## Package ‘FLORAL’

May 11, 2023

**Type** Package  
**Title** Fit Log-Ratio Lasso Regression for Compositional Data  
**Version** 0.1.0  
**Date** 2023-05-09  
**Description** Log-ratio Lasso regression for continuous, binary, and survival outcomes with compositional features. See Fei and others (2023) <doi:10.1101/2023.05.02.538599>.  
**License** GPL (>= 3)  
**URL** https://vdblab.github.io/FLORAL/  
**BugReports** https://github.com/vdblab/FLORAL/issues  
**Depends** R (>= 3.5.0)  
**biocViews**  
**Imports** Rcpp (>= 1.0.9), stats, survival, ggplot2, survcomp, reshape, dplyr, glmnet, caret, grDevices, utils, mvtnorm  
**LinkingTo** Rcpp, RcppArmadillo, RcppProgress  
**RoxygenNote** 7.2.3  
**Encoding** UTF-8  
**Suggests** covr, knitr, rmarkdown, spelling, testthat (>= 3.0.0), patchwork  
**Language** en-US  
**Config/testthat/edition** 3  
**VignetteBuilder** knitr  
**NeedsCompilation** yes  
**Author** Teng Fei [aut, cre, cph] (<https://orcid.org/0000-0001-7888-1715>), Tyler Funnell [aut] (<https://orcid.org/0000-0003-1612-5644>), Nicholas Waters [aut] (<https://orcid.org/0000-0002-9035-2143>), Sandeep Raj [aut] (<https://orcid.org/0000-0003-4629-0528>)  
**Maintainer** Teng Fei <feit1@mskcc.org>  
**Repository** CRAN  
**Date/Publication** 2023-05-11 08:50:22 UTC
**R topics documented:**

- FLORAL .................................................. 2
- simu ..................................................... 4

**Index** .................................................. 6

---

**FLORAL**

*Fit Log-ratio lasso regression for compositional covariates*

**Description**

Conduct log-ratio lasso regression for continuous, binary and survival outcomes.

**Usage**

```r
FLORAL(
  x,
  y,
  family = "gaussian",
  longitudinal = FALSE,
  id = NULL,
  tobs = NULL,
  failcode = NULL,
  length.lambda = 100,
  lambda.min.ratio = NULL,
  mu = 1,
  ncv = 5,
  intercept = FALSE,
  foldid = NULL,
  step2 = TRUE,
  progress = TRUE,
  plot = TRUE
)
```

**Arguments**

- `x`: Count data matrix, where rows specify subjects and columns specify features. If `x` contains longitudinal data, the rows must be sorted in the same order of the subject IDs used in `y`.
- `y`: Outcome. For a continuous or binary outcome, `y` is a vector. For survival outcome, `y` is a Surv object.
- `family`: Available options are gaussian, binomial, cox, finegray.
- `longitudinal`: TRUE or FALSE, indicating whether longitudinal data matrix is specified for input `x`. (Still under development. Please use with caution)
- `id`: If `longitudinal` is TRUE, `id` specifies subject IDs corresponding to the rows of input `x`.
**tobs**  If `longitudinal` is TRUE, tobs specifies time points corresponding to the rows of input x.

**failcode**  If `family = finegray`, failcode specifies the failure type of interest. This must be a positive integer.

**length.lambda**  Number of penalty parameters used in the path

**lambda.min.ratio**  Ratio between the minimum and maximum choice of lambda. Default is NULL, where the ratio is chosen as 1e-2.

**mu**  Value of penalty for the augmented Lagrangian

**ncv**  Number of cross-validation runs. Use NULL if cross-validation is not wanted.

**intercept**  TRUE or FALSE, indicating whether an intercept should be estimated.

**foldid**  A vector of fold indicator. Default is NULL.

**step2**  TRUE or FALSE, indicating whether a second-stage feature selection for specific ratios should be performed for the features selected by the main lasso algorithm. Will only be performed if cross validation is enabled.

**progress**  TRUE or FALSE, indicating whether printing progress bar as the algorithm runs.

**plot**  TRUE or FALSE, indicating whether returning plots of model fitting.

**Value**

A list with path-specific estimates (beta), path (lambda), and others. Details can be found in README.md.

**Author(s)**

Teng Fei. Email: feit1@mskcc.org

**References**


**Examples**

```r
set.seed(23420)

# Continuous outcome
dat <- simu(n=50, p=30, model="linear")
fit <- FLORAL(dat$xcount, dat$y, family="gaussian", progress=FALSE, step2=TRUE)
```
Simulate data following log-ratio model

Description
Simulate a dataset from log-ratio model.

Usage
```r
simu(
  n = 100,
  p = 200,
  model = "linear",
  weak = 4,
  strong = 6,
  weaksize = 0.125,
  strongsize = 0.25,
  pct.sparsity = 0.5,
  rho = 0,
  intercept = FALSE
)
```

Arguments

- `n`: An integer of sample size
- `p`: An integer of number of features (taxa).
- `model`: Type of models associated with outcome variable, can be "linear", "binomial", "cox", or "finegray".
- `weak`: Number of features with weak effect size.
- `strong`: Number of features with strong effect size.
- `weaksize`: Actual effect size for weak effect size. Must be positive.
- `strongsize`: Actual effect size for strong effect size. Must be positive.
- `pct.sparsity`: Percentage of zero counts for each sample.
- `rho`: Parameter controlling the correlated structure between taxa. Ranges between 0 and 1.
- `intercept`: Boolean. If TRUE, then a random intercept will be generated in the model. Only works for linear or binomial models.

Value
A list with simulated count matrix `xcount`, log1p-transformed count matrix `x`, outcome (continuous `y`, continuous centered `y0`, binary `y`, or survival `t`, `d`), true coefficient vector `beta`, list of non-zero features `idx`, value of intercept `intercept` (if applicable).
`simu`

**Author(s)**

Teng Fei. Email: feit1@mskcc.org

**Examples**

```r
set.seed(23420)
dat <- simu(n=50,p=30,model="linear")
```
Index

FLORAL, 2

simu, 4