Package: prime (via r-universe)

October 22, 2026

Type Package

Title Proportional Inclusive Marginal Effects

Version 0.5.5

Description Implements a model-free variable importance measure based on proportional marginal effects estimated from noisy input-output data. See Cho and Joseph (2026) <doi:10.48550/arXiv.2610.17053> for details.

License MIT + file LICENSE

Encoding UTF-8

URL https://github.com/ChoYoungSeo/prime-importance-r

BugReports https://github.com/ChoYoungSeo/prime-importance-r/issues

Depends R (>= 4.2.0)

Imports FNN, MASS, parallel, Rcpp

LinkingTo Rcpp, RcppArmadillo

Suggests testthat (>= 3.0.0), twinning

Config/testthat/edition 3

NeedsCompilation yes

Author Youngseo Cho [aut, cre, cph], V. Roshan Joseph [aut, cph]

Maintainer Youngseo Cho <josun@gatech.edu>

Repository https://cran.r-universe.dev

Date/Publication 2026-10-22 19:23:43 UTC

RemoteUrl https://github.com/cran/prime

RemoteRef HEAD

RemoteSha 457262bf8021a7690bdf754013b44140fe4f9cf5

Contents

prime-package 2
plot.prime_fit 3
prime 4
prime_screen 6

Index **8**

`prime-package` *PRIME: Proportional Inclusive Marginal Effects*

Description

The **prime** package implements the proportional inclusive marginal effects of Cho and Joseph (2026), which estimates model-free variable importance measures from noisy input-output data. It also contains a screening method that can be used for model-free variable selection.

Details

PRIME estimates explained-variance subset values with local-linear neighborhoods, constructs a BIC-gated ridge-start candidate set, applies paired querywise forward and backward calibration, and evaluates proportional marginal effects (Herin et al. 2024) from noise-adjusted risks on the retained coordinates.

The main interface is `prime()`, which runs coordinate selection, subset-value estimation, and proportional allocation. The supporting interface is `prime_screen()` for coordinate selection and diagnostic traces without the allocation step. Use `summary(fit)` or `summary(screen)` for a concise result summary, and `plot(fit)` to display variable importance.

Exact subset enumeration grows exponentially with the retained dimension. The ambient dimension affects screening, while the selected dimension determines the final allocation cost. Twinning (Vakayil and Joseph 2022) option is provided to reduce the cost with large-scale data. Fitted objects have concise print, summary, and plot methods.

Acknowledgments

This research is supported by a U.S. National Science Foundation grant DMS-2310637.

Author(s)

Youngseo Cho and V. Roshan Joseph

References

Herin, M., Il Idrissi, M., Chabridon, V., and Iooss, B. (2024). Proportional marginal effects for global sensitivity analysis. *SIAM/ASA Journal on Uncertainty Quantification*, 12(2), 667–692. doi:10.1137/22M153032X.

Cho, Y. and Joseph, V. R. (2026). Measuring Variable Importance Using Proportional Inclusive Marginal Effects. *arXiv:2610.17053*. doi:10.48550/arXiv.2610.17053.

Vakayil, A. and Joseph, V. R. (2022). "Data Twinning". Statistical Analysis and Data Mining: The ASA Data Science Journal, 15, 598-610.

See Also

`prime`, `plot.prime_fit`, `prime_screen`

plot.prime_fit *Plot PRIME Variable Importance*

Description

Displays PRIME importance from a fitted result. Retained coordinates are highlighted and excluded coordinates remain visible at zero.

Usage

```
## S3 method for class 'prime_fit'
plot(x, y = NULL, max_display = 20L,
  main = NULL, xlab = "PRIME importance",
  col = "#087F8C", zero_col = "#A5ADB4", ...)
```

Arguments

x	A fitted `prime_fit` object returned by `prime()`, such as `obj` in the example below.
y	Optional argument retained for compatibility with the generic `plot(x, y, ...)` interface; no second input is needed.
max_display	Maximum number of coordinates to display.
main	Optional plot title.
xlab	Horizontal-axis label.
col	Color for retained coordinates.
zero_col	Color for excluded coordinates.
...	Additional graphical parameters passed to `plot.default()`.

Value

The fitted object, invisibly.

See Also

prime

Examples

```
set.seed(5)
X <- matrix(runif(960, -1, 1), nrow = 120)
colnames(X) <- paste0("X", seq_len(ncol(X)))
y <- X[, 1] - X[, 2] + X[, 5]^2 + rnorm(120, sd = 0.08)

obj <- prime(X, y)
plot(obj)
```

`prime` *Proportional Inclusive Marginal Effects*

Description

Estimates Proportional Inclusive Marginal Effects of Cho and Joseph (2026).

Usage

```
prime(X, y, rescale = TRUE, twin_r = NULL, lambda = 1e-6,
  neighbor_eps = 0, query_block_size = 256L)
```

Arguments

`X`	Numeric predictor matrix or data frame with unique column names.
`y`	Finite numeric response vector with length `nrow(X)`.
`rescale`	Logical; standardize predictors before distance calculations.
`twin_r`	Optional integer inverse Data-Twinning split ratio. `NULL` uses every observation as a query location.
`lambda`	Nonnegative slope-only numerical nugget used in local-linear fits.
`neighbor_eps`	Nonnegative relative Euclidean distance tolerance. Zero uses exact indexed search. Positive values enable approximate neighbors; the distance tolerance does not bound allocation or selection error.
`query_block_size`	Positive query batch size. The native engine streams individual neighborhoods and checks for interrupts at these intervals.

Details

The function executes BIC-gated ridge-start screening, query-calibrated forward and backward reduction, and proportional allocation (Herin et al. 2024) on the retained total-effect set function. Please see Cho and Joseph (2026) for details.

Exact allocation requires 2^d subset values and $d2^{d-1}$ projection edges for d retained coordinates. The edge-index capacity is checked before subset evaluation; dimensions 28 and above are rejected by the current 32-bit projection indices. Lower dimensions can still be costly in memory and runtime.

Search builds an index in each evaluated coordinate subset and retains all observations in its neighborhood pool. If observations are equally distant from a query point, exact search orders them by their original row numbers. This tie-breaking rule can choose different neighbors from earlier versions that used the FNN package. Query contributions are retained for at most two subset fits; allocation caches scalar risks. Setting `neighbor_eps > 0` makes approximation explicit in `settings`. The default preserves exact distances and the neighborhood size rule.

For retained dimension d, the local neighborhood size is

$$K_d = \left\lceil \max\{1.5\sqrt{n},\, 3d\} \right\rceil ,$$

subject to the local-fit feasibility cap. The local nugget is applied to slopes only. Every observation is used as a query by default; Data Twinning reduces the outer query average without removing observations from the neighbor pool.

Exact allocation evaluates 2^d subset values and therefore scales exponentially with the retained dimension.

Use `summary(fit)` to see retained coordinates, importance estimates, explained variation, and projection diagnostics.

Value

An object of class `prime_fit` with components:

- `call`: matched function call;
- `effects`: full-length, zero-padded PRIME allocation;
- `reduced_effects`: allocation on retained coordinates;
- `selected` and `selected_names`: retained coordinates;
- `screen`: complete `prime_screen` result;
- `subset_values`: raw and projected subset values with complement and noise-adjusted risks;
- `signal`, `noise_floor`, and `risk_full`: estimated explained variation and full retained-set risk;
- `projection`: monotone-projection diagnostics;
- `settings`: fixed numerical and statistical settings.

If screening retains no coordinate, `effects` is a zero vector and `subset_values` is `NULL`, indicating that a normalized allocation is unavailable.

References

Herin, M., Il Idrissi, M., Chabridon, V., and Iooss, B. (2024). Proportional marginal effects for global sensitivity analysis. *SIAM/ASA Journal on Uncertainty Quantification*, 12(2), 667–692.

Cho, Y. and Joseph, V. R. (2026). Measuring Variable Importance Using Proportional Inclusive Marginal Effects. *arXiv:2610.17053*. doi:10.48550/arXiv.2610.17053.

See Also

`plot.prime_fit`, `prime_screen`

Examples

```
set.seed(5)
X <- matrix(runif(960, -1, 1), nrow = 120)
colnames(X) <- paste0("X", seq_len(ncol(X)))
y <- X[, 1] - X[, 2] + X[, 5]^2 + rnorm(120, sd = 0.08)

fit <- prime(X, y)
fit$selected_names
round(fit$effects, 3)
```

```
summary(fit)
plot(fit)
```

`prime_screen` *Nonparametric variable selection*

Description

This function implements the nonparametric model-free variable selection technique in Cho and Joseph (2026).

Usage

```
prime_screen(X, y, rescale = TRUE, twin_r = NULL, lambda = 1e-6,
  neighbor_eps = 0, query_block_size = 256L)
```

Arguments

`X`	Numeric predictor matrix or data frame with unique column names.
`y`	Finite numeric response vector with length `nrow(X)`.
`rescale`	Logical; standardize predictors before distance calculations.
`twin_r`	Optional integer inverse Data-Twinning split ratio. `NULL` uses every observation as a query location.
`lambda`	Nonnegative slope-only numerical nugget used in local-linear fits.
`neighbor_eps`	Nonnegative relative Euclidean distance tolerance. Zero uses exact indexed search. Positive values enable approximate neighbors; the distance tolerance does not bound allocation or selection error.
`query_block_size`	Positive query batch size. The native engine streams individual neighborhoods and checks for interrupts at these intervals.

Details

It constructs a response-relevant candidate set from a BIC-selected prefix of a standardized ridge ordering and a query-calibrated local-linear forward path, and then applies sequential query-calibrated backward reduction. Please see Cho and Joseph (2026) for details.

Search builds an index in each evaluated coordinate subset and retains all observations in its neighbor pool. If observations are equally distant from a query point, exact search orders them by their original row numbers. This tie-breaking rule can choose different neighbors from earlier versions that used the FNN package. Query contributions are retained for at most two subset fits; allocation caches scalar risks. Setting `neighbor_eps > 0` makes approximation explicit in `settings`. The default preserves exact distances and the neighborhood size rule.

The ridge ordering is searched through $\lceil 1.5 \log\{\max(2, p)\}\rceil$ coordinates, including the intercept-only model, and BIC selects the initial prefix. The local path adds the coordinate with the largest

explained-variance gain when its paired querywise calibrated gain is positive. Backward reduction sequentially removes zero calibrated deletion losses and recomputes all projected neighborhoods after every removal.

Ridge ranking and BIC use a centered unit-variance response internally. The reported BIC omits the response-scale constant shared by all candidate models; ridge-path RSS remains in original response units. Paired comparison scales are computed in response-variance units and returned in original units, avoiding overflow or underflow of their empirical variances under changes of units.

Use `summary(screen)` to see the ridge seed, forward path, retained coordinates, and backward-reduction counts.

Value

An object of class `prime_screen` with components:

- `call`: matched function call;
- `candidates` and `candidate_names`: inclusive forward set;
- `selected` and `selected_names`: backward-cleaned set;
- `forward`: ridge seed, local path, stopping reason, and trace;
- `backward`: retained set and deletion-comparison trace;
- `settings`: neighborhood, query, nugget, and calibration rules.

References

Cho, Y. and Joseph, V. R. (2026). Measuring Variable Importance Using Proportional Inclusive Marginal Effects. *arXiv:2610.17053*. doi:10.48550/arXiv.2610.17053.

See Also

`prime`

Examples

```
set.seed(5)
X <- matrix(runif(960, -1, 1), nrow = 120)
colnames(X) <- paste0("X", seq_len(ncol(X)))
y <- X[, 1] - X[, 2] + X[, 5]^2 + rnorm(120, sd = 0.08)

screen <- prime_screen(X, y)
screen$selected_names
summary(screen)
screen$forward$trace
screen$backward$trace
```

Index

plot.prime_fit, *2*, 3, *5*
prime, *2*, *3*, 4, *7*
prime-package, 2
prime_screen, *2*, *5*, 6