

Package ‘CAM’

March 5, 2015

Type Package

Title Causal Additive Model (CAM)

Version 1.0

Date 2015-03-03

Author Jonas Peters and Jan Ernest

Encoding UTF-8

Depends glmnet, mboost, Matrix, parallel, mgcv

Maintainer Jonas Peters <jonas.peters@tuebingen.mpg.de>

Description The code takes an $n \times p$ data matrix and fits a Causal Additive Model (CAM) for estimating the causal structure of the underlying process. The output is a $p \times p$ adjacency matrix (a one in entry (i,j) indicates an edge from i to j). Details of the algorithm can be found in: P. Bühlmann, J. Peters, J. Ernest: “CAM: Causal Additive Models, high-dimensional order search and penalized regression”, Annals of Statistics 42:2526-2556, 2014.

License FreeBSD

NeedsCompilation no

Repository CRAN

Date/Publication 2015-03-05 22:01:05

R topics documented:

CAM-package	2
CAM	2
computeScoreMat	4
computeScoreMatParallel	6
selGam	6
selGamBoost	6
selLasso	7
selLm	7
selLmBoost	7
updateScoreMat	7

Index	9
--------------	----------

 CAM-package

Causal Additive Model (package)

Description

fits a causal additive model using the CAM algorithm, see references below

Details

Package: CAM
 Type: Package
 Version: 1.0
 Date: 2015-03-03
 License: GPL

Fitting of a causal additive model. The algorithm can be called with CAM.

Author(s)

Jonas Peters and Jan Ernest

References

P. Bühlmann, J. Peters, J. Ernest: "CAM: Causal Additive Models, high-dimensional order search and penalized regression", *Annals of Statistics* 42:2526-2556, 2014.

 CAM

Causal Additive Model

Description

fits a causal additive model using the CAM algorithm, see references below

Usage

```
CAM(X, scoreName = "SEMGAM", parsScore = list(numBasisFcts = 10), numCores = 1,
    maxNumParents = min(dim(X)[2] - 1, round(dim(X)[1]/20)), output = FALSE,
    variableSel = FALSE, variableSelMethod = selGamBoost,
    variableSelMethodPars = list(atLeastThatMuchSelected = 0.02,
    atMostThatManyNeighbors = 10), pruning = FALSE, pruneMethod = selGam,
    pruneMethodPars = list(cutOffPVal = 0.001, numBasisFcts = 10), intervData = FALSE,
    intervMat = NA)
```

Arguments

<code>X</code>	n x p matrix of training inputs (n data points, p dimensions)
<code>scoreName</code>	specifies the model type which is used to compute the score. Default is "SEMGAM" which assumes a generalized additive model class. Other options include "SEMLIN" which fits a linear model.
<code>parsScore</code>	additional parameters can be supported to the score function.
<code>numCores</code>	specifies the number of cores that can be used for computation.
<code>maxNumParents</code>	specifies the maximal number of parents that are allowed in the model.
<code>output</code>	shall output be printed to the console (TRUE/FALSE)
<code>variableSel</code>	specifies whether initial variable selection (Step 1 of CAM algorithm) shall be performed (TRUE) or not (FALSE). Initial variable selection reduces the number of possible parents for a given node and therefore enables computing the causal structure for large p.
<code>variableSelMethod</code>	specifies the method that is used for variable selection. Default is <code>selGamBoost</code> which uses the <code>gamboost</code> function from <code>mboost</code> package. Other options include: <code>selGam</code> (<code>gam()</code> from <code>mgcv</code>), <code>selLm</code> based on linear regression, <code>selLasso</code> based on Lasso regression from package <code>glmnet</code> .
<code>variableSelMethodPars</code>	optional parameters to modify settings of the selection method.
<code>pruning</code>	specifies whether pruning (Step 3 of CAM algorithm) shall be performed (TRUE) or not (FALSE). Pruning reduces the number of edges in the estimated causal structure.
<code>pruneMethod</code>	specifies the method used for the pruning step. Default is <code>selGAM</code> which is based on the <code>gam()</code> function from the <code>mgcv</code> package.
<code>pruneMethodPars</code>	optional parameters to tune the pruning step.
<code>intervData</code>	boolean that indicates whether we use interventional data.
<code>intervMat</code>	the matrix <code>intervMat</code> has the same dimension as <code>X</code> . entry <code>(i,j) == TRUE</code> indicates that in experiment <code>i</code> , variable <code>j</code> has been intervened on.

Details

The code fits a CAM model. See the references below for more details. Identifiability results for the model class can be found in

J. Peters, J. Mooij, D. Janzing, B. Schölkopf: Causal Discovery with Continuous Additive Noise Models, *JMLR* 15:2009-2053, 2014.

Value

list of attributes of the final estimated causal structure

<code>Adj</code>	adjacency matrix of estimated causal graph
<code>Score</code>	Total edge score of estimated graph
<code>timesVec</code>	Vector containing various time measurements for execution times of the individual steps of the CAM algorithm

Author(s)

Jonas Peters <jonas.peters@tuebingen.mpg.de> and Jan Ernest <ernest@stat.math.ethz.ch>

References

P. B\u00fchlmann, J. Peters, J. Ernest: CAM: Causal Additive Models, high-dimensional Order Search and Penalized Regression *Annals of Statistics* 42:2526-2556, 2014.

Examples

```
n <- 500
eps1<-rnorm(n)
eps2<-rnorm(n)
eps3<-rnorm(n)
eps4<-rnorm(n)

x2 <- 0.5*eps2
x1 <- 0.9*sign(x2)*(abs(x2)^(0.5))+0.5*eps1
x3 <- 0.8*x2^2+0.5*eps3
x4 <- -0.9*sin(x3) - abs(x1) + 0.5*eps4

X <- cbind(x1,x2,x3,x4)

trueDAG <- cbind(c(0,1,0,0),c(0,0,0,0),c(0,1,0,0),c(1,0,1,0))
## x4 <- x3 <- x2 -> x1 -> x4
## adjacency matrix:
## 0 0 0 1
## 1 0 1 0
## 0 0 0 1
## 0 0 0 0

estDAG <- CAM(X, scoreName = "SEMGAM", numCores = 1, output = TRUE, variableSel = FALSE,
              pruning = TRUE, pruneMethod = selGam, pruneMethodPars = list(cutOffPVal = 0.001))

cat("true DAG:\n")
show(trueDAG)

cat("estimated DAG:\n")
show(estDAG$Adj)
```

computeScoreMat

auxiliary file for CAM: Computes the initial score matrix.

Description

auxiliary file to CAM. Computes the initial score matrix.

Usage

```
computeScoreMat(X, scoreName, numParents, output, numCores, selMat, parsScore,
                intervMat, intervData)
```

Arguments

X	n x p matrix of training inputs (n data points, p dimensions)
scoreName	specifies the model type which is used to compute the score. Default is "SEMGAM" which assumes a generalized additive model class. Other options include "SEMLIN" which fits a linear model.
numParents	indicates how many parents we consider. If numParents = 1 (default), then the score matrix is of dimension (p-1) x p. If numParents = 2, then the score matrix is of dimension (p-1)(p-2) x p and so on
output	boolean indicating whether information about the progress is written to the console.
numCores	specifies the number of cores that can be used for computation.
selMat	indicating the possible parent relationships.
parsScore	additional parameters can be supported to the score function.
intervMat	the matrix intervMat has the same dimension as X. entry (i,j) == TRUE indicates that in experiment i, variable j has been intervened on.
intervData	boolean that indicates whether we use interventional data.

Value

A list with elements	
scoreMat	The score matrix. scoreMat[i,j] contains the gain in score if we consider i being a parent of j
rowParents	Contains the row names of the score matrix. Only relevant if numParents > 1.
scoreEmptyNodes	Vector containing the scores of each node in the empty graph without any edges.

Note

This is an auxiliary file for CAM.

Author(s)

J. Peters (jonas.peters@tuebingen.mpg.de) and J. Ernest (ernest@stat.math.ethz.ch)

References

P. Buhlmann, J. Peters, J. Ernest: CAM: Causal Additive Models, high-dimensional Order Search and Penalized Regression *Annals of Statistics* 42:2526-2556, 2014.

See Also

[CAM](#)

computeScoreMatParallel

auxiliary file for CAM: Computes the initial score matrix.

Description

This is an auxiliary file for the function CAM in package CAM.

Details

This is an auxiliary file for the function CAM in package CAM.

Note

This is an auxiliary file for the function CAM in package CAM.

Author(s)

Jonas Peters <jonas.peters@tuebingen.mpg.de> and Jan Ernest <ernest@stat.math.ethz.ch>

References

P. Bühlmann, J. Peters, J. Ernest: CAM: Causal Additive Models, high-dimensional Order Search and Penalized Regression *Annals of Statistics* 42:2526-2556, 2014.

See Also

[CAM](#)

selGam

auxiliary file for CAM: selection based on gam (mgcv)

Description

auxiliary file for CAM: selection based on gam (mgcv)

selGamBoost

auxiliary file for CAM: selection based on gam boost (mboost)

Description

auxiliary file for CAM: selection based on gam boost (mboost)

selLasso	<i>auxiliary file for CAM: selection based on lasso (glmnet)</i>
----------	--

Description

auxiliary file for CAM: selection based on lasso (glmnet)

selLm	<i>auxiliary file for CAM: selection based on lm</i>
-------	--

Description

auxiliary file for CAM: selection based on lm

selLmBoost	<i>auxiliary file for CAM: selection based on lm (mboost)</i>
------------	---

Description

auxiliary file for CAM: selection based on lm (mboost)

updateScoreMat	<i>auxiliary file for CAM: Updates the score matrix after having added edge i -> j to the graph.</i>
----------------	---

Description

auxiliary file for CAM. Updates the score matrix after having added edge i -> j to the graph.

Usage

```
updateScoreMat(scoreMat, X, scoreName, i, j, scoreNodes, Adj, output, numCores,  
               maxNumParents, parsScore, intervMat, intervData)
```

Arguments

scoreMat	the current score matrix that has to be updated
X	n _x p matrix of training inputs (n data points, p dimensions)
scoreName	specifies the model type which is used to compute the score. Default is "SEMGAM" which assumes a generalized additive model class. Other options include "SEMLIN" which fits a linear model.
i	starting point of the edge i->j that has been added
j	end point of the edge i->j that has been added
scoreNodes	vector containing current scores of each node
Adj	adjacency matrix of the graph
output	boolean indicating whether information about the progress is written to the console.
numCores	specifies the number of cores that can be used for computation.
maxNumParents	specifies the maximal number of parents that are allowed in the model.
parsScore	additional parameters can be supported to the score function.
intervMat	the matrix <code>intervMat</code> has the same dimension as X. entry (i,j) == TRUE indicates that in experiment i, variable j has been intervened on.
intervData	boolean that indicates whether we use interventional data.

Value

scoreMat	the updated score matrix.
----------	---------------------------

Note

This is an auxiliary file for CAM.

Author(s)

J. Peters (jonas.peters@tuebingen.mpg.de) and J. Ernest (ernest@stat.math.ethz.ch)

References

P. Buhlmann, J. Peters, J. Ernest: CAM: Causal Additive Models, high-dimensional Order Search and Penalized Regression *Annals of Statistics* 42:2526-2556, 2014.

See Also

[CAM](#)

Index

*Topic **Additive Noise Models**

CAM, [2](#)

*Topic **Causality**

CAM, [2](#)

*Topic **Regression**

CAM, [2](#)

*Topic **Restricted Structural
Equation Model**

CAM, [2](#)

CAM, [2](#), [5](#), [6](#), [8](#)

CAM-package, [2](#)

computeScoreMat, [4](#)

computeScoreMatParallel, [6](#)

selGam, [6](#)

selGamBoost, [6](#)

selLasso, [7](#)

selLm, [7](#)

selLmBoost, [7](#)

updateScoreMat, [7](#)