

Package ‘JOPT’

May 20, 2025

Type Package

Title J-optimal Subdata Selection

Version 0.1.0

Author Alvaro Cia-Mina <aciamina@unav.es>

Maintainer Alvaro Cia-Mina <aciamina@unav.es>

Description Implements the J-optimal Subdata Selection for linear regression.

License GPL

Encoding UTF-8

LazyData true

Imports Matrix,
latex2exp

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

Contents

create_model_function	1
jseq	2
run_efficiency_comparison_example	4
Index	5

create_model_function	<i>Create a Model Function from Expressions</i>
-----------------------	-------------------------------------------------

Description

This function takes a vector of mathematical expressions (as character strings) and generates a function that, given an input vector ‘x’, computes the specified expressions and returns the results as a column matrix.

Usage

```
create_model_function(expressions)
```

Arguments

expressions A character vector of mathematical expressions to define the model. Each expression should be valid R code and reference elements of 'x' (e.g., "x[1]", "x[2]^2").

Value

A function that takes an input vector 'x' and evaluates the expressions in 'expressions', returning a column matrix of the results.

Examples

```
# Define the model expressions
expressions <- c("1", "x[1]", "x[1]*x[2]^2")

# Create the model function
model_function <- create_model_function(expressions)

# Test the model function with an input vector
input_vector <- c(2, 3) # x[1] = 2, x[2] = 3
result <- model_function(input_vector)
print(result)
```

jseq

*J-Optimal Subsample Selection***Description**

This function implements the J-optimal subsample selection method, as described in Cia-Mina et al. (2025). It takes a dataset of covariates x , a subsample proportion α (between 0 and 1), and a vector defining a regression model. Additional parameters can be specified to control the selection process.

Usage

```
jseq(
  x,
  alpha,
  model_vec,
  k0 = 5 * length(model_vec),
  q = 5/8,
  gamma = 1/10,
  eps1 = 0
)
```

Arguments

x A dataset (data frame) containing the covariates for the regression model.

alpha A numeric value between 0 and 1 specifying the subsample proportion.

<code>model_vec</code>	A character vector defining the regression model. Each element should represent a term in the model, written as an expression involving <code>x</code> . For example, <code>"1"</code> for the intercept, <code>"x[1]"</code> for the first covariate, or <code>"x[1]*x[2]^2"</code> for an interaction term.
<code>k0</code>	An integer specifying the initial size of the subsample. Defaults to <code>5*length(model_vec)</code> .
<code>q</code>	A numeric value between 0.5 and 1. Defaults to 5/8.
<code>gamma</code>	A numeric value between 0 and <code>q-0.5</code> . Defaults to 1/10.
<code>eps1</code>	A small positive value. Defaults to 0.

Details

The J-optimal subsample selection algorithm selects a subset of observations from the dataset `x` that optimizes the statistical efficiency of the model defined by `model_vec`. For technical details, refer to Cia-Mina et al. (2025).

Value

A list with the following components:

<code>x_j</code>	A subsample of <code>x</code> containing the selected observations (rows) according to J-optimality.
<code>idx</code>	A vector of indices corresponding to the selected rows of <code>x</code> .

Examples

```
# Example 1: Bivariate regression
set.seed(123)
x1 <- runif(1e3, min = -1, max = 1)
x2 <- runif(1e3, min = -1, max = 1)
x <- data.frame(x1 = x1, x2 = x2)
model_vec <- c("1", "x[1]", "x[2]", "x[1]*x[2]", "x[1]^2", "x[2]^2")
result <- jseq(x, 0.3, model_vec)

# Plot the full dataset and the selected subsample
plot(x$x1, x$x2, col = "black", pch = 16, cex = 0.7, xlab = "x1", ylab = "x2")
points(result$x_j$x1, result$x_j$x2, col = "red", pch = 16, cex = 0.7)
title(main = "J-OPT", line = 1)

# Example 2: Univariate regression
set.seed(123)
x <- data.frame(x = rnorm(1e4))
model_vec <- c("1", "x[1]", "x[1]^2")
result <- jseq(x, 0.3, model_vec)

# Plot the density of the selected subsample
plot(density(result$x_j$x, bw = 2 / 100, kernel = "epanechnikov"),
     ylab = "", lwd = 1.7, xlim = c(-3.5, 3.5), main = "", xlab = "")
result$idx
```

`run_efficiency_comparison_example`*Run Efficiency Comparison Example*

Description

This function provides an example of running an efficiency comparison using the package's capabilities. The function contains the example code previously executed at the top level.

Usage

```
run_efficiency_comparison_example()
```

Examples

```
# To run the example:  
run_efficiency_comparison_example()
```

Index

`create_model_function`, [1](#)

`jseq`, [2](#)

`run_efficiency_comparison_example`, [4](#)