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 enviornment, but is not guaranteed under Windows.

Details

Package:	FGSG
Type:	Package
Version:	1.0
Date:	2014-04-10
License:	GPL-2

Author(s)

Xiaotong Shen, Yiwen Sun

Maintainer: Yiwen Sun <sunxx847@umn.edu>

References

[1]S.Kim and E.Xing. Statisctical estimation of correlated genome associations to a quantitative trait network. PLoS genetics, 5(8):e1000587, 2009

[2]S.Yang, L.Yuan, Y.Lai, X.Shen, P.Wonka, and J.Ye. Feature grouping and selection over an undirected graph. KDD, 2012.

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 solves

$$min 1/2||Ax - y||^2 + \lambda_1||x||_1 + \lambda_2 \sum_{(i,j)inE} w_(i,j)|x_i - r_(i,j)x_j|$$

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

S.Kim and E.Xing. Statisctical estimation of correlated genome associations to a quantitative trait network. PLoS genetics, 5(8):e1000587, 2009

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 solves

$$min 1/2||Ax - y||^2 + \lambda_1||x||_1 + \lambda_2 \sum_{(i,j)inE} w_(i,j)max|x_i|, |x_j|$$

Usage

```
goscar(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 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	s1 The l_1 regularization parameter, $s1 >= 0$.
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

S.Yang, L.Yuan, Y.Lai, X.Shen, P.Wonka, and J.Ye. Feature grouping and selection over an undirected graph. KDD, 2012.

Examples

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

$$min 1/2||Ax - y||^2 + \lambda_1||x||_1 + \lambda_2 \sum_{(i,j)inE} w_(i,j)||x_i| - |x_j||$$

Usage

```
ncFGS(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 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	s1 The l_1 regularization parameter, $s1 >= 0$.
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

S.Yang, L.Yuan, Y.Lai, X.Shen, P.Wonka, and J.Ye. Feature grouping and selection over an undirected graph. KDD, 2012.

Examples

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

---

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 solves

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

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 $r = sign(Rwt)$.

Usage

```
ncTF(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 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

S.Yang, L.Yuan, Y.Lai, X.Shen, P.Wonka, and J.Ye. Feature grouping and selection over an undirected graph. KDD, 2012.

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)
```

---

ncTFGS *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

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

where $J_\tau(x) = min(x/\tau, 1)$ is a surrogate of the L0 norm.

Usage

```
ncTFGS(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 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	s1 The l_1 regularization parameter, $s1 >= 0$.
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

S.Yang, L.Yuan, Y.Lai, X.Shen, P.Wonka, and J.Ye. Feature grouping and selection over an undirected graph. KDD, 2012.

Examples

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

ncTL *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 solves

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

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 $r = sign(Rwt)$.

Usage

```
ncTL(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 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	s1 The l_1 regularization parameter, $s1 >= 0$.
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

S.Yang, L.Yuan, Y.Lai, X.Shen, P.Wonka, and J.Ye. Feature grouping and selection over an undirected graph. KDD, 2012.

Examples

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

---

ncTLF *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 solves

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

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 $r = sign(Rwt)$.

Usage

```
ncTLF(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 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	s1 The l_1 regularization parameter, $s1 >= 0$.
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

S.Yang, L.Yuan, Y.Lai, X.Shen, P.Wonka, and J.Ye. Feature grouping and selection over an undirected graph. KDD, 2012.

Examples

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