

Package ‘FindIt’

July 2, 2014

Version 0.3

Date 2013-01-31

Title R Package for Finding Heterogeneous Treatment Effects

Author Marc Ratkovic <ratkovic@princeton.edu>, Kosuke Imai <kimai@princeton.edu>,

Maintainer Marc Ratkovic <ratkovic@princeton.edu>

Depends R (>= 2.15.0), glmnet, lars

Imports glmnet, lars

Description FindIt is an R package that implements the heterogeneous treatment effect estimation procedure proposed by Imai and Ratkovic (2013). The proposed method is applicable, for example, when selecting a small number of most (or least) efficacious treatments from a large number of alternative treatments as well as when identifying subsets of the population who benefit (or are harmed by) a treatment of interest. The method adapts the Support Vector Machine classifier by placing separate LASSO constraints over the pre-treatment parameters and causal heterogeneity parameters of interest. This allows for the qualitative distinction between causal and other parameters, thereby making the variable selection suitable for the exploration of causal heterogeneity.

LazyLoad yes

LazyData yes

License GPL (>= 2)

Repository CRAN

Date/Publication 2013-02-01 07:53:25

NeedsCompilation no

R topics documented:

FindIt	2
GerberGreen	5
LaLonde	7
makellway	7
maketwoway	8

Index	10
--------------	-----------

FindIt	<i>FindIt for Estimating Heterogeneous Treatment Effects</i>
--------	--

Description

FindIt takes a binary outcome and a set of pre-treatment covariates, and returns a model with the most predictive treatment-covariate interactions returned.

Usage

```
FindIt(y,X.c, treat ,type="single", rescale.c=FALSE, search.lambdas=TRUE,
       lambdas=NULL, wts=1, scale.c=1, scale.t=1, n.highlow=10, fit.glmnet=TRUE)
```

Arguments

y	A vector of outcomes that takes values of +1 or -1.
treat	A vector of treatment assignments that takes values of 0 or 1. If type="multiple", a matrix of multiple treatments, as may arise in a factorial design. This matrix should take values 0 and 1.
type	"single" for interactions between a single variable and multiple covariates; "multiple" if treat is a matrix of multiple treatments.
X.c	A covariate matrix.
rescale.c	Should X.c be rescaled internally? Defaults to FALSE; see maketwoway and the example below for implementing our function that automatically generates two-way interactions.
search.lambdas	Whether to search for the tuning parameters for the LASSO constraints. If FALSE, lambdas must be supplied.
lambdas	Tuning parameters to be given to FindIt; only used if search.lambdas=FALSE
wts	An optional set of scaling weights.
scale.c	A set of weights for recaling the pre-treatment covariates. maketwoway is useful for generating these.
scale.t	A set of scaling weights for the treatment covariates.
n.highlow	When type is "single," returns the indices for observations with the highest and lowest treatment effect estimates.
fit.glmnet	Whether to fit using the coordinate descent method in glmnet (TRUE) or the regularization path method of LARS (FALSE).

Details

Implements the alternating line search algorithm for estimating the tuning parameters, as described in Imai and Ratkovic (2012).

Value

coefs	A named vector of coefficients
coefs.orig	Coefficients on the original scale, if scale.c and scale.t was used
fit	The fitted values on an SVM scale
highlow	Treatment effect estimates and indices for observations with the 10 highest and lowest treatment effects
scale	Vector of weights for scaling the coefficients
names.out	Names of hte coefficients
y	The vector of observed outcomes
X.c	Matrix of pre-treatment covariates
X.t	Matrix of treatments or treatment-covariate interactions
GCV	GCV statistic at the minimum
ATE	For a single treatment, the estimated ATE. For multiple treatments, the estimated ATE for each unique treatment combination
lambdas	Tuning parameters used for the fit
n.highlow	Number of indices for assessing observations with highest and lowest effect estimates.

Author(s)

Marc Ratkovic and Kosuke Imai.

References

Imai, Kosuke and Marc Ratkovic. “Estimating Treatment Effect Heterogeneity in Randomized Program Evaluation.” Forthcoming, Annals of Applied Statistics. <http://www.princeton.edu/~ratkovic/svm.pdf>

Examples

```
###These examples reproduce the results in Imai and Ratkovic (forthcoming).  
  
###  
### Example: heterogeneous effect estimation in the National Supported Work Study Data  
###  
  
##Load the LaLonde data  
data(LaLonde)  
attach(LaLonde)
```

```

y<- 2*(re78>re75)-1

#Generate independent variables.
white<-(black==0&hisp==0)
log.re75<-log(1+re75)
u75<-re75==0

#Generate matrices of main effects and scaling factors. Center on weighted means, then interact.
X.orig<-cbind(age,educ,black,hisp,white,marr,nodegr,log.re75,u75)

#Construct named matrix of all sensible two-way interactions.
X<-maketwoway(X.orig)
wts.extrap<-psid.wts/mean(psid.wts)
X2<-maketwoway(X.orig,wts=wts.extrap)

#Run to find the LASSO parameters
#F1<-FindIt(y,X.c=X$X,treat, scale.c=X$scale.X, search.lambdas=TRUE, fit.glmnet=TRUE) #Run to find the LASSO parameters
#F2<-FindIt(y=y, X.c=X2$X, treat=treat, type="single", rescale.c=F, scale.c=X2$scale.X, search.lambdas=TRUE, wts=wts.extrap)

#Fit with uncovered lambda parameters.
F1<-FindIt(y,X.c=X$X,treat, scale.c=X$scale.X, search.lambdas=FALSE,lambdas=c(-4.000 , -4.127),fit.glmnet=TRUE )
F2<-FindIt(y=y, X.c=X2$X, treat=treat, type="single", rescale.c=FALSE, scale.c=X2$scale.X, search.lambdas=FALSE)

#Returns a summary of coefficients and model fit.
summary(F1)
summary(F2)
#Return the estimated ATE.
F1$ATE
F2$ATE

#Returns the ten highest and lowest treatment effects, as well as the covariate
#profiles on the original covariate scale (X.orig)
highlow(F1, X0=X.orig)

detach(LaLonde)

## Not run:
###
### Example: heterogeneous effect estimation in the New Haven Get-Out-the-Vote Experiment
###

data(GerberGreen)

#Subset to single family households
Data1<-subset(Data1,persons==1)
attach(Data1)

#Make matrix of treatments.
X.lin<-cbind(persngrp,phnscrip,mailings,appeal)
X.t.0<-makeallway(X.lin)
colnames.t<-colnames(X.t.0)

```

```

names.keep<-c("persngrp_1",paste("mailings_",1:3,sep=""), paste("phnscrip_",1:6,sep=""))
keeps<-unique(unlist(sapply(names.keep,F=function(x) grep(x,colnames.t))))
X.t<-X.t.0[,keeps]
rm(X.t.0)

#Generate outcome variable and matrix of pre-treatment covariates.
y<-(2*voted98-1)
X.orig<-cbind(age,majorpty,vote96.1,vote96.0)
X<-maketwaway(X.orig,center=T)

#Run to search for lambdas.
F3<-FindIt(y=y, X.c=X$X, treat=X.t, type="multiple", rescale.c=FALSE , scale.c=X$s, search.lambdas=TRUE,fit.glm)

#Fit, given selected lambdas.
#F3<-FindIt(y=y,X.c=X$X, treat=X.t, type="multiple", rescale.c=FALSE, scale.c=X$s, search.lambdas=FALSE, #lambdas)

#Returns coefficient estimates.
summary(F3)

#Returns treatment effect estimates.
F3$ATE

## End(Not run)

```

GerberGreen

Data from the 1998 New Haven Get-Out-the-Vote Experiment

Description

This data set contains the full and most recent corrected data from the field experiment analyzed in Gerber and Green (2000).

Usage

LaLonde

Format

A data frame consisting of 23 columns and 29,380 observations.

Source

A data set with columns:
persons: Number persons in household
ward: Ward of residence
question: Asked to commit to voting
mailgrp: Sent mail

phonegrp: Phone batch 1
 persngrp: Personal contact attempted
 appeal: Content of message
 1: Civic Duty
 2: Neighborhood Solidarity
 3: Close Election
 contact: Personal contact occurred
 mailings: Number of mailings sent
 age: Age of respondent
 majorpty: Democratic or Republican
 vote96.0: Abstained in 1996
 vote96.1: Voted in 1996
 mailcall: Phone batch 2
 voted98: Voted in 1998
 phnscrpt: Script read to phone respondents
 1: Civic-Blood
 2: Civic
 3: Civic or Blood-Civic
 4: Neighbor
 5: Neighbor or Civic-Neighbor
 6: Close
 dis.mc: Contacted by phone in batch 2
 dis.phn: Contacted by phone in batch 1
 phn.c: Contacted by phone
 phntrt1: Phone contact attempted (no blood or blood-civic)
 phntrt2: Phone contact attempted (no blood)
 phn.c1: Contact occurred in phntrt1
 phn.c2: Contact occurred in phntrt2
 Note: Descriptions come from the description of dataset GerberGreenImai in package Matching.

References

- Gerber, A. S. and Green, D. P. (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review* 94, 3, 653-663.
- Imai, K. (2005). Do get-out-the-vote calls reduce turnout?: The importance of statistical methods for field experiments. *American Political Science Review* 99, 2, 283-300.

LaLonde

National Supported Work Study Experimental Data

Description

This data set gives the outcomes as well as treatment assignments and covariates for the National Supported Work Study, as analyzed in LaLonde (1986).

Usage

LaLonde

Format

A data frame consisting of 5 columns (including a treatment assignment vector) and 2787 observations.

Source

Data from the National Supported Work Study. A benchmark matching dataset. Columns consist of an indicator for whether the observed unit was treated; an indicator for whether the individual received the treatment; age in years; schooling in years; indicators for black and Hispanic; an indicator for marriage status, one of married; an indicator for no high school degree; reported earnings in 1975 and 1978; and extrapolation weights to the 1978 Panel Study for Income Dynamics dataset. 1974 and 1975 earnings are pre-treatment. 1978 earnings is taken as the outcome variable.

References

LaLonde, R.J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* 76, 4, 604-620.

makellway

Constructing named all possible interactions from a given set of factors.

Description

makeallway takes a data frame of factors, each denoting a treatment level, and returns a matrix of 0's and 1's for all possible interactions.

Usage

```
makeallway(X, wts = 1)
```

Arguments

X	A matrix of data frame of treatments, each column interpreted as a factor.
wts	An optional set of extrapolation weights.

Details

Useful for constructing a matrix for `FindIt` with `type="multiple"`. Takes a data frame of factors and returns a design matrix with all possible interactions. A named matrix of 0's and 1's, for every possible interaction, is returned. Pairwise linearly dependent columns and columns of all 0's or 1's are eliminated.

Value

X	A named matrix of 0's and 1's.
---	--------------------------------

Author(s)

Marc Ratkovic and Kosuke Imai.

Examples

```
#See the help page for FindIt() for an example.
```

maketwoway

Constructing a named matrix of all two-way interactions.

Description

`maketwoway` takes a set of pre-treatment covariates and returns all sensible two-way interactions.

Usage

```
maketwoway(X, wts = 1, center = T)
```

Arguments

X	A matrix of pre-treatment covariates.
wts	An optional set of extrapolation weights.
center	Whether the columns should be centered before interacting. Set to false with a factorial treatment design.

Details

Useful for constructing a matrix for `FindIt`. Takes a matrix and returns a matrix of two-way interactions. A named matrix with an intercept, all two-way interactions and square terms are returned. If `center=TRUE`, interactions are formed by (1) centering, (2) scaling to sd 1, and (3) interacting. Non-sensible interactions, such as interacting dummy variables for male and female, as well as squared terms for dummy variables, are not returned. A rescaling vector is returned, so that the columns of the standardized matrix can be returned to their original scale.

Value

<code>X</code>	A matrix with all linear terms, squared terms, and two-way interactions from the original matrix.
<code>scale.X</code>	A vector of weights for reutrning <code>X</code> to its original scale.

Author(s)

Marc Ratkovic and Kosuke Imai.

Examples

```
#See the help page for FindIt() for an example.
```

Index

*Topic **datasets**

GerberGreen, [5](#)

LaLonde, [7](#)

FindIt, [2](#)

GerberGreen, [5](#)

highlow (FindIt), [2](#)

LaLonde, [7](#)

makeallway (makellway), [7](#)

makellway, [7](#)

maketwoway, [8](#)

summary.FindIt (FindIt), [2](#)