag (OOB) Error Estimator

Each tree trained on bootstrap sample.

Probability a sample not selected: $\left(1 - \frac{1}{n}\right)^n \approx e^{-1} \approx 0.37$

So ~37% samples unused per tree.

OOB prediction:

- Predict sample using only trees that did not see it.
- Compute average error.

OOB error approximates test error.

4. Why Does Feature Randomness Decorrelate Trees?

At each split:

- Only subset of features considered.

Thus:

- Different trees choose different splits.
- Correlation between trees reduces.
- Lower correlation → lower ensemble variance.

5. Random Forest vs Gradient Boosting (Mathematically)

Random Forest: $f(x) = \frac{1}{B} \sum f_b(x)$

Trees trained independently.

Gradient Boosting: $f_m(x) = f_{m-1}(x) + \gamma_m h_m(x)$

Sequential residual fitting.

RF reduces variance.
Boosting reduces bias.

6. When Does Random Forest Fail?

1. Extremely noisy labels.
2. Very small datasets.
3. Strong extrapolation tasks.
4. Very high memory constraints.
5. When smooth continuous function required.

7. Computational Complexity vs Single Tree

Single tree: $O(pn\log n)$

Random forest: $O(Bpn\log n)$

Inference: $O(B \cdot depth)$

Linear increase with number of trees.

8. Explain Consistency of Random Forest

Under conditions:

- Number of trees $\to \infty$
- Subsampling controlled
- Trees not too deep

Random Forest is consistent: $\hat{f}(x) \to E[Y \mid X = x]$

Proof relies on partition shrinking with increasing data.

9. Why Are Forests Robust to Noise?

Reasons:

1. Averaging reduces variance.
2. Random feature selection prevents overfitting to noise.
3. Bootstrap sampling smooths noise effects.

10. How Do Forests Handle High-Dimensional Data?

- Feature subsampling reduces curse of dimensionality.
- Trees ignore irrelevant features automatically.
- Works well when p >> n.

However:

Too many irrelevant features can still degrade performance.