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 x p data matrix and fits a Causal Additive Model (CAM) for estimating the causal structure of the underlying process. The output is a p x 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
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

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(cutOffFVal = 0.001, numBasisFcts = 10), intervData = FALSE, intervMat = NA)
Arguments

X  nxp 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.

parsScore  additional parameters can be supported to the score function.

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.

output  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.

variableSelMethod  specifies the method that is used for variable selection. Default is selGamBoost which uses the gamboost function from mboost package. Other options include: selGam (gam() from mgcv), selLm based on linear regression, selLasso based on Lasso regression from package glmnet.

variableSelMethodPars  optional parameters to modify settings of the selection method.

pruning  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.

pruneMethod  specifies the method used for the pruning step. Default is selGAM which is based on the gam() function from the mgcv package.

pruneMethodPars  optional parameters to tune the pruning step.

intervData  boolean that indicates whether we use interventional data.

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.

Details

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


Value

list of attributes of the final estimated causal structure

Adj  adjacency matrix of estimated causal graph

Score  Total edge score of estimated graph

timesVec  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. Bühlmann, J. Peters, J. Ernest: CAM: Causal Additive Models, high-dimensional Order Search and Penalized Regression

Examples

```r
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 = "SEMgm", numCores = 1, output = TRUE, variableSel = FALSE,
pruning = TRUE, pruneMethod = selGam, pruneMethodPars = list(cutOffFPVal = 0.001))

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

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

computeScoreMat

**auxiliary file for CAM**: Computes the initial score matrix.

Description

auxiliary file to CAM. Computes the initial score matrix.
Usage

```r
computeScoreMat(xL, scoreNameL, numParentsL, outputL, numCoresL, selMatL, parsScoreL,
intervMatL, intervDataL)
```

Arguments

- **x**
  - nxp 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.

- **scoreEmtpyNodes**
  - 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


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**


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

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

**selLm**  
*auxiliary file for CAM: selection based on lm*

**Description**  
auxiliary file for CAM: selection based on lm

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

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

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

**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, numCols, maxNumParents, parsScore, interMat, interData)
Arguments

- **scoreMat**: the current score matrix that has to be updated
- **X**: nxp 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 "SEM-LIN" 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 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

- **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


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