Variable Importance in High-Dimensional and Correlated Settings

Angel D Reyero Lobo[1,2] , Pierre Neuvial[1] and Bertrand Thirion[2]
Email: angel.reyero-lobo@inria.fr
[1] Institut de Mathématiques de Toulouse
[2] MIND, Inria Paris-Saclay

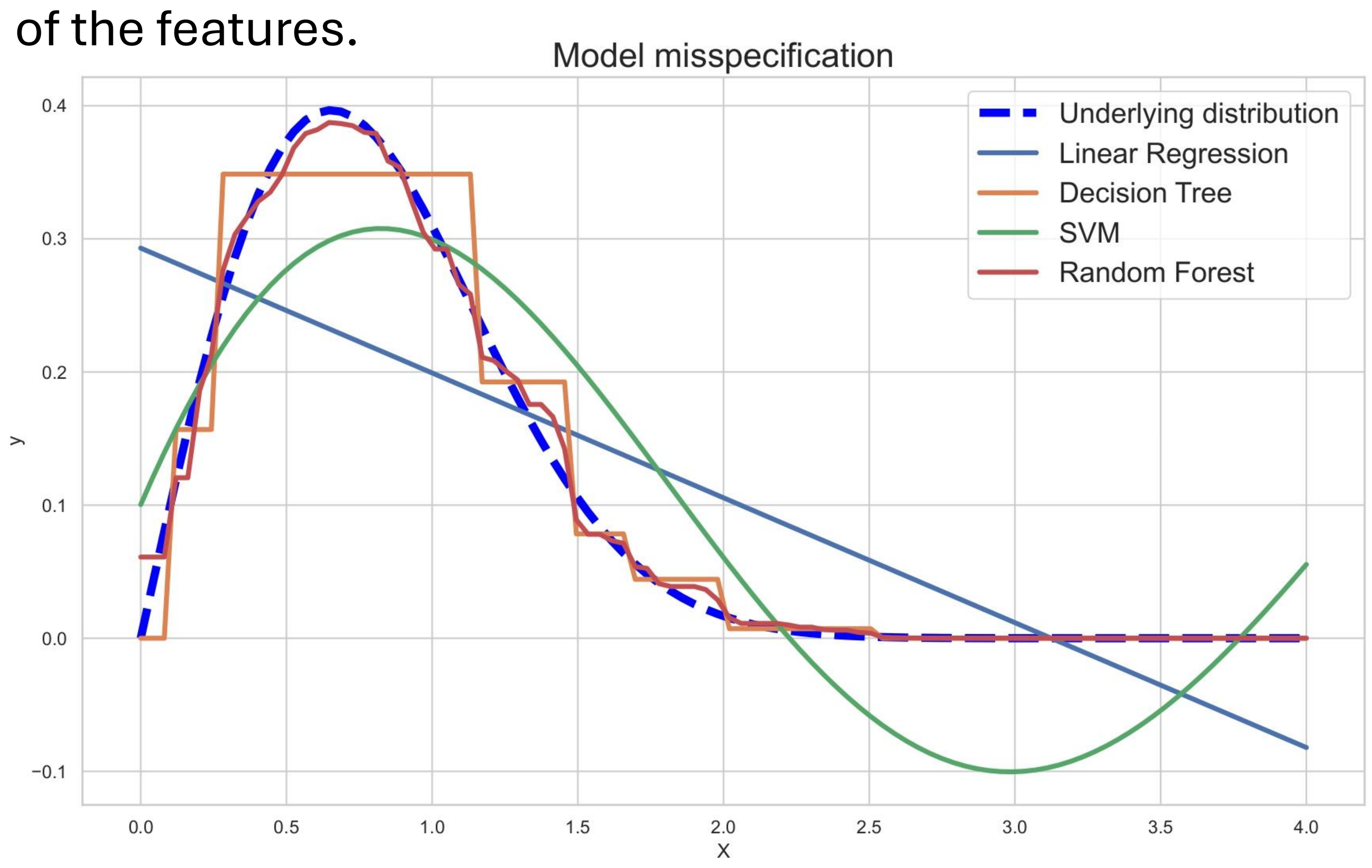
Introduction

- **Goal:** Evaluate the contribution of each feature in input X to the output variable Y.
- Their relationship can be complex, so we use a machine learning model to understand it:

$$\widehat{m} \in \underset{f\in\mathcal{F}}{\operatorname{argmin}}\widehat{\mathbb{E}}\left[(f(X)-Y)^2\right].$$

Motivation

- Oversimplified models, while transparent, fail to capture the underlying distribution and, consequently, the intrinsic value of the features.



- Moreover, we cannot directly compare the importance of variables specific to each model between them!
- We need a **model-agnostic** approach, i.e. an interpretation method applicable to any regression technique.

Total Sobol Index

- From **Global Sensitivity Analysis**, the Sobol index is a model-agnostic statistical functional of interest that quantifies the difference in precision between the model excluding the group of covariates $s \subset \{1\dots p\}$ and the complete model:

$$\psi_s(P) := \frac{\mathbb{E}_P\left[\left(Y-m_{P,s}(X^{-s})\right)^2\right]}{\mathrm{Var}_P(Y)} - \frac{\mathbb{E}_P\left[\left(Y-m_P(X)\right)^2\right]}{\mathrm{Var}_P(Y)},$$

where $m_P(x) := \mathbb{E}_P\left[Y|X=x\right]$ and $m_{P,s}(x^{-s}) := \mathbb{E}_P\left[Y|X_{-s}=x^{-s}\right]$.
- See Bénard et al. (2022).

Estimation

- Naive idea: **plug-in** estimate.

$$\widehat{\psi}_s(P) := \frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}\frac{\left(y_i-\widehat{m}_{P,s}(x_i^{-s})\right)^2}{\widehat{\mathrm{Var}}_P(Y)} - \frac{\left(y_i-\widehat{m}_P(x_i)\right)^2}{\widehat{\mathrm{Var}}_P(Y)}.$$

  - It needs a bias correction (Williamson et al. (2021)).
  - It is **computationally intensive**: we need to refit a $m_{P_0,s}$ for each group **s**!

- **Refitting** is expensive but **predicting** is cheap.

(Conditional) Permutation Feature Importance

- **Permutation Feature Importance (PFI):** It measures the predictive capability of the j-th covariate by breaking its relationship with the output.
- In practice, the j-th column of the test matrix is permuted:

$$\widehat{\chi}_j^{\mathrm{PFI}} := \frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}\Big((y_i-\widehat{m}(x_i^{(j)}))^2-(y_i-\widehat{m}(x_i))^2\Big).$$

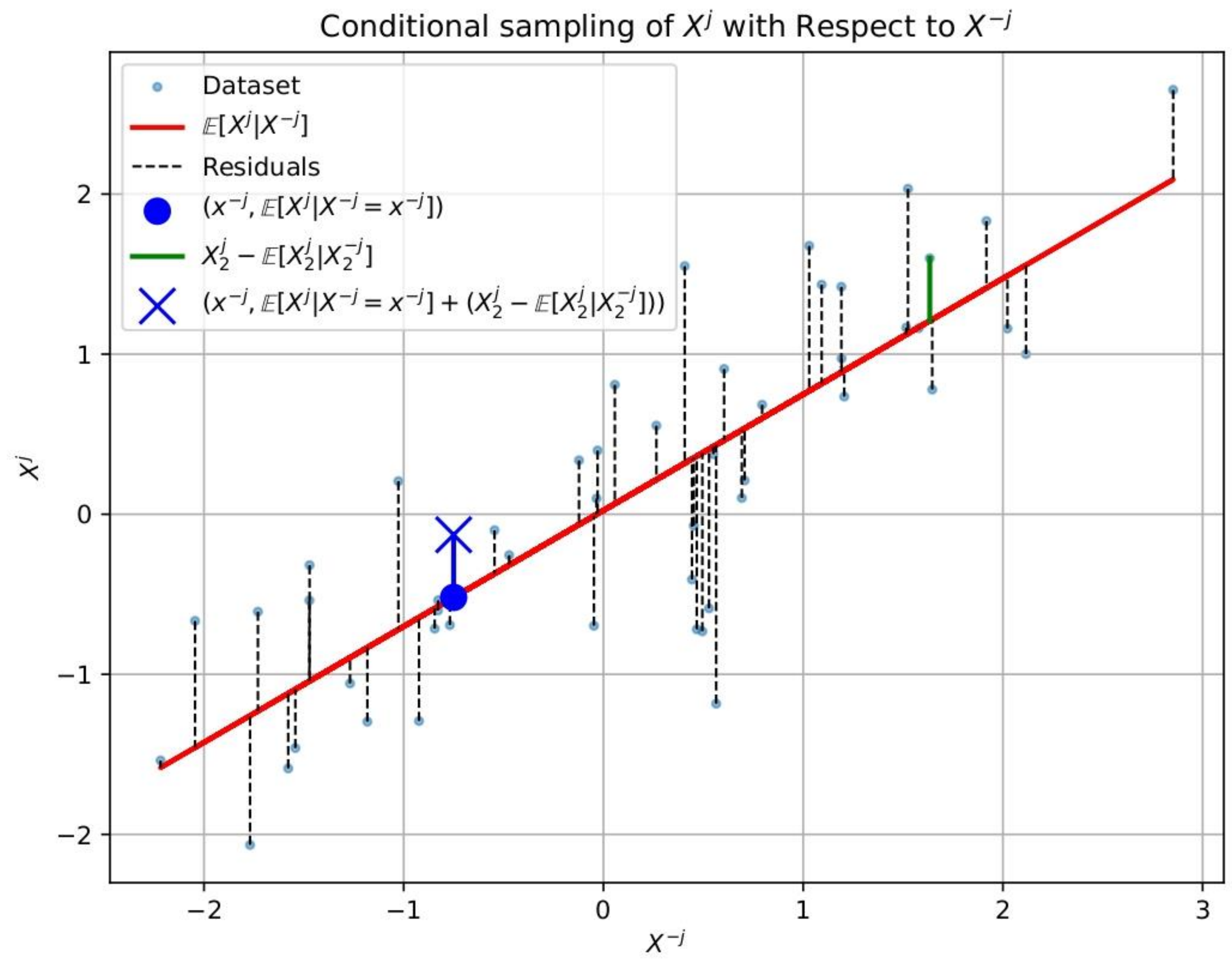
- It also breaks the relationship with the rest of the input covariates, inducing **extrapolation** bias.

- **Conditional Permutation Importance (CPI):** introduced by Chamma et al.(NeurIPS 2023), the j-th covariate is *conditionally* permuted on the rest of the input covariates, breaking only its relationship with the output variable:

$$\widehat{\chi}_j^{\mathrm{CPI}} := \frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}\Big((y_i-\widehat{m}(\widetilde{x}_i^{(j)}))^2-(y_i-\widehat{m}(x_i))^2\Big).$$

An idea of conditional sampling/ permutation

- Under Gaussianity assumption, the conditional distribution $X^j|X^{-j}=x^{-j}$ is i.i.d. to $\widetilde{X}^{(j)}|X^{-j}=x^{-j}\sim\nu_{-j}(x^{-j})+\left(X_2^j-\nu_{-j}(X_2^{-j})\right)$ where $\nu_{-j}(x^{-j}) := \mathbb{E}\left[X^j|X^{-j}=x^{-j}\right]$ and $X_2 \overset{\text{iid}}{\sim} X$.
- In practice, we regress X^j given X^{-j} to obtain $\widehat{\nu}_{-j}$. Then, to sample from $\widetilde{X}^{(j)}|X^{-j}=x^{-j}$ we first predict $\widehat{\nu}_{-j}(x^{-j})$ and finally we add a random residual $X_2^j-\widehat{\nu}_{-j}(X_2^{-j})$.



Discussion

- **Aim:** a **computationally feasible**, **statistically valid** and **model-agnostic** variable importance measure that works in **highly-dimensional** and **correlated** settings.
- How can we relate the total Sobol index with CPI?
- How can we generalize the conditional sampling?
- What assumptions are necessary about the model estimate $\widehat{m}$ to provide information about the intrinsic variable importance?
- Does the total Sobol index capture all the interactions between the covariates, or do we need to move forward with Shapley values?

Selected references:

- Clément Bénard, Sébastien da Veiga, and Erwan Scornet. Mda for random forests: inconsistency, and a practical solution via the sobol-mda, 2022.
- Williamson BD, Gilbert PB, Carone M, Simon N. Nonparametric variable importance assessment using machine learning techniques. Biometrics. 2021; 77: 9–22.
- Chamma, A., Engemann, D. A., and Thirion, B. (2023). Statistically valid variable importance assessment through conditional permutations.