

Class activity – Generalized Random Forests (GRF)

Please write your full names here:

Student 1:

Student 2:

1. The GRF framework defines a target parameter $\theta(x)$ at covariate value x as the solution to a local moment condition

$$\mathbb{E}[\psi_{\theta(x), \nu(x)}(O_i) \mid X_i = x] = 0,$$

where $\nu(x)$ is a nuisance function and O_i is the observable for unit i .

- (a) Explain in your own words what this moment condition means and why it generalizes standard regression forests.

.....

.....

.....

.....

.....

.....

2. In GRF the forest defines adaptive weights $\alpha_i(x)$ for each training unit i relative to a query point x . The prediction for $\theta(x)$ is written as

$$\hat{\theta}(x) = \sum_{i=1}^n \alpha_i(x) \hat{\theta}_i,$$

or more precisely via a weighted estimating-equation solution using the $\alpha_i(x)$.

- (a) Describe how the weights $\alpha_i(x)$ are constructed from the forest (hint: tree leaves, neighbor counts).
- (b) Why does thinking of these weights as “adaptive nearest-neighbor weights” help understand the method’s behavior in high dimensions?

.....

.....

.....

.....

.....

.....

3. The GRF paper explains how the prediction step works: after training, for a new point x , each tree assigns x to a leaf, training units that ended up in the same leaf get a “vote” / contribution. Then the forest weights are normalized across trees. Suppose a forest of B trees yields for a test point x counts $c_i(x)$ for each training example i (how many trees placed i and x in the same leaf). Then one may define

$$\alpha_i(x) = \frac{c_i(x)}{\sum_{j=1}^n c_j(x)}.$$

- (a) Explain how this weight definition reproduces the kernel-weighted intuition of GRF.
- (b) Give one potential drawback when many $c_i(x)$ are zero or extremely small, and what practical mitigation the paper suggests (or you would suggest).

.....

.....

.....

.....

.....

4. Theoretical results in the GRF paper show that under regularity conditions

$$\sqrt{n}(\hat{\theta}(x) - \theta(x)) \xrightarrow{d} \mathcal{N}(0, V(x)),$$

and the package provides an estimator $\hat{V}(x)$ to form confidence intervals.

- (a) Why is asymptotic normality important for the applied econometrician using GRF?
- (b) Suppose you compute $\hat{\theta}(x) \pm 1.96 \sqrt{\hat{V}(x)}$. What assumptions underlie the validity of this interval? What happens if one of these assumptions fails?

.....

.....

.....

.....

.....