Package ‘FGSG’

February 25, 2015

**Title** Feature Grouping and Selection Over an Undirected Graph

**Version** 1.0.2

**Author** Xiaotong Shen, Yiwen Sun, Julie Langou

**Maintainer** Yiwen Sun <sunxx847@umn.edu>

**Description** Implement algorithms for feature grouping and selection over an undirected graph, solves problems like graph fused lasso, graph OSCAR and so on.

**License** GPL-2

**Note** The header file blaswrap.h, f2c.h and fgsg.h are from the VisualStudio library created by Julie Langou.

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2015-02-25 10:32:12

**R topics documented:**

- FGSG-package ......................................................... 2
- gflasso ............................................................... 2
- goscar ............................................................... 4
- ncFGS ............................................................... 5
- ncTF ................................................................. 6
- ncTFGS ............................................................. 7
- ncTL ................................................................. 9
- ncTLF .............................................................. 10

**Index** 12
FGSG-package

Feature grouping and selection over an undirected graph

Description

FGSG package implements algorithms for feature grouping and selection over an undirected graph. This package can work under Linux environment, but is not guaranteed under Windows.

Details

<table>
<thead>
<tr>
<th>Package:</th>
<th>FGSG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type:</td>
<td>Package</td>
</tr>
<tr>
<td>Version:</td>
<td>1.0</td>
</tr>
<tr>
<td>Date:</td>
<td>2014-04-10</td>
</tr>
<tr>
<td>License:</td>
<td>GPL-2</td>
</tr>
</tbody>
</table>

Author(s)

Xiaotong Shen, Yiwen Sun
Maintainer: Yiwen Sun <sunxx847@umn.edu>

References


gflasso

Graph Fused Lasso (FGSG)

Description

Given $A = a_1, \ldots, a_n$, the response $y$, and a set of edges $E$, this function aims to solve

$$
\min \frac{1}{2} ||Ax - y||^2 + \lambda_1 ||x||_1 + \lambda_2 \sum_{(i,j) \in E} w(i,j)|x_i - r(i,j)x_j| \n$$

where $w(i,j)$ is the weight of the edge $(i, j)$, and $r(i,j)$ is the sign of the correlation between features $a_i$ and $a_j$. The weight and sign can be specified in $Rwt$: $w = |Rwt|$, and $r = sign(Rwt)$. 
Usage

gflasso(A, y, tp, s1, s2, RmaxIter = 100,
RaMaxIter = 1000, Rrho = 5, Rtau = 0.15,
Rwt = rep(1, length(tp)), Rtol = 0.001,
RaTol = 0.001, x0 = rep(0, ncol(A)))

Arguments

A A The data matrix of size $n \times p$, each row corresponds to one sample.
y y The response vector of length n.
tp tp The edges vector of length $2g$ (eg. (1,2,3,4) means an edge between 1 and 2, and an edge between 3 and 4, $g=2$ is the number of edges).
s1 s1 The $l_1$ regularization parameter, $s1 = 0$.
s2 s2 The grouping penalty parameter, $s2 = 0$.
RmaxIter RmaxIter The maximum number of iterations in DC programming (default 100).
RaMaxIter RaMaxIter The maximum number of iterations in ADMM (default 1000).
Rrho Rrho The dual update length of ADMM (default 5).
Rtau Rtau The tuning parameter for non-convex penalty (default 0.15).
Rwt Rwt The weight and signs of edges (default rep(1,g)).
Rtol Rtol The tolerance for convergence in DC programming (default 1e-3).
RaTol RaTol The tolerance for convergence in ADMM (default 1e-3).
x0 x0 The returned weight vector (default rep(0,p)).

Value

Returned value x0 is the solution to the optimization problem.

Author(s)

Yiwen Sun

References


Examples

A<-matrix(rnorm(25),5,5)
y<-rnorm(5)
.tp<-c(1,2,2,3,3,4,4,5)
gflasso(A,y,tp,0,0)
goscar  

Graph OSCAR (FGSG)

Description

Given $A = a_1, \ldots, a_n$, the response $y$, and a set of edges $E$, this function aims to solve

$$\min \frac{1}{2}||Ax - y||^2 + \lambda_1||x||_1 + \lambda_2 \sum_{(i,j) \in E} w(i,j) \max|x_i|, |x_j|$$

Usage

```r
goscar(A, y, tp, s1, s2, RmaxIter = 100,
        RmaxIter = 1000, Rrho = 5, Rtau = 0.15,
        Rwt = rep(1, length(tp)), Rtol = 0.001,
        RaTol = 0.001, x0 = rep(0, ncol(A)))
```

Arguments

- **A**  
The data matrix of size $n \times p$, each row corresponds to one sample.
- **y**  
The response vector of length $n$.
- **tp**  
The edges vector of length $2g$ (e.g., (1,2,3,4) means an edge between 1 and 2, and an edge between 3 and 4, $g=2$ is the number of edges).
- **s1**  
The $l_1$ regularization parameter, $s1 >= 0$.
- **s2**  
The grouping penalty parameter, $s2 >= 0$.
- **RmaxIter**  
The maximum number of iterations in DC programming (default 100).
- **RaMaxIter**  
The maximum number of iterations in ADMM (default 1000).
- **Rrho**  
The dual update length of ADMM (default 5).
- **Rtau**  
The tuning parameter for non-convex penalty (default 0.15).
- **Rwt**  
The weight and signs of edges (default rep(1,g)).
- **Rtol**  
The tolerance for convergence in DC programming (default 1e-3).
- **RaTol**  
The tolerance for convergence in ADMM (default 1e-3).
- **x0**  
The returned weight vector (default rep(0,p)).

Value

Returned value $x0$ is the solution to the optimization problem.

Author(s)

Yiwen Sun
References


Examples

```r
A <- matrix(rnorm(25), 5, 5)
y <- rnorm(5)
 tp <- c(1, 2, 3, 4, 5)
goscar(A, y, tp, 0, 0)
```

ncFGS

Non Convex Feature Grouping and Selection (FGSG)

Description

Given $A = a_1, \ldots, a_n$, the response $y$, and a set of edges $E$, this function aims to solves

$$
\text{min} \frac{1}{2} ||Ax - y||^2 + \lambda_1 ||x||_1 + \lambda_2 \sum_{(i, j) \in E} w(i, j)||x_i - x_j||
$$

Usage

```r
cncFGS(A, y, tp, s1, s2, RmaxIter = 100,
RaMaxIter = 1000, Rrho = 5, Rtau = 0.15,
Rwt = rep(1, length(tp)), Rtol = 0.001,
RaTol = 0.001, x0 = rep(0, ncol(A)))
```

Arguments

- **A** - The data matrix of size $n \times p$, each row corresponds to one sample.
- **y** - The response vector of length $n$.
- **tp** - The edges vector of length $2g$ (e.g., (1, 2, 3, 4) means an edge between 1 and 2, and an edge between 3 and 4, $g=2$ is the number of edges).
- **s1** - The $l_1$ regularization parameter, $s1 >= 0$.
- **s2** - The grouping penalty parameter, $s2 >= 0$.
- **RmaxIter** - The maximum number of iterations in DC programming (default 100).
- **RaMaxIter** - The maximum number of iterations in ADMM (default 1000).
- **Rrho** - The dual update length of ADMM (default 5).
- **Rtau** - The tuning parameter for non-convex penalty (default 0.15).
- **Rwt** - The weight and signs of edges (default rep(1, g)).
- **Rtol** - The tolerance for convergence in DC programming (default 1e-3).
- **RaTol** - The tolerance for convergence in ADMM (default 1e-3).
- **x0** - The returned weight vector (default rep(0, p)).
Value

Returned value $x_0$ is the solution to the optimization problem.

Author(s)

Yiwen Sun

References


Examples

```r
A <- matrix(rnorm(25), 5, 5)
y <- rnorm(5)
.tp <- c(1, 2, 2, 3, 3, 4, 4, 5)
ncFGS(A, y, tp, 0)
```

ncTF  

**Non Convex Truncated Fused Feature Grouping and Selection (FGSG)**

Description

Given $A = a_1, \ldots, a_n$, the response $y$, and a set of edges $E$, this function aims to solve

$$
\min \frac{1}{2} ||Ax - y||^2 + \lambda_1 ||x||_1 + \lambda_2 \sum_{(i,j) \in E} w(i,j) J_\tau(x_i - r_{(i,j)}x_j)
$$

where $J_\tau(x) = \min(x/\tau, 1)$ is a surrogate of the $L0$ norm, $w(i, j)$ is the weight of the edge $(i, j)$, and $r_{(i,j)}$ is the sign of the correlation between features $a_i$ and $a_j$. The weight and sign can be specified in $Rwt$: $w = |Rwt|$, and $r = \text{sign}(Rwt)$.

Usage

```r
ncTF(A, y, tp, s1, s2, RmaxIter = 100,
RmaxIter = 1000, Rh = 5, Rtau = 0.15,
Rwt = rep(1, length(tp)), Rtol = 0.001,
RaTol = 0.001, xo = rep(0, ncol(A)))
```

Arguments

- **A**  
The data matrix of size $n \times p$, each row corresponds to one sample.

- **y**  
The response vector of length $n$.

- **tp**  
The edges vector of length $2g$ (e.g., (1,2,3,4) means an edge between 1 and 2, and an edge between 3 and 4, $g=2$ is the number of edges).

- **s1**  
The $l_1$ regularization parameter, $s1 >= 0$. 

- **RmaxIter**  
The maximum number of iterations.

- **Rh**  
The relaxation parameter.

- **Rtau**  
The threshold parameter.

- **Rwt**  
The weight vector.

- **Rtol**  
The relative tolerance parameter.

- **RaTol**  
The absolute tolerance parameter.

- **xo**  
The initial value for the optimization.
s2  s2 Tge grouping penatly parameter, s2 > = 0.
RmaxIter  RmaxIter The maximum number of iterations in DC programming (default 100).
RaMaxIter  RaMaxIter The maximum number of iterations in ADMM (default 1000).
Rrho  Rrho The dual update length ofor ADMM (default 5).
Rtau  Rtau The tuning parameter for non-convex penalty (default 0.15).
Rwt  Rwt The weight and signs of edges (default rep(1,g)).
Rtol  Rtol The tolerance for convergence in DC programming (default 1e-3).
RaTol  RaTol The tolerance for convergence in ADMM (default 1e-3).
x0  x0 The returned weight vector (default rep(0,p)).

Value

Returned value x0 is the solution to the optimizaiton problem.

Author(s)

Yiwen Sun

References


Examples

A<-matrix(rnorm(25),5,5)
y<-rnorm(5)
tp<-c(1,2,2,3,3,4,4,5)
ncTF(A,y,tp,0,0)

tcTFS
Non Convex Truncated Feature Grouping and Selection (FGSG)

Description

Given $A = a_1, \ldots, a_n$, the response $y$, and a set of edges $E$, this function aims to solves

$$
\min \left\{ 1/2 \|Ax - y\|^2 + \lambda_1 \sum_i \left( J_\tau(|x_i|) \right) + \lambda_2 \sum_{(i,j) \in E} w(i,j)J_\tau(|x_i| - |x_j|) \right\}
$$

where $J_\tau(x) = \min(x/\tau, 1)$ is a surrogate of the L0 norm.
Usage
ncTFGS(A, y, tp, s1, s2, RmaxIter = 100,
RmaxIter = 1000, Rho = 5, Rtau = 0.15,
Rwt = rep(1, length(tp)), Rtol = 0.001,
Rtol = 0.001, x0 = rep(0, ncol(A)))

Arguments
A A The data matrix of size \( n \times p \), each row corresponds to one sample.
y y The response vector of length \( n \).
tp tp The edges vector of length \( 2g \) (eg. (1,2,3,4) means an edge between 1 and 2, and an edge between 3 and 4, \( g \) is the number of edges).
s1 s1 The \( l_1 \) regularization parameter, \( s_1 \geq 0 \).
s2 s2 The grouping penalty parameter, \( s_2 \geq 0 \).
RmaxIter RmaxIter The maximum number of iterations in DC programming (default 100).
RmaxIter RmaxIter The maximum number of iterations in ADMM (default 1000).
Rho Rho The dual update length of ADMM (default 5).
Rtau Rtau The tuning parameter for the non-convex penalty (default 0.15).
Rwt Rwt The weight and signs of edges (default rep(1, g)).
Rtol Rtol The tolerance for convergence in DC programming (default 1e-3).
Rtol Rtol The tolerance for convergence in ADMM (default 1e-3).
x0 x0 The returned weight vector (default rep(0, p)).

Value
Returned value \( x_0 \) is the solution to the optimization problem.

Author(s)
Yiwen Sun

References

Examples
A<-matrix(rnorm(25),5,5)
y<-rnorm(5)
.tp<-c(1,2,3,4,5)
ncTFGS(A,y,tp,0,0)
Non Convex Truncated L1 Feature Grouping and Selection (FGSG)

Description

Given \( A = a_1, \ldots, a_n \), the response \( y \), and a set of edges \( E \), this function aims to solve
\[
\min \frac{1}{2} ||Ax - y||^2 + \lambda_1 \sum_{i} (J_{\tau}(|x_i|)) + \lambda_2 \sum_{(i,j) \in E} w(i,j)|x_i - r_{(i,j)}x_j|
\]

where \( J_{\tau}(x) = \min(x/\tau, 1) \) is a surrogate of the L0 norm, \( w(i,j) \) is the weight of the edge \( (i,j) \), and \( r_{(i,j)} \) is the sign of the correlation between features \( a_i \) and \( a_j \). The weight and sign can be specified in \( Rwt: w = |Rwt| \), and \( r = \text{sign}(Rwt) \).

Usage

\[
\text{ncTL}(A, y, tp, s1, s2, RmaxIter = 100, \\
RaMaxIter = 1000, Rrho = 5, Rtau = 0.15, \\
Rwt = \text{rep}(1, \text{length(tp)}), Rtol = 0.001, \\
RaTol = 0.001, x0 = \text{rep}(0, \text{ncol(A)}))
\]

Arguments

- \( A \): The data matrix of size \( n \times p \), each row corresponds to one sample.
- \( y \): The response vector of length \( n \).
- \( tp \): The edges vector of length \( 2*g \) (eg. \( (1,2,3,4) \) means an edge between 1 and 2, and an edge between 3 and 4, \( g=2 \) is the number of edges).
- \( s1 \): The \( l_1 \) regularization parameter, \( s1 \geq 0 \).
- \( s2 \): The grouping penalty parameter, \( s2 \geq 0 \).
- \( RmaxIter \): The maximum number of iterations in DC programming (default 100).
- \( RaMaxIter \): The maximum number of iterations in ADMM (default 1000).
- \( Rrho \): The dual update length of ADMM (default 5).
- \( Rtau \): The tuning parameter for non-convex penalty (default 0.15).
- \( Rwt \): The weight and signs of edges (default \( \text{rep}(1,g) \)).
- \( Rtol \): The tolerance for convergence in DC programming (default \( 1e-3 \)).
- \( RaTol \): The tolerance for convergence in ADMM (default \( 1e-3 \)).
- \( x0 \): The returned weight vector (default \( \text{rep}(0,p) \)).

Value

Returned value \( x0 \) is the solution to the optimization problem.
Non Convex Truncated L1 and Fused Feature Grouping and Selection (FGSG)

**Description**

Given $A = a_1, \ldots, a_n$, the response $y$, and a set of edges $E$, this function aims to solve

$$
\min \frac{1}{2} ||Ax - y||^2 + \lambda_1 \sum_i J_\tau(|x_i|) + \lambda_2 \sum_{(i,j) \in E} w(i,j) J_\tau(|x_i - r(i,j)x_j|)
$$

where $J_\tau(x) = \min(x/\tau, 1)$ is a surrogate of the L0 norm, $w(i,j)$ is the weight of the edge $(i,j)$, and $r(i,j)$ is the sign of the correlation between features $a_i$ and $a_j$. The weight and sign can be specified in $Rwt$: $w = |Rwt|$, and $r = \text{sign}(Rwt)$.

**Usage**

```r
ncTLF(A, y, tp, s1, s2, RmaxIter = 100,
RaMaxIter = 1000, Rrho = 5, Rtol = 0.15,
Rwt = rep(1, length(tp)), Rtol = 0.001,
RaTol = 0.001, x0 = rep(0, ncol(A)))
```

**Arguments**

- **A** The data matrix of size $n \times p$, each row corresponds to one sample.
- **y** The response vector of length $n$.
- **tp** The edges vector of length $2g$ (e.g., $(1,2,3,4)$ means an edge between 1 and 2, and an edge between 3 and 4, $g=2$ is the number of edges).
- **s1** The $l_1$ regularization parameter, $s1 >= 0$.
- **s2** The grouping penalty parameter, $s2 >= 0$. 
ncTLF

RmaxIter  RmaxIter The maximum number of iterations in DC programming (default 100).
RaMaxIter  RaMaxIter The maximum number of iterations in ADMM (default 1000).
Rrho  Rrho The dual update length of or ADMM (default 5).
Rtau  Rtau The tuning parameter for non-convex penalty (default 0.15).
Rwt  Rwt The weight and signs of edges (default rep(1,g)).
Rtol  Rtol The tolerance for convergence in DC programming (default 1e-3).
RaTol  RaTol The tolerance for convergence in ADMM (default 1e-3).
x0  x0 The returned weight vector (default rep(0,p)).

Value

Returned value x0 is the solution to the optimization problem.

Author(s)

Yiwen Sun

References


Examples

A<-matrix(rnorm(25),5,5)
y<-rnorm(5)
tp<-c(1,2,3,3,4,4,5)
ncTLF(A,y,tp,0,0)
Index

FGSG (FGSG-package), 2
FGSG-package, 2

gflasso, 2
goscar, 4

ncFGS, 5
ncTF, 6
ncTFGS, 7
ncTL, 9
ncTLF, 10