Package 'NNS'

February 2, 2020

Type Package

Title Nonlinear Nonparametric Statistics

Version 0.4.9
Date 2020-01-29
Author Fred Viole
Maintainer Fred Viole <ovvo.financial.systems@gmail.com></ovvo.financial.systems@gmail.com>
Description Nonlinear nonparametric statistics using partial moments. Partial moments are the elements of variance and asymptotically approximate the area of f(x). These robust statistics provide the basis for nonlinear analysis while retaining linear equivalences. NNS offers: Numerical integration, Numerical differentiation, Clustering, Correlation, Dependence, Causal analysis, ANOVA, Regression, Classification, Seasonality, Autoregressive modeling, Normalization and Stochastic dominance. All routines based on: Viole, F. and Nawrocki, D. (2013), Nonlinear Nonparametric Statistics: Using Partial Moments (ISBN: 1490523995).
License GPL-3
BugReports https://github.com/OVVO-Financial/NNS/issues
LazyData TRUE
RoxygenNote 6.1.1
Depends R ($>= 3.3.0$), doParallel
Imports data.table, rgl, stringr, dtw
Suggests knitr, rmarkdown
VignetteBuilder knitr
R topics documented:
Co.LPM Co.UPM D.LPM D.UPM dy.dx dy.d_ LPM LPM.ratio LPM.VaR NNS.ANOVA NNS.ARMA
1

Co.LPM

	NNS.ARMA.optim	14
	NNS.boost	16
	NNS.caus	18
	NNS.cor	19
	NNS.dep	20
	NNS.dep.base	22
	•	22
	NNS.diff	23
		24
	NNS.FSD	25
		26
	NNS.norm	27
	NNS.part	28
	NNS.PDF	29
		30
	NNS.SD.efficient.set	33
		34
	NNS.SSD	35
		36
	NNS.stack	36
		39
		40
		40
	NNS.VAR	41
	PM.matrix	43
	UPM	44
		45
	UPM.VaR	46
Index		47

Co.LPM

Co-Lower Partial Moment (Lower Left Quadrant 4)

Description

This function generates a co-lower partial moment for between two equal length variables for any degree or target.

Usage

```
Co.LPM(degree.x, degree.y, x, y, target.x = mean(x),
  target.y = mean(y))
```

Arguments

```
degree.x integer; Degree for variable X. (degree.x = 0) is frequency, (degree.x = 1) is area.

degree.y integer; Degree for variable Y. (degree.y = 0) is frequency, (degree.y = 1) is area.

x a numeric vector.
```

Co.UPM 3

у	a numeric vector of equal length to x.
target.x	numeric; Typically the mean of Variable X for classical statistics equivalences, but does not have to be. (Vectorized)
target.y	numeric; Typically the mean of Variable Y for classical statistics equivalences, but does not have to be. (Vectorized)

Value

Co-LPM of two variables

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

Examples

```
set.seed(123)
x <- rnorm(100) ; y <- rnorm(100)
Co.LPM(0, 0, x, y, mean(x), mean(y))</pre>
```

Co.UPM

Co-Upper Partial Moment (Upper Right Quadrant 1)

Description

This function generates a co-upper partial moment between two equal length variables for any degree or target.

Usage

```
Co.UPM(degree.x, degree.y, x, y, target.x = mean(x),
  target.y = mean(y))
```

Arguments

degree.x	integer; Degree for variable X. (degree $. x = 0$) is frequency, (degree $. x = 1$) is area.
degree.y	integer; Degree for variable Y. (degree.y = 0) is frequency, (degree.y = 1) is area.
x	a numeric vector.
у	a numeric vector of equal length to x.
target.x	numeric; Typically the mean of Variable X for classical statistics equivalences, but does not have to be. (Vectorized)
target.y	numeric; Typically the mean of Variable Y for classical statistics equivalences, but does not have to be. (Vectorized)

D.LPM

Value

Co-UPM of two variables

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

Examples

```
set.seed(123)
x <- rnorm(100) ; y <- rnorm(100)
Co.UPM(0,0,x,y,mean(x),mean(y))</pre>
```

D.LPM

Divergent-Lower Partial Moment (Lower Right Quadrant 3)

Description

This function generates a divergent lower partial moment between two equal length variables for any degree or target.

Usage

```
D.LPM(degree.x, degree.y, x, y, target.x = mean(x), target.y = mean(y))
```

Arguments

degree.x	integer; Degree for variable X. (degree $. x = 0$) is frequency, (degree $. x = 1$) is area.
degree.y	integer; Degree for variable Y. (degree.y = 0) is frequency, (degree.y = 1) is area.
x	a numeric vector.
у	a numeric vector of equal length to x.
target.x	numeric; Typically the mean of Variable X for classical statistics equivalences, but does not have to be. (Vectorized)
target.y	numeric; Typically the mean of Variable Y for classical statistics equivalences, but does not have to be. (Vectorized)

Value

Divergent LPM of two variables

Author(s)

Fred Viole, OVVO Financial Systems

D.UPM 5

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

Examples

```
set.seed(123)
x <- rnorm(100) ; y <- rnorm(100)
D.LPM(0, 0, x, y, mean(x), mean(y))</pre>
```

D.UPM

Divergent-Upper Partial Moment (Upper Left Quadrant 2)

Description

This function generates a divergent upper partial moment between two equal length variables for any degree or target.

Usage

```
D.UPM(degree.x, degree.y, x, y, target.x = mean(x), target.y = mean(y))
```

Arguments

degree.x	integer; Degree for variable X. (degree $.x = 0$) is frequency, (degree $.x = 1$) is area.
degree.y	integer; Degree for variable Y. (degree.y = 0) is frequency, (degree.y = 1) is area.
X	a numeric vector.
У	a numeric vector of equal length to x.
target.x	numeric; Typically the mean of Variable X for classical statistics equivalences, but does not have to be. (Vectorized)
target.y	numeric; Typically the mean of Variable Y for classical statistics equivalences, but does not have to be. (Vectorized)

Value

Divergent UPM of two variables

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

```
set.seed(123)
x <- rnorm(100) ; y <- rnorm(100)
D.UPM(0, 0, x, y, mean(x), mean(y))</pre>
```

dy.dx

dy.dx	Partial Derivative dy/dx

Description

Returns the numerical partial derivate of y wrt x for a point of interest.

Usage

```
dy.dx(x, y, eval.point = median(x), deriv.order = 1, h = 0.05,
  deriv.method = "FS")
```

Arguments

X	a numeric vector.
У	a numeric vector.
eval.point	numeric; x point to be evaluated. Defaults to (eval.point = median(x)). Set to (eval.point = "overall") to find an overall partial derivative estimate.
deriv.order	numeric options: $(1, 2)$; 1 (default) for first derivative. For second derivative estimate of $f(x)$, set (deriv.order = 2).
h	numeric [0,]; Percentage step used for finite step method. Defaults to h = .05 representing a 5 percent step from the value of the independent variable.
deriv.method	method of derivative estimation, options: ("NNS", "FS"); Determines the partial derivative from the coefficient of the NNS.reg output when (deriv.method = "NNS") or generates a partial derivative using the finite step method (deriv.method = "FS") (Defualt).

Value

Returns the value of the partial derivative estimate for the given order.

Note

If a vector of derivatives is required, ensure (deriv.method = "FS").

Author(s)

Fred Viole, OVVO Financial Systems

References

```
Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995
```

 $\label{thm:comparametric} Vinod, H. and Viole, F. (2017) "Nonparametric Regression Using Clusters" \verb| https://link.springer.com/article/10.1007/s10614-017-9713-5| | Clusters | https://link.springer.com/article/10.1007/s10614-017-9713-017-9713-017-9713-017-9713-017-9713-017-9713-017-9713-017-9713-017-9713-017-9713-017-9713-017-9713-017-9713-$

dy.d_ 7

Examples

```
## Not run:
x <- seq(0, 2 * pi, pi / 100) ; y <-sin(x)
dy.dx(x, y, eval.point = 1.75)

# Vector of derivatives
dy.dx(x, y, eval.point = c(1.75, 2.5), deriv.method = "FS")
## End(Not run)</pre>
```

dy.d_

Partial Derivative dy/d_[wrt]

Description

Returns the numerical partial derivate of y with respect to [wrt] any regressor for a point of interest. Finite difference method is used with NNS.reg estimates as f(x + h) and f(x - h) values.

Usage

```
dy.d_(x, y, wrt, eval.points = "median", folds = 5, mixed = FALSE,
    plot = FALSE, messages = TRUE)
```

Arguments

x a numeric matrix or data frame.

y a numeric vector with compatible dimsensions to x.

wrt integer; Selects the regressor to differentiate with respect to.

eval.points numeric or options: ("mean", median", "last", "all"); Regressor points to be eval-

uated. (eval.points = "median") (default) to find the average partial derivative at the median of the variable with respect to. Set to (eval.points = "last") to find the average partial derivative at the last observation of the variable with respect to (relevant for time-series data). Set to (eval.points="mean") to find the average partial derivative at the mean of the variable with respect to. Set to (eval.points = "all") to find the overall partial derivative at every observa-

tion of the variable with respect to.

folds integer; 5 (default) Sets the number of folds in the NNS.stack procedure for

optimal n.best parameter.

mixed logical; FALSE (default) If mixed derivative is to be evaluated, set (mixed =

TRUE).

plot logical; FALSE (default) Set to (plot = TRUE) to view plot. Default setting is

(noise.reduction = "mean").

messages logical; TRUE (default) Prints status messages of cross-validation on n.best pa-

rameter for NNS.reg.

Value

Returns:

- dy.d_(...)\$"First Derivative" the 1st derivative
- dy.d_(...)\$"Second Derivative" the 2nd derivative
- dy.d_(...)*"Mixed Derivative" the mixed derivative (for two independent variables only).

8 LPM

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

Examples

```
## Not run:
set.seed(123) ; x_1 <- runif(100) ; x_2 <- runif(100) ; y <- x_1 ^ 2 * x_2 ^ 2
B <- cbind(x_1, x_2)

## To find average partial derivative of y wrt 1st regressor, only supply 1 value in [eval.points]
dy.d_(B, y, wrt = 1, eval.points = c(.5))

dy.d_(B, y, wrt = 1, eval.points = mean(B[, 1]))

## To find derivatives of y wrt 1st regressor and specified 2nd regressor
dy.d_(B, y, wrt = 1, eval.points = c(.5, .5))

## Known function analysis: [y = a ^ 2 * b ^ 2]
x_1 <- seq(0, 1, .1) ; x_2 <- seq(0, 1, .1)
B <- expand.grid(x_1, x_2) ; y <- B[, 1] ^ 2 * B[, 2] ^ 2
dy.d_(B, y, wrt = 1, eval.points = c(.5, .5))

## End(Not run)</pre>
```

LPM

Lower Partial Moment

Description

This function generates a univariate lower partial moment for any degree or target.

Usage

```
LPM(degree, target, variable)
```

Arguments

degree integer; (degree = 0) is frequency, (degree = 1) is area.

target numeric; Typically set to mean, but does not have to be. (Vectorized)

variable a numeric vector.

Value

LPM of variable

Author(s)

Fred Viole, OVVO Financial Systems

LPM.ratio 9

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

Examples

```
set.seed(123)
x <- rnorm(100)
LPM(0, mean(x), x)</pre>
```

LPM.ratio

Lower Partial Moment RATIO

Description

This function generates a standardized univariate lower partial moment for any degree or target.

Usage

```
LPM.ratio(degree, target, variable)
```

Arguments

degree integer; (degree = 0) is frequency, (degree = 1) is area.

target numeric; Typically set to mean, but does not have to be. (Vectorized)

variable a numeric vector.

Value

Standardized LPM of variable

Author(s)

Fred Viole, OVVO Financial Systems

References

```
Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995
```

Viole, F. (2017) "Continuous CDFs and ANOVA with NNS" https://ssrn.com/abstract=3007373

```
set.seed(123)
x <- rnorm(100)
LPM.ratio(0, mean(x), x)

## Not run:
## Empirical CDF (degree = 0)
lpm_cdf <- LPM.ratio(0, sort(x), x)
plot(sort(x), lpm_cdf)</pre>
```

10 LPM.VaR

```
## Continuous CDF (degree = 1)
lpm_cdf_1 <- LPM.ratio(1, sort(x), x)
plot(sort(x), lpm_cdf_1)

## Joint CDF
x <- rnorm(5000); y <- rnorm(5000)
plot3d(x, y, Co.LPM(0, 0, sort(x), sort(y), x, y), col = "blue", xlab = "X", ylab = "Y",
zlab = "Probability", box = FALSE)

## End(Not run)</pre>
```

LPM.VaR

LPM VaR

Description

Generates a value at risk (VaR) based on the Lower Partial Moment ratio.

Usage

```
LPM.VaR(percentile, degree, x)
```

Arguments

percentile numeric [0, 1]; The percentile for left-tail VaR.

degree integer; (degree = 0) for discrete distributions, (degree = 1) for continuous

distributions.

x a numeric vector.

Value

Returns a numeric value representing the point at which "percentile" of the area of x is above.

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

```
set.seed(123)
x <- rnorm(100)
## For 95th percentile VaR (left-tail)
LPM.VaR(0.95, 0, x)</pre>
```

NNS.ANOVA 11

Description

Analysis of variance (ANOVA) based on lower partial moment CDFs for multiple variables. Returns a degree of certainty the difference in sample means is zero, not a p-value.

Usage

```
NNS.ANOVA(control, treatment, confidence.interval = 0.95,
  tails = "Both", pairwise = FALSE, plot = TRUE, binary = TRUE)
```

Arguments

control a numeric vector, matrix or data frame. treatment NULL (default) a numeric vector, matrix or data frame. confidence.interval numeric [0, 1]; The confidence interval surrounding the control mean when (binary = TRUE). Defaults to (confidence.interval = 0.95). tails options: ("Left", "Right", "Both"). tails = "Both"(Default) Selects the tail of the distribution to determine effect size. logical; FALSE (defualt) Returns pairwise certainty tests when set to pairwise pairwise = TRUE. logical; TRUE (default) Returns the boxplot of all variables along with grand plot mean identification. When (binary = TRUE), returns the boxplot of both variables along with grand mean identification and confidence interval thereof. binary logical; TRUE (default) Selects binary analysis between a control and treatment variable.

Value

For (binary = FALSE) returns the degree certainty the difference in sample means is zero [0, 1]. For (binary = TRUE) returns:

- "Control Mean"
- "Treatment Mean"
- "Grand Mean"
- "Control CDF"
- "Treatment CDF"
- "Certainty" the certainty of the same population statistic
- "Lower Bound Effect" and "Upper Bound Effect" the effect size of the treatment for the specified confidence interval

Author(s)

Fred Viole, OVVO Financial Systems

12 NNS.ARMA

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

Examples

```
### Binary analysis and effect size
set.seed(123)
x <- rnorm(100); y <- rnorm(100)
NNS.ANOVA(control = x, treatment = y)
### Two variable analysis with no control variable
A \leftarrow cbind(x, y)
NNS.ANOVA(A)
### Multiple variable analysis with no control variable
set.seed(123)
x \leftarrow rnorm(100); y \leftarrow rnorm(100); z \leftarrow rnorm(100)
A \leftarrow cbind(x, y, z)
NNS.ANOVA(A)
```

NNS.ARMA

NNS ARMA

Description

Autoregressive model incorporating nonlinear regressions of component series.

Usage

```
NNS.ARMA(variable, h = 1, training.set = NULL,
  seasonal.factor = TRUE, weights = NULL, best.periods = 2,
 negative.values = FALSE, method = "nonlin", dynamic = FALSE,
 plot = TRUE, seasonal.plot = TRUE, intervals = FALSE,
  ncores = NULL)
```

Arguments

variable a numeric vector.

integer; 1 (default) Number of periods to forecast.

training.set numeric; NULL (defualt) Sets the number of variable observations

(variable[1 : training.set]) to monitor performance of forecast over in-

sample range.

seasonal.factor

logical or integer(s); TRUE (default) Automatically selects the best seasonal lag from the seasonality test. To use weighted average of all seasonal lags set to (seasonal.factor = FALSE). Otherwise, directly input known frequency integer lag to use, i.e. (seasonal.factor = 12) for monthly data. Multiple frequency integers can also be used, i.e. (seasonal.factor = c(12, 24, 36))

weights

numeric; NULL (default) sets the weights of the seasonal.factor vector when specified as integers. If (weights = NULL) each seasonal.factor is weighted

on its NNS.seas result and number of observations it contains.

NNS.ARMA 13

best.periods integer; [2] (default) used in conjuction with (seasonal.factor = FALSE), uses

the best.periods number of detected seasonal lags instead of ALL lags when

(seasonal.factor = FALSE).

negative.values

logical; FALSE (default) If the variable can be negative, set to (negative.values = TRUE). If there are negative values within the variable, negative.values will

automatically be detected.

method options: ("lin", "nonlin", "both"); "nonlin" (default) To select the regression

type of the component series, select (method = "both") where both linear and nonlinear estimates are generated. To use a nonlineaer regression, set to (method

= "nonlin"); to use a linear regression set to (method = "lin").

dynamic logical; FALSE (default) To update the seasonal factor with each forecast point,

set to (dynamic = TRUE). The default is (dynamic = FALSE) to retain the origi-

nal seasonal factor from the inputted variable for all ensuing h.

plot logical; TRUE (default) Returns the plot of all periods exhibiting seasonality and

the variable level reference in upper panel. Lower panel returns original data

and forecast.

seasonal.plot logical; TRUE (default) Adds the seasonality plot above the forecast. Will be set

to FALSE if no seasonality is detected or seasonal.factor is set to an integer

value.

intervals logical; FALSE (default) Plots the surrounding forecasts around the final esti-

mate when (intervals = TRUE) and (seasonal.factor = FALSE). There are

no other forecasts to plot when a single seasonal. factor is selected.

ncores integer; value specifying the number of cores to be used in the parallelized pro-

cedure. If NULL (default), the number of cores to be used is equal to half the

number of cores of the machine - 1.

Value

Returns a vector of forecasts of length (h).

Note

For monthly data series, increased accuracy may be realized from forcing seasonal factors to multiples of 12. For example, if the best periods reported are: $\{37, 47, 71, 73\}$ use (seasonal.factor = c(36, 48, 72)).

(seasonal.factor = FALSE) can be a very comutationally expensive exercise due to the number of seasonal periods detected.

If error encountered when (seasonal.factor = TRUE):

"NaNs produced Error in seq.default(length(variable)+1,1,-lag[i]) : wrong sign in 'by' argument"

use the combination of (seasonal.factor = FALSE, best.periods = 1).

Author(s)

Fred Viole, OVVO Financial Systems

14 NNS.ARMA.optim

References

```
Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments"
https://www.amazon.com/dp/1490523995
Viole, F. (2019) "Forecasting Using NNS" https://ssrn.com/abstract=3382300
```

Examples

```
## Nonlinear NNS.ARMA using AirPassengers monthly data and 12 period lag
NNS.ARMA(AirPassengers, h = 45, training.set = 100, seasonal.factor = 12, method = "nonlin")
## Linear NNS.ARMA using AirPassengers monthly data and 12, 24, and 36 period lags
NNS.ARMA(AirPassengers, h = 45, training.set = 120, seasonal.factor = c(12, 24, 36), method = "lin")
## Nonlinear NNS.ARMA using AirPassengers monthly data and 2 best periods lag
NNS.ARMA(AirPassengers, h = 45, training.set = 120, seasonal.factor = FALSE, best.periods = 2)
## End(Not run)
```

NNS.ARMA.optim

NNS ARMA Optimizer

Description

Wrapper function for optimizing any combination of a given seasonal.factor vector in NNS.ARMA. Minimum sum of squared errors (forecast-actual) is used to determine optimum across all NNS.ARMA methods.

Usage

```
NNS.ARMA.optim(variable, training.set, seasonal.factor,
  negative.values = FALSE, obj.fn = expression(sum((predicted -
  actual)^2)), objective = "min", linear.approximation = TRUE,
 print.trace = TRUE, ncores = NULL)
```

mize the objective function obj.fn.

Arguments

objective

variable a numeric vector. numeric; NULL (defualt) Sets the number of variable observations training.set seasonal.factor integers; Multiple frequency integers considered for NNS.ARMA model, i.e. (seasonal.factor = c(12,24,36))negative.values logical; FALSE (default) If the variable can be negative, set to (negative.values expression; expression(sum((predicted -actual)^2)) (default) Sum of squared obj.fn errors is the default objective function. Any expression() using the specific terms predicted and actual can be used. options: ("min", "max") "min" (default) Select whether to minimize or maxiNNS.ARMA.optim 15

linear.approximation

logical; TRUE (default) Uses the best linear output from NNS.reg to generate a nonlinear and mixture regression for comparison. FALSE is a more exhaustive search over the objective space.

print.trace

logical; TRUE (defualt) Prints current iteration information. Suggested as backup in case of error, best parameters to that point still known and copyable!

ncores

integer; value specifying the number of cores to be used in the parallelized procedure. If NULL (default), the number of cores to be used is equal to half the number of cores of the machine.

Value

Returns a list containing:

- \$period a vector of optimal seasonal periods
- \$weights the optimal weights of each seasonal period between an equal weight or NULL weighting
- \$obj. fn the objective function value
- \$method the method identifying which NNS.ARMA method was used.
- \$bias.shift a numerical result of the overall bias of the optimum objective function result. To be added to the final result when using the NNS.ARMA with the derived parameters.

Note

- The number of combinations will grow prohibitively large, they should be kept as small as possible. seasonal.factor containing an element too large will result in an error. Please reduce the maximum seasonal.factor.
- If variable cannot logically assume negative values, then the \$bias.shift must be limited to 0 via a pmax(0,...) call.

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

```
## Nonlinear NNS.ARMA period optimization using 2 yearly lags on AirPassengers monthly data
## Not run:
nns.optims <- NNS.ARMA.optim(AirPassengers[1:132], training.set = 120,
seasonal.factor = seq(12, 24, 6))

## Then use optimal parameters in NNS.ARMA to predict 12 periods in-sample.
## Note the {$bias.shift} usage in the {NNS.ARMA} function:
nns.estimates <- NNS.ARMA(AirPassengers, h = 12, training.set = 132,
seasonal.factor = nns.optims$periods, method = nns.optims$method) + nns.optims$bias.shift
## If variable cannot logically assume negative values</pre>
```

NNS.boost

```
nns.estimates <- pmax(0, nns.estimates)
## End(Not run)</pre>
```

NNS.boost

NNS Boost

Description

Ensemble method for classification using the predictions of the NNS multivariate regression NNS.reg collected from uncorrelated feature combinations.

Usage

```
NNS.boost(IVs.train, DV.train, IVs.test, type = NULL,
  representative.sample = FALSE, depth = NULL, n.best = NULL,
  learner.trials = NULL, epochs = NULL, CV.size = 0.25,
  ts.test = NULL, folds = 5, threshold = NULL,
  obj.fn = expression(sum((predicted - actual)^2)), objective = "min",
  extreme = FALSE, feature.importance = TRUE, status = TRUE,
  ncores = NULL)
```

Arguments

CV.size

IVs.train	a matrix or data frame of variables of numeric or factor data types.
DV.train	a numeric or factor vector with compatible dimsensions to (IVs.train).
IVs.test	a matrix or data frame of variables of numeric or factor data types with compatible dimsensions to ($IVs.train$).
type	NULL (default). To perform a classification of disrete integer classes from factor target variable (DV.train), set to (type = "CLASS"), else for continuous (DV.train) set to (type = NULL).
representative.	sample
	logical; FALSE (default) Reduces observations of IVs.train to a set of representative observations per regressor.
depth	options: (integer, NULL, "max"); NULL (default) Specifies the order parameter in the NNS.reg routine, assigning a number of splits in the regressors. (depth = "max") will be significantly faster, but increase the variance of results.
n.best	integer; NULL (default) Sets the number of nearest regression points to use in weighting for multivariate regression at $sqrt(\# of regressors)$. Analogous to k in a k Nearest Neighbors algorithm. If NULL, determines the optimal clusters via the NNS.stack procedure.
learner.trials	integer; NULL (default) Sets the number of trials to obtain an accuracy threshold level. Number of observations in the training set is the default setting.
epochs	integer; $2*length(DV.train)$ (default) Total number of feature combinations to run.

numeric [0, 1]; (CV.size = .25) (default) Sets the cross-validation size. De-

faults to 0.25 for a 25 percent random sampling of the training set.

NNS.boost 17

ts.test	integer; NULL (default) Sets the length of the test set for time-series data; typically 2*h parameter value from NNS.ARMA or double known periods to forecast.
folds	integer; 5 (default) Sets the number of folds in the NNS.stack procedure for optimal n.best parameter.

numeric; NULL (default) Sets the obj. fn threshold to keep feature combinations.

obj.fn expression; expression(sum((predicted -actual)^2)) (default) Sum of squared

errors is the default objective function. Any expression() using the specific terms predicted and actual can be used. Automatically selects an accuracy

measure when (type = "CLASS").

objective options: ("min", "max") "max" (default) Select whether to minimize or maxi-

mize the objective function obj.fn.

extreme logical; FALSE (default) Uses the maximum (minimum) threshold obtained

from the learner.trials, rather than the upper (lower) quintile level for max-

imization (minimization) objective.

feature.importance

threshold

logical; TRUE (default) Plots the frequency of features used in the final estimate.

status logical; TRUE (default) Prints status update message in console.

ncores integer; value specifying the number of cores to be used in the parallelized pro-

cedure. If NULL (default), the number of cores to be used is equal to the number

of cores of the machine - 1.

Value

Returns a vector of fitted values for the dependent variable test set \$results, and the final feature loadings \$feature.weights.

Note

Like a logistic regression, the (type = "CLASS") setting is not necessary for target variable of two classes e.g. [0, 1].

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. (2016) "Classification Using NNS Clustering Analysis" https://ssrn.com/abstract=2864711

```
## Using 'iris' dataset where test set [IVs.test] is 'iris' rows 141:150.
## Not run:
a <- NNS.boost(iris[1:140, 1:4], iris[1:140, 5],
IVs.test = iris[141:150, 1:4],
epochs = 100, learner.trials = 100,
type = "CLASS")
## Test accuracy
mean( a$results == as.numeric(iris[141:150, 5]))</pre>
```

NNS.caus

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

NNS.caus

NNS Causation

Description

Returns the causality from observational data between two variables.

Usage

```
NNS.caus(x, y, factor.2.dummy = FALSE, tau = 0, plot = FALSE)
```

Arguments

x a numeric vector, matrix or data frame.

y NULL (default) or a numeric vector with compatible dimsensions to x.

factor.2.dummy logical; FALSE (default) Automatically augments variable matrix with numerical dummy variables based on the levels of factors. Includes dependent variable y.

tau options: ("cs", "ts", integer); 0 (default) Number of lagged observations to consider (for time series data). Otherwise, set (tau = "cs") for cross-sectional data. (tau = "ts") automatically selects the lag of the time series data, while (tau = [integer]) specifies a time series lag.

plot logical; FALSE (default) Plots the raw variables, tau normalized, and cross-normalized variables.

Value

Returns the directional causation $(x \longrightarrow y)$ or $(y \longrightarrow x)$ and net quantity of association. For causal matrix, directional causation is returned as ([column variable] \longrightarrow [row variable]). Negative numbers represent causal direction attributed to [row variable].

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

```
## Not run:
## x causes y...
set.seed(123)
x <- rnorm(1000) ; y <- x ^ 2
NNS.caus(x, y, tau = "cs")</pre>
```

NNS.cor

```
## Causal matrix without per factor causation
NNS.caus(iris, tau = 0)
## Causal matrix with per factor causation
NNS.caus(iris, factor.2.dummy = TRUE, tau = 0)
## End(Not run)
```

NNS.cor

NNS Correlation

Description

Returns the nonlinear correlation between two variables based on higher order partial moment matrices measured by frequency or area.

Usage

```
NNS.cor(x, y = NULL, order = NULL, degree = NULL)
```

Arguments

x a numeric vector, matrix or data frame.

y NULL (default) or a numeric vector with compatible dimsensions to x.

order integer; Controls the level of quadrant partitioning. Defualts to (order = NULL).

Errors can generally be rectified by setting (order = 1).

degree integer; (degree = 0) is frequency based correlations, while (degree = 1) is

for area based correlations. Defaults to (degree = 0) for smaller number of

observations.

Value

Returns nonlinear correlation coefficient between two variables, or nonlinear correlation matrix for matrix input.

Note

p-values and confidence intervals can be obtained from sampling random permutations of y_p and running NNS. $dep(x,y_p)$ to compare against a null hypothesis of 0 correlation or independence between x,y.

See NNS.dep for examples.

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

20 NNS.dep

Examples

```
## Not run:
set.seed(123)
## Pairwise Correlation
x <- rnorm(100); y <- rnorm(100)
NNS.cor(x, y)
## Correlation Matrix
x <- rnorm(100); y <- rnorm(100); z <- rnorm(100)
B \leftarrow cbind(x, y, z)
NNS.cor(B)
## End(Not run)
```

NNS.dep

NNS Dependence

Description

Returns the dependence and nonlinear correlation between two variables based on higher order partial moment matrices measured by frequency or area.

Usage

```
NNS.dep(x, y = NULL, order = NULL, degree = NULL,
 print.map = FALSE, ncores = NULL)
```

Arguments

a numeric vector, matrix or data frame. NULL (default) or a numeric vector with compatible dimsensions to x. У integer; Controls the level of quadrant partitioning. Defaults to (order = NULL). order Errors can generally be rectified by setting (order = 1). Will not partition further if less than 4 observations exist in a quadrant. integer; Defaults to NULL to allow number of observations to be "degree" degree determinant. print.map logical; FALSE (default) Plots quadrant means. integer; value specifying the number of cores to be used in the parallelized proncores

cedure. If NULL (default), the number of cores to be used is equal to the number

of cores of the machine - 1.

Value

Returns the bi-variate "Correlation" and "Dependence" or correlation / dependence matrix for matrix input.

Note

p-values and confidence intervals can be obtained from sampling random permutations of y_p and running NNS.dep(x,y_p) to compare against a null hypothesis of 0 correlation or independence between x, y.

NNS.dep 21

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

```
## Not run:
set.seed(123)
x <- rnorm(100); y <- rnorm(100)
NNS.dep(x, y)
## Correlation / Dependence Matrix
x \leftarrow rnorm(100); y \leftarrow rnorm(100); z \leftarrow rnorm(100)
B \leftarrow cbind(x, y, z)
NNS.dep(B)
## p-values for [NNS.dep]
x \leftarrow seq(-5, 5, .1); y \leftarrow x^2 + rnorm(length(x))
nns\_cor\_dep \leftarrow NNS.dep(x, y, print.map = TRUE)
nns_cor_dep
\#\# Create permutations of y
y_p <- replicate(1000, sample.int(length(y)))</pre>
## Generate new correlation and dependence measures on each new permutation of y
nns.mc \leftarrow apply(y_p, 2, function(g) NNS.dep(x, y[g]))
## Store results
cors <- unlist(lapply(nns.mc, "[[", 1))</pre>
deps <- unlist(lapply(nns.mc, "[[", 2))</pre>
## View results
hist(cors)
hist(deps)
## Left tailed correlation p-value
cor_p_value <- LPM(0, nns_cor_dep$Correlation, cors)</pre>
cor_p_value
## Right tailed correlation p-value
cor_p_value <- UPM(0, nns_cor_dep$Correlation, cors)</pre>
cor_p_value
## Confidence Intervals
## For 95th percentile VaR (both-tails) see [LPM.VaR] and [UPM.VaR]
## Lower CI
LPM.VaR(.975, 0, cors)
## Upper CI
UPM.VaR(.975, 0, cors)
```

NNS.dep.hd

```
## Left tailed dependence p-value
dep_p_value <- LPM(0, nns_cor_dep$Dependence, deps)
dep_p_value

## Right tailed dependence p-value
dep_p_value <- UPM(0, nns_cor_dep$Dependence, deps)
dep_p_value

## End(Not run)</pre>
```

NNS.dep.base

NNS Dependence Base

Description

Internal function for NNS dependence NNS.dep parallel instances.

Usage

```
NNS.dep.base(x, y = NULL, order = NULL, degree = NULL,
    print.map = FALSE)
```

Arguments

X	from NNS.part
у	from NNS.part
order	from NNS.part
degree	from NNS.part
print.map	from NNS.part

Value

Returns NNS dependence.

NNS.dep.hd

NNS Co-Partial Moments Higher Dimension Dependence

Description

Determines higher dimension dependence coefficients based on degree 0 co-partial moments.

Usage

```
NNS.dep.hd(x, plot = FALSE, independence.overlay = FALSE)
```

NNS.diff 23

Arguments

x a numeric matrix or data frame.

plot logical; FALSE (default) Generates a 3d scatter plot with regression points using plot3d.

independence.overlay

logical; FALSE (default) Creates and overlays independent Co.LPM and Co.UPM regions to visually reference the difference in dependence from the data.frame of variables being analyzed. Under independence, the light green and red shaded areas would be occupied by green and red data points respectively.

Value

- \$actual.observations Number of Co.LPM and Co.UPM observations.
- \$independent.null Expected number of Co.LPM and Co.UPM observations under the null hypothesis of independence.
- \$Dependence Multivariate nonlinear dependence coefficient [0,1]

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. (2016) "Beyond Correlation: Using the Elements of Variance for Conditional Means and Probabilities" http://ssrn.com/abstract=2745308.

Examples

```
set.seed(123) x \leftarrow rnorm(1000); y \leftarrow rnorm(1000); z \leftarrow rnorm(1000) A \leftarrow data.frame(x, y, z) NNS.dep.hd(A, plot = TRUE, independence.overlay = TRUE)
```

NNS.diff

NNS Numerical Differentiation

Description

Determines numerical derivative of a given function using projected secant lines on the y-axis. These projected points infer finite steps h, in the finite step method.

Usage

```
NNS.diff(f, point, h = 0.1, tol = 1e-10, print.trace = FALSE)
```

24 NNS.distance

Arguments

f an expression or call or a formula with no lhs.

point numeric; Point to be evaluated for derivative of a given function f.

h numeric [0, ...]; Initial step for secant projection. Defaults to (h = 0.1).

tol numeric; Sets the tolerance for the stopping condition of the inferred h. Defualts

to (tol = 1e-10).

print.trace logical; FALSE (default) Displays each iteration, lower y-intercept, upper y-

intercept and inferred h.

Value

Returns a matrix of values, intercepts, derivatives, inferred step sizes for multiple methods of estimation.

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

Examples

```
f <- function(x) sin(x) / x
NNS.diff(f, 4.1)</pre>
```

NNS.distance

NNS Distance

Description

Internal function for NNS multivariate regression NNS.reg parallel instances.

Usage

```
NNS.distance(rpm, dist.estimate, type, k)
```

Arguments

rpm REGRESSION.POINT.MATRIX from NNS.reg

dist.estimate Vector to generate distances from.

type "L1", "L2" or "DTW" k n.best from NNS.reg

Value

Returns sum of weighted distances.

NNS.FSD 25

NNS . FSD NNS FSD Test

Description

Bi-directional test of first degree stochastic dominance using lower partial moments.

Usage

```
NNS.FSD(x, y, type = "discrete")
```

Arguments

x a numeric vector.

y a numeric vector.

type options: ("discrete", "continuous"); "discrete" (default) selects the type of CDF.

Value

Returns one of the following FSD results: "X FSD Y", "Y FSD X", or "NO FSD EXISTS".

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2016) "LPM Density Functions for the Computation of the SD Efficient Set." Journal of Mathematical Finance, 6, 105-126. http://www.scirp.org/Journal/PaperInformation.aspx?PaperID=63817.

Viole, F. (2017) "A Note on Stochastic Dominance." https://ssrn.com/abstract=3002675.

```
set.seed(123)
x <- rnorm(100) ; y <- rnorm(100)
NNS.FSD(x, y)</pre>
```

NNS.FSD.uni

NNS.FSD.uni

NNS FSD Test uni-directional

Description

Uni-directional test of first degree stochastic dominance using lower partial moments used in SD Efficient Set routine.

Usage

```
NNS.FSD.uni(x, y, type = "discrete")
```

Arguments

x a numeric vector.y a numeric vector.

type options: ("discrete", "continuous"); "discrete" (default) selects the type of

CDF.

Value

```
Returns (1) if "X FSD Y", else (0).
```

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2016) "LPM Density Functions for the Computation of the SD Efficient Set." Journal of Mathematical Finance, 6, 105-126. http://www.scirp.org/Journal/PaperInformation.aspx?PaperID=63817.

Viole, F. (2017) "A Note on Stochastic Dominance." https://ssrn.com/abstract=3002675.

```
set.seed(123)
x <- rnorm(100) ; y <- rnorm(100)
NNS.FSD.uni(x, y)</pre>
```

NNS.norm 27

|--|

Description

Normalizes a matrix of variables based on nonlinear scaling normalization method.

Usage

```
NNS.norm(A, linear = FALSE, chart.type = NULL, location = "topleft")
```

Arguments

A a numeric matrix or data frame.

linear logical; FALSE (default) Performs a linear scaling normalization, resulting in equal means for all variables.

chart.type options: ("l", "b"); NULL (default). Set (chart.type = "l") for line, (chart.type

= "b") for boxplot.

location Sets the legend location within the plot, per the x and y co-ordinates used in base

graphics legend.

Value

Returns a data.frame of normalized values.

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

```
set.seed(123)
x <- rnorm(100) ; y<-rnorm(100)
A <- cbind(x, y)
NNS.norm(A)</pre>
```

28 NNS.part

Description

Creates partitions based on partial moment quadrant means, iteratively assigning identifications to observations based on those quadrants (unsupervised partitional and hierarchial clustering method). Basis for correlation NNS.cor, dependence NNS.dep, regression NNS.reg routines.

Usage

```
NNS.part(x, y, Voronoi = FALSE, type = NULL, order = NULL,
  obs.req = 8, min.obs.stop = TRUE, noise.reduction = "off")
```

Arguments

iguitettes				
	x	a numeric vector.		
	У	a numeric vector with compatible dimsensions to x.		
	Voronoi	logical; FALSE (default) Displays a Voronoi type diagram using partial moment quadrants.		
	type	NULL (default) Controls the partitioning basis. Set to (type = "XONLY") for X-axis based partitioning. Defaults to NULL for both X and Y-axis partitioning.		
	order	integer; Number of partial moment quadrants to be generated. (order = "max") will institute a perfect fit.		
	obs.req	integer; (8 default) Required observations per cluster where quadrants will not be further partitioned if observations are not greater than the entered value. Reduces minimum number of necessary observations in a quadrant to 1 when (obs.req = 1).		
	min.obs.stop	logical; TRUE (default) Stopping condition where quadrants will not be further partitioned if a single cluster contains less than the entered value of obs.req.		

the method of determing regression points options: ("mean", "median", "mode", "off"); (noise.reduction = "mean") uses means for partitions. (noise.reduction = "median") uses medians instead of means for partitions, while (noise.reduction = "mode") uses modes instead of means for partitions. Defaults to (noise.reduction

= "off") where an overall central tendency measure is used.

Value

Returns:

noise.reduction

- "dt" a data.table of x and y observations with their partition assignment "quadrant" in the 3rd column and their prior partition assignment "prior.quadrant" in the 4th column.
- "regression.points" the data.table of regression points for that given (order = ...).
- "order" the order of the final partition given "min.obs.stop" stopping condition.

Author(s)

Fred Viole, OVVO Financial Systems

NNS.PDF 29

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

Examples

```
set.seed(123)
x <- rnorm(100) ; y <- rnorm(100)
NNS.part(x, y)

## Data.table of observations and partitions
NNS.part(x, y, order = 1)$dt

## Regression points
NNS.part(x, y, order = 1)$regression.points

## Voronoi style plot
NNS.part(x, y, Voronoi = TRUE)

## Examine final counts by quadrant
DT <- NNS.part(x, y)$dt
DT[ , counts := .N, by = quadrant]
DT</pre>
```

NNS.PDF

NNS PDF

Description

This function generates an empirical PDF using continuous CDFs from LPM.ratio.

Usage

```
NNS.PDF(variable, degree = 1, target = NULL, bins = NULL,
plot = TRUE)
```

Arguments

variable a numeric vector.

degree integer; (degree = 0) is frequency, (degree = 1) (default) is area.

target a numeric range of values [a,b] where a < b. NULL (default) uses the variable

observations.

bins numeric; NULL (default) Selects number of observations as default bins.

plot logical; plots PDF.

Value

Returns a data.table containing the intervals used and resulting PDF of the variable.

Author(s)

Fred Viole, OVVO Financial Systems

NNS.reg

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

Examples

```
set.seed(123)
x <- rnorm(100)
NNS.PDF(x)
## Custom target range
NNS.PDF(x, target = c(-5, 5))</pre>
```

NNS.reg

NNS Regression

Description

Generates a nonlinear regression based on partial moment quadrant means.

Usage

```
NNS.reg(x, y, factor.2.dummy = TRUE, order = NULL, stn = 0.95,
   dim.red.method = NULL, tau = NULL, type = NULL, point.est = NULL,
   location = "top", return.values = TRUE, plot = TRUE,
   plot.regions = FALSE, residual.plot = TRUE, std.errors = FALSE,
   confidence.interval = NULL, threshold = 0, n.best = NULL,
   noise.reduction = "off", dist = "L2", ncores = NULL,
   multivariate.call = FALSE)
```

Arguments

x a vector, matrix or data frame of variables of numeric or factor data types.

y a numeric or factor vector with compatible dimsensions to x.

factor.2.dummy logical; TRUE (default) Automatically augments variable matrix with numerical

dummy variables based on the levels of factors.

order integer; Controls the number of partial moment quadrant means. Users are en-

couraged to try different (order = \dots) integer settings with (noise.reduction

= "off"). (order = "max") will force a limit condition perfect fit.

stn numeric [0, 1]; Signal to noise parameter, sets the threshold of (NNS.dep) which

reduces ("order") when (order = NULL). Defaults to 0.96 to ensure high de-

pendence for higher ("order") and endpoint determination.

dim.red.method options: ("cor", "NNS.dep", "NNS.caus", "all", NULL) method for determin-

ing synthetic X* coefficients. Selection of a method automatically engages the dimension reduction regression. The default is NULL for full multivariate regression. (dim.red.method = "NNS.dep") uses NNS.dep for nonlinear dependence weights, while (dim.red.method = "NNS.caus") uses NNS.caus for causal weights. (dim.red.method = "cor") uses standard linear correlation for weights. (dim.red.method = "all") averages all methods for further feature

engineering.

NNS.reg 31

tau options("ts", NULL); NULL(default) To be used in conjuction with (dim.red.method

= "NNS.caus") or (dim.red.method = "all"). If the regression is using time-

series data, set (tau = "ts") for more accurate causal analysis.

type NULL (default). To perform a classification, set to (type = "CLASS"). Like a

logistic regression, it is not necessary for target variable of two classes e.g. [0,

1].

point.est a numeric or factor vector with compatible dimsensions to x. Returns the fitted

value y.hat for any value of x.

location Sets the legend location within the plot, per the x and y co-ordinates used in base

graphics legend.

return.values logical; TRUE (default), set to FALSE in order to only display a regression plot

and call values as needed.

plot logical; TRUE (default) To plot regression.

plot.regions logical; FALSE (default). Generates 3d regions associated with each regression

point for multivariate regressions. Note, adds significant time to routine.

residual.plot logical; TRUE (default) To plot y.hat and Y.

std.errors logical; FALSE (default) To provide standard errors of each linear segment in the

"Fitted.xy" output.

confidence.interval

numeric $[0,\,1]$; NULL (default) Plots the associated confidence interval with the

estimate and reports the standard error for each individual segment.

threshold numeric [0, 1]; (threshold = 0) (default) Sets the threshold for dimension re-

duction of independent variables when (dim.red.method) is not NULL.

n.best integer; NULL (default) Sets the number of nearest regression points to use in

weighting for multivariate regression at sqrt(# of regressors). (n.best = "all") will select and weight all generated regression points. Analogous to k in a k Nearest Neighbors algorithm. Different values of n.best are tested using

cross-validation in NNS.stack.

noise.reduction

the method of determing regression points options: ("mean", "median", "mode", "off"); In low signal:noise situations, (noise.reduction = "mean") uses means for NNS.dep restricted partitions, (noise.reduction = "median") uses medians instead of means for NNS.dep restricted partitions, while (noise.reduction

= "mode") uses modes instead of means for NNS.dep restricted partitions. (noise.reduction

= "off") uses an overall central tendency measure for partitions.

dist options:("L1", "L2", "DTW", "FACTOR") the method of distance calculation;

Selects the distance calculation used. dist = "L2" (default) selects the Euclidean distance and (dist = "L1") seclects the Manhattan distance; (dist = "DTW") selects the dynamic time warping distance; (dist = "FACTOR") uses a

frequency.

ncores integer; value specifying the number of cores to be used in the parallelized pro-

cedure. If NULL (default), the number of cores to be used is equal to the number

of cores of the machine - 1.

multivariate.call

Internal parameter for multivariate regressions.

Value

UNIVARIATE REGRESSION RETURNS THE FOLLOWING VALUES:

NNS.reg

- "R2" provides the goodness of fit;
- "SE" returns the overall standard error of the estimate between y and y.hat;
- "Prediction.Accuracy" returns the correct rounded "Point.est" used in classifications versus the categorical y;
- "derivative" for the coefficient of the x and its applicable range;
- "Point.est" for the predicted value generated;
- "regression.points" provides the points used in the regression equation for the given order of partitions;
- "Fitted.xy" returns a data.table of x, y, y.hat, resid, NNS.ID, gradient;

MULTIVARIATE REGRESSION RETURNS THE FOLLOWING VALUES:

- "R2" provides the goodness of fit;
- "equation" returns the numerator of the synthetic X* dimension reduction equation as a data.table consisting of regressor and its coefficient. Denominator is simply the length of all coefficients > 0, returned in last row of equation data.table.
- "x.star" returns the synthetic X* as a vector;
- "rhs.partitions" returns the partition points for each regressor x;
- "RPM" provides the Regression Point Matrix, the points for each x used in the regression equation for the given order of partitions;
- "Point.est" returns the predicted value generated;
- "Fitted.xy" returns a data.table of x,y, y.hat, gradient, and NNS.ID.

Note

Please ensure point.est is of compatible dimensions to x, error message will ensue if not compatible.

Author(s)

Fred Viole, OVVO Financial Systems

References

```
Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995
```

Vinod, H. and Viole, F. (2017) "Nonparametric Regression Using Clusters" https://link.springer.com/article/10.1007/s10614-017-9713-5

Vinod, H. and Viole, F. (2018) "Clustering and Curve Fitting by Line Segments" https://www.preprints.org/manuscript/201801.0090/v1

```
## Not run:
set.seed(123)
x <- rnorm(100) ; y <- rnorm(100)
NNS.reg(x, y)
## Manual {order} selection
NNS.reg(x, y, order = 2)</pre>
```

NNS.SD.efficient.set 33

```
## Maximum {order} selection
NNS.reg(x, y, order = "max")
## x-only paritioning (Univariate only)
NNS.reg(x, y, type = "XONLY")
## For Multiple Regression:
x <- cbind(rnorm(100), rnorm(100), rnorm(100)); y <- rnorm(100)
NNS.reg(x, y, point.est = c(.25, .5, .75))
## For Multiple Regression based on Synthetic X* (Dimension Reduction):
x \leftarrow cbind(rnorm(100), rnorm(100), rnorm(100)); y \leftarrow rnorm(100)
NNS.reg(x, y, point.est = c(.25, .5, .75), dim.red.method = "cor")
## IRIS dataset examples:
# Dimension Reduction:
NNS.reg(iris[,1:4], iris[,5], dim.red.method = "cor", order = 5)
# Dimension Reduction using causal weights:
NNS.reg(iris[,1:4], iris[,5], dim.red.method = "NNS.caus", order = 5)
# Multiple Regression:
NNS.reg(iris[,1:4], iris[,5], order = 2, noise.reduction = "off")
# Classification:
NNS.reg(iris[,1:4], iris[,5], point.est = iris[1:10, 1:4], type = "CLASS")$Point.est
## To call fitted values:
x <- rnorm(100); y <- rnorm(100)
NNS.reg(x, y)$Fitted
## To call partial derivative (univariate regression only):
NNS.reg(x, y)$derivative
## End(Not run)
```

NNS.SD.efficient.set NNS SD Efficient Set

Description

Determines the set of stochastic dominant variables for various degrees.

Usage

```
NNS.SD.efficient.set(x, degree, type = "discrete")
```

Arguments

```
x a numeric matrix or data frame.

degree numeric options: (1, 2, 3); Degree of stochastic dominance test from (1, 2 or 3).

type options: ("discrete", "continuous"); "discrete" (default) selects the type of CDF.
```

34 NNS.seas

Value

Returns set of stochastic dominant variable names.

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2016) "LPM Density Functions for the Computation of the SD Efficient Set." Journal of Mathematical Finance, 6, 105-126. http://www.scirp.org/Journal/PaperInformation.aspx?PaperID=63817.

Viole, F. (2017) "A Note on Stochastic Dominance." https://ssrn.com/abstract=3002675.

Examples

```
set.seed(123)
x <- rnorm(100) ; y<-rnorm(100) ; z<-rnorm(100)
A <- cbind(x, y, z)
NNS.SD.efficient.set(A, 1)</pre>
```

NNS.seas

NNS Seasonality Test

Description

Seasonality test based on the coefficient of variation for the variable and lagged component series. A result of 1 signifies no seasonality present.

Usage

```
NNS.seas(variable, modulo = NULL, mod.only = TRUE, plot = TRUE)
```

Arguments

variable a numeric vector.

modulo integer(s); NULL (default) Used to find the nearest multiple(s) in the reported seasonal period.

mod.only logical; codeTRUE (default) Limits the number of seasonal periods returned to the specified modulo.

plot logical; TRUE (default) Returns the plot of all periods exhibiting seasonality and

the variable level reference.

Value

Returns a matrix of all periods exhibiting less coefficient of variation than the variable with "all.periods"; and the single period exhibiting the least coefficient of variation versus the variable with "best.period"; as well as a vector of "periods" for easy call into NNS.ARMA.optim. If no seasonality is detected, NNS.seas will return ("No Seasonality Detected").

NNS.SSD 35

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

Examples

```
set.seed(123)
x <- rnorm(100)

## To call strongest period based on coefficient of variation:
NNS.seas(x, plot = FALSE)$best.period

## Using modulos for logical seasonal inference:
NNS.seas(x, modulo = c(2,3,5,7), plot = FALSE)</pre>
```

NNS.SSD

NNS SSD Test

Description

Bi-directional test of second degree stochastic dominance using lower partial moments.

Usage

```
NNS.SSD(x, y)
```

Arguments

```
x a numeric vector.
y a numeric vector.
```

Value

Returns one of the following SSD results: "X SSD Y", "Y SSD X", or "NO SSD EXISTS".

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2016) "LPM Density Functions for the Computation of the SD Efficient Set." Journal of Mathematical Finance, 6, 105-126. http://www.scirp.org/Journal/PaperInformation.aspx?PaperID=63817.

```
set.seed(123)
x <- rnorm(100) ; y <- rnorm(100)
NNS.SSD(x, y)</pre>
```

36 NNS.stack

NNS.SSD.uni

NNS SSD Test uni-directional

Description

Uni-directional test of second degree stochastic dominance using lower partial moments used in SD Efficient Set routine.

Usage

```
NNS.SSD.uni(x, y)
```

Arguments

```
x a numeric vector.
y a numeric vector.
```

Value

```
Returns (1) if "X SSD Y", else (0).
```

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2016) "LPM Density Functions for the Computation of the SD Efficient Set." Journal of Mathematical Finance, 6, 105-126. http://www.scirp.org/Journal/PaperInformation.aspx?PaperID=63817.

Examples

```
set.seed(123)
x <- rnorm(100) ; y <- rnorm(100)
NNS.SSD.uni(x, y)</pre>
```

NNS.stack

NNS Stack

Description

Prediction model using the predictions of the NNS base models NNS.reg as features (i.e. meta-features) for the stacked model.

Usage

```
NNS.stack(IVs.train, DV.train, IVs.test = NULL, type = NULL,
  obj.fn = expression(sum((predicted - actual)^2)), objective = "min",
  dist = "L2", CV.size = NULL, ts.test = NULL, folds = 5,
  order = NULL, norm = NULL, method = c(1, 2),
  dim.red.method = "cor", status = TRUE, ncores = NULL)
```

NNS.stack 37

Arguments

IVs. train a vector, matrix or data frame of variables of numeric or factor data types.

DV. train a numeric or factor vector with compatible dimsensions to (IVs.train).

IVs. test a vector, matrix or data frame of variables of numeric or factor data types with

compatible dimsensions to (IVs.train).

type NULL (default). To perform a classification of disrete integer classes from fac-

tor target variable (DV.train), set to (type = "CLASS"), else for continuous (DV.train) set to (type = NULL). Like a logistic regression, this setting is not

necessary for target variable of two classes e.g. [0, 1].

obj.fn expression(sum((predicted -actual)^2)) (default) Sum of squared

errors is the default objective function. Any expression() using the specific

terms predicted and actual can be used.

objective options: ("min", "max") "min" (default) Select whether to minimize or maxi-

mize the objective function obj.fn.

dist options:("L1", "L2", "DTW", "FACTOR") the method of distance calculation;

Selects the distance calculation used. dist = "L2" (default) selects the Euclidean distance and (dist = "L1") seclects the Manhattan distance; (dist = "DTW") selects the dynamic time warping distance; (dist = "FACTOR") uses a

frequency.

CV. size numeric [0, 1]; NULL (default) Sets the cross-validation size if (IVs.test =

NULL). Defaults to 0.25 for a 25 percent random sampling of the training set

under (CV.size = NULL).

ts.test integer; NULL (default) Sets the length of the test set for time-series data; typ-

ically 2*h parameter value from NNS.ARMA or double known periods to fore-

cast.

folds integer; folds = 5 (default) Select the number of cross-validation folds.

order integer; NULL (default) Sets the order for NNS.reg, where (order = "max") is

the k-nearest neighbors equivalent.

norm options: ("std", "NNS", NULL); NULL (default) 3 settings offered: NULL, "std",

and "NNS". Selects the norm parameter in NNS.reg.

method numeric options: (1, 2); Select the NNS method to include in stack. (method =

1) selects NNS.reg; (method = 2) selects NNS.reg dimension reduction regression. Defaults to method = c(1,2), including both NNS regression methods in

the stack.

dim.red.method options: ("cor", "NNS.dep", "NNS.caus", "all") method for determining syn-

thetic X^* coefficients. (dim.red.method = "cor") (default) uses standard linear correlation for weights. (dim.red.method = "NNS.dep") uses NNS.dep for nonlinear dependence weights, while (dim.red.method = "NNS.caus") uses NNS.caus for causal weights. (dim.red.method = "all") averages all methods are considered as a constant of the constant of t

ods for further feature engineering.

status logical; TRUE (default) Prints status update message in console.

ncores integer; value specifying the number of cores to be used in the parallelized sub-

routine NNS.reg. If NULL (default), the number of cores to be used is equal to

the number of cores of the machine - 1.

38 NNS.stack

Value

Returns a vector of fitted values for the dependent variable test set for all models.

- "NNS.reg.n.best" returns the optimum "n.best" paramater for the NNS.reg multivariate regression. "SSE.reg" returns the SSE for the NNS.reg multivariate regression.
- "OBJfn.reg" returns the obj.fn for the NNS.reg regression.
- "NNS.dim.red.threshold" returns the optimum "threshold" from the NNS.reg dimension reduction regression.
- "OBJfn.dim.red" returns the obj.fn for the NNS.reg dimension reduction regression.
- "reg" returns NNS.reg output.
- "dim. red" returns NNS.reg dimension reduction regression output.
- "stack" returns the output of the stacked model.

Note

- Like a logistic regression, the (type = "CLASS") setting is not necessary for target variable of two classes e.g. [0, 1].
- Missing data should be handled prior as well using na.omit or complete.cases on the full dataset.

If error received:

```
"Error in is.data.frame(x) : object 'RP' not found" reduce the CV.size.
```

Author(s)

Fred Viole, OVVO Financial Systems

References

```
Viole, F. (2016) "Classification Using NNS Clustering Analysis" https://ssrn.com/abstract=2864711
```

```
## Using 'iris' dataset where test set [IVs.test] is 'iris' rows 141:150.
## Not run:
NNS.stack(iris[1:140, 1:4], iris[1:140, 5], IVs.test = iris[141:150, 1:4], type = "CLASS")

## Using 'iris' dataset to determine [n.best] and [threshold] with no test set.
NNS.stack(iris[ , 1:4], iris[ , 5], type = "CLASS")

## Selecting NNS.reg and dimension reduction techniques.
NNS.stack(iris[1:140, 1:4], iris[1:140, 5], iris[141:150, 1:4], method = c(1, 2), type = "CLASS")
## End(Not run)
```

NNS.term.matrix 39

NNS.term.matrix

NNS Term Matrix

Description

Generates a term matrix for text classification use in NNS.reg.

Usage

```
NNS.term.matrix(x, oos = NULL, names = FALSE)
```

Arguments

X	Text A two column dataset should be used. Concatenate text from original sources to comply with format. Also note the possibility of factors in "DV", so "as.numeric(as.character())" is used to avoid issues.
00S	Out-of-sample text dataset to be classified.
names	Column names for "IV" and "oos". Defaults to FALSE.

Value

Returns the text as independent variables "IV" and the classification as the dependent variable "DV". Out-of-sample independent variables are returned with "OOS".

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

40 NNS.TSD.uni

NNS.TSD

NNS TSD Test

Description

Bi-directional test of third degree stochastic dominance using lower partial moments.

Usage

```
NNS.TSD(x, y)
```

Arguments

```
x a numeric vector.y a numeric vector.
```

Value

Returns one of the following TSD results: "X TSD Y", "Y TSD X", or "NO TSD EXISTS".

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2016) "LPM Density Functions for the Computation of the SD Efficient Set." Journal of Mathematical Finance, 6, 105-126. http://www.scirp.org/Journal/PaperInformation.aspx?PaperID=63817.

Examples

```
set.seed(123)
x <- rnorm(100) ; y <- rnorm(100)
NNS.TSD(x, y)</pre>
```

NNS.TSD.uni

NNS TSD Test uni-directional

Description

Uni-directional test of third degree stochastic dominance using lower partial moments used in SD Efficient Set routine.

Usage

```
NNS.TSD.uni(x, y)
```

NNS.VAR 41

Arguments

```
x a numeric vector.
y a numeric vector.
```

Value

```
Returns (1) if "X TSD Y", else (0).
```

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2016) "LPM Density Functions for the Computation of the SD Efficient Set." Journal of Mathematical Finance, 6, 105-126. http://www.scirp.org/Journal/PaperInformation.aspx?PaperID=63817.

Examples

```
set.seed(123)
x <- rnorm(100) ; y <- rnorm(100)
NNS.TSD.uni(x, y)</pre>
```

NNS. VAR

NNS VAR

Description

Nonparametric vector autoregressive model incorporating NNS.ARMA estimates of variables into NNS.reg for a multi-variate time-series forecast.

Usage

```
NNS.VAR(variables, h, tau = 1, dim.red.method = "cor",
  obj.fn = expression(sum((predicted - actual)^2)), objective = "min",
  status = TRUE, ncores = NULL)
```

Arguments

variables a numeric matrix or data.frame of contemporaneous time-series to forecast.

h integer; 1 (default) Number of periods to forecast.

tau positive integer [> 0]; 1 (default) Number of lagged observations to consider for the time-series data.

dim.red.method options: ("cor", "NNS.dep", "NNS.caus", "all") method for reducing regressors via NNS.stack. (dim.red.method = "cor") (default) uses standard linear correlation for dimension reduction in the lagged variable matrix. (dim.red.method = "NNS.dep") uses NNS.dep for nonlinear dependence weights, while (dim.red.method = "NNS.caus") uses NNS.caus for causal weights. (dim.red.method = "all") averages all methods for further feature engineering.

42 NNS.VAR

obj.fn	expression; expression(sum((predicted -actual)^2)) (default) Sum of squared errors is the default objective function. Any expression() using the specific terms predicted and actual can be used.
objective	options: ("min", "max") "min" (default) Select whether to minimize or maximize the objective function obj.fn.
status	logical; TRUE (default) Prints status update message in console.
ncores	integer; value specifying the number of cores to be used in the parallelized sub- routine NNS.ARMA.optim. If NULL (default), the number of cores to be used is equal to the number of cores of the machine - 1.

Value

Returns the following matrices of forecasted variables:

- "eelevant_variables" Returns the relevant variables from the dimension reduction step.
- "univariate" Returns the univariate NNS.ARMA forecasts.
- "multivariate" Returns the multi-variate NNS.reg forecasts.
- "ensemble" Returns the ensemble of both "univariate" and "multivariate" forecasts.

Note

dim.red.method = "cor" is significantly faster than the other methods, but comes at the expense of ignoring possible nonlinear relationships between lagged variables.

Author(s)

Fred Viole, OVVO Financial Systems

References

```
Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

Viole, F. (2019) "Multi-variate Time-Series Forecasting: Nonparametric Vector Autoregression Using NNS" https://ssrn.com/abstract=3489550

Viole, F. (2019) "Forecasting Using NNS" https://ssrn.com/abstract=3382300

Vinod, H. and Viole, F. (2017) "Nonparametric Regression Using Clusters" https://link.springer.com/article/10.1007/s10614-017-9713-5

Vinod, H. and Viole, F. (2018) "Clustering and Curve Fitting by Line Segments" https://www.preprints.org/manuscript/201801.0090/v1
```

```
## Not run:
set.seed(123)
x <- rnorm(100) ; y <- rnorm(100) ; z <- rnorm(100)
A <- cbind(x = x, y = y, z = z)
NNS.VAR(A, h = 12, tau = 4, status = TRUE)
## End(Not run)</pre>
```

PM.matrix 43

PM.matrix	Partial Moment Matrix	

Description

This function generates a co-partial moment matrix for the specified co-partial moment.

Usage

```
PM.matrix(LPM.degree, UPM.degree, target = "mean", variable,
    pop.adj = FALSE)
```

Arguments

LPM.degree	integer; Degree for variable below target deviations. (degree = 0) is frequency, (degree = 1) is area.	
UPM.degree	integer; Degree for variable above target deviations. (degree = 0) is frequency, (degree = 1) is area.	
target	numeric; Typically the mean of Variable X for classical statistics equivalences, but does not have to be. (Vectorized) (target = "mean") (default) will set the target as the mean of every variable.	
variable	a numeric matrix or data.frame.	
pop.adj	logical; FALSE (default) Adjusts the sample co-partial moment matrices for population statistics.	

Value

Matrix of partial moment quadrant values (CUPM, DUPM, DLPM, CLPM), and overall covariance matrix. Uncalled quadrants will return a matrix of zeros.

Note

For divergent asymmetical "D.LPM" and "D.UPM" matrices, matrix is D.LPM(column, row, ...).

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

Viole, F. (2017) "Bayes' Theorem From Partial Moments" https://ssrn.com/abstract=3457377

44 UPM

Examples

```
set.seed(123)
x <- rnorm(100) ; y <- rnorm(100) ; z <- rnorm(100)
A <- cbind(x,y,z)
PM.matrix(LPM.degree = 1, UPM.degree = 1, target = "mean", variable = A)

## Use of vectorized numeric targets (target_x, target_y, target_z)
PM.matrix(LPM.degree = 1, UPM.degree = 1, target = c(0, 0.15, .25), variable = A)

## Calling Individual Partial Moment Quadrants
cov.mtx <- PM.matrix(LPM.degree = 1, UPM.degree = 1, target = "mean", variable = A)
cov.mtx$cupm

## Full covariance matrix
cov.mtx$cov.matrix</pre>
```

UPM

Upper Partial Moment

Description

This function generates a univariate upper partial moment for any degree or target.

Usage

```
UPM(degree, target, variable)
```

Arguments

degree integer; (degree = 0) is frequency, (degree = 1) is area.

target numeric; Typically set to mean, but does not have to be. (Vectorized)

variable a numeric vector.

Value

UPM of variable

Author(s)

Fred Viole, OVVO Financial Systems

References

```
Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995
```

```
set.seed(123)
x <- rnorm(100)
UPM(0, mean(x), x)</pre>
```

UPM.ratio 45

UPM.ratio

Upper Partial Moment RATIO

Description

This function generates a standardized univariate upper partial moment for any degree or target.

Usage

```
UPM.ratio(degree, target, variable)
```

Arguments

degree integer; (degree = 0) is frequency, (degree = 1) is area.

target numeric; Typically set to mean, but does not have to be. (Vectorized)

variable a numeric vector.

Value

Standardized UPM of variable

Author(s)

Fred Viole, OVVO Financial Systems

References

Viole, F. and Nawrocki, D. (2013) "Nonlinear Nonparametric Statistics: Using Partial Moments" https://www.amazon.com/dp/1490523995

```
set.seed(123)
x <- rnorm(100)
UPM.ratio(0, mean(x), x)

## Joint Upper CDF
## Not run:
x <- rnorm(5000) ; y <- rnorm(5000)
plot3d(x, y, Co.UPM(0, 0, sort(x), sort(y), x, y), col = "blue", xlab = "X", ylab = "Y",
zlab = "Probability", box = FALSE)

## End(Not run)</pre>
```

46 UPM.VaR

Description

Generates an upside value at risk (VaR) based on the Upper Partial Moment ratio

Usage

```
UPM.VaR(percentile, degree, x)
```

Arguments

percentile numeric [0, 1]; The percentile for right-tail VaR.

degree integer; (degree = 0) for discrete distributions, (degree = 1) for continuous

distributions.

x a numeric vector.

Value

Returns a numeric value representing the point at which "percentile" of the area of x is below.

```
set.seed(123)
x <- rnorm(100)
## For 95th percentile VaR (right-tail)
UPM.VaR(0.95, 0, x)</pre>
```

Index

```
Co.LPM, 2, 23
                                                     PM.matrix, 43
Co. UPM, 3, 23
                                                    UPM, 44
complete.cases, 38
                                                     UPM. ratio, 45
D.LPM, 4
                                                    UPM. VaR, 46
D. UPM, 5
data.frame, 27
data.table, 28, 32
dy.d_{-}, 7
dy.dx, 6
legend, 27, 31
LPM, 8
LPM. ratio, 9, 29
LPM. VaR, 10
na.omit, 38
NNS. ANOVA, 11
NNS. ARMA, 12, 14, 15, 17, 37, 41, 42
NNS.ARMA.optim, 14, 34, 42
NNS.boost, 16
NNS.caus, 18, 30, 37, 41
NNS.cor, 19, 28
NNS.dep, 19, 20, 22, 28, 30, 31, 37, 41
NNS.dep.base, 22
NNS.dep.hd, 22
NNS.diff, 23
NNS.distance, 24
NNS.FSD, 25
NNS.FSD.uni, 26
NNS.norm, 27
NNS.part, 22, 28
NNS.PDF, 29
NNS. reg, 6, 7, 16, 24, 28, 30, 36–39, 41, 42
NNS.SD.efficient.set, 33
NNS. seas, 12, 34
NNS.SSD, 35
NNS.SSD.uni, 36
NNS. stack, 7, 16, 17, 31, 36, 41
NNS.term.matrix, 39
NNS.TSD, 40
NNS.TSD.uni, 40
NNS. VAR, 41
plot3d, 23
```