Scio-machine.com

Forecasting software

2.0

Scio-machine.com is an easy to use and powerful software.

No installation, no programming and a single interface.

Forecasts based on multi-variable time series, or not.

Two neural networks in parallel.

Secure data.

Makes these 2 experiments right now

go to
https://machine1.scio-machine.com/engine.html

load
the demo database from
Demos & Help

The price of the AUD/NZD currency pair, 2 days before.

The hypotenuse of a right-angled triangle with no formula.

Before you start

2 notions to know

The stationary process of a data series characterizes the mathematical predictability of its evolution.
A random sequence of numbers will have a zero degree of stationary process.
An arithmetic sequence such as (1, 2, 3, 4 ... 12, 13) will have a strong degree of stationary process.
Other data sequences may have several stationary processes running in a partial or complex manner.

The dependence of the future on the past characterizes the interval between cause and effect.
An arithmetic sequence such as (1, 2, 3, 4 ... 12, 13) will have a dependence of the future on the past tense of 1.
A random sequence of numbers will have a dependency of the future to the past tense of 0.

Both concepts can be used intuitively or empirically without the need for mathematical knowledge.

Diagram of the learning process
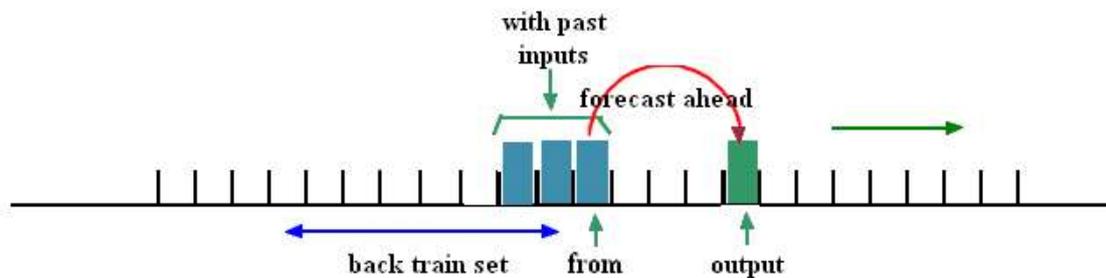
Learning consists of taking into account the data (inputs) in the zone with past in order to predict a data (output) in the zone output.
from is the head of the zone with past.
forecast ahead is the distance between from and output.
back train test is a back area allowing additional testing at each stage.
Learning is done in several steps (forecast to run) by advancing all the zones of a record at the same time.
During the learning process output cannot exceed the last record in the table.
During the forecast process, from cannot exceed the last record in the table.
All parameters are detailed later in this document.

In conclusion

If the stationary process is null or strong, the learning can start from anywhere in the series.
Otherwise, the learning will have to be done as close as possible to the forecast to be made, generally at the end of the series.
The distance between from and output (forecast ahead) should not be greater than the length of time the future is dependent on the past.
with past should not be greater than the duration of dependence of the future on the past +1.

Experiment 1

Forecast the price of the AUD/NZD currency pair.
Table to be used: *forex.*

Analysis
+ It's a time series with no strong stationary process but supposedly non-zero.
+ The dependence of the future on the past does not exceed 2 trading sessions.
+ The forecast value is the closing price.
+ The data used will be the historical values of several currency pairs available in the table.

Parameterization

|	
inputs #1 and #2	AUDCADOpen... NZDCHFVolume
output #1 and #2	AUDNZDClose
from54	*Here*
	from is set to 54 for the pedagogical interest, it would be placed as far as possible away towards the end in a real situation.
with past	3
	Duration of *dependence +1*
forecast ahead	2
	Duration of dependency
G	75
back train set	1
	If the dependency time is short and/or the stationarity is badly established the additional tests may be out of context.
forecasts to run	1 (i.e. 1+1)
deep	5
	Could be greater.
0/1 view	no
real plots to display	20
to do	*learn* for the learning phase
	forcast o*nly for* the forecasting phase, with:
	<u>forecast to run</u> = 3 (i.e. 3+1)

Results
On the graph obtained at the end of the forecasts we can see that the first two forecasts are in fact the "restitution" of the learning, <u>from</u> not having been moved. As for the last two forecasts, they are the pure fruit of neural networks and data known <u>inputs</u> 2 previous quotations. We can clearly see the anticipation of the end of the fall and an increase in prices. In a real context, the known data would not exist beyond record 57 (in blue and purple).
The two forecasts diverge somewhat despite having been obtained with the same data and settings. This is due to the fact that the genetic technique of forming the two networks leads them to differ. The green line represents the average of the two forecasts.
It is very likely that the resulting networks do not make good predictions at other locations in the series due to the fact that the series does not have a strong stationary process.

Graph obtained at the end of the course

learn 2 steps ahead

forecast 2 steps ahead
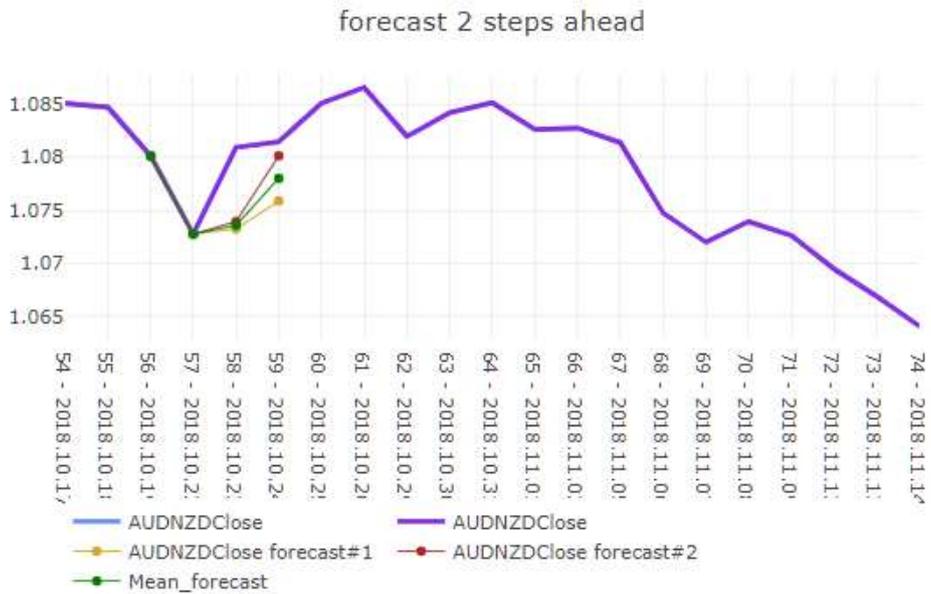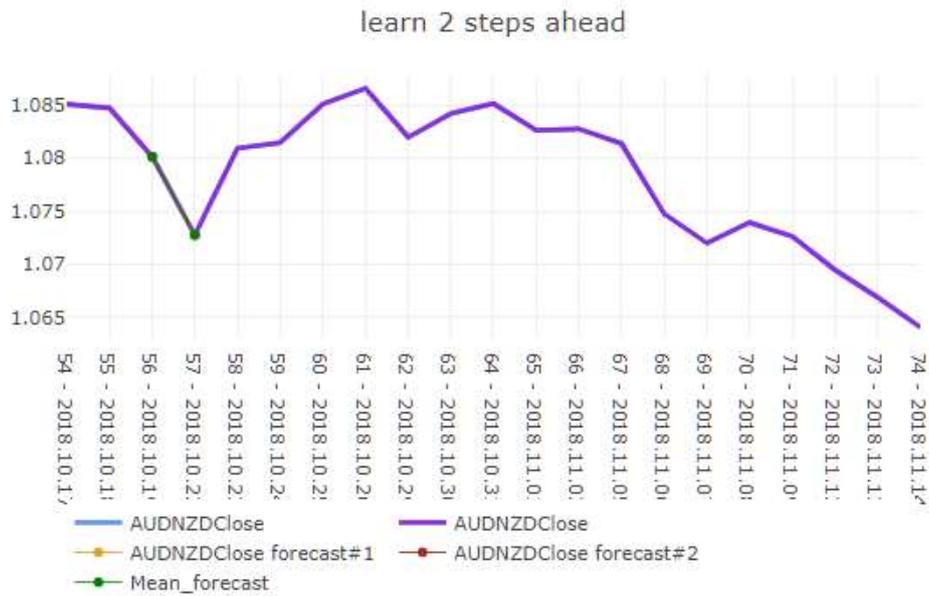
Graph obtained at the end of the forecast

Experiment 2

Give the length of the hypotenuse of a right-angled triangle along its other two sides.
Table to be used: *pithagoras.*
Each record contains the lengths of the three sides of a right-angled triangle.
The records have no particular order.

Analysis
+ The stationarity process is null
+ The dependency of the future on the past is 0.
+ The target value is the hypotenuse.
+ The data used are the lengths of the other two sides.

Parameterization

inputs #1 and #2	AB , AC
output #1 and #2	hypotenuse
from	25
	from can start anywhere.
with past	1
	i.e. containing only the from record.
forecast ahead	0
	No dependency of the future on the past, the variable to be expected is the one that is in the current from record.
G	15
back train set	20
	The back train set is important in series with zero stationarity processes.
forecasts to run	10 (i.e. 10+1)
	Could be bigger to refine learning.
deep	15
	Could be bigger to refine learning.
0/1 view	no
real plots to display	45
to do	*learn* for the learning process
	forecast only for the forecasting process, with:
	from = 250
	forecast to run = 40 (i.e. 40+1)

Results
The forecast graph shows that a very small amount of learning was sufficient to obtain consistent forecasts, although they lack mathematical accuracy. Indeed, Scio-machine.com does not build a mathematical model. Further learning could have resulted in even more accurate forecasts.

The two forecasts diverge despite having been obtained with the same data and settings. This is due to the fact that the genetic technique of formation of the two networks leads them to differ. The green line represents the average of the two predictions.
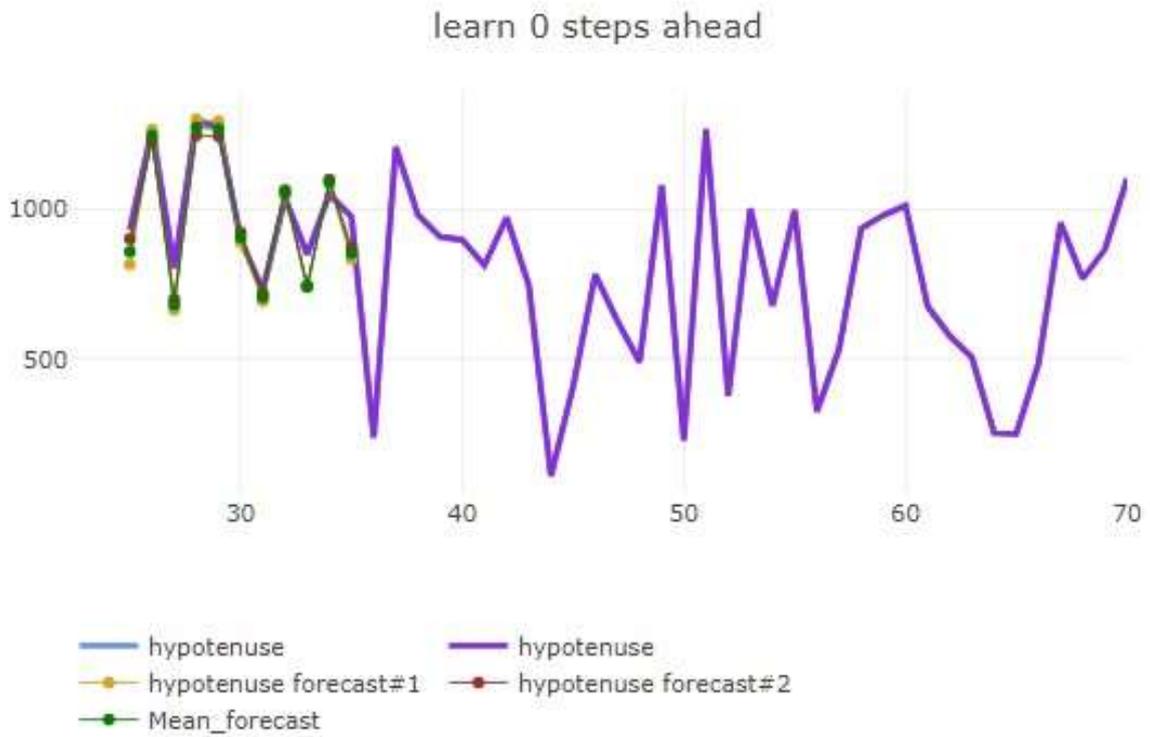
Profiling is a tool for defining a profile according to constraints.
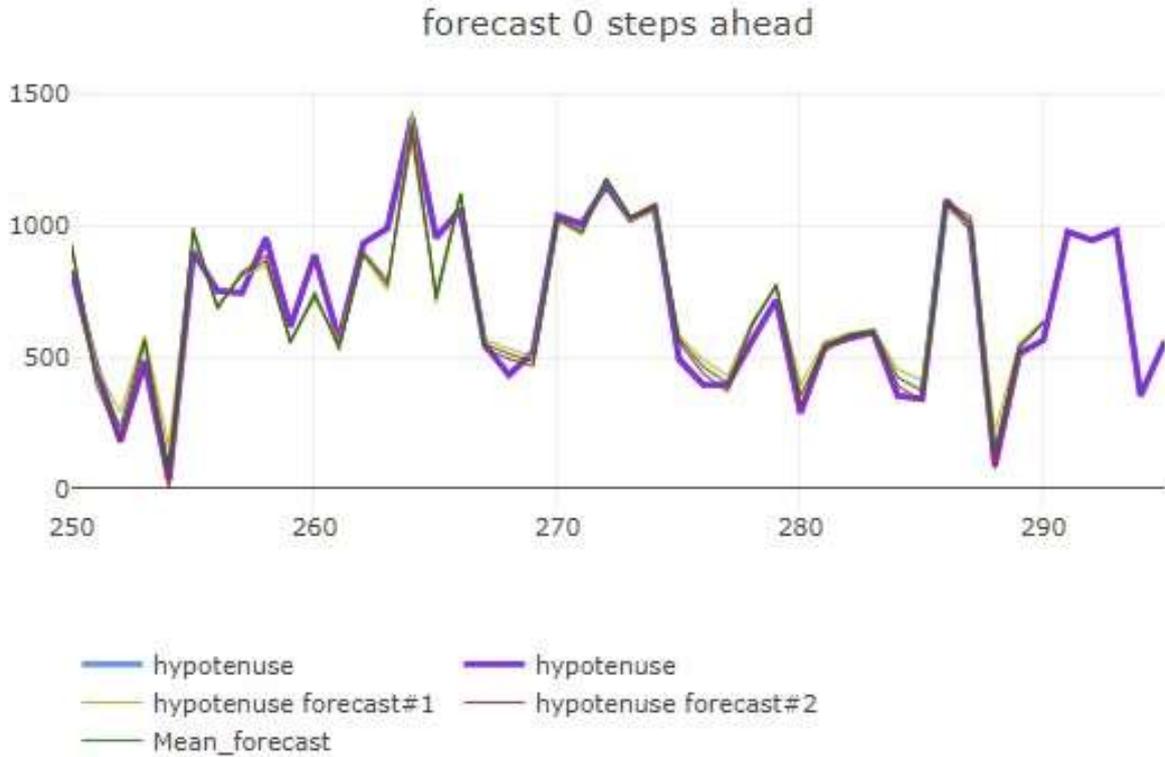Here the selected constraints were hypotenuse 1000 and AC 900.
The minimum and maximum limits of the values of the input variables are defined by their respective minimum and maximum in the series.
There is possible profiling for output #1 and #2.

Graph obtained at the end of the course

learn 0 steps ahead

Legend:
- hypotenuse
- hypotenuse
- hypotenuse forecast#1
- hypotenuse forecast#2
- Mean_forecast

Graph obtained at the end of the forecast

forecast 0 steps ahead

Legend:
- hypotenuse
- hypotenuse
- hypotenuse forecast#1
- hypotenuse forecast#2
- Mean_forecast

Profiling

A single interface for
displaying of results
profiling
the parameter area
the console

Displaying the results

The graph has an icon menu at the top right allowing the graph to be saved as a png file. The data displayed in the graph are dynamic and can be hidden/shown. The chart data can be edited, analyzed, processed, saved and shared with **plotly Chart Studio,** with or without an account (an internet connection must be established to transfer the data to the plot.ly site).
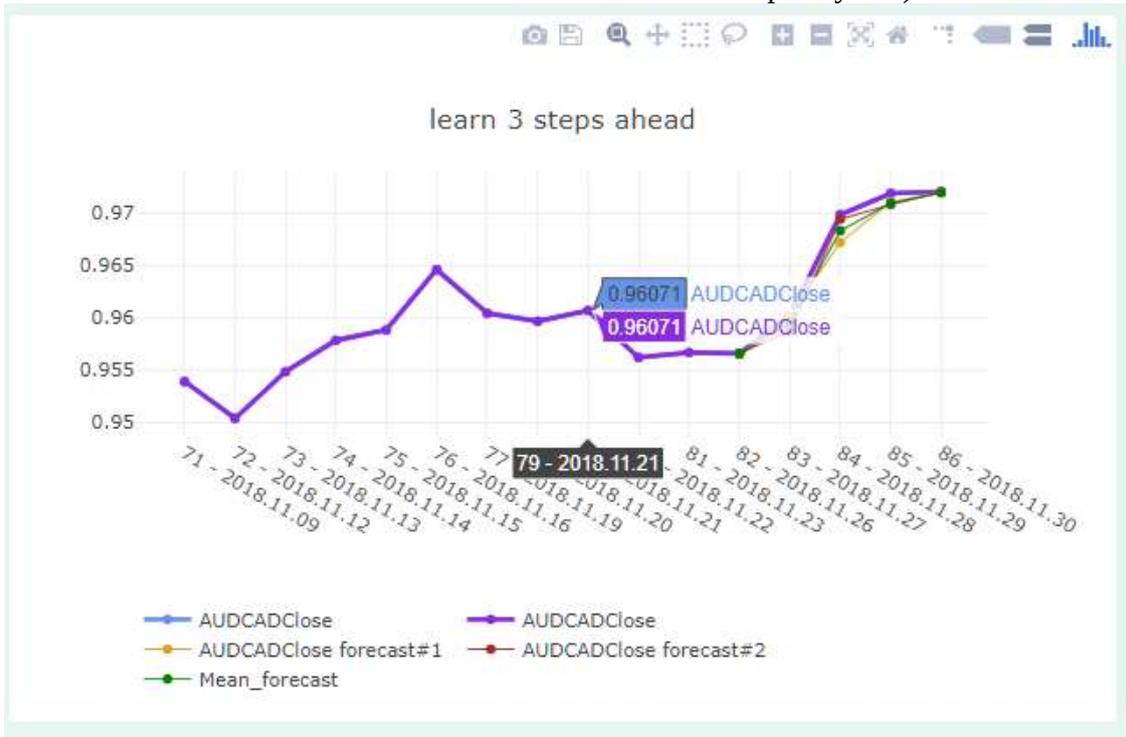
Profiling

PROFILER#1 and #2 allow, after learning, to dynamically modify the values of the input variables in order to obtain relevant profiles. The characteristics of the obtained profiles can be saved in CSV files.

The parameterization

DATABASE / TABLE

database | table
Choisir un fichier | SCIO-MAC…E.sqlite | forex

FORECAST #1 FIELDS

inputs | output | ready
AUDCADOpen
AUDCADHigh
AUDCADLow
AUDCADClose
AUDCADVolume
AUDCHFOpen
AUDCHFHigh
AUDCHFLow

output: AUDCADClose validate ✓

FORECAST #2 FIELDS

inputs | output | ready
AUDCADOpen
AUDCADHigh
AUDCADLow
AUDCADClose
AUDCADVolume
AUDCHFOpen
AUDCHFHigh
AUDCHFLow

output: AUDCADClose validate ✓

FORECAST PARAMETERS

from | with past | forecast ahead
79 | Min | Max | 3 | Min | Max | 3 | Min | Max

LEARNING PARAMETERS

G | back train set | forecasts to run
45 | Min | Max | 1 | Min | Max | 4 | Min | Max

3 | deep

DISPLAY PARAMETERS

O/1 view

real plots to display
15 | Min | Max

no

RUNNING PARAMETERS

to do

learning

save report

HELP

Demos & Help

The parameters

database: to load the Sqlite database to use.

table: to select the table (data table) to be used.

inputs #1 and #2: to select one or more variables (table columns) to be taken into account for the calculation of the forecasts.

output #1 and #2: to select the target variable to be predicted.

NB: The inputs and outputs can be set up identically or differently for each of the two networks.

ready: to validate the choices.

from: to indicate from which record (row in the data table) the learnings or predictions will be made. from is to be interpreted as a pointer whose position increments at each step of the process. The input data comes from the record pointed by from and possibly from the previous record (see with past).
It should be noted that the first record starts at zero.

with past: to extend the input data to be taken into account by using the data in the records preceding from.
Exemples:
1 indicates to consider as inputs only the data of the record from.
3 indicates to consider the data of the record from and those of the two previous records.

forecast ahead: to indicate how far away the output prediction is.
Examples:
0 indicates to predict the output variable in the from record.
3 indicates to predict the output variable which is three records ahead.

G : to indicate the amount of calculation done at each stage of learning.
Larger numbers lead to a more refined result but consume more time.

back train set: to indicate the number of trainings to be done at each stage.
Examples:
1 indicates to do only one learning, from the current position of from.
2 indicates to do two learning operations, one from the current position of from, the second from the previous record.
A greater number of training sessions makes it possible to refine the forecasts but increases the calculation time.
Increasing the number of apprentices can also be counterproductive.

forecasts to run: to indicate the number of forecasts to be calculated, plus 1.
Examples:
1 indicates to calculate a forecast from from and then another one forward by one record.
3 indicates to calculate 4 forecasts by advancing from one record at a time.
NB: minimum is two steps (1 + 1) to form a segment.

<u>deep</u>: to indicate the level of complexity of the two neural networks.

A higher level of complexity allows for more combinations.

This makes sense if the number of <u>input</u> variables is large and/or the forecast is difficult to find.

Increasing the complexity of neural networks too much can be counterproductive in learning.

Memory usage and computing time increase with the level of complexity and a browser crash can occur during the learning process.

<u>real plots to display</u>: to indicate the number of known values of the two <u>output</u> variables to be displayed, starting from <u>from</u>.

Examples:

15 indicates to display the fifteen values of the two ouput from <u>from</u>.

If the two <u>output</u> variables are identical, their display will be merged into a single curve.

<u>0/1 view</u>: to indicate to display all values in real or normalized mode between 0 and 1.

Standard mode allows visual comparison of values with a large deviation.

<u>to do</u>: to indicate the action to be taken.

+learning+ to start the learning phase.

+forecast+ to start the forecasting phase (after a learning phase).

+display only real+ to display only the known values of the two <u>output</u> variables.

+learn more+ to do a new learning phase without resetting both neural networks.

Ancillary functions

<u>save report</u>: produces a report containing the graph and the settings.

<u>Demo & Help:</u> download the demo database and this documentation.

The console

start: to start the selected operation to do.

stop: to interrupt the operation. The interruption of a learning phase leaves both neural networks in the state they are in at the time of the interruption. They can therefore still be used for forecasting.

On loading, the console displays the license used and its validity period.

The progress of the learning phases is displayed learning by learning.

The last value indicated in each cartridge reports the learning process. It is purely indicative.
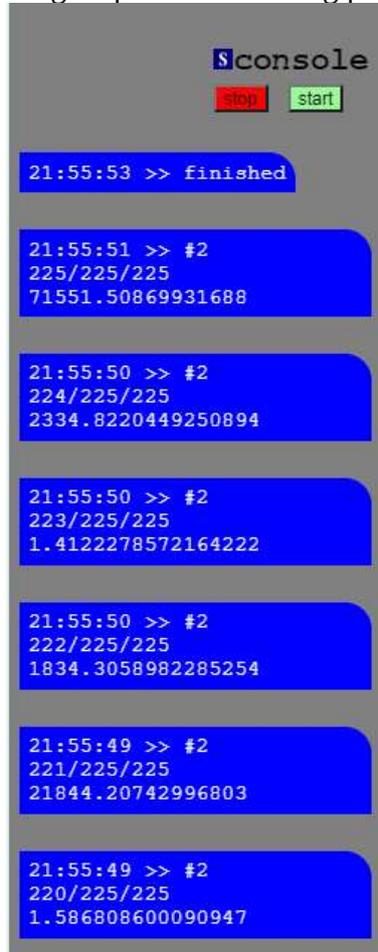
S console

stop   start

21:55:53 >> finished

21:55:51 >> #2
225/225/225
71551.50869931688

21:55:50 >> #2
224/225/225
2334.8220449250894

21:55:50 >> #2
223/225/225
1.4122278572164222

21:55:50 >> #2
222/225/225
1834.3058982285254

21:55:49 >> #2
221/225/225
21844.20742996803

21:55:49 >> #2
220/225/225
1.586808600090947

Data preparation

A simple approach is proposed here.
Free third-party software to use : DB Browser for SQLite https://sqlitebrowser.org/
This software creates and manages SQLite databases in a simple way.

<u>Step 1</u> Collect the data in CSV tabular form, with the first row containing the variable names. Only numeric data will be taken into account.

<u>Step 2</u> Import the CSV data table(s) into a Sqlite database with the DB Browser for SQLite software.

<u>Step 3</u> Modify if necessary the type of numerical data (Modify table) and save the database file.

To be respected

- The numerical data must be in Anglo-Saxon format (the point as the decimal delimiter).

 - The type of all numeric data in the SQLite database must be in "REAL" format.
Modify the type if necessary (Modify table).

- Variable names, including non-numeric variable names, must contain only alphanumeric characters. Spaces and other characters are prohibited and cause a failure.

Good to know

- If a data table contains a time series, the last row must be the most recent.

- If a column is named "Scio_X" its content (text, date, numerical value...) will be added in the abscissa of the graph. Useful for a better reading of the graph.

- A zero value will be interpreted as 0 (zero). A mean value should be preferred to replace insignificant null values.

- There is no need to standardize the data.

-When transforming textual data into numerical data, the proximity of the textual values must be reflected in the numerical values. For example, in a series of colours, "blue" and "purple" are close, but translated numerically by their alphabetical order they will be far apart.

Data security

The data used does not pass through the Internet. Once Scio-machine.com has been loaded on the browser the operations will be done locally, the internet connection can even be disconnected.

User licenses

There are two user licenses.

Community channel for personal or educational use. Accessible free of charge from any compatible browser (Chrome, Firefox...) at http://scio-machine.com. Limited use in number of input variables.

Corporate channel For professional use. Accessible only on Windows with a license.
Launching is done with the executable "Scio-machine-Launcher.exe" to be downloaded at http://scio-machine.com.
Not limited use.
One of these two compatible browsers, Firefox or Chrome, must be installed on the computer.
A same license can be deployed on several Windows workstations, without any network or domain constraints.

Limitations

Profiling can only be done on learning phase done with with past = 1

Neural networks cannot be saved for re-use.

Scio-machine.com is not...
Scio-machine.com is not a statistical software.
Scio-machine.com is not a mathematical regression calculation software.
Scio-machine.com is not mathematical modeling software.

Contact

service@scio-machine.com