Package 'vscc'

July 2, 2014

Type Package

Title Variable selection for clustering and classification

Version 0.2

Date 2013-11-16

Author Jeffrey L. Andrews, Paul D. McNicholas

Maintainer Jeffrey L. Andrews <jeffrey.andrews@macewan.ca>

Description Performs variable selection/feature reduction under a clustering or classification framework. In particular, it can be used in an automated fashion using mixture model-based methods (tEIGEN and MCLUST are currently supported).

License GPL (>= 2)

Imports teigen, mclust

NeedsCompilation no

Repository CRAN

Date/Publication 2013-11-17 08:24:55

R topics documented:

vscc-package *Variable selection for clustering and classification*

Description

Performs variable selection under a clustering or classification framework. Automated implementation using model-based clustering is based on `teigen` version 2.0 and `mclust` version 4.0; issues *may* arise when using different versions.

Details

Package:	vscc
Type:	Package
Version:	0.2
Date:	2013-11-16
License:	GPL>=2

Author(s)

Jeffrey L. Andrews and Paul D. McNicholas

Maintainer: Jeffrey L. Andrews <jeffrey.andrews@macewan.ca>

References

See `citation("vscc")`.

See Also

`vscc`

plot.vscc *Plotting for vscc objects*

Description

Dedicated plot function for objects of class vscc.

Usage

```
## S3 method for class 'vscc'
plot(x, ...)
```

Arguments

x	An object of class vscc.
...	Further arguments to be passed on

Details

Provides a scatterplot matrix of the selected variables with colours corresponding to each group.

Author(s)

Jeffrey L. Andrews

See Also

vscc

Examples

```
require("mclust")
data(banknote)
bankrun <- vscc(banknote[,-1])
plot(bankrun)
```

---

print.vscc *Printing for vscc*

Description

Dedicated print function for objects of class vscc.

Usage

```
## S3 method for class 'vscc'
print(x, ...)
```

Arguments

x	An object of class vscc
...	Further arguments to be passed on

Details

Same as summary.

Author(s)

Jeffrey L. Andrews

See Also

summary.vscc, vscc

Examples

```
require("mclust")
data(banknote)
vscc(banknote[,-1])
```

---

summary.vscc *Summary for VSCC*

Description

Dedicated summary function for objects of class vscc

Usage

```
## S3 method for class 'vscc'
summary(object, ...)
```

Arguments

object	An object of class vscc
...	Additional arguments to be passed

Author(s)

Jeffrey L. Andrews

See Also

vscc

Examples

```
require("mclust")
data(banknote)
summary(vscc(banknote[,-1]))
```

vscc *Variable selection for clustering and classification*

Description

Performs variable selection under a clustering or classification framework. Automated implementation using model-based clustering is based on `teigen` version 2.0 and `mclust` version 4.0; issues *may* arise when using different versions.

Usage

```
vscc(x, G=1:9, automate = "mclust", initial = NULL, train = NULL, forcereduction = FALSE)
```

Arguments

`x`	Data frame or matrix to perform variable selection on
`G`	Vector for the number of groups to consider during initialization and/or post-selection analysis. Default is 1-9.
`automate`	Character string (`"teigen"`, `"mclust"` (default), or NULL only) indicating which mixture model family to implement as initialization and/or post-selection analysis. If NULL, the function assumes manual operation of the algorithm (meaning an initial clustering vector must be given, and no post-selection analysis is performed).
`initial`	Optional vector giving the initial clustering.
`train`	Optional vector of training data (for classification framework).
`forcereduction`	Logical indicating if the full data set should be considered (FALSE) when selecting the 'best' variable subset via total model uncertainty. Not used if `automate=NULL`.

Value

`selected`	A list containing the subsets of variables selected for each relation. Each set is numbered according to the number in the exponential of the relationship. For instance, `vscc_object$selected[[3]]` corresponds to the variable subset selected by the cubic relationship.
`family`	The family used as initialization and/or post selection. (Same as user input `automate`, and can be `NULL`).
`wss`	The within-group variance associated with each variable from the full data set.

The remaining values are provided as long as `automate` is not `NULL`:

`topselected`	The best variable subset according to the total model uncertainty.
`initialrun`	Results from the initialization; an object of class `teigen` or `mclust`.
`bestmodel`	Results from the best model on the selected variable subset; an object of class `teigen` or `mclust`.

`chosenrelation`	Numeric indication of the relationship chosen according to total model uncertainty. The number corresponds to exponent in the relationship: for instance, a value of '4' suggests the quartic relationship. If the value `"Full dataset"` is given, then the unreduced data provides the best model uncertainty; can be avoided by specifying `forcereduction=TRUE` in the function call.
`uncertainty`	Total model uncertainty associated with the best relationship.
`allmodelfit`	List containing the results (`teigen` or `mclust` objects) from the post-selection analysis on each variable subset. Number corresponds to the exponent in the relationship. For instance, `vscc_object$allmodelfit[[1]]` gives the results from the analysis on the variables selected by the linear relationship.

Author(s)

Jeffrey L. Andrews, Paul D. McNicholas

References

See `citation("vscc")` for the variable selection references. See also `citation("teigen")` and `citation("mclust")` if using those families of models via the `automate` call.

See Also

`teigen`, `Mclust`

Examples

```
require("mclust")
data(banknote) #Load data
head(banknote[,-1]) #Show preview of full data set
bankrun <- vscc(banknote[,-1])
head(bankrun$topselected) #Show preview of selected variables
table(banknote[,1], bankrun$initialrun$classification) #Clustering results on full data set
table(banknote[,1], bankrun$bestmodel$classification) #Clustering results on reduced data set
```

Index