Package 'SVMMatch'

February 19, 2015

Type Package

Title Causal Effect Estimation and Diagnostics with Support Vector Machines

Version 1.1

Date 2015-02-05

Author Marc Ratkovic

Maintainer Marc Ratkovic <ratkovic@princeton.edu>

Description Causal effect estimation in observational data often requires identifying a set of untreated observations that are comparable to some treated group of interest. This package provides a suite of functions for identifying such a set of observations and for implementing standard and new diagnostics tools. The primary function, svmmatch(), uses support vector machines to identify a region of common support between treatment and control groups. A sensitivity analysis, balance checking, and assessment of the region of overlap between treated and control groups is included. The Bayesian implementation allows for recovery of uncertainty estimates for the treatment effect and all other parameters.

License GPL (>= 2)

Depends R (>= 3.0.2)

Imports Rcpp (>= 0.11.0)

LinkingTo Rcpp, RcppArmadillo

NeedsCompilation yes

Repository CRAN

Date/Publication 2015-02-08 09:24:56

R topics documented:

SVMMatch-package *Title: Causal effect estimation and diagnostics with support vector machines.*

Description

SVMMatch identifies the region of common support between a set of treated and control units in observational data. Using the observations in this region, a set of balancing weights and a treatment effect are estimated. The method, described in Ratkovic (2014), adapts the support vector machine technology in order to estimate these balancing weights, using a Bayesian implementation so as to give uncertainty effects both in treatment assignment and effect estimation.

Details

Package:	SVMMatch
Type:	Package
Version:	1.0
Date:	2015-02-05
License:	GPL (>= 2)

The method implements the matching algorithm through the main function, svmmatch.

A series of diagnostics are implemented. The function balance() assesses the posterior density of covariate imbalance; effect() returns the posterior estimate of the treatment effect; sensitivity() assesses the effect estimate's sensitivity to unobserved confounders; control.overlap() returns the posterior density of number of control observations returned in matching; and treatment.overlap() examines difficult-to-match treated observations.

Author(s)

Marc Ratkovic
Maintainer: Marc Ratkovic <ratkovic@princeton.edu>

References

Ratkovic, Marc. 2014. "Balancing within the Margin: Causal Effect Estimation with Support Vector Machines." Working paper.

autocorr *Autocorrelation in estimated coefficients.*

Description

autocorr tests the autocorrelation in the coefficients in an SVMMatch object.

Usage

```
autocorr(obj)
```

Arguments

obj	A fitted SVMMatch object.

Details

Returns the autocorrelation in the posterior means of the coefficients of an SVMMatch object.

Author(s)

Marc Ratkovic

References

Ratkovic, Marc. 2014. "Balancing within the Margin: Causal Effect Estimation with Support Vector Machines." Working paper.

See Also

svmmatch

Examples

```
## Not run:
##See svmmatch() for a full implementation
##Load data
data("LaLonde")
Data1<-LaLonde
Data1<-Data1[Data1$exper==0|Data1$treat==1,]
attach(Data1)

##Format X matrix
varnames<-c("age","educ","black","married","nodegr","hisp",
"re75","re74")
X<-cbind(Data1[,varnames],Data1$re75==0,Data1$re74==0)
X<-as.matrix(X)
```

```
##Fit model
set.seed(1)
m1.param<-svmmatch(treat, X, dv=re78, burnin=100, gibbs=100, thin=5)

##Assess autocorrelation
ac1<-autocorr(m1.param)

## End(Not run)
```

---

balance *Assessing balance when using SVMMatch.*

Description

balance returns a diagnostic plot of covariate balance as well as an object containing the posterior estimate of covariate imbalance.

Usage

```
balance(treat, X, obj, plot.it=TRUE, sd.plot=.2, color=TRUE)
```

Arguments

treat	A binary vector indicating treatment status.
X	A covariate matrix.
obj	A fitted SVMMatch object.
plot.it	Whether to return a figure illustrating balancing in the raw and balanced data. TRUE or FALSE.
sd.plot	Number of standard deviations at which to include dashed vertical lines in the figure.
color	Whether to plot in color or black and white. TRUE or FALSE.

Details

Assess the covariate balance using weights from an SVMMatch object.

Value

balance	Posterior density of imbalance, by covariate.

Author(s)

Marc Ratkovic

References

Ratkovic, Marc. 2014. "Balancing within the Margin: Causal Effect Estimation with Support Vector Machines." Working paper.

See Also

svmmatch

Examples

```
## Not run:
##See svmmatch() for a full implementation
##Load data
data("LaLonde")
Data1<-LaLonde
Data1<-Data1[Data1$exper==0|Data1$treat==1,]
attach(Data1)

##Format X matrix
varnames<-c("age","educ","black","married","nodegr","hisp",
"re75","re74")
X<-cbind(Data1[,varnames],Data1$re75==0,Data1$re74==0)
X<-as.matrix(X)

##Fit model
set.seed(1)
m1.param<-svmmatch(treat, X, dv=re78, burnin=100, gibbs=100, thin=5)

##Assess balance
bal1<-balance(treat, X, m1.param)


## End(Not run)
```

---

bayesmatch_cpp *Rcpp implementation for Bayesian SVM.*

Description

bayesmatch_cpp is the Rcpp object that fits the Bayesian SVM. Primarily for internal use.

Usage

```
bayesmatch_cpp(X0, boldX0, lambda0, treat0, total_gibbs0, thin0,
                 param0, dv0, nu0, burnin0)
```

Arguments

X0	Scaled covariate matrix.
boldX0	Scaled and transformed covariate matrix.
lambda0	Vector of initial weights for each observation
treat0	Binary treatment vector.
total_gibbs0	Number of posterior draws to save.
thin0	Number of draws to thin.
param0	No longer used. Set to 0.
dv0	Outcome variable.
nu0	Initial estimate for prior weight.
burnin0	Number of posterior draws discarded.

Details

Internal function.

Author(s)

Marc Ratkovic

References

Ratkovic, Marc. 2014. "Balancing within the Margin: Causal Effect Estimation with Support Vector Machines." Working paper.

See Also

svmmatch

Examples

```
## See svmmatch() for implementation.
```

---

control.overlap *Assessing the number of control observations used in estimating the treatment effect.*

Description

control.overlap returns a diagnostic plot showing the posterior mass over the number of control observations estimated in the common support region between the treatment and control groups.

Usage

```
control.overlap(obj, color=TRUE, label.main="Assessing Control Overlap",
label.x="Size of Control Set", label.y="Mass")
```

Arguments

obj	A fitted SVMMatch object.
color	Whether to plot in color or black and white. TRUE or FALSE.
label.main	Main title for figure
label.x	X-axis label.
label.y	Y-axis label.

Details

Gives the posterior mass of control observations that fall in the common support region between treated and untreated obsevations. These are the only observations used in constructing the treatment effect.

Value

counts	Table containing posterior mass of number untreated observations.

Author(s)

Marc Ratkovic

References

Ratkovic, Marc. 2014. "Balancing within the Margin: Causal Effect Estimation with Support Vector Machines." Working paper.

See Also

svmmatch

Examples

```
## Not run:
##See svmmatch() for a full implementation
##Load data
data("LaLonde")
Data1<-LaLonde
Data1<-Data1[Data1$exper==0|Data1$treat==1,]
attach(Data1)

##Format X matrix
varnames<-c("age","educ","black","married","nodegr","hisp",
"re75","re74")
```

```
X<-cbind(Data1[,varnames],Data1$re75==0,Data1$re74==0)
X<-as.matrix(X)

##Fit model
set.seed(1)
m1.param<-svmmatch(treat, X, dv=re78, burnin=100, gibbs=100, thin=5)

##Assess ignorability:
#Assess control overlap--are there enough
#controls to balance the treated?
cont1<-control.overlap(m1.param)


## End(Not run)
```

`effect` *Posterior density of the treatment effect estimate from an SVMMatch object.*

Description

`effect` characterizes the posterior density of the treatment effect, as estimated through SVMMatch.

Usage

```
effect(obj, color=TRUE, quant=c(0.025,0.975), legend.pos="topleft",
label.main="Posterior Density of Effect Estimate", label.x="Outcome",
label.y="Density")
```

Arguments

`obj`	A fitted SVMMatch object.
`color`	Whether to plot in color or black and white. TRUE or FALSE.
`quant`	Quantiles of the effect to be marked on the figure.
`legend.pos`	Where to place the margin. See the help file for `legend`.
`label.main`	Main title for figure
`label.x`	X-axis label.
`label.y`	Y-axis label.

Details

Characterizes and plots the posterior density of the treatment effect, given an SVMMatch object. At each posterior draw, a set of balancing weights are constructed, as described in Ratkovic (2014). A treatment effect is estimated for each posterior draw of balancing weights, and the posterior density returned as a figure.

Value

balance Posterior density of imbalance, by covariate.

Author(s)

Marc Ratkovic

References

Ratkovic, Marc. 2014. "Balancing within the Margin: Causal Effect Estimation with Support Vector Machines." Working paper.

See Also

svmmatch, legend

Examples

```
## Not run:
##See svmmatch() for a full implementation
##Load data
data("LaLonde")
Data1<-LaLonde
Data1<-Data1[Data1$exper==0|Data1$treat==1,]
attach(Data1)

##Format X matrix
varnames<-c("age","educ","black","married","nodegr","hisp",
"re75","re74")
X<-cbind(Data1[,varnames],Data1$re75==0,Data1$re74==0)
X<-as.matrix(X)

##Fit model
set.seed(1)
m1.param<-svmmatch(treat, X, dv=re78, burnin=100, gibbs=100, thin=5)

##Summarize treatment effect
effect1<-effect(m1.param)


## End(Not run)
```

LaLonde *LaLonde Data for Covariate Balancing Propensity Score*

Description

This data set gives the outcomes a well as treatment assignments and covariates for the econometric evaluation of training programs in LaLonde (1986). Taken from package CBPS.

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 in the experimental subset; 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; and reported earnings in 1974, 1975, and 1978. 1974 and 1975 earnings are pre-treatment. 1978 earnings is taken as the outcome variable. Note: This data and description were originally prepared for the CBPS package.

References

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

`sensitivity` *Sensitivity analysis for SVMMatch.*

Description

`sensitivity` assesses the sensitivity of an effect estimate to an omitted confounder.

Usage

```
sensitivity(obj, seq.eval=seq(-1,1,.1), quant.eval=c(0.025,0.5,0.975),
color=TRUE, legend.pos="topleft", label.main="Sensitivity Analysis",
label.x="Sensitivity Parameter",  label.y="Outcome")
```

Arguments

`obj`	A fitted SVMMatch object.
`seq.eval`	Values at which to set the omitted confounder, in the range [-1, +1].
`quant.eval`	Values at which to plot the posterior density as a function of the omitted confounder. By default, a solid line is drawn through the posterior medians, with dashed lines at the 2.5th and 97.5th percentiles.
`color`	Whether to plot in color or black and white. TRUE or FALSE.
`legend.pos`	Where to place the margin. See the help file for `legend`.

label.main	Main title for figure
label.x	X-axis label.
label.y	Y-axis label.

Details

Conducts a sensitivity analysis using an SVMMatch object. An unoberseved parameter, u, that predicts the treatment assignment is introduced and varied between -1 and 1. For each value of u, balancing weights are constructed and the posterior density of the effect estimate recalcluated, with u=0 returning the results from the original fit. The figure gives the researcher a sense as to how sensitive the effect estimate is to omitted confounders.

Value

sens.mat	A matrix of the posterior estimates as a function of the unobserved confounder.

Author(s)

Marc Ratkovic

References

Ratkovic, Marc. 2014. "Balancing within the Margin: Causal Effect Estimation with Support Vector Machines." Working paper.

See Also

svmmatch, legend

Examples

```
## Not run:
##See svmmatch() for a full implementation
##Load data
data("LaLonde")
Data1<-LaLonde
Data1<-Data1[Data1$exper==0|Data1$treat==1,]
attach(Data1)

##Format X matrix
varnames<-c("age","educ","black","married","nodegr","hisp",
"re75","re74")
X<-cbind(Data1[,varnames],Data1$re75==0,Data1$re74==0)
X<-as.matrix(X)

##Fit model
set.seed(1)
m1.param<-svmmatch(treat, X, dv=re78, burnin=100, gibbs=100, thin=5)

##Sensitivity analysis (Takes a little longer)
sens1<-sensitivity(m1.param)
```

```
## End(Not run)
```

---

svmmatch *SVMMatch for Causal Effect Estimation*

Description

svmmatch estimates balancing weights in the presence of non-random treatment assignment. The method allows the researcher to assess the two crucial ignorability assumptions: a test of sensitivity to omitted confounders and a test of common support between treated and control. A fully Bayesian implementation allows characterization of the posterior of the treatment effect. The method is implemented in C++ through Rcpp, greatly speeding calculation time.

Usage

```
svmmatch(treat, X, burnin = 100, gibbs = 200, thin = 2, dv = NULL)
```

Arguments

treat	A binary vector indicating treatment status.
X	A covariate matrix.
burnin	Number of burnin draws for the Gibbs sampler.
gibbs	Number of posterior draws to be saved.
thin	Number of chains when thinning. If burnin = 10, gibbs = 20, and thin=30,then 500 = 10 x 30 + 20 x 30 samples will be drawn, but only 20 returned.
dv	A dependent variable of interest. Optional.

Details

Fits SVMMatch to the data, returning balancing weights and a posterior density of the estimated effect.

Value

effect	Posterior density of treatment effect on dv.
beta	Matrix of posterior draws for coefficients.
margin	Binary vector for posterior estimate of marginal observations.
bal.wts	Balancing weights.
X.scale	Scaled covariate matrix. Used internally.
X.orig	Original matrix of covariates.
treat	Treatment vector.
dv	The dependent variable.

Author(s)

Marc Ratkovic

References

Ratkovic, Marc. 2014. "Balancing within the Margin: Causal Effect Estimation with Support Vector Machines." Working paper.

See Also

balance, effect, sensitivity, control.overlap, treatment.overlap, LaLonde

Examples

```
##Load data
data("LaLonde")
Data1<-LaLonde
Data1<-Data1[Data1$exper==0|Data1$treat==1,]
attach(Data1)

##Format X matrix
varnames<-c("age","educ","black","married","nodegr","hisp",
"re75","re74")
X<-cbind(Data1[,varnames],Data1$re75==0,Data1$re74==0)
X<-as.matrix(X)

##Fit model
set.seed(1)
m1.param<-svmmatch(treat, X, dv=re78, burnin=100, gibbs=100, thin=5)

##Assess balance
bal1<-balance(treat, X, m1.param)

##Summarize treatment effect
effect1<-effect(m1.param)

## Not run:
##Sensitivity analysis (Takes a little longer)
sens1<-sensitivity(m1.param)

## End(Not run)

##Assess ignorability:
#Assess control overlap--are there enough
#controls to balance the treated?
cont1<-control.overlap(m1.param)

#Assess treatment overlap--are there treated
#values that can't be matched?
treat1<-treatment.overlap(m1.param)
```

```
##Assess autocorrelation
ac1<-autocorr(m1.param)
```

`treatment.overlap` *Exploring hard-to-match treated observations.*

Description

`treatment overlap` characterizes treated observations which have no natural matches in the data because they fall outside the common support region.

Usage

```
treatment.overlap(obj, color=TRUE, thresh=.95)
```

Arguments

`obj`	A fitted SVMMatch object.
`color`	Whether to plot in color or black and white. TRUE or FALSE.
`thresh`	Proportion of time a treated observation must fall outside the common support region in order to be considered unmatched.

Details

SVMMatch estimates a region of common support between treated and untreated observations. This function identifies and characterizes the treated observations that regularly fall outside the common support region. Having too many of these observations makes estimation of the treatment effect casts doubt on whether the average treatment effect is identified in the data.

Value

`no.overlap`	A binary vector the length of the number of treated observations taking a value of one if the observation falls outside the region of common support at some rate above the threshold.
`logit`	A logistic regression using the original covariates in the SVMMatch object to predict whether an observation falls outside the common support region.

Author(s)

Marc Ratkovic

References

Ratkovic, Marc. 2014. "Balancing within the Margin: Causal Effect Estimation with Support Vector Machines." Working paper.

See Also

svmmatch

Examples

```
## Not run:
##See svmmatch() for a full implementation
##Load data
data("LaLonde")
Data1<-LaLonde
Data1<-Data1[Data1$exper==0|Data1$treat==1,]
attach(Data1)

##Format X matrix
varnames<-c("age","educ","black","married","nodegr","hisp",
"re75","re74")
X<-cbind(Data1[,varnames],Data1$re75==0,Data1$re74==0)
X<-as.matrix(X)

##Fit model
set.seed(1)
m1.param<-svmmatch(treat, X, dv=re78, burnin=100, gibbs=100, thin=5)

#Assess treatment overlap--are there treated
#values that can't be matched?
treat1<-treatment.overlap(m1.param)


## End(Not run)
```

Index