Model-Free Variable Importance

Zhao Lyu, Beining Wu

November 21, 2023

THE UNIVERSITY OF CHICAGO

Leave-One-Covariate-Out (LOCO) [LGR+18]

- Set up:
  - A training data index set: $\mathcal{I}_1 \subseteq \{1, \ldots, n\}$
  - Training data: $\{(X_i, Y_i)\}_{i \in \mathcal{I}_1}$, where $X_i \in \mathbb{R}^d$, $Y_i \in \mathbb{R}$
- Predictive Goal: Use the training data to construct $\widehat{\mu}$, our estimate of the mean function $\mu(x) = \mathbb{E}(Y \mid X = x)$

$$\widehat{\mu} = \mathcal{A}(\{(X_i, Y_i)\}_{i \in \mathcal{I}_1})$$

- Selective Goal: Use the training data to select covariates important to the prediction

$$\widehat{\mathcal{I}} = \mathcal{A}'(\{(X_i, Y_i)\}_{i \in \mathcal{I}_1}) \subseteq [d]$$

LOCO-continued

- New data: $\{(X_i(-j), Y_i)\}_{i \in \mathcal{I}_1}$, where $X_i(-j) = (X_i(1), \ldots, X_i(j-1), X_i(j+1), \ldots, X_i(d) \in \mathbb{R}^{d-1}$
- New estimate: $\widehat{\mu}_{(-j)} = \mathcal{A}(\{(X_i(-j), Y_i)\}_{i \in \mathcal{I}_1})$
- Excess prediction error of covariate j at (X_{n+1}, Y_{n+1}), a new i.i.d draw:

$$\Delta_j(X_{n+1}, Y_{n+1}) = |Y_{n+1} - \widehat{\mu}_{(-j)}(X_{n+1})| - |Y_{n+1} - \widehat{\mu}(X_{n+1}| \tag{1}$$

- Interpretation: an increase in prediction error due to not having covariate j in the data set.
- A QUESTION: could $\Delta_j(X_{n+1}, Y_{n+1})$ be negative? How to interpret this?

Local measure of Variable Importance

- Intuition: If covariate j is important, Δ_j should be large
- Method: Construct a valid prediction interval for $\Delta_j(X_{n+1}, Y_{n+1})$:

$$W_j(x) = \{|y - \widehat{\mu}_{-j}(x)| - |y - \widehat{\mu}(x)| \ : \ y \in C(x)\}, \qquad (2)$$

  where C is a conformal prediction set with $\mathbb{P}(Y_{n+1} \in C(X_{n+1}) \geq 1 - \alpha$.
- Validity: $\forall j \in [d]$,

$$\mathbb{P}(\Delta_j(X_{n+1}, Y_{n+1}) \in W_j(X_{n+1})) \geq 1 - \alpha. \qquad (3)$$

Demo 1: independent X_i in low-dim LOCO via local measure

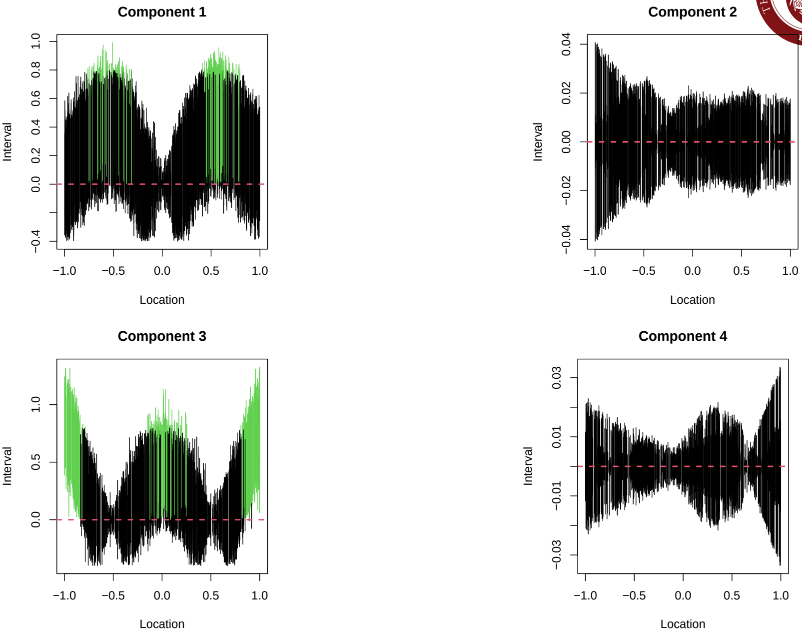
- Let $X_i \sim \text{Unif}[-1,1]^d$ with $d = 4$.
- Let 1st and 3rd covariates be the important ones by constructing

$$\mu(x) = \sum_{i=1}^{4} f_i(x(i)),$$

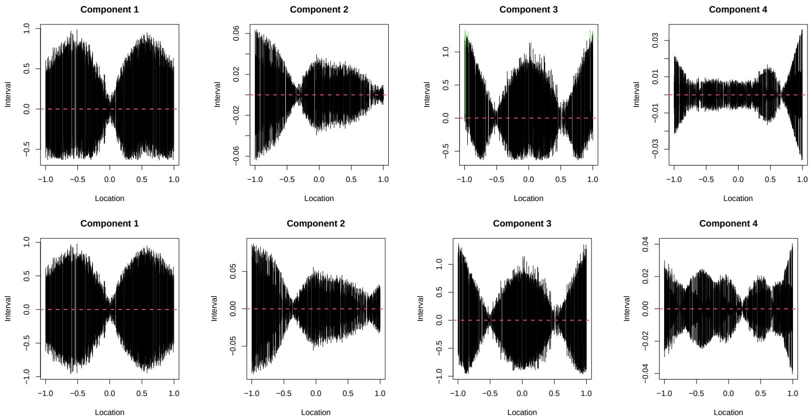
where $f_1(t) = \sin(\pi t), f_3(t) = \cos(\pi t)$ and $f_2 = f_4 = 0$.
- Set the response for $i \in [n]$, and set $n = 1000$

$$Y_i = \mu(X_i) + \epsilon_i$$

Demo 1-continued: $\epsilon_i \sim N(0, 0.1)$



Demo 1-continued: if there comes more noise



Demo 2: correlated X_i in low-dim LOCO via local measure

- Let $X \sim \mathrm{N}(0, \Sigma)$ with $d = 4$ and $\Sigma = \begin{bmatrix} 1 & \rho & 0 & 0 \\ \rho & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.
- Let 1st and 3rd covariates be the important ones by constructing

$$\mu(x) = \sum_{i=1}^{4} f_i(x(i)),$$

where $f_1(t) = 100t^2, f_3(t) = 50|t|$ and $f_2 = f_4 = 0$.

- Set the response for $i \in [n]$, and set $n = 1000$

$$Y_i = \mu(X_i) + \epsilon_i, \epsilon_i \sim \mathrm{N}(0, 0.1)$$

Demo 2-continued: changes on ρ

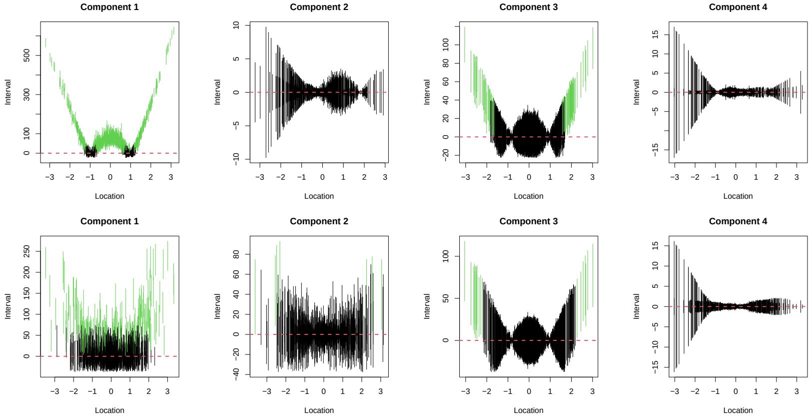


Figure: When $\rho = 0.1$ and $\rho = 0.9$

Conclusion from Demo 1 & 2

- Question: when LOCO work when LOCO does not work?
- (Incomplete) answer:
  - Signal to Noise Ratio(SNR): in Demo 1, when SNR is small (i.e. noise is relatively large), LOCO fails to identify the important covariates: $X_j \not\perp\!\!\!\perp Y \mid X_{-j}$
  - Suprious correlation: in Demo 2, when null covariates are strongly correlated with important covariates, LOCO fails: $X_j \perp\!\!\!\perp Y \mid X_{-j}$

Global measure of Variable Importance

Reference

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman, *Distribution-free predictive inference for regression*, Journal of the American Statistical Association **113** (2018), no. 523, 1094–1111.