

# CRE: An R package for interpretable discovery and inference of heterogeneous treatment effects

Riccardo Cadei<sup>1,2\*</sup>, Naeem Khoshnevis<sup>1\*</sup>, Kwonsang Lee<sup>1</sup>, Daniela Maria Garcia<sup>1</sup>, and Falco J. Bargagli Stoffi<sup>1</sup>

<sup>1</sup> Department of Biostatistics, Harvard School of Public Health <sup>2</sup> Department of Computer and Communication Science, EPFL \* These authors contributed equally.

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

## Software

- Review [↗](#)
- Repository [↗](#)
- Archive [↗](#)

Editor: [Open Journals](#) [↗](#)

## Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

In health and social sciences, it is critically important to identify subgroups of the study population where a treatment has notable heterogeneity in the causal effects with respect to the average treatment effect (ATE). The bulk of heterogeneous treatment effect (HTE) literature focuses on two major tasks ([Dwivedi et al., 2020](#)): (i) estimating HTEs by examining the conditional average treatment effect (CATE); (ii) discovering subgroups of a population characterized by HTE.

Several methodologies have been proposed for both tasks, but providing interpretability in the results is still an open challenge. Interpretability is a non-mathematical concept, yet it is often defined as the degree to which a human can understand the cause of a decision ([Kim et al., 2016](#); [Lakkaraju et al., 2016](#); [Miller, 2019](#); [Wang & Rudin, 2022](#)). Honest Causal Tree ([Athey & Imbens, 2016](#)) fits this definition perfectly, but despite its high interpretability, it tends to be highly unstable and to find an oversimplified representation of treatment heterogeneity ([Bargagli-Stoffi et al., 2022](#)). To accommodate these shortcomings, Bargagli-Stoffi et al. (2023) proposed Causal Rule Ensemble, a new method for HTE characterization in terms of decision rules, via an extensive exploration of heterogeneity patterns by an ensemble-of-trees approach, enforcing high stability in the discovery. CRE is an R Package providing a flexible implementation of the Causal Rule Ensemble algorithm.

## Algorithm

Causal Rule Ensemble relies on the Treatment Effect linear decomposition assumption, characterizing the Conditional Average Treatment Effect (CATE) by  $M + 1$  distinct contributions:

$$\tau(x) = \mathbb{E}[\tau_i | X_i = x] = \bar{\tau} + \sum_{m=1}^M \alpha_m \cdot r_m(x)$$

where  $\bar{\tau}$  is the ATE, and for each  $m$  in  $\{1, \dots, M\}$ ,  $r_m$  is an interpretable decision rule characterizing a specific subset of the covariate space, and  $\alpha_m$  is the corresponding Additive Average Treatment Effect (AATE). CRE procedure is divided into two steps, discovery and estimation, and each observation is used for only one of the two steps (honest splitting). During the discovery step, CRE retrieves the  $M$  decision rules characterizing the heterogeneity in the treatment effect. A set of candidate decision rules is extracted by an ensemble of trees trained by a *fit-the-fit* procedure to model some Individual Treatment Effect (ITE) estimates, and among these, only a simple and robust subset of rules is selected for the linear decomposition by the Stability Selection algorithm via LASSO. During the estimation step, CRE estimates the ATE and AATEs, by the normal equations to model some ITE estimates. In both steps, CRE is agnostic concerning the method used for ITE estimation.

## Usage

CRE is available both on [CRAN](#) and [GitHub](#) and can be installed and loaded into the R session using:

```
install.packages("CRE")
library("CRE")
```

`generate_cre_dataset()` is a flexible synthetic dataset generator, which can be used for simulations before applying CRE to real-world observational data sets.

```
dataset <- generate_cre_dataset(n = 2000,
                                rho = 0,
                                n_rules = 4,
                                p = 10,
                                effect_size = 2,
                                binary_covariates = TRUE,
                                binary_outcome = FALSE,
                                confounding = "no")
```

```
y <- dataset$y
z <- dataset$z
X <- dataset$X
```

We propose here three examples of how to run the Causal Rule Ensemble algorithm by the CRE package.

**Example 1.** Running Causal Rule Ensemble with default parameters described in Bargagli-Stoffi et al. (2023).

```
cre_results <- cre(y, z, X)
```

**Example 2.** Running Causal Rule Ensemble with customized ITE estimator.

```
ite_pred <- ... # personalized ite estimation
cre_results <- cre(y, z, X, ite = ite_pred)
```

**Example 3.** Running Causal Rule Ensemble with customized parameters.

```
method_params <- list(ratio_dis = 0.25,
                      ite_method_dis = "aipw",
                      ps_method_dis = "SL.xgboost",
                      oreg_method_dis = "SL.xgboost",
                      ite_method_inf = "aipw",
                      ps_method_inf = "SL.xgboost",
                      oreg_method_inf = "SL.xgboost")

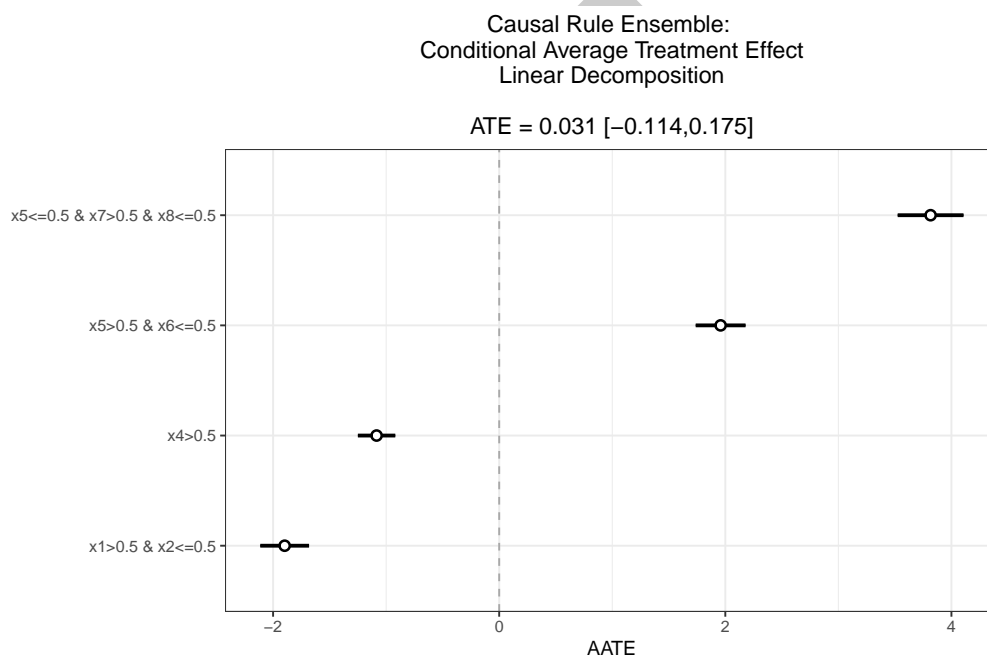
hyper_params <- list(intervention_vars = c("x1", "x2", "x3", "x4"),
                    offset = NULL,
                    ntrees_rf = 20,
                    ntrees_gbm = 20,
                    node_size = 20,
                    max_nodes = 5,
                    max_depth = 3,
                    t_decay = 0.025,
                    t_ext = 0.025,
                    t_corr = 1,
                    t_pvalue = 0.05,
                    replace = FALSE,
                    stability_selection = TRUE,
                    cutoff = 0.8,
```

```
pfer = 0.1,  
penalty rl = 1)
```

```
cre_results <- cre(y, z, X, method_params, hyper_params)
```

50 The results are collected in an S3 object containing: the number of decision rules extracted  
51 at each step (M), the data.frame of the CATE decomposition estimates with correspond-  
52 ing uncertainty quantification (CATE), the list of selected parameters (method\_params and  
53 hyper\_params), and the predicted ITEs (ite\_pred).

54 `summarize()` and `print()` display a summary of these results, and `plot()` visualizes the CATE  
55 decomposition estimates in a range bar plot. Figure 1 reports an example of the proposed  
56 results visualization for Example 1.



**Figure 1:** Visualization of Causal Rule Ensemble HTE linear decomposition for Example 1. For each decision rule discovered, the corresponding AATE estimate with 95% confidence interval is reported in a range bar plot. The decision rules are ordered from the most vulnerable (high AATE) to the least, and the ATE is reported on top of the plot.

57 Online documentation for the package can be found at <https://nsaph-software.github.io/CRE/>.

## 58 Acknowledgements

59 This work was partially funded by the following grants: NIH: R01ES026217, R01MD012769,  
60 R01ES028033, 1R01ES030616, 1R01AG066793, 1R01MD016054-01A1, R01AG066793-02S1  
61 and R01ES028033-03S1; Sloan Foundation: G-2020-13946.

## 62 References

63 Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects.  
64 *Proceedings of the National Academy of Sciences*, 113(27), 7353–7360.

- 65 Bargagli-Stoffi, F. J., Cadei, R., Lee, K., & Dominici, F. (2023). Causal rule ensemble:  
66 Interpretable Discovery and Inference of Heterogeneous Treatment Effects. *arXiv Preprint*  
67 *arXiv:2009.09036*.
- 68 Bargagli-Stoffi, F. J., De Witte, K., & Gnecco, G. (2022). Heterogeneous causal effects with  
69 imperfect compliance: A bayesian machine learning approach. *The Annals of Applied*  
70 *Statistics*, 16(3), 1986–2009.
- 71 Dwivedi, R., Tan, Y. S., Park, B., Wei, M., Horgan, K., Madigan, D., & Yu, B. (2020).  
72 Stable discovery of interpretable subgroups via calibration in causal studies. *International*  
73 *Statistical Review*, 88, 135–178.
- 74 Kim, B., Khanna, R., & Koyejo, O. O. (2016). Examples are not enough, learn to criti-  
75 cize! Criticism for interpretability. *Advances in Neural Information Processing Systems*,  
76 2280–2288.
- 77 Lakkaraju, H., Bach, S. H., & Leskovec, J. (2016). Interpretable decision sets: A joint framework  
78 for description and prediction. *Proceedings of the 22nd ACM SIGKDD International*  
79 *Conference on Knowledge Discovery and Data Mining*, 1675–1684.
- 80 Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences.  
81 *Artificial Intelligence*, 267, 1–38.
- 82 Wang, T., & Rudin, C. (2022). Causal rule sets for identifying subgroups with enhanced  
83 treatment effects. *INFORMS Journal on Computing*.