



Honest Causal Tree

Presenter: FU Zixuan & SONG Shuo

Date: 2025-07-24

Preliminaries

- Object: $\tau(x)$.
- Goal:
 - How **large** is the effect τ for a specific x ?
 - How **different** is the effect across x ?
- Use:
 - Improvement on **treatment design** itself.
 - Improvement of **assignment** of treatment.

On a side note

- Causal Trees: focuses on the heterogeneity of treatment effects. How different is subgroup l_1 from subgroup l_2 ? Which group is benefiting the most?
- Causal Forests: focuses on the precision. personalized treatment effect for each individual x_i . Can be used as input into estimation of lower dimensional objects, such as to estimate optimal policy.

Trees: an overview of regression trees

“ The best time to plant a tree was 20 years ago. The second best time is now. ”

1. **Build a tree** (partition) π square based on data.
2. Given the tree, **construct an estimator** $\hat{\mu}$ by calculating the mean outcome in each leaf in the tree (box in the square).

Trees: a walkthrough of the Conventional

Sample splitting

1. Split the data into training $S^{tr, tr}$, and validation set $S^{tr, cv}$ and test set S^{te} .
2. Hold out the test set. This is not going to be used in following steps except the final evaluation.

Tree growing

1. Grow the tree π_0 on the entire training set S^{tr} fold by minimizing the MSE until the leaves contain a minimum number of samples.

$$\widehat{\text{MSE}}(S^{tr}, S^{tr}) = \frac{1}{|S^{tr}|} \sum_{i \in S^{tr}} (Y_i - \hat{\mu}_{\pi}(X_i; S^{tr}))^2$$

where

$$\hat{\mu}_{\pi}(x; S^{tr}) = \frac{1}{|S_l^{tr}|} \sum_{i \in S_l^{tr}} Y_i, \quad x \in l$$

Tree pruning

1. Get a set of effective α_k to prune the tree π_0 .
2. For one α_k . Split the training set into two parts: $S^{tr,tr}$ and $S^{tr,cv}$. Find the pruned tree π^* that minimizes the following.

$$C_{\alpha_k}(\pi') = \frac{1}{|S^{tr,cv}|} \sum_{i \in S^{tr,cv}} (Y_i - \hat{\mu}_{\pi'}(X_i; S^{tr,tr}))^2 + \alpha_k |\pi'|$$

3. Pick the α_k (along with its corresponding pruned tree π_{α_k}) that gives the smallest cost-complexity measure.

Final evaluation

1. Construct the final estimator $\hat{\mu}_{\pi^*}(x; S^{tr})$ based on the optimal tree $\pi^* = \pi_{\alpha_k}$
2. Evaluate the estimator $\hat{\mu}_{\pi^*}$ on the hidden test set.

A conundrum

We shift our focus from predicting outcome Y_i to predicting effect τ_i . However, while Y_i is observed in the data, the τ_i is not!

Therefore, we need to adapt the evaluation criteria from $\widehat{\text{MSE}}$ to $\widehat{\text{EMSE}}$.

Adapt the criterion

$$\widehat{\text{MSE}}_{\pi}(S_1, S_2) = \frac{1}{|S_1|} \sum_{i \in S_1} (Y_i - \hat{\mu}_{\pi}(X_i; S_2))^2$$

$$\text{MSE}_{\pi}(S_1) = \frac{1}{|S_1|} \sum_{i \in S_1} \mathbb{E}_{S_2} (Y_i - \hat{\mu}_{\pi}(X_i; S_2))^2$$

In the case of conventional,

$$\text{MSE}_{\pi} = \mathbb{E}_S \left\{ \frac{1}{|S|} \sum_{i \in S} (Y_i - \hat{\mu}_{\pi}(X_i; S))^2 \right\} = \mathbb{E}(Y_i^2) - 2\mathbb{E}(Y_i \hat{\mu}_{\pi}(X_i; S)) + \mathbb{E}(\hat{\mu}_{\pi}(X_i; S))^2$$

$$\begin{aligned}\text{EMSE}_\pi &= \mathbb{E}_{S_1, S_2} \left(\frac{1}{|S_1|} \sum_{i \in S_1} (Y_i - \hat{\mu}_\pi(X_i; S_2))^2 \right) \\ &= \mathbb{V}_{S_1, S_2}(\hat{\mu}_\pi(X_i; S_2)) - \mathbb{E}_{S_1}(\mu_\pi(X_i))^2 + \mathbb{E}_{S_1}(Y_i^2)\end{aligned}$$

Why different set?

We want to make use the independence between sets.

First

$$\begin{aligned} & \mathbb{E} \left\{ (Y_i - \mu_\pi(X_i) + \mu_\pi(X_i) - \hat{\mu}_\pi(X_i; S_2))^2 \right\} \\ &= \mathbb{E} \left\{ (Y_i - \mu_\pi(X_i))^2 \right\} + \mathbb{E} \left\{ (\mu_\pi(X_i) - \hat{\mu}_\pi(X_i; S_2))^2 \right\} \\ & \quad + \mathbb{E}_{S_1, S_2} \left\{ (Y_i - \mu_\pi(X_i))(\mu_\pi(X_i) - \hat{\mu}_\pi(X_i; S_2)) \right\} \end{aligned}$$

The third term is zero because

$$\mathbb{E}_{S_1} \{ (Y_i - \mu_\pi(X_i))(\mu_\pi(X_i) - \mathbb{E}_{S_2} \hat{\mu}_\pi(X_i; S_2)) \} = 0$$

where we make use of the independence between S_1 and S_2 !

The second term is $\mathbb{E}_{S_2} \{ \mathbb{V}_{S_1}(\hat{\mu}_\pi(X_i; S_2)) \}$

The first term can be simplified to

$$\begin{aligned} & \mathbb{E}_{S_1} \{ (Y_i^2) - 2Y_i\mu_\pi(X_i) + \mu_\pi(X_i)^2 \} \\ &= \mathbb{E}_{S_1} \{ Y_i^2 - 2\mathbb{E}(Y_i|X_i)\mu_\pi(X_i) + \mu_\pi(X_i)^2 \} = \mathbb{E}_{S_1} \{ Y_i^2 - \mu_\pi(X_i)^2 \} \end{aligned}$$

Putting it together

We estimate EMSE by the estimating the two components.

$$\mathbb{E}_{S_2}\{\mathbb{V}_{S_1}(\hat{\mu}_{\pi}(X_i; S_2))\}, \mathbb{E}_{S_1}[\mu_{\pi}(X_i)^2]$$

Therefore, we can estimate the EMSE by

$$\widehat{\text{EMSE}}_{\pi}(S^{tr}, S^{est})$$

Therefore, in the tree growing step, we replace

$$\widehat{\text{MSE}}_{\pi}(S^{tr}, S^{tr}) \quad \text{by} \quad \widehat{\text{EMSE}}_{\pi}(S^{est}, S^{tr})$$

Similarly in the tree pruning step, we replace

$$C_{\alpha_k}(\pi') = \widehat{\text{MSE}}_{\pi}(S^{tr,cv}, S^{tr,tr}) + \alpha_k |\pi'|$$

by

$$C_{\alpha_k}(\pi') = \widehat{\text{EMSE}}_{\pi}(S^{tr,cv}, S^{est}) + \alpha_k |\pi'|$$

The conundrum?

We adapt the criterion because we don't observe the treatment effect τ_i directly. But how does it solve the problem?

Then

$$\widehat{\text{EMSE}}_{\tau}(S^{\text{tr}}, S^{\text{est}}, \Pi) \equiv \frac{1}{N_{\text{tr}}} \sum_{i \in S^{\text{tr}}} \hat{\tau}^2(X_i; S^{\text{tr}}, \Pi) \\ - \left(\frac{1}{N_{\text{tr}}} + \frac{1}{N_{\text{est}}} \right) \cdot \sum_{\ell \in \Pi} \left(\frac{S_{\text{treat}}^{2, \text{tr}}(\ell)}{p} + \frac{S_{\text{control}}^{2, \text{tr}}(\ell)}{1 - p} \right)$$

$$\widehat{\text{EMSE}}_{\mu}(S^{\text{tr}}, N^{\text{est}}, \Pi) \equiv \frac{1}{N_{\text{tr}}} \sum_{i \in S^{\text{tr}}} \hat{\mu}^2(X_i; S^{\text{tr}}, \Pi) \\ - \left(\frac{1}{N_{\text{tr}}} + \frac{1}{N_{\text{est}}} \right) \cdot \sum_{\ell \in \Pi} S_{\text{tr}}^2(\ell(x; \Pi))$$

Other methods

1. Single trees (S-learner): does not split on treatment...
2. Two trees (T-learner): split on different features therefore can not compare...
3. Transform the LHS variable from outcome into treatment effect
4. Other splitting criteria:
 1. split on the outcome
 2. split based on some T statistics.

Related literature

Thanks!

