pycast Documentation

Release 0.0.7-prealpha

Christian Schwarz

November 18, 2012

CONTENTS

1	pycast.common	1
2	TimeSeries	3
3	Smoothing Methods	9
4	Forecasting Methods	11
5	Error Measures	13
6	Optimization Methods	15
7	Indices and tables	17
Ру	thon Module Index	19
Index		21

CHAPTER

ONE

PYCAST.COMMON

```
pycast.common.profileme.profileMe
    alias of _ProfileDecorator

pycast.common.helper.linear_interpolation (first, last, steps)
```

Interpolates all missing values using linear interpolation.

Parameters

- **first** (*Numeric*) Start value for the interpolation.
- last (Numeric) End Value for the interpolation
- **steps** (*Integer*) Number of missing values that have to be calculated.

Returns Returns a list of floats containing only the missing values.

Return type List

TIMESERIES

class pycast.common.timeseries.**TimeSeries** (isNormalized=False, isSorted=False) A TimeSeries instance stores all relevant data for a real world time series. Warning TimeSeries instances are NOT threadsafe. add (otherTimeSeries) Creates a new TimeSeries instance containing the data of self and otherTimeSeries. Parameters otherTimeSeries (TimeSeries) - TimeSeries instance that will be merged with Returns Returns a new TimeSeries instance containing the data entries of self and otherTime-Series. **Return type** TimeSeries **__eq**__(otherTimeSeries) Returns if self and the other TimeSeries are equal. TimeSeries are equal to each other if: • they contain the same number of entries • each data entry in one TimeSeries is also member of the other one. Parameters otherTimeSeries (TimeSeries) - TimeSeries instance that is compared with self. **Returns** True if the TimeSeries objects are equal, False otherwise. Return type Boolean _getitem__(index) Returns the item stored at the TimeSeries index-th position. **Parameters index** (*Integer*) – Position of the element that should be returned. Starts at 0 **Returns** Returns a list containing [timestamp, data] lists. Return type List **Raise** Raises an IndexError if the index is out of range. init (isNormalized=False, isSorted=False) Initializes the TimeSeries.

Parameters

- isNormalized (Boolean) Within a normalized TimeSeries, all data points have the same temporal distance to each other. When this is True, the memory consumption of the TimeSeries might be reduced. Also algorithms will probably run faster on normalized TimeSeries. This should only be set to True, if the TimeSeries is realy normalized! TimeSeries normalization can be forced by executing TimeSeries.normalize().
- **isSorted** (*Boolean*) If all data points added to the time series are added in their ascending temporal order, this should set to True.

___iter__()

Returns an iterator that can be used to iterate over the data stored within the TimeSeries.

Returns Returns an iterator for the TimeSeries.

Return type Iterator

__len__()

Returns the number of data entries stored in the TimeSeries.

Returns Returns an Integer representing the number on data entries stored within the Time-Series.

Return type Integer

__setitem__(index, value)

Sets the item at the index-th position of the TimeSeries.

Parameters

- index (Integer) Index of the element that should be set.
- value (List) A list of the form [timestamp, data]

Raise Raises an IndexError if the index is out of range.

__str__()

Returns a string representation of the TimeSeries.

Returns

Returns a string representing the TimeSeries in the format:

"TimeSeries([timestamp, data], [timestamp, data], [timestamp, data])".

Return type String

__weakref_

list of weak references to the object (if defined)

add entry(timestamp, data, format=None)

Adds a new data entry to the TimeSeries.

Parameters

- **timestamp** Time stamp of the data. This has either to be a float representing the UNIX epochs or a string containing a timestamp in the given format.
- data (Numeric) Actual data value.
- **format** (*String*) Format of the given timestamp. This is used to convert the timestamp into UNIX epochs, if set. For valid examples take a look into the time.strptime() documentation.

apply (method)

Applies the given ForecastingAlgorithm or SmoothingMethod from the pycast.methods module to the TimeSeries.

Parameters method (*BaseMethod*) – Method that should be used with the TimeSeries. For more information about the methods take a look into their corresponding documentation.

classmethod convert_epoch_to_timestamp (timestamp, format)

Converts the given float representing UNIX-epochs into an actual timestamp.

Parameters

- **timestamp** (*Float*) Timestamp as UNIX-epochs.
- **format** (*String*) Format of the given timestamp. This is used to convert the timestamp from UNIX epochs. For valid examples take a look into the time.strptime() documentation.

Returns Returns the timestamp as defined in format.

Return type String

classmethod convert_timestamp_to_epoch (timestamp, format)

Converts the given timestamp into a float representing UNIX-epochs.

Parameters

- **timestamp** (*String*) Timestamp in the defined format.
- **format** (*String*) Format of the given timestamp. This is used to convert the timestamp into UNIX epochs. For valid examples take a look into the time.strptime() documentation.

Returns Returns an float, representing the UNIX-epochs for the given timestamp.

Return type Float

classmethod from_json (json, format=None)

Creates a new TimeSeries instance from the given json string.

Parameters

- **json** (*String*) JSON string, containing the time series data. This should be a string created by TimeSeries.to_json().
- **format** (*String*) Format of the given timestamps. This is used to convert the timestamps into UNIX epochs, if set. For valid examples take a look into the time.strptime() documentation.

Returns Returns a TimeSeries instance containing the data.

Return type TimeSeries

Warning This is probably an unsafe version! Only use it with JSON strings created by TimeSeries.to_json(). All assumtions regarding normalization and sort order will be ignored and set to default.

classmethod from_twodim_list (datalist, format=None)

Creates a new TimeSeries instance from the data stored inside a two dimensional list.

Parameters

- **datalist** (*List*) List containing multiple iterables with at least two values. The first item will always be used as timestamp in the predefined format, the second represents the value. All other items in those sublists will be ignored.
- **format** (*String*) Format of the given timestamp. This is used to convert the timestamp into UNIX epochs, if necessary. For valid examples take a look into the time.strptime() documentation.

Returns Returns a TimeSeries instance containing the data from datalist.

Return type TimeSeries

initialize_from_sql_cursor(sqlcursor, format=None)

Initializes the TimeSeries's data from the given SQL cursor.

Parameters

- **sqlcursor** (*SQLCursor*) Cursor that was holds the SQL result for any given "SELECT timestamp, value, ... FROM ..." SQL query. Only the first two attributes of the SQL result will be used.
- **format** (*String*) Format of the given timestamp. This is used to convert the timestamp into UNIX epochs, if set. For valid examples take a look into the time.strptime() documentation.

Returns Returns the number of entries added to the TimeSeries.

Return type Integer

is normalized()

Returns if the TimeSeries is normalized.

Returns Returns True if the TimeSeries is normalized. False otherwise.

Return type Boolean

is sorted()

Returns if the TimeSeries is sorted.

Returns Returns True if the TimeSeries is sorted ascending, False in all other cases.

Return type Boolean

normalize (normalizationLevel='minute', fusionMethod='average', interpolationMethod='linear') Normalizes the TimeSeries data points.

If this function is called, the TimeSeries gets ordered ascending automatically. The new timestamps will represent the center of each time bucket. Within a normalized TimeSeries, the temporal distance between two consecutive data points is constant.

Parameters

- **normalizationLevel** (*String*) Level of normalization that has to be applied. The available normalization levels are defined in timeseries. NormalizationLevels.
- **fusionMethod** (*String*) Normalization method that has to be used if multiple data entries exist within the same normalization bucket. The available methods are defined in timeseries. FusionMethods.
- interpolationMethod (String) Interpolation method that is used if a data entry at a specific time is missing. The available interpolation methods are defined in timeseries. InterpolationMethods.

Raise Raises a ValueError if a normalizationLevel, fusionMethod or interpolationMethod hanve an unknown value.

sort timeseries (ascending=True)

Sorts the data points within the TimeSeries according to their occurence inline.

Parameters ascending (*Boolean*) — Determines if the TimeSeries will be ordered ascending or decending. If this is set to decending once, the ordered parameter defined in TimeSeries. init () will be set to False FOREVER.

Returns Returns self for convenience.

Return type TimeSeries

sorted_timeseries (ascending=True)

Returns a sorted copy of the TimeSeries, preserving the original one.

As an assumtion this new TimeSeries is not ordered anymore if a new value is added.

Parameters ascending (*Boolean*) – Determines if the TimeSeries will be ordered ascending or decending.

Returns Returns a new TimeSeries instance sorted in the requested order.

Return type TimeSeries

to_gnuplot_datafile (datafilepath, format=None)

Dumps the TimeSeries into a gnuplot compatible data file.

Parameters

- **datafilepath** (*String*) Path used to create the file. If that file already exists, it will be overwritten!
- **format** (*String*) Format of the timestamp. This is used to convert the timestamp from UNIX epochs, if set. For valid examples take a look into the time.strptime() documentation.

Returns Returns True if the data could be written, False otherwise.

Return type Boolean

to_json(format=None)

Returns a JSON representation of the TimeSeries data.

Parameters format (*String*) – Format of the given timestamp. This is used to convert the timestamp into UNIX epochs, if set. For valid examples take a look into the time.strptime() documentation.

Returns Returns a basestring, containing the JSON representation of the current data stored within the TimeSeries.

Return type String

to_twodim_list (format=None)

Serializes the TimeSeries data into a two dimensional list of [timestamp, value] pairs.

Parameters format (*String*) – Format of the timestamp. This is used to convert the timestamp from UNIX epochs, if set. For valid examples take a look into the time.strptime() documentation.

Returns Returns a two dimensional list containing [timestamp, value] pairs.

Return type List

SMOOTHING METHODS

class pycast.methods.SimpleMovingAverage (windowsize=5)

Implements the simple moving average.

The SMA algorithm will calculate the average value at time t based on the datapoints between [t - floor(windowsize / 2), t + floor(windowsize / 2)].

Explanation: http://en.wikipedia.org/wiki/Moving_average

__init__ (windowsize=5)

Initializes the SimpleMovingAverage.

Parameters windowsize (*Integer*) – Size of the SimpleMovingAverages window.

Raise Raises a ValueError if windowsize is an even or not larger than zero.

execute (timeSeries)

Creates a new TimeSeries containing the SMA values for the predefined windowsize.

Parameters timeSeries (*TimeSeries*) – The TimeSeries used to calculate the simple moving average values.

Returns TimeSeries object containing the smooth moving average.

Return type TimeSeries

Note This implementation aims to support independent for loop execution.

FORECASTING METHODS

class pycast.methods.ExponentialSmoothing(smoothingFactor=0.1, valuesToForecast=1)
 Implements an exponential smoothing algorithm.

Explanation: http://www.youtube.com/watch?v=J4iODLa9hYw

__init__ (smoothingFactor=0.1, valuesToForecast=1)

Initializes the ExponentialSmoothing.

Parameters

- **smoothingFactor** (*Float*) Defines the alpha for the ExponentialSmoothing. Valid values are in (0.0, 1.0).
- valuesToForecast (Integer) Number of values that should be forecasted.

Raise Raises a ValueError when smoothingFactor has an invalid value.

execute (timeSeries)

Creates a new TimeSeries containing the smoothed and forcasted values.

Returns TimeSeries object containing the smoothed TimeSeries, including the forecasted values.

Return type TimeSeries

Note The first normalized value is chosen as the starting point.

class pycast.methods.HoltMethod(smoothingFactor=0.1, trendSmoothingFactor=0.5, valuesToFore-cast=1)

Implements the Holt algorithm.

Explanation: http://en.wikipedia.org/wiki/Exponential_smoothing#Double_exponential_smoothing

__init__ (smoothingFactor=0.1, trendSmoothingFactor=0.5, valuesToForecast=1) Initializes the HoltMethod.

Parameters

- **smoothingFactor** (*Float*) Defines the alpha for the ExponentialSmoothing. Valid values are in (0.0, 1.0).
- **trendSmoothingFactor** (*Float*) Defines the beta for the HoltMethod. Valid values are in (0.0, 1.0).
- valuesToForecast (Integer) Defines the number of forecasted values that will be part of the result.

Raise Raises a ValueError when smoothingFactor or trendSmoothingFactor has an invalid value.

execute (timeSeries)

Creates a new TimeSeries containing the smoothed values.

Returns TimeSeries object containing the smoothed TimeSeries, including the forecasted values.

Return type TimeSeries

Note The first normalized value is chosen as the starting point.

ERROR MEASURES

class pycast.errors.BaseErrorMeasure(minimalErrorCalculationPercentage=60)

Baseclass for all error measures.

__init__ (minimalErrorCalculationPercentage=60)

Initializes the error measure.

Parameters minimalErrorCalculationPercentage (*Integer*) – The number of entries in an original TimeSeries that have to have corresponding partners in the calculated TimeSeries. Corresponding partners have the same time stamp. Valid values are in [0.0, 100.0].

Raise Raises a ValueError if minimalErrorCalculationPercentage is not in [0.0, 100.0].

__weakref_

list of weak references to the object (if defined)

get_error (startingPercentage=0.0, endPercentage=100.0)

Calculates the error for the given interval (startingPercentage, endPercentage) between the TimeSeries given during BaseErrorMeasure.initialize().

Parameters

- **startingPercentage** (*Float*) Defines the start of the interval. This has to be a value in [0.0, 100.0]. It represents the value, where the error calculation should be started. 25.0 for example means that the first 25% of all calculated errors will be ignored.
- **endPercentage** (*Float*) Defines the end of the interval. This has to be a value in [0.0, 100.0]. It represents the value, after which all error values will be ignored. 90.0 for example means that the last 10% of all local errors will be ignored.

Returns Returns a float representing the error.

Return type Float

Raise Raises a ValueError in one of the following cases:

- startingPercentage not in [0.0, 100.0]
- endPercentage not in [0.0, 100.0]
- endPercentage < startingPercentage

Raise Raises a StandardError if BaseErrorMeasure.initialize() was not successfull before.

 $\verb"initialize" (original Time Series", calculated Time Series")$

Initializes the ErrorMeasure.

During initialization, all BaseErrorMeasure.local errors() are calculated.

Parameters

- **originalTimeSeries** (*TimeSeries*) TimeSeries containing the original data.
- **calculatedTimeSeries** (*TimeSeries*) TimeSeries containing calculated data. Calculated data is smoothed or forecasted data.

Returns Return True if the error could be calculated, False otherwise based on the minimalErrorCalculationPercentage.

Return type Boolean

local_error (originalValue, calculatedValue)

Calculates the error between the two given values.

Parameters

- originalValue (Numeric) Value of the original data.
- calculated Value (Numeric) Value of the calculated TimeSeries that corresponds to original Value.

Returns Returns the error measure of the two given values.

Return type Numeric

Raise Raises a NotImplementedError if the child class does not overwrite this method.

class pycast.errors.MeanSquaredError (minimalErrorCalculationPercentage=60)

Implements the mean squared error measure.

Explanation: http://en.wikipedia.org/wiki/Mean squared error

local_error (originalValue, calculatedValue)

Calculates the error between the two given values.

Parameters

- **originalValue** (*Numeric*) Value of the original data.
- calculated Value (*Numeric*) Value of the calculated TimeSeries that corresponds to original Value.

Returns Returns the error measure of the two given values.

Return type Numeric

class pycast.errors.**SymmetricMeanAbsolutePercentageError** (*minimalErrorCalculationPercentage*=60) Implements the symmetric mean absolute percentage error with a boarder of 200%.

Explanation: http://monashforecasting.com/index.php?title=SMAPE (Formula (3))

If the calculated value and the original value are equal, the error is 0.

local_error (originalValue, calculatedValue)

Calculates the error between the two given values.

Parameters

- originalValue (Numeric) Value of the original data.
- **calculated Value** (*Numeric*) Value of the calculated TimeSeries that corresponds to original Value.

Returns Returns the error measure of the two given values.

Return type Numeric

OPTIMIZATION METHODS

class pycast.optimization.BaseOptimizationMethod(errorMeasureClass, precision=-1)
 Baseclass for all optimization methods.
___init___(errorMeasureClass, precision=-1)
 Initializes the optimization method.

Parameters

- errorMeasureClass (BaseErrorMeasure) Error measure class from pycast.errors
- **precision** (*Integer*) Defines the accuracy for parameter tuning in 10^precision. This parameter has to be an integer in [-10, 0].

Raise Raises a TypeError if errorMeasureClass is not a valid class. Valid classes are derived from pycast.errors.BaseErrorMeasure.

Raise Raises a ValueError if precision is not in [-10, 0].

__weakref_

list of weak references to the object (if defined)

optimize (timeSeries, forecastingMethods=[])

Runs the optimization on the given TimeSeries.

Parameters

- timeSeries (*TimeSeries*) TimeSeries instance that requires an optimized forecast.
- $\bullet \ \ \textbf{forecastingMethods} \ (\textit{List}) \text{List of forecastingMethods that will be used for optimization}. \\$

Returns Returns the optimzed forecasting method with the smallest error.

Return type (BaseForecastingMethod, Dictionary)

Raise Raises a ValueError ValueError if no forecastingMethods is empty.

class pycast.optimization.**GridSearch** (*errorMeasureClass*, *precision=-1*) Implements the grid search method for parameter optimization.

GridSearch is the brute force method.

optimize (timeSeries, forecastingMethods=[])

Runs the optimization of the given TimeSeries.

Parameters

- timeSeries (TimeSeries) TimeSeries instance that requires an optimized forecast.
- **forecastingMethods** (*List*) List of forecastingMethods that will be used for optimization.

Returns Returns the optimzed forecasting method with the smallest error.

Return type BaseForecastingMethod, Dictionary

Raise Raises a ValueError ValueError if no forecastingMethods is empty.

optimize_forecasting_method(timeSeries, forecastingMethod)

Optimizes the parameters for the given timeSeries and forecastingMethod.

Parameters

- timeSeries (*TimeSeries*) TimeSeries instance, containing hte original data.
- **forecastingMethod** (*BaseForecastingMethod*) ForecastingMethod that is used to optimize the parameters.

Todo Errorclass for calculation

Todo percentage for start_error_measure, end_error_measure

Todo Definition of the result that will be returned.

CHAPTER

SEVEN

INDICES AND TABLES

- genindex
- modindex
- search

PYTHON MODULE INDEX

р

pycast.common.helper, 1
pycast.common.profileme, 1
pycast.common.timeseries, 1

20 Python Module Index

INDEX

Symbols	C		
add() (pycast.common.timeseries.TimeSeries method), 3eq() (pycast.common.timeseries.TimeSeries method), 3	convert_epoch_to_timestamp() (py- cast.common.timeseries.TimeSeries class method), 5 convert_timestamp_to_epoch() (py-		
getitem() (pycast.common.timeseries.TimeSeries method), 3	cast.common.timeseries.TimeSeries class method), 5		
init() (pycast.common.timeseries.TimeSeries method), 3init() (pycast.errors.BaseErrorMeasure method), 13init() (pycast.methods.ExponentialSmoothing method), 11init() (pycast.methods.HoltMethod method), 11init() (pycast.methods.SimpleMovingAverage method), 9init() (pycast.optimization.BaseOptimizationMethod method), 15iter() (pycast.common.timeseries.TimeSeries method), 4len() (pycast.common.timeseries.TimeSeries method), 4setitem() (pycast.common.timeseries.TimeSeries	execute() (pycast.methods.ExponentialSmoothing method), 11 execute() (pycast.methods.HoltMethod method), 11 execute() (pycast.methods.SimpleMovingAverage method), 9 ExponentialSmoothing (class in pycast.methods), 11 F from_json() (pycast.common.timeseries.TimeSeries class method), 5 from_twodim_list() (pycast.common.timeseries.TimeSeries class method), 5		
method), 4str() (pycast.common.timeseries.TimeSeries method), 4weakref (pycast.common.timeseries.TimeSeries attribute), 4weakref (pycast.errors.BaseErrorMeasure attribute),	G get_error() (pycast.errors.BaseErrorMeasure method), 13 GridSearch (class in pycast.optimization), 15 H		
13weakref (pycast.optimization.BaseOptimizationMethol	dHoltMethod (class in pycast.methods), 11		
attribute), 15 A add_entry() (pycast.common.timeseries.TimeSeries method), 4	initialize() (pycast.errors.BaseErrorMeasure method), 13 initialize_from_sql_cursor() (py- cast.common.timeseries.TimeSeries method),		
apply() (pycast.common.timeseries.TimeSeries method), 4 BaseErrorMeasure (class in pycast.errors), 13 BaseOptimizationMethod (class in pycast.optimization), 15	is_normalized() (pycast.common.timeseries.TimeSeries method), 6 is_sorted() (pycast.common.timeseries.TimeSeries method), 6		

```
L
linear_interpolation() (in module pycast.common.helper),
local_error() (pycast.errors.BaseErrorMeasure method),
local_error() (pycast.errors.MeanSquaredError method),
local_error() (pycast.errors.SymmetricMeanAbsolutePercentageError
         method), 14
M
MeanSquaredError (class in pycast.errors), 14
Ν
normalize()
                 (pycast.common.timeseries.TimeSeries
         method), 6
0
optimize() (pycast.optimization.BaseOptimizationMethod
         method), 15
optimize() (pycast.optimization.GridSearch method), 15
optimize_forecasting_method()
                                                  (py-
         cast.optimization.GridSearch method), 16
Р
profileMe (in module pycast.common.profileme), 1
pycast.common.helper (module), 1
pycast.common.profileme (module), 1
pycast.common.timeseries (module), 1
S
SimpleMovingAverage (class in pycast.methods), 9
sort_timeseries() (pycast.common.timeseries.TimeSeries
         method), 6
sorted_timeseries()
                                                  (py-
         cast.common.timeseries.TimeSeries method),
SymmetricMeanAbsolutePercentageError (class in py-
         cast.errors), 14
Т
TimeSeries (class in pycast.common.timeseries), 3
to_gnuplot_datafile()
                                                  (py-
         cast.common.timeseries.TimeSeries method),
to_json()
                 (pycast.common.timeseries.TimeSeries
         method), 7
to_twodim_list() (pycast.common.timeseries.TimeSeries
         method), 7
```

22 Index