Data Mining


DATA
MINING
PATTERNS
LEARNING
CUSTOMERS
ANALYSIS
SET
DECISION
PEOPLE
CUSTOMER
MACHINE
BUSINESS
OFFERS
INFORMATION
DATABASE
STRUCTURES
IDENTIFY
VISUALIZATION
E-MAIL
FINALLY
MARKETING
SOPHISTICATED
ALGORITHMS
DETECTION
OFFER
ARTIFICIAL
RULE
CITATION
NEW
SEND
WITHOUT
MIGHT
USE
STEP
GROUPS
TASK
DESIRED
WITHIN
LEARNED
EXAMPLE
LARGE
TEST
MARKET
E-MAILS
BASKET
APPLICATIONS
PREDICT
LIKELY
DETERMINE
METHODS
MODEL
INTELLIGENCE
TARGET
PRODUCT
PRESENT
USING
TRAINED
DISCOVERY
SYSTEM
MANY
PRE-PROCESSING
COMMON
HELP
INVOLVES
NUMBER
MODELING
KNOWLEDGE
PROCESS
ASSOCIATION
NEEDED
RECORDS
MANAGEMENT
ALGORITHM
MODELS
TERM

INDEX

1. Preparing data
2. Windowing
3. Model training

1. Preparing data

1. Preparing data

http://www.learnpredictiveanalytics.com/download.html

1. Preparing data

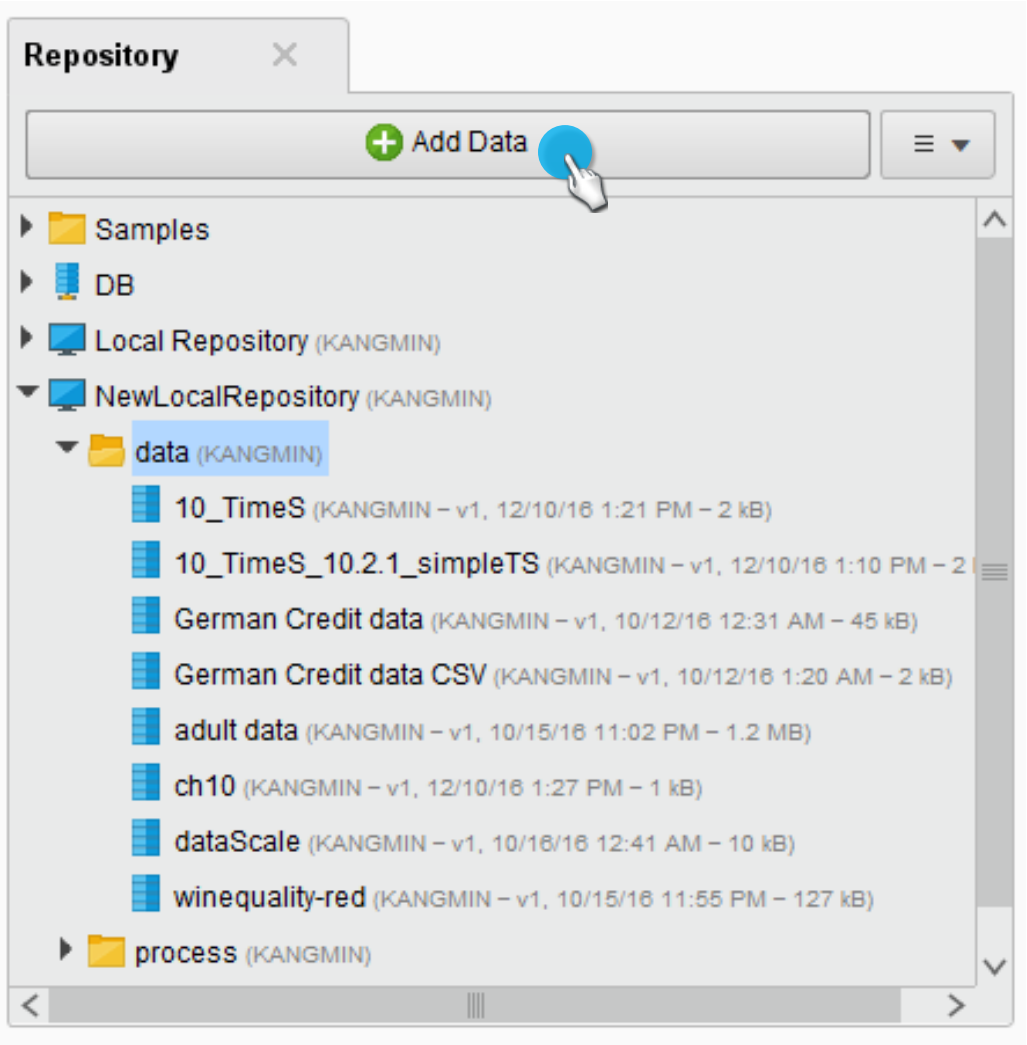
Install RapidMiner Software

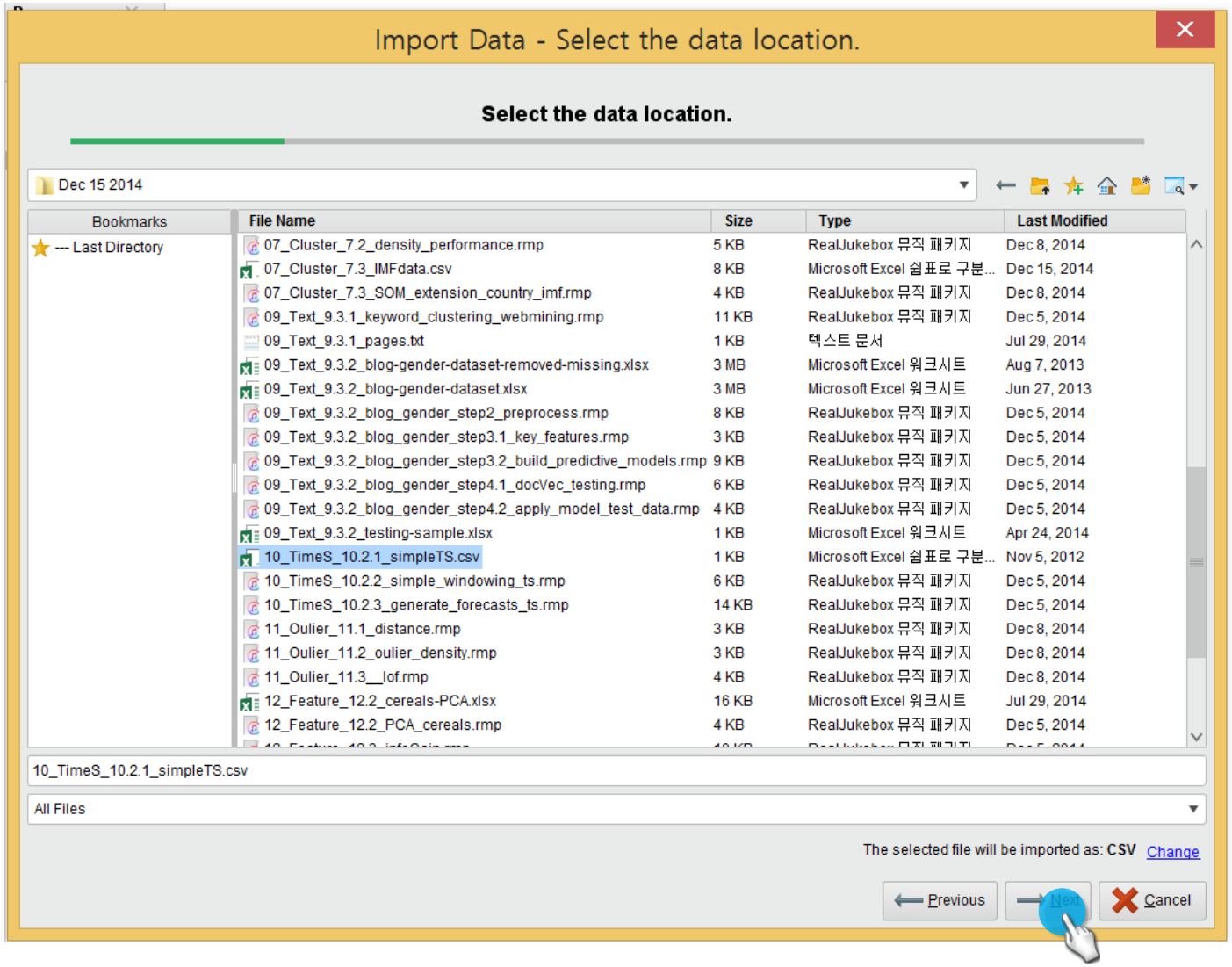
You can download the implementation files covered in book chapters along with the RapidMiner processes (.rmp files) and supporting data files for the examples described in the book. You may download all the files and recreate them using RapidMiner on your computer. You need to "Import Process File" to bring in the process into RapidMiner and then link the data files to their corresponding locations in your desktop with RapidMiner.

RM Process files and data sets

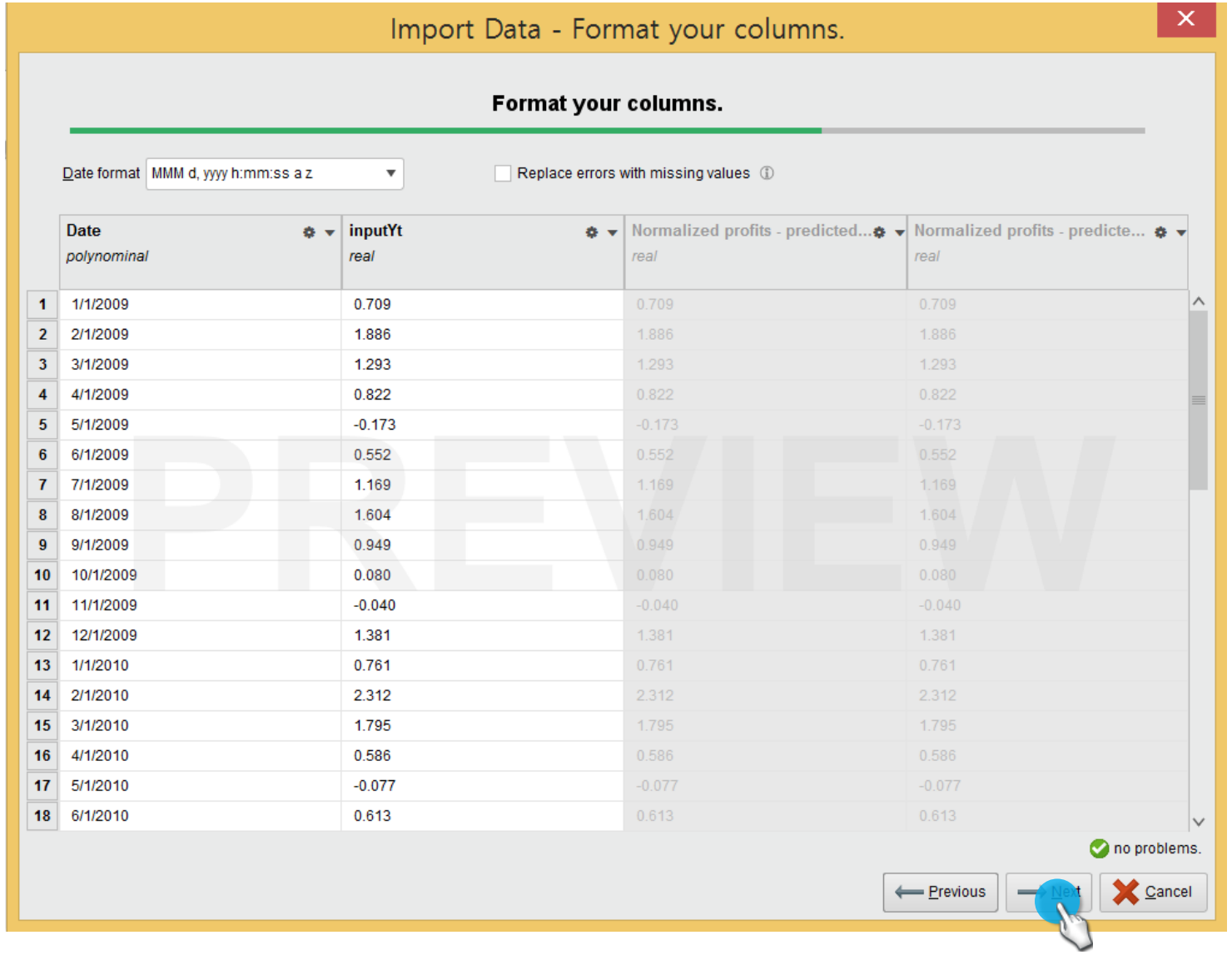
Download File

1. Preparing data



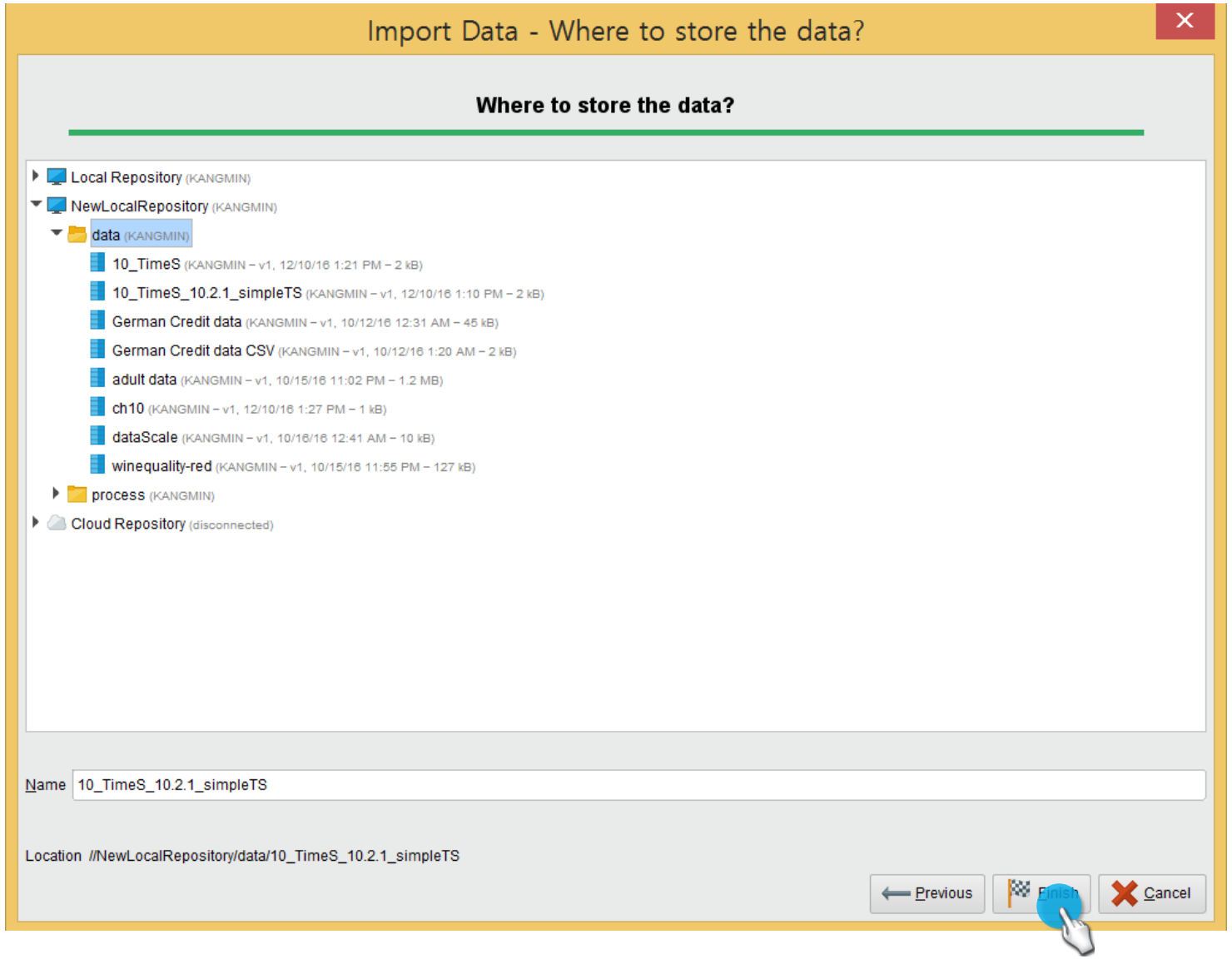


1. Preparing data



	Date *polynominal*	inputYt *real*	Normalized profits - predicted... *real*	Normalized profits - predicte... *real*
1	1/1/2009	0.709	0.709	0.709
2	2/1/2009	1.886	1.886	1.886
3	3/1/2009	1.293	1.293	1.293
4	4/1/2009	0.822	0.822	0.822
5	5/1/2009	-0.173	-0.173	-0.173
6	6/1/2009	0.552	0.552	0.552
7	7/1/2009	1.169	1.169	1.169
8	8/1/2009	1.604	1.604	1.604
9	9/1/2009	0.949	0.949	0.949
10	10/1/2009	0.080	0.080	0.080
11	11/1/2009	-0.040	-0.040	-0.040
12	12/1/2009	1.381	1.381	1.381
13	1/1/2010	0.761	0.761	0.761
14	2/1/2010	2.312	2.312	2.312
15	3/1/2010	1.795	1.795	1.795
16	4/1/2010	0.586	0.586	0.586
17	5/1/2010	-0.077	-0.077	-0.077
18	6/1/2010	0.613	0.613	0.613

1. Preparing data



2. Windowing

1.1. Windowing – Retrieve operator

Retrieve ch10

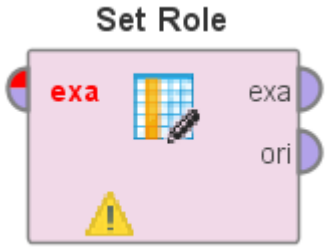
ExampleSet (50 examples, 0 special attributes, 2 regular attributes)

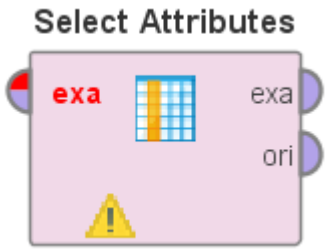
Row No.	Date	inputYt
1	1/1/2009	0.709
2	2/1/2009	1.886
3	3/1/2009	1.293
4	4/1/2009	0.822
5	5/1/2009	-0.173
6	6/1/2009	0.552
7	7/1/2009	1.169
8	8/1/2009	1.604
9	9/1/2009	0.949
10	10/1/2009	0.080

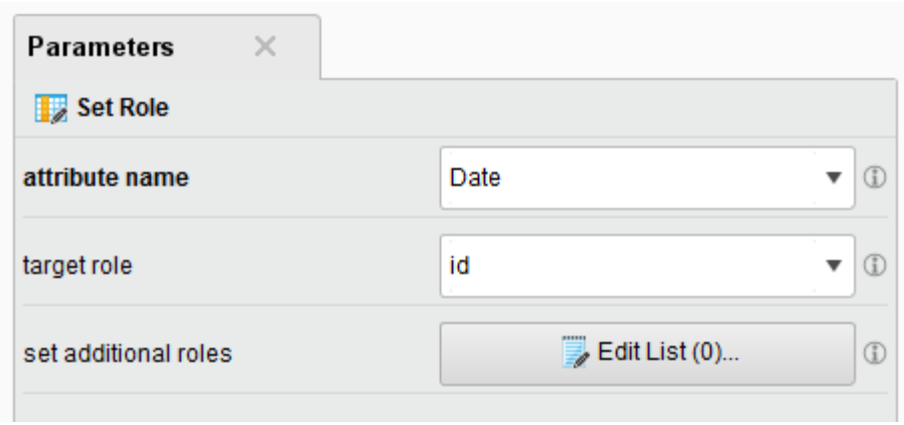
⋮

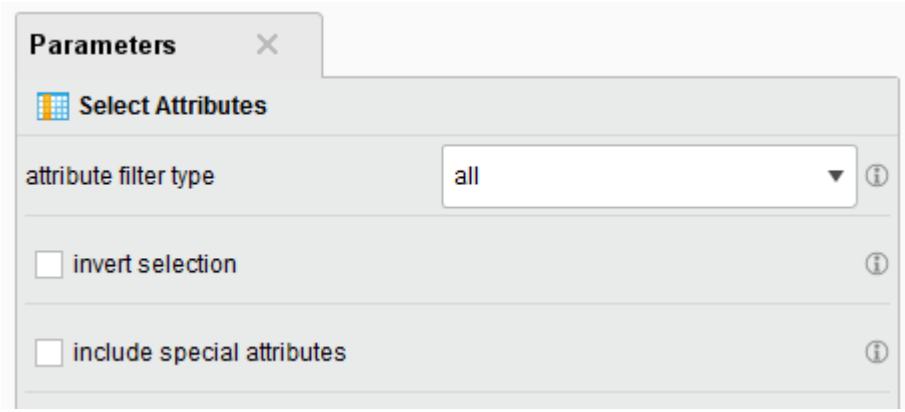
50

1.1. Windowing – Set Role and Select Attributes operator

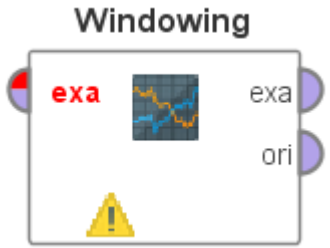


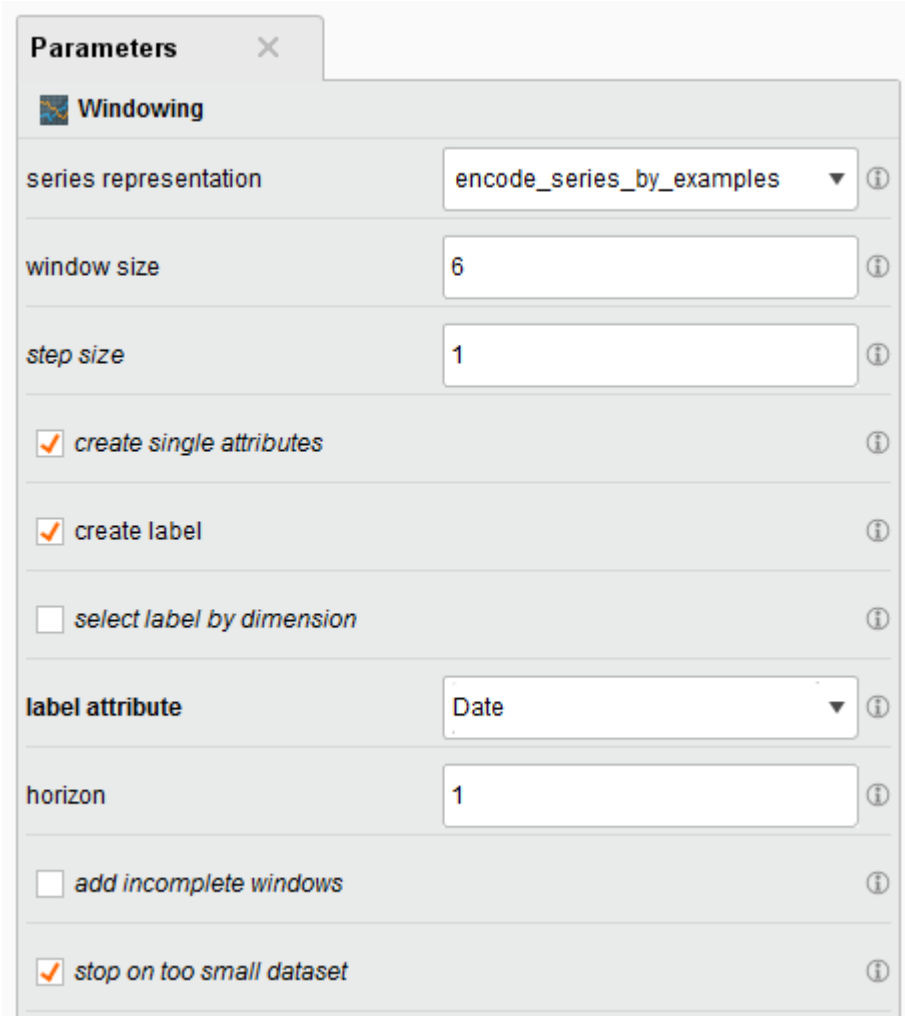




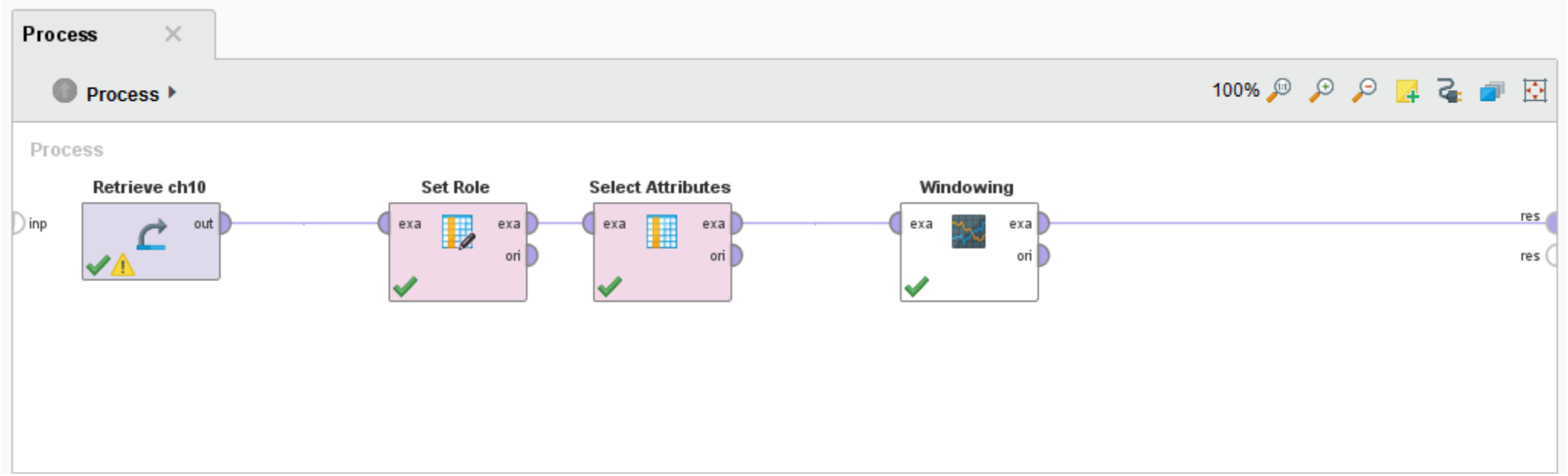


1.1. Windowing – Windowing operator





1.1. Windowing – process



1.2.1. Windowing – example set

ExampleSet (44 examples, 2 special attributes, 6 regular attributes)

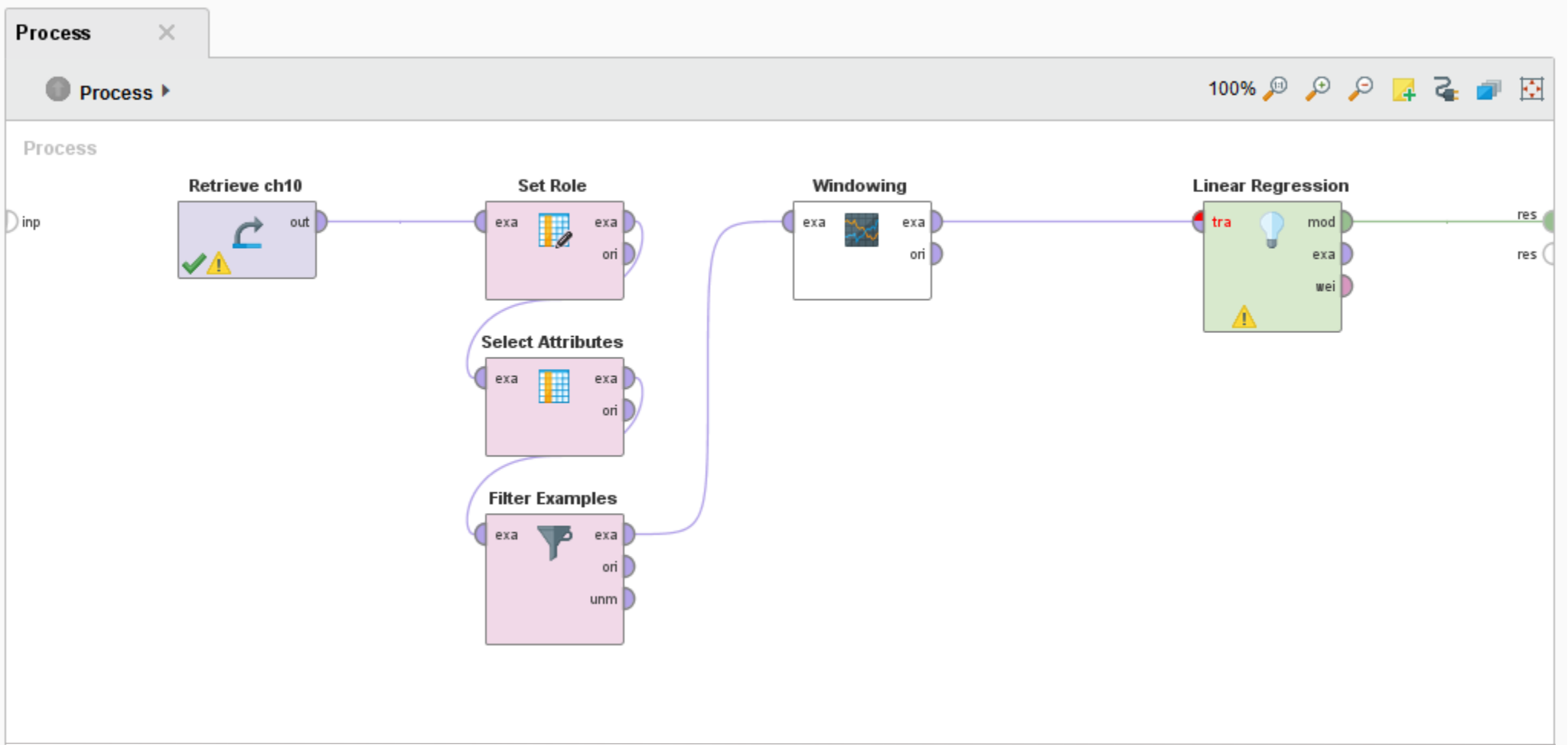
Row No.	Date	label	inputYt-5	inputYt-4	inputYt-3	inputYt-2	inputYt-1	inputYt-0
1	6/1/2009	1.169	0.709	1.886	1.293	0.822	-0.173	0.552
2	7/1/2009	1.604	1.886	1.293	0.822	-0.173	0.552	1.169
3	8/1/2009	0.949	1.293	0.822	-0.173	0.552	1.169	1.604
4	9/1/2009	0.080	0.822	-0.173	0.552	1.169	1.604	0.949
5	10/1/2009	-0.040	-0.173	0.552	1.169	1.604	0.949	0.080
6	11/1/2009	1.381	0.552	1.169	1.604	0.949	0.080	-0.040
7	12/1/2009	0.761	1.169	1.604	0.949	0.080	-0.040	1.381
8	1/1/2010	2.312	1.604	0.949	0.080	-0.040	1.381	0.761
9	2/1/2010	1.795	0.949	0.080	-0.040	1.381	0.761	2.312
10	3/1/2010	0.586	0.080	-0.040	1.381	0.761	2.312	1.795
11	4/1/2010	-0.077	-0.040	1.381	0.761	2.312	1.795	0.586
12	5/1/2010	0.613	1.381	0.761	2.312	1.795	0.586	-0.077
13	6/1/2010	1.845	0.761	2.312	1.795	0.586	-0.077	0.613
14	7/1/2010	1.984	2.312	1.795	0.586	-0.077	0.613	1.845
15	8/1/2010	1.861	1.795	0.586	-0.077	0.613	1.845	1.984
16	9/1/2010	0.661	0.586	-0.077	0.613	1.845	1.984	1.861
17	10/1/2010	0.692	-0.077	0.613	1.845	1.984	1.861	0.661
18	11/1/2010	1.108	0.613	1.845	1.984	1.861	0.661	0.692
19	12/1/2010	1.688	1.845	1.984	1.861	0.661	0.692	1.108
20	1/1/2011	2.167	1.984	1.861	0.661	0.692	1.108	1.688

1.2.2. Windowing – example set

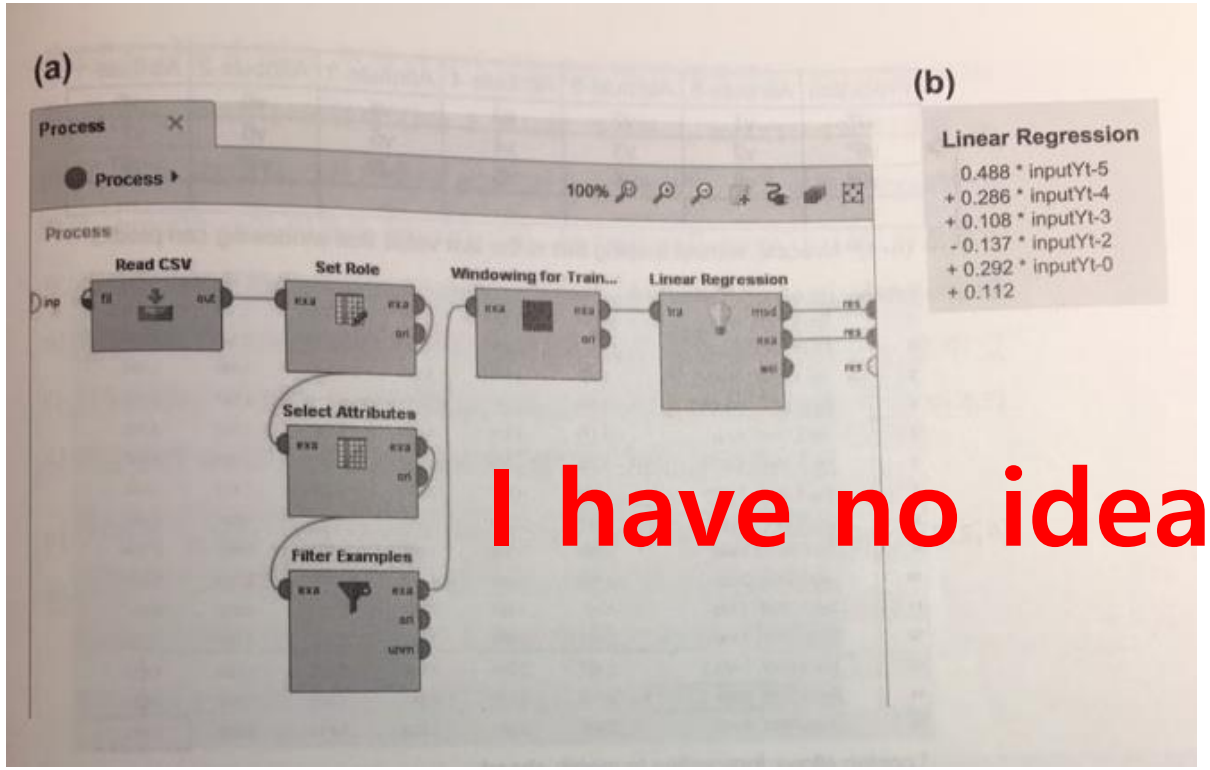
21	2/1/2011	2.295	1.861	0.661	0.692	1.108	1.688	2.167
22	3/1/2011	1.577	0.661	0.692	1.108	1.688	2.167	2.295
23	4/1/2011	0.601	0.692	1.108	1.688	2.167	2.295	1.577
24	5/1/2011	1.201	1.108	1.688	2.167	2.295	1.577	0.601
25	6/1/2011	2.466	1.688	2.167	2.295	1.577	0.601	1.201
26	7/1/2011	2.497	2.167	2.295	1.577	0.601	1.201	2.466
27	8/1/2011	2.245	2.295	1.577	0.601	1.201	2.466	2.497
28	9/1/2011	1.179	1.577	0.601	1.201	2.466	2.497	2.245
29	10/1/2011	1.119	0.601	1.201	2.466	2.497	2.245	1.179
30	11/1/2011	1.934	1.201	2.466	2.497	2.245	1.179	1.119
31	12/1/2011	?	2.466	2.497	2.245	1.179	1.119	1.934
32	1/1/2012	?	2.497	2.245	1.179	1.119	1.934	?
33	2/1/2012	?	2.245	1.179	1.119	1.934	?	?
34	3/1/2012	?	1.179	1.119	1.934	?	?	?
35	4/1/2012	?	1.119	1.934	?	?	?	?
36	5/1/2012	?	1.934	?	?	?	?	?
37	6/1/2012	?	?	?	?	?	?	?
38	7/1/2012	?	?	?	?	?	?	?
39	8/1/2012	?	?	?	?	?	?	?
40	9/1/2012	?	?	?	?	?	?	?
41	10/1/2012	?	?	?	?	?	?	?
42	11/1/2012	?	?	?	?	?	?	?
43	12/1/2012	?	?	?	?	?	?	?
44	1/1/2013	?	?	?	?	?	?	?

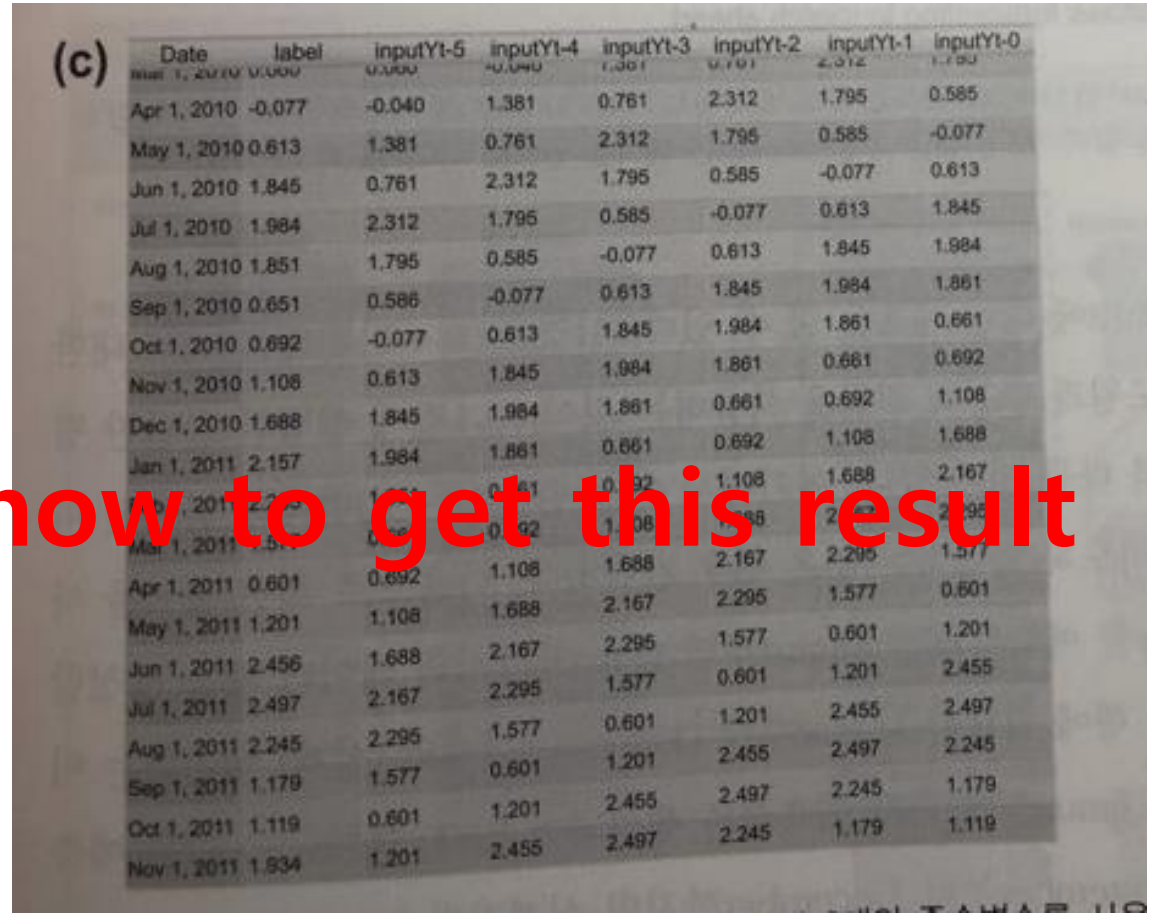
3. Model training

3. Model training – process



3. Model training – process





Date	label	inputYt-5	inputYt-4	inputYt-3	inputYt-2	inputYt-1	inputYt-0
Apr 1, 2010	-0.077	-0.040	1.381	0.761	2.312	1.795	0.585
May 1, 2010	0.613	1.381	0.761	2.312	1.795	0.585	-0.077
Jun 1, 2010	1.845	0.761	2.312	1.795	0.585	-0.077	0.613
Jul 1, 2010	1.984	2.312	1.795	0.585	-0.077	0.613	1.845
Aug 1, 2010	1.851	1.795	0.585	-0.077	0.613	1.845	1.984
Sep 1, 2010	0.651	0.586	-0.077	0.613	1.845	1.984	1.861
Oct 1, 2010	0.692	-0.077	0.613	1.845	1.984	1.861	0.661
Nov 1, 2010	1.108	0.613	1.845	1.984	1.861	0.661	0.692
Dec 1, 2010	1.688	1.845	1.984	1.861	0.661	0.692	1.108
Jan 1, 2011	2.157	1.984	1.861	0.661	0.692	1.108	1.688
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.108	1.688	2.167
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Apr 1, 2011	0.601	0.692	1.108	1.688	2.167	2.295	1.577
May 1, 2011	1.201	1.108	1.688	2.167	2.295	1.577	0.601
Jun 1, 2011	2.456	1.688	2.167	2.295	1.577	0.601	1.201
Jul 1, 2011	2.497	2.167	2.295	1.577	0.601	1.201	2.455
Aug 1, 2011	2.245	2.295	1.577	0.601	1.201	2.455	2.497
Sep 1, 2011	1.179	1.577	0.601	1.201	2.455	2.497	2.245
Oct 1, 2011	1.119	0.601	1.201	2.455	2.497	2.245	1.179
Nov 1, 2011	1.934	1.201	2.455	2.497	2.245	1.179	1.119

I have no idea how to get this result

3. Model training – process

Prediction	Attribute-6	Attribute-5	Attribute-4	Attribute-3	Attribute-2	Attribute-1
v7*	v1	v2	v3	v4	v5	v6
v8*	v2	v3	v4	v5	v6	v7*
v9*	v3	v4	v5	v6	v7*	v8*
...	...	...	...	...	...	...

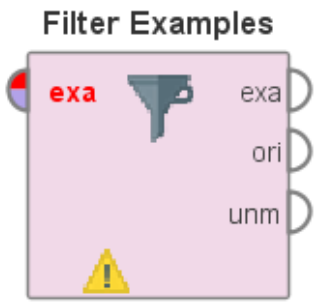
(n+1)th forecast: without looping this is the last value that windowing can predict.

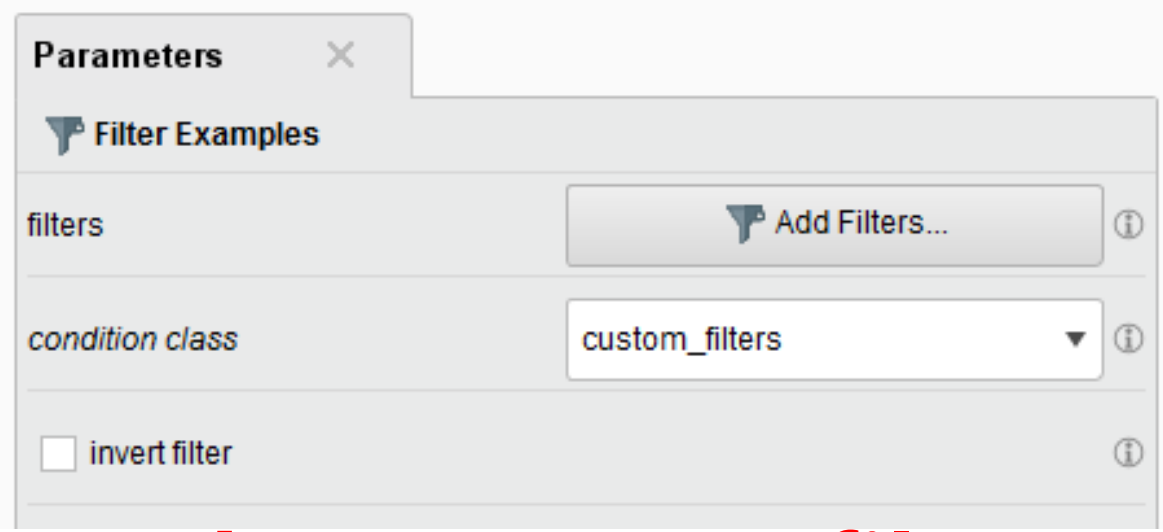
Row No.	Date	prediction(label)	inputM-5	inputM-4	inputM-3	inputM-2	inputM-1	inputM-0
1	Nov 1, 2011	1.694	1.201	2.466	2.497	2.245	1.179	1.119
2	Dec 1, 2011	2.597	2.466	2.497	2.245	1.179	1.119	1.694
3	Jan 1, 2012	2.693	2.497	2.245	1.179	1.119	1.694	2.597
4	Feb 1, 2012	2.196	2.245	1.179	1.119	1.694	2.597	2.693
5	Mar 1, 2012	1.457	1.179	1.119	1.694	2.597	2.693	2.196
6	Apr 1, 2012	1.457	1.119	1.694	2.597	2.693	2.196	1.457
7	Ma[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
8	Jun 1, 2012	2.784	2.597	2.693	2.196	1.457	1.457	2.087
9	Jul 1, 2012	2.807	2.693	2.196	1.457	1.457	2.087	2.784
10	Aug 1, 2012	2.265	2.196	1.457	1.457	2.087	2.784	2.807
11	Sep 1, 2012	1.720	1.457	1.457	2.087	2.784	2.807	2.265
12	Oct 1, 2012	1.816	1.457	2.087	2.784	2.807	2.265	1.720
13	Nov 1, 2012	2.433	2.087	2.784	2.807	2.265	1.720	1.816
14	Dec 1, 2012	2.974	2.784	2.807	2.265	1.720	1.816	2.433
15	Jan 1, 2013	2.911	2.807	2.265	1.720	1.816	2.433	2.974

Looping allows forecasting to march ahead.

I have no idea how to get this result

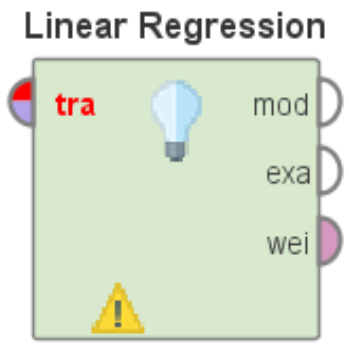
3. Model training – Filter Examples operator

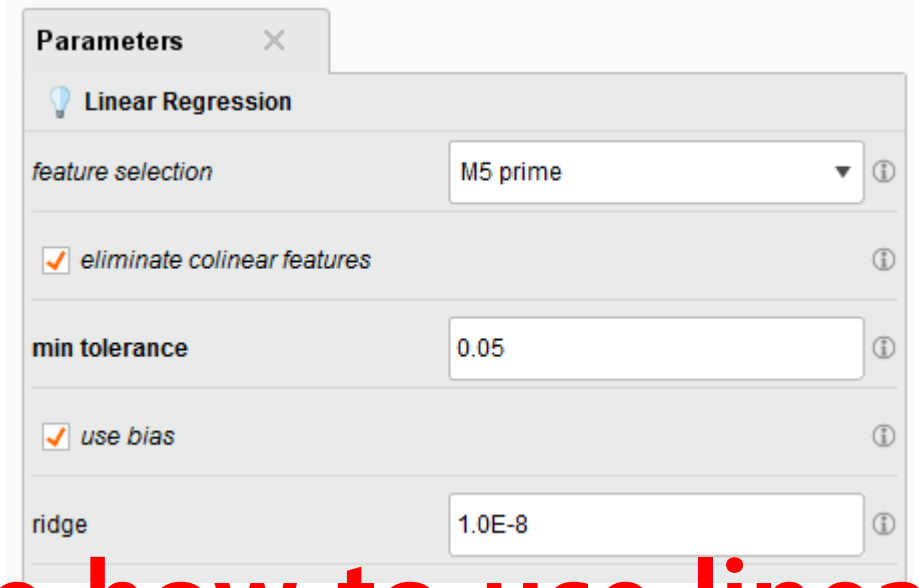




Please teach me how to use filter examples operator

3. Model training – Linear Regression operator





Please teach me how to use linear regression operator