Overview of the Simulation Function

Parameters of interest:

- k : Number of studies
- n_k : Number of observations within each study
- p : Number of features
- $\boldsymbol{\mu_x}$: a matrix $k \times p$ that contains mean value for covariates to be generated from multi-variate distribution
- $\boldsymbol{\Sigma_x}$: a $p \times p$ variance-covariance matrix assigned to generate covariates from multi-variate distribution
- m : Number of mixing component
- $\boldsymbol{\pi}$: an $m \times 1$ vector that contain the probability value of each component to be selected (sum of the component will be 1)
- $\boldsymbol{\mu_\beta}$: an $m \times p$ matrix that contains mean value for coefficients to be generated from specified distribution
- $\boldsymbol{\Sigma_\beta}$: an $m \times p$ variance-covariance matrix to generate coeffcients

Learners in the Analysis

- Elastic Net (`glmnet`)
- Neural Network (`monmlp`)
- Gradient Boosting (`gbm`)
- Bagged CART (`treebag`)
- Random Forest (`ranger`)

Weights Used in Comparison - Overview

Consider K training datasets and V validation datasets, within which we have an outcome Y_k of size n_k that corresponded sets of predictors $\boldsymbol{X_k}$ of dimension $n_k \times p, k = 1, 2, ..., K + V$.

Note

Let the prediction function trained on dataset k using learner ℓ to be $\hat{Y}_k^\ell(\boldsymbol{X})$. A linear Single-Study Learners with weights w_k^ℓ has the form

$$\hat{Y}^\ell(\boldsymbol{X_k}) = \sum_{k=1}^{K} w_k^\ell \hat{Y}_k^\ell(\boldsymbol{X_k}) \tag{1}$$

The expected performance in the validation dataset for learner ℓ can be shown as:

$$\mathcal{U} = \sum_{k=K+1}^{V} U(\hat{Y}^\ell(\boldsymbol{X_k}) - Y_k) \tag{2}$$

Weights Used in Comparison -Stacking Regression

- Minimizing the least squares distance between $\hat{Y}(\boldsymbol{X})$ and $E(Y|\boldsymbol{X})$ in a in a hypothetical $(K+i)$th study, for $i = 1, ..., V$
- Let $\boldsymbol{Y} = (\boldsymbol{Y_1'}, ..., \boldsymbol{Y_K'})'_{n\times 1}$ be the vector of the true outcome in training datasets
- Let $\tilde{\boldsymbol{Y}} = [\hat{\boldsymbol{Y}}_{\boldsymbol{1}}, ..., \hat{\boldsymbol{Y}}_{\boldsymbol{K}}]_{n\times K}$ be the matrix of predicted values that are trained by K subsets of the training samples ($n = \sum_{k=1}^{K} n_k$).
- Assuming $\tilde{\boldsymbol{Y}}$ is invertible, the optimal weights can be shown as:

$$\boldsymbol{w_s} = (\tilde{\boldsymbol{Y}}' \tilde{\boldsymbol{Y}})^{-1} \tilde{\boldsymbol{Y}}' \boldsymbol{Y} \tag{3}$$

  where $\boldsymbol{w_s}$ is a $K-$dimensional vector of coefficients (computed via nonnegative least squares as given by the `nnls` R package)
- Let $\boldsymbol{Y_v^*} = [\boldsymbol{Y_{v,1}^*}, ..., \boldsymbol{Y_{v,K}^*}]_{n_v\times K}$ be the matrix of the predicted value for the v_{th} validation sets generated from K training sets
- The final predicted value given by stacking weights for the v_{th} validation dataset will be:

$$\hat{\boldsymbol{Y}}_{\boldsymbol{v,s}} = \boldsymbol{Y_v^*} \boldsymbol{w_s} \tag{4}$$

Weights Used in Comparison - Stacking with Zero

After obtaining the predicted outcome matrix $\tilde{\boldsymbol{Y}}_{n\times K}$, set all predicted values that use in-sample prediction to be 0, where the $\tilde{\boldsymbol{Y}}_{n\times K}$ now becomes

$$\boldsymbol{T} = \begin{bmatrix} 0 & \hat{Y}_{21} & \hat{Y}_{31} & \cdots & \hat{Y}_{K1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \hat{Y}_{2n_1} & \hat{Y}_{3n_1} & \cdots & \hat{Y}_{Kn_1} \\ \hat{Y}_{1(n_1+1)} & 0 & \hat{Y}_{3(n_1+1)} & \cdots & \hat{Y}_{K(n_1+1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{Y}_{1(n_1+n_2)} & 0 & \hat{Y}_{3(n_1+n_2)} & \cdots & \hat{Y}_{K(n_1+n_2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{Y}_{1n} & \hat{Y}_{2n} & \hat{Y}_{3n} & \cdots & 0 \end{bmatrix}_{n\times K}$$

With the new matrix $\boldsymbol{T}$, and assuming $\boldsymbol{T}$ is invertible, the optimal weights can be shown as the $K-$dimensional vector of coefficients to be

$$\boldsymbol{w}_z^* = (\boldsymbol{T}'\boldsymbol{T})^{-1}\boldsymbol{T}'\boldsymbol{Y} \tag{5}$$

Weights Used in Comparison - Stacking with Zero

Re-scale the coefficients

For $\boldsymbol{w}_z^* = [w_{z1}^*, ..., w_{zk}^*]'$, re-scale the coefficients based on study-specific sample size

$$w_{zk} = (1 - \frac{n_k}{n}) w_{zk}^* \qquad k = 1, 2, ..., K$$

Let $\boldsymbol{w}_z = [w_1, w_2, ..., w_k]'$
The final predicted value now has the form:

$$\hat{\boldsymbol{Y}}_{\boldsymbol{v},\boldsymbol{z}} = \boldsymbol{Y}_{\boldsymbol{v}}^* \boldsymbol{w}_{\boldsymbol{z}} \tag{6}$$

Weights Used in Comparison - Stacking with Intercept

Let $\hat{\boldsymbol{Y}}* = [\mathbf{1}, -\mathbf{1}, \tilde{\boldsymbol{Y}}] = [\mathbf{1}, -\mathbf{1}, \hat{\boldsymbol{Y}}_{\mathbf{1}}, ..., \hat{\boldsymbol{Y}}_{\boldsymbol{K}}]$
Assuming $\hat{\boldsymbol{Y}}*$ is invertible, this $(K+2)$-dimensional vector of coefficients will be

$$\boldsymbol{w_I} = (\hat{\boldsymbol{Y}}*' \hat{\boldsymbol{Y}}*)^{-1} \hat{\boldsymbol{Y}}*' \boldsymbol{Y} = [w_{I0}, w_{I1}, w_{I2}, ..., w_{I(K+2)}]$$

where

$$\begin{aligned} \hat{Y}_j &= (w_{I0} - w_{I1}) + w_{I2} \hat{Y}_{1j} + w_{I3} \hat{Y}_{2j} + ... + w_{I(K+1)} \hat{Y}_{Kj} \\ &= (w_{I0} - w_{I1}) + \sum_{q=2}^{K+1} w_{Iq} \hat{Y}_{(q-1)j} \quad j = 1, ..., n \end{aligned} \tag{7}$$

Weights Used in Comparison - General Weighting methods

- **Merged:** Combine all the datasets into one large group and train a single prediction function using selected learners
- **Simple Average:** Take the mean value of predicted outcome from K studies as the final prediction, $w = \frac{1}{K}$
- **Elastic Net:** Motivated by the concept of regularized regression to regress $\tilde{\boldsymbol{Y}}$ onto $\boldsymbol{Y}$ with 5-fold cross-validation