Tree Ensemble Explainability through the Hoeffding Functional Decomposition and TreeHFD Algorithm

NeurIPS 2025

Clément Bénard

Thales cortAIx-Labs, Palaiseau, France

Hoeffding Functional Decomposition (HFD)

- ν is a real function defined on a bounded subset of $\mathbb{R}^p$.
- $\mathbf{X} = (X^{(1)}, \ldots, X^{(p)})$ is a random input vector of dimension p.
- Hoeffding functional decomposition:

$$\begin{aligned}\nu(\mathbf{X}) =& \nu_1(X^{(1)}) + \ldots + \nu_p(X^{(p)}) \\ &+ \nu^{(1,2)}(X^{(1)}, X^{(2)}) + \ldots + \nu^{(p-1,p)}(X^{(p-1)}, X^{(p)}) \\ &+ \sum_{\substack{J \subset \{1,\ldots,p\} \\ \text{s.t.} |J| \geq 3}} \nu^{(J)}(X^{(J_1)}, X^{(J_2)}, \ldots).\end{aligned}$$

- Properties : uniqueness, pure interactions, sparsity, causal variable selection

Hoeffding Functional Decomposition (HFD)

Theorem (Hoeffding Decomposition (Stone, 1994; Hooker, 2007))

If the input vector $\boldsymbol{X}$ *admits a positive bounded density on* $[0,1]^p$*, and* ν *is a square-integrable real function defined on* $[0,1]^p$*, then there exists a unique set of functions* $\{\nu^{(J)}\}_{J\in\mathcal{P}_p}$*, such that*

$$\nu(\boldsymbol{X}) = \sum_{J\in\mathcal{P}_p} \nu^{(J)}(\boldsymbol{X}^{(J)}),$$

and for all $J \in \mathcal{P}_p$*,* $I \subset J$ *with* $I \neq J$*,* $\mathbb{E}[\nu^{(J)}(\boldsymbol{X}^{(J)})|\boldsymbol{X}^{(I)}] = 0$*.*

HFD Illustration

Example of the HFD with the California Housing dataset.

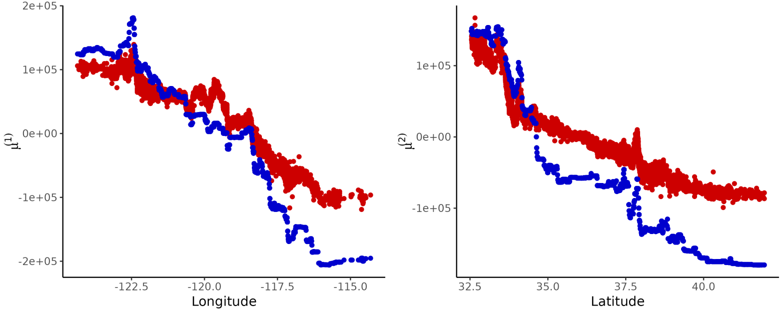

Figure: For the "California Housing" dataset, main effects of "Longitude" and "Latitude" in the decompositions of respectively TreeHFD in blue and TreeSHAP with interactions in red.

State-of-the-art:

- For independent inputs, the HFD has closed formula.
- For dependent inputs with known distribution, the HFD can be estimated with tree ensembles (Lengerich et al., 2020).
- Open problem for dependent inputs without the input distribution.

Objective: Estimate the HFD in standard machine learning settings, using tree ensembles (gradient boosted trees (Friedman, 2001), XGBoost (Chen and Guestrin, 2016), random forests (Breiman, 2001))

Main idea: Discretization of the input distribution across the Cartesian tree partitions.

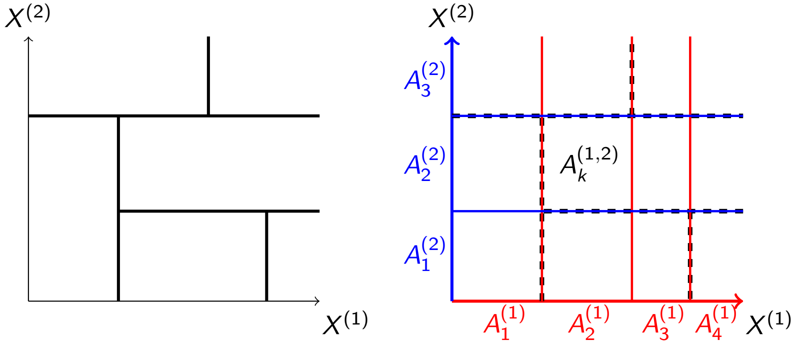

Figure: Example of the partition of $[0,1]^2$ by a tree T_ℓ (left side), and the associated Cartesian tree partitions $\mathcal{A}_\ell^{(1)} = \{A_1^{(1)}, A_2^{(1)}, A_3^{(1)}, A_4^{(1)}\}$, $\mathcal{A}_\ell^{(2)} = \{A_1^{(2)}, A_2^{(2)}, A_3^{(2)}\}$, and $\mathcal{A}_\ell^{(1,2)}$ (right side).

HFD for Tree Ensembles

- HFD can be adapted for tree ensembles, with a discretization of the orthogonality constraints across the Cartesian tree partitions.
- Strong theoretical properties for hierarchical orthogonality, sparsity, causal variable selection, and accuracy.
- TreeHFD algorithm minimizes a lost function for each tree to estimate the HFD.

Analytical Example:

- Gaussian random vector $\mathbf{X}$ with $p = 6$ (correlation $\rho = 1/2$)
- $Y = m(\mathbf{X}) + \varepsilon \quad (\varepsilon \sim \mathcal{N}(0, 0.5^2))$
- $m(\mathbf{X}) = \sin(2\pi X^{(1)}) + X^{(1)}X^{(2)} + X^{(3)}X^{(4)}$

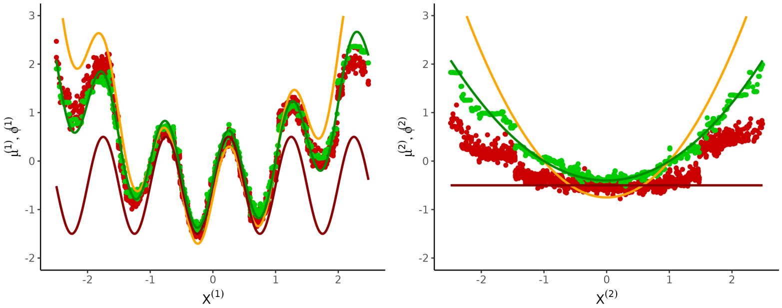

Figure: Main effects of the decompositions for $X^{(1)}$ and $X^{(2)}$. Solid lines provide the theoretical functions, with the HFD in green, int. SHAP in red, and obs. SHAP in orange. Green and red points are respectively the values provided by TreeHFD and TreeSHAP with interactions for `xgboost`.

TreeHFD for the California Housing dataset versus TreeSHAP.

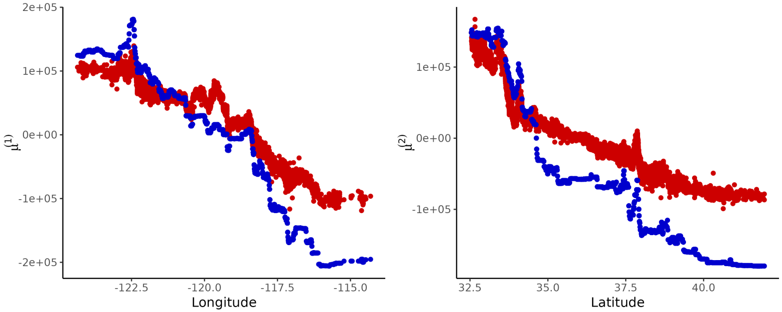

Figure: For the "California Housing" dataset, main effects of "Longitude" and "Latitude" in the decompositions of respectively TreeHFD in blue and TreeSHAP with interactions in red.

Conclusion

- TreeHFD is an accurate estimate of the HFD based on tree ensembles.
- Open source Python package for XGBoost: **treehfd**
- **treehfd** is available on pypi and github: `https://github.com/ThalesGroup/treehfd`.

L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, New York, 2016. ACM.

J.H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, pages 1189–1232, 2001.

Giles Hooker. Generalized functional anova diagnostics for high-dimensional functions of dependent variables. *Journal of Computational and Graphical Statistics*, 16:709–732, 2007.

Benjamin Lengerich, Sarah Tan, Chun-Hao Chang, Giles Hooker, and Rich Caruana. Purifying interaction effects with the functional anova: An efficient algorithm for recovering identifiable additive models. In *International Conference on Artificial Intelligence and Statistics*, pages 2402–2412. PMLR, 2020.

Charles J Stone. The use of polynomial splines and their tensor products in multivariate function estimation. *The Annals of Statistics*, 22:118–171, 1994.