

Scio-machine.com

Logiciel de prévisions

2.0

Scio-machine.com est un logiciel facile à utiliser et performant.

Pas d'installation, pas de programmation et une interface unique.

Des prévisions à partir de séries multi-variables temporelles, ou non.

Deux réseaux neuronaux en parallèle.

Des données sécurisées.

**Faites l'expérience
tout de suite**

allez sur

<https://machine1.scio-machine.com/engine.html>

téléchargez

la base de données de démonstration à
Demos & Help

faites ces deux expériences

Le cours de la paire de devises AUD/NZD, 2 jours avant.

L'hypoténuse d'un triangle rectangle, sans formule.

Avant de commencer

2 notions à connaître

Le processus stationnaire d'une série de données caractérise la prévisibilité mathématique de son évolution.

Une suite aléatoire de nombres aura un degré de processus stationnaire nul.

Une suite arithmétique comme (1, 2, 3, 4... 12, 13) aura un degré de processus stationnaire fort.

D'autres suites de données peuvent avoir plusieurs processus stationnaires fonctionnant de manière partielle ou complexe.

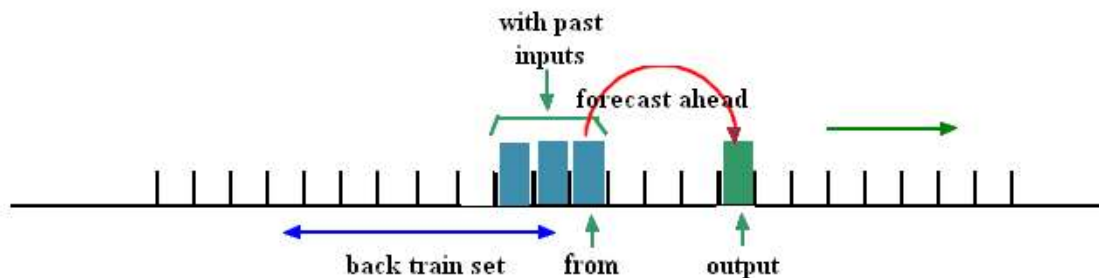
Le dépendance du futur au passé caractérise l'intervalle entre la cause et effet.

Une suite arithmétique comme (1, 2, 3, 4... 12, 13) aura une dépendance du futur au passé de 1.

Une suite aléatoire de nombres aura une dépendance du futur au passé de 0.

Ces deux notions pourront être utilisées de manière intuitive ou empirique sans nécessité de connaissances mathématiques.

Schéma du processus d'apprentissage



L'apprentissage consiste à prendre en compte les données (inputs) se trouvant dans la zone with past pour prévoir une donnée (output) se trouvant dans la zone output.

from est la tête de la zone with past.

forecast ahead est l'éloignement entre from et output.

back train test est une zone arrière permettant de faire des test supplémentaires à chaque étape.

L'apprentissage se fait en plusieurs étapes (forecast to run) en avançant toutes les zones d'un enregistrement à la fois.

Lors du processus d'apprentissage output ne peut dépasser le dernier enregistrement de la table.

Lors du processus de prévision from ne peut dépasser le dernier enregistrement de la table.

Tous les paramètres sont détaillés plus loin dans ce document.

En conclusion

Si le processus stationnaire est nul ou fort, l'apprentissage pourra être fait depuis n'importe quel endroit de la série.

Autrement, l'apprentissage devra être fait au plus proche de la prévision à faire, généralement en fin de série.

L'éloignement entre from et output (forecast ahead) ne devraient pas être supérieur à la durée de dépendance du futur au passé.

with past ne devrait pas être supérieur à la durée de dépendance du futur au passé +1.

Expérience 1

Prévoir le cours de la paire de devises AUD/NZD.

Table à utiliser : *forex*.

Analyse

- + C'est une série temporelle sans processus stationnaire fort mais supposé non nul.
- + La dépendance du futur au passé ne dépasse pas 2 séances de cotation.
- + La valeur recherchée est le cours à la clôture.
- + Les données utilisées seront les historiques de plusieurs paires de devises disponibles dans la table.

Paramétrage

<u>inputs</u> #1 et #2	AUDCADOpen ... NZDCHFVolume
<u>output</u> #1 et #2	AUDNZDClose
<u>from</u>	54 <i>Ici <u>from</u> est placé à 54 pour l'intérêt pédagogique, il serait placé le plus loin possible vers la fin dans une situation réelle.</i>
<u>with past</u>	3 <i>Durée de dépendance +1</i>
<u>forecast ahead</u>	2 <i>Durée de dépendance</i>
<u>G</u>	75
<u>back train set</u>	1 Si la durée de dépendance est courte et/ou la stationnarité mal établie les tests supplémentaires risquent d'être hors contexte.
<u>forecasts to run deep</u>	1 (soit 1+1) 5 Un nombre plus important pourrait être adopté.
<u>0/1 view</u>	no
<u>real plots to display to do</u>	20 <i>learn</i> pour la phase d'apprentissage <i>forecast only</i> pour la phase de prévisions, avec : <u>forecast to run</u> = 3 (soit 3+1)

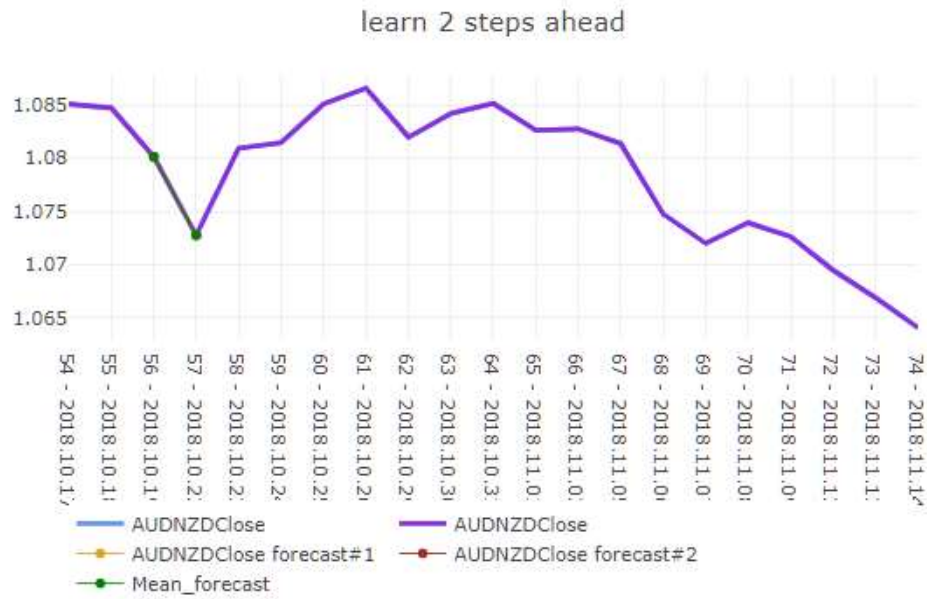
Résultats

Sur le graphique obtenu en fin de prévisions on constate que les deux premières prévisions sont en fait la "restitution" de l'apprentissage, from n'ayant pas été déplacé. Quant aux deux dernières prévisions, elles sont le pur fruit des réseaux de neurones et des données connues inputs 2 cotations antérieures. On constate bien l'anticipation de la fin de la baisse et une remontée des cours. Dans un contexte réel, les données connues n'existeraient pas au-delà de l'enregistrement 57 (en bleu et violet).

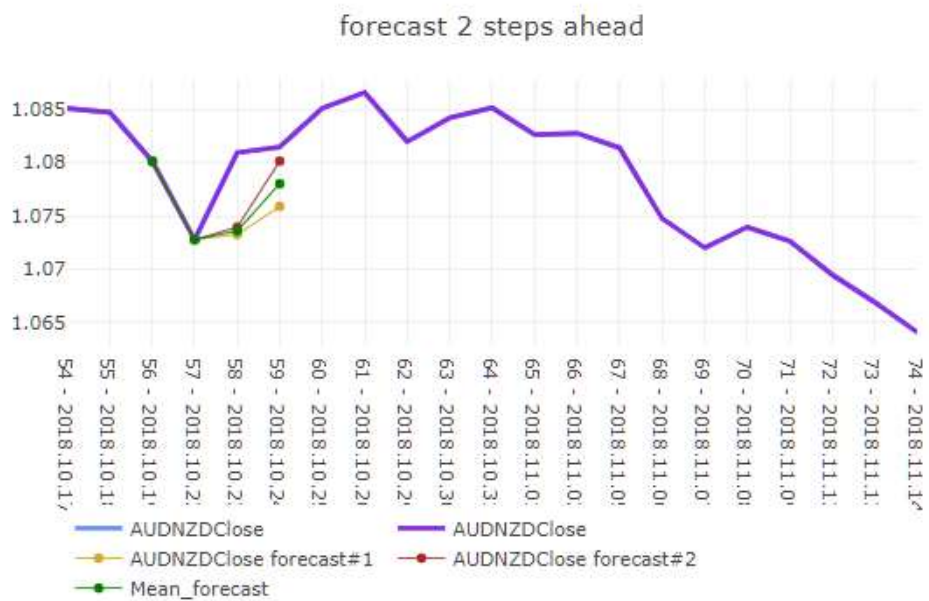
Les deux prévisions divergent quelque peu malgré avoir été obtenues avec les mêmes données et paramètres. Cela est dû au fait que la technique génétique de formation des deux réseaux les amène à se distinguer. La ligne verte représente la moyenne des deux prévisions.

Il est fort probable que les réseaux obtenus ne fassent pas de bonnes prévisions à d'autres endroits de la série dû au fait que la série n'a pas de processus stationnaire fort.

Graphique obtenu en fin d'apprentissage



Graphique obtenu en fin de prévisions



Expérience 2

Donner la longueur de l'hypoténuse d'un triangle rectangle selon ses deux autres côtés.

Table à utiliser : *pithagoras*.

Chaque enregistrement contient les longueurs des trois côtés d'un triangle rectangle. Les enregistrements n'ont pas d'ordre particulier.

Analyse

- + Le processus de stationnarité est nul
- + La dépendance du futur au passé est 0.
- + La valeur recherchée l'hypoténuse.
- + Les données utilisées sont les longueurs des deux autres côtés.

Paramétrage

<u>inputs #1 et #2</u>	AB, AC
<u>output #1 et #2</u>	hypoténuse
<u>from</u>	25
	<u>from</u> peut débuter n'importe où.
<u>with past</u>	1
	C'est-à-dire ne contenant que l'enregistrement <u>from</u> .
<u>forecast ahead</u>	0
	Pas de dépendance du futur au passé, la variable à prévoir est celle se trouvant dans l'enregistrement courant de <u>from</u> .
<u>G</u>	15
<u>back train set</u>	20
	Le <u>back train set</u> est important dans les séries à processus de stationnarité nul.
<u>forecasts to run</u>	10 (soit 10+1)
	Pourrait être plus grand pour affiner l'apprentissage.
<u>deep</u>	15
	Pourrait être plus grand pour affiner l'apprentissage.
<u>0/1 view</u>	no
<u>real plots to display</u>	45
<u>to do</u>	<i>learn</i> pour la phase d'apprentissage
	<i>forecast only</i> pour la phase de prévision, avec :
	<u>from</u> = 250
	<u>forecast to run</u> = 40 (soit 1+1)

Résultats

Le graphique des prévisions montre qu'une très petite quantité d'apprentissages a suffi pour obtenir des prévisions cohérentes, bien que celles-ci n'ont pas la précision mathématique. En effet, Scio-machine.com ne construit pas de modèle mathématique. Un apprentissage plus approfondi aurait pu donner des prévisions encore plus fines.

Les deux prévisions divergent malgré avoir été obtenues avec les mêmes données et paramétrages. Cela est dû au fait que la technique génétique de formation des deux réseaux les amène à se distinguer. La ligne verte représente la moyenne des deux prévisions.

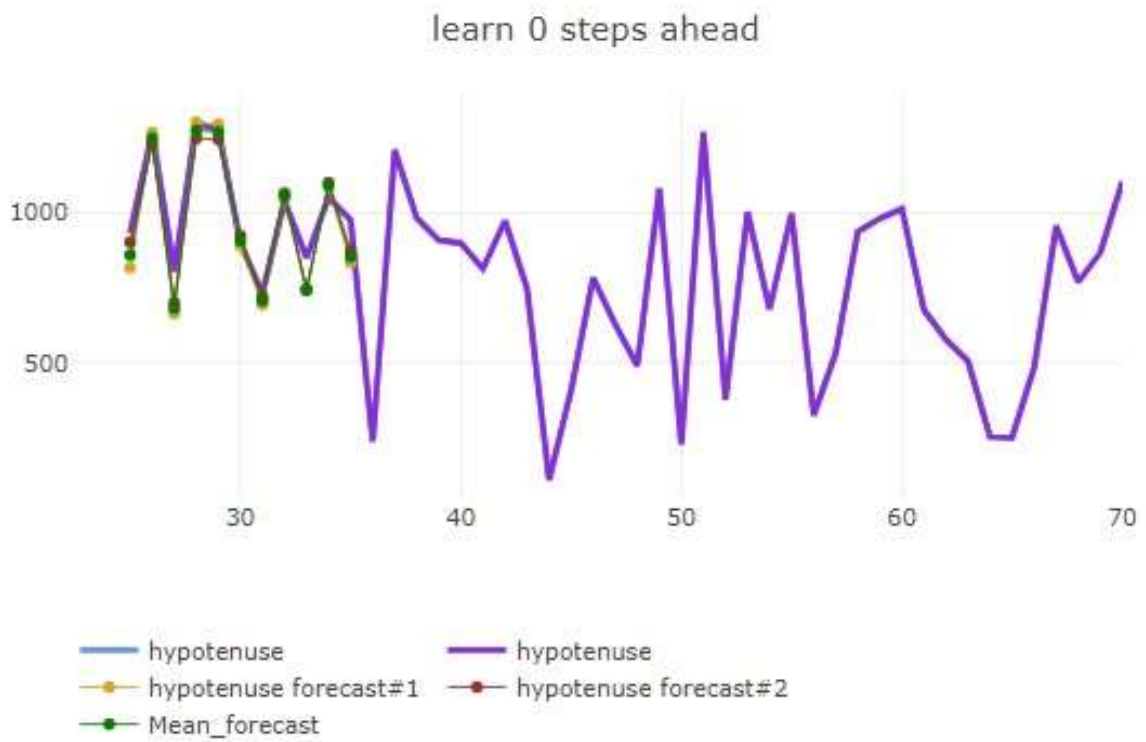
Le profilage est un outil permettant de définir un profil selon des contraintes.

Ici les contraintes choisies étaient hypoténuse 1000 et AC 900.

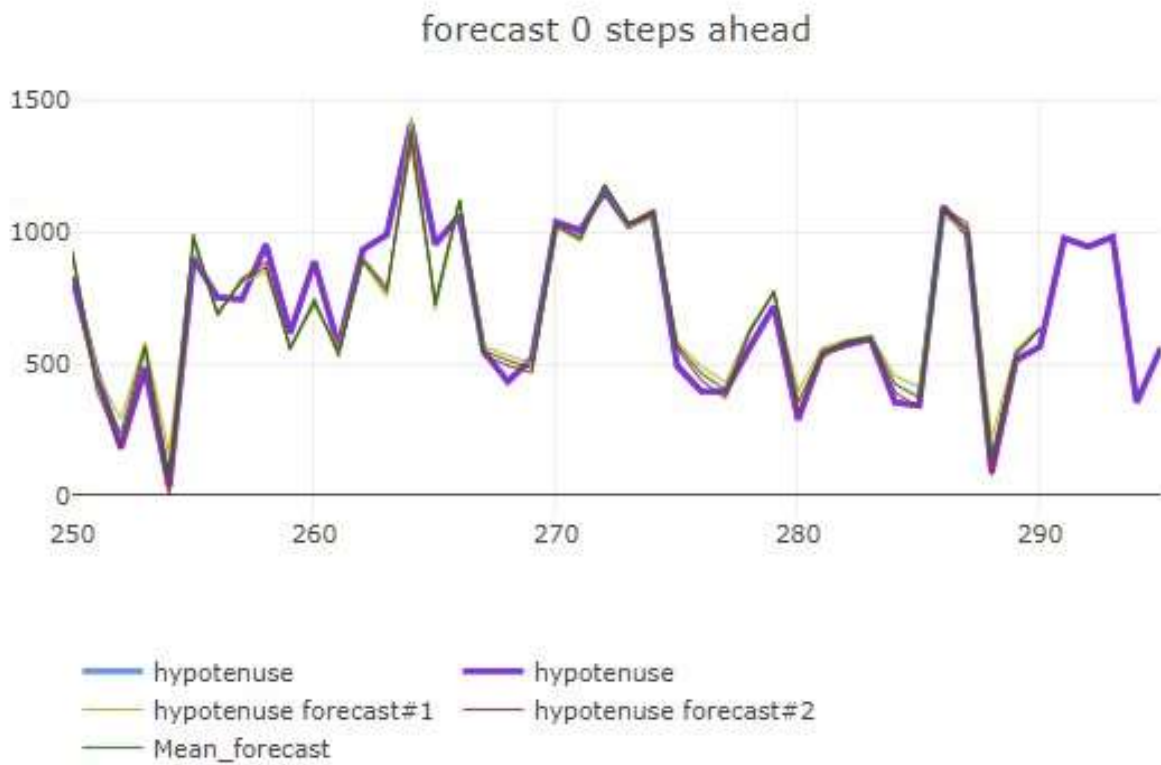
Les bornes minimum et maximum des valeurs des variables inputs sont définies par leurs minimum et maximum respectifs dans la série.

Il y a un profilage possible pour output #1 et #2..

Graphique obtenu en fin d'apprentissage



Graphique obtenu en fin de prévisions



Profilage

PROFILER #1

hypotenuse
1001.6835

AB

391.6600

Your comment

AC

898.7000

Your comment

Save (CSV)

Une interface unique

l'affichage des résultats

le profilage

la zone de paramétrage

la console d'information

L'affichage des résultats

le graphique dispose d'un menu d'icônes en haut à droite permettant la sauvegarde du graphique au format png. Les données affichées dans le graphique sont dynamiques et peuvent être masquées/affichées. Les données du graphique peuvent être éditées, analysées, traitées, sauvegardées et partagées avec **plotly Chart Studio**, avec ou sans compte (une connexion internet doit être établie pour le transfert des données au site plot.ly).



Le profilage

PROFILER#1 et #2 permettent après apprentissage de modifier dynamiquement les valeurs des variables inputs en vue d'obtenir des profils pertinents. Les caractéristiques des profils obtenus peuvent être sauvegardés dans des fichiers CSV.

The image shows a software interface for 'PROFILER #1'. At the top, a cyan bar displays 'AUDCADClose' with a value of '0.9710'. Below this, five adjustable parameters are listed, each with a slider and a text input field for a comment:

- AUDCADOpen**: Value 0.9450, Comment: Your comment
- AUDCADHigh**: Value 0.9450, Comment: Your comment
- AUDCADLow**: Value 0.9401, Comment: Your comment
- AUDCADClose**: Value 0.9424, Comment: Your comment
- AUDCADVolume**: Value 89270.0000, Comment: Your comment

At the bottom of the yellow area is a 'Save (CSV)' button. Below the yellow area is a pink bar labeled 'PROFILER #2'.

Le paramétrage

DATABASE / TABLE

database SCIO-MAC...E.sqlite table

FORECAST #1 FIELDS

inputs	output	ready
<input type="text" value="AUDCADOpen"/> <input type="text" value="AUDCADHigh"/> <input type="text" value="AUDCADLow"/> <input type="text" value="AUDCADClose"/> <input type="text" value="AUDCADVolume"/> <input type="text" value="AUDCHFOpen"/> <input type="text" value="AUDCHFHigh"/> <input type="text" value="AUDCHFLow"/>	<input type="text" value="AUDCADClose"/>	<input type="text" value="validate"/>

FORECAST #2 FIELDS

inputs	output	ready
<input type="text" value="AUDCADOpen"/> <input type="text" value="AUDCADHigh"/> <input type="text" value="AUDCADLow"/> <input type="text" value="AUDCADClose"/> <input type="text" value="AUDCADVolume"/> <input type="text" value="AUDCHFOpen"/> <input type="text" value="AUDCHFHigh"/> <input type="text" value="AUDCHFLow"/>	<input type="text" value="AUDCADClose"/>	<input type="text" value="validate"/>

FORECAST PARAMETERS

from <input type="text" value="79"/> <input type="text" value="Min"/> <input type="text" value="Max"/>	with past <input type="text" value="3"/> <input type="text" value="Min"/> <input type="text" value="Max"/>	forecast ahead <input type="text" value="3"/> <input type="text" value="Min"/> <input type="text" value="Max"/>
--	--	---

LEARNING PARAMETERS

G <input type="text" value="45"/> <input type="text" value="Min"/> <input type="text" value="Max"/>	back train set <input type="text" value="1"/> <input type="text" value="Min"/> <input type="text" value="Max"/>	forecasts to run <input type="text" value="4"/> <input type="text" value="Min"/> <input type="text" value="Max"/>
<input type="text" value="3"/> deep		

DISPLAY PARAMETERS

real plots to display <input type="text" value="15"/> <input type="text" value="Min"/> <input type="text" value="Max"/>	<input type="text" value="0/1 view"/> <input type="text" value="no"/>
---	---

RUNNING PARAMETERS

<input type="text" value="to do"/> <input type="text" value="learning"/>
<input type="text" value="save report"/>

HELP

Les paramètres

database: pour charger la base de données Sqlite à utiliser.

table: pour sélectionner la table (tableau de données) à utiliser.

inputs #1 et #2: pour sélectionner une ou plusieurs variables (colonnes de la table) à prendre en compte pour le calcul des prévisions.

output #1 et #2: pour sélectionner la variable à prévoir.

NB: Les inputs et output peuvent être paramétrés de manière identique ou différente pour chacun des deux réseaux.

ready: pour valider les choix.

from: pour indiquer à partir de quel enregistrement (rangée dans le tableau de données) les apprentissages ou prévisions seront faites. from doit être interprété comme un pointeur dont la position s'incrémente à chaque étape du processus. Les données inputs proviennent de l'enregistrement pointé par from et éventuellement de ceux qui le précèdent (voir with past). Il est à noter que le premier enregistrement commence à zéro.

with past: pour étendre les données inputs à prendre en compte en utilisant celles se trouvant dans les enregistrements qui précèdent from.

Exemples:

1 indique de ne prendre en compte en inputs que les données de l'enregistrement from.

3 indique de prendre en compte les données de l'enregistrement from et celles des deux enregistrements précédents.

forecast ahead: pour indiquer l'éloignement de la prédiction output.

Exemples:

0 indique de prédire la variable output qui se trouve dans l'enregistrement from.

3 indique de prédire la variable output qui se trouve trois enregistrements plus loin.

G: pour indiquer la quantité de calcul faite à chaque étape de l'apprentissage.

Un nombre plus grand induit un résultat plus affiné mais consomme plus de temps.

back train set: pour indiquer le nombre d'apprentissages à faire à chaque étape.

Exemples:

1 indique de ne faire qu'un seul apprentissage, depuis de la position courante de from.

2 indique de faire deux d'apprentissages, un depuis la position courante de from, le second depuis l'enregistrement précédent.

Un nombre d'apprentissages plus important permet d'affiner les prévisions mais augmente le temps de calcul.

Augmenter le nombre d'apprentissages peut aussi s'avérer contre productif.

forecasts to run: pour indiquer le nombre de prévisions à calculer, plus 1.

Exemples:

1 indique de calculer une prévision depuis from puis une autre en avançant d'un enregistrement.

3 indique de calculer 4 prévisions en faisant avancer from d'un enregistrement à chaque fois.

NB: minimum est de deux étapes pour constituer un segment.

deep: pour indiquer le niveau de complexité des deux réseaux neuronaux.

Un niveau de complexité plus important permet plus de combinaisons. Cela est judicieux si le nombre de variables inputs est important et/ou la prévision difficile à trouver.

Augmenter trop la complexité des réseaux neuronaux peut s'avérer contre productif en apprentissage. La mémoire utilisée et le temps de calcul augmentent avec le niveau de complexité et un crash du navigateur peut survenir pendant le processus d'apprentissage.

real plots to display: pour indiquer le nombre de valeurs connues des deux variables output à afficher, à partir de from.

Exemples:

15 indique d'afficher les quinze valeurs des deux output à partir de from.

Si les deux variables output sont identiques leur affichage se confondra en une seule courbe.

0/1 view: pour indiquer d'afficher toutes les valeurs en mode réel ou normalisé entre 0 et 1.

Le mode normalisé permet de comparer visuellement des valeur réelles ayant un grand écart.

to do: pour indiquer l'action à faire.

+*learning*+ pour lancer la phase d'apprentissage.

+*forecast*+ pour lancer la phase des prévisions (après une phase d'apprentissage).

+*display only real*+ pour n'afficher que les valeurs connues des deux variables output.

+*learn more*+ pour faire une nouvelle phase d'apprentissage sans réinitialisation des deux réseaux neuronaux.

Fonctions annexes

save report: produit un rapport contenant le graphique et les paramétrages.

Demo & Help: télécharge la base de données de démonstration ainsi que cette documentation.

La console

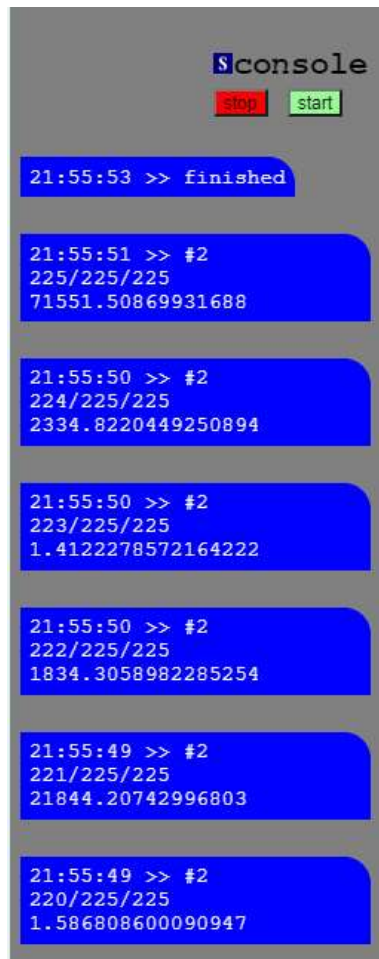
start: pour lancer l'opération choisie à to do.

stop: pour interrompre l'opération. L'interruption d'une phase d'apprentissage laisse les deux réseaux neuronaux dans l'état où ils se trouvent au moment de l'interruption. Ils peuvent donc toujours être utilisés pour des prévisions.

Au chargement la console affiche la licence utilisée et sa durée de validité.

La progression des phases d'apprentissage est affichée apprentissage par apprentissage.

La dernière valeur indiquée dans chaque cartouche rend compte de l'apprentissage. Elle est purement indicative.



The screenshot shows the 'Sconsole' interface with a 'stop' button and a 'start' button. Below these are several blue rounded rectangular cards, each containing a timestamp, a status, and numerical data. The cards are arranged vertically, with the top card showing 'finished' and the others showing '#2' and various numbers.

```
Sconsole
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
```

Préparations des données

Une démarche simple est ici proposée.

Logiciel tiers libre à utiliser : DB Browser for SQLite <https://sqlitebrowser.org/>

Ce logiciel crée et gère les bases de données SQLite de manière simple.

Etape 1 Rassembler les données sous formes tableaux CVS, la première rangée contenant le nom des variables. Seules les données numériques seront prises en compte.

Etape 2 Importer le ou les tableaux de données CVS dans une base de donnée Sqlite avec le logiciel DB Browser for SQLite.

Etape 3 Modifier si nécessaire le type des données numériques (Modifier table) et enregistrer le fichier de la base de donnée.

A respecter

- Les données numériques doivent être au format anglo-saxon (le point comme délimiteur des décimales).
- Le type de toutes les données numériques dans la base SQLite doit être au format "REEL". Modifier le type si nécessaire (Modifier table).
- Les noms des variables, y compris non numériques, ne doivent contenir que des caractères alphanumériques. Les espaces et autres caractères sont à prohiber et provoque un blocage.

Bon à savoir

- Si un tableau de données contient une série temporelle, la dernière rangée doit être la plus récente.
- Si une colonne est nommée "Scio_X" son contenu (texte, date, valeur numérique...) sera rajouté dans l'abscisse du graphique. Utile pour une meilleure lecture du graphique.
- Une valeur nulle sera interprétée comme 0 (zéro). Il conviendra de préférer une valeur moyenne en remplacement des valeurs nulles non significatives.
- Il n'est pas nécessaire de normaliser les données.
- En cas de transformation de données textuelles en données numériques, la proximité des valeurs textuelles doit se refléter dans les valeurs numériques. Par exemple, dans une série de couleurs, "bleu" et "violet" sont proches, mais traduit numériquement par leur ordre alphabétique elles seront éloignées.

Sécurité des données

Les données utilisées ne transitent pas sur le réseau internet. Une fois Scio-machine.com chargé sur le navigateur les opérations seront faites localement, la liaison internet peut même être déconnectée.

Licences d'utilisation

Il existe deux licences d'utilisation.

Community channel pour un usage personnel ou éducatif. Accessible gratuitement depuis tout navigateur compatible (Chrome, Firefox...) à l'adresse <http://scio-machine.com>. Usage limité en nombre de variables inputs.

Corporate channel pour un usage professionnel. Accessible uniquement sur Windows avec une licence. Le lancement se fait avec l'exécutable "Scio-machine-Launcher.exe" à télécharger à l'adresse <http://scio-machine.com>.

Usage non limité.

L'un de ces deux navigateurs compatibles, Firefox ou Chrome, doit être installé sur l'ordinateur. Une même licence Corporate channel peut être déployée sur plusieurs postes Windows, sans contrainte de réseaux ou de domaine.

Limitations

Le profilage ne peut se faire que sur des apprentissages avec with past = 1

Les réseaux neuronaux ne peuvent pas être sauvegardés en vue de leur réemploi.

Scio-machine.com n'est pas...

Scio-machine.com n'est pas un logiciel de statistique.

Scio-machine.com n'est pas un logiciel de calcul de régression mathématique.

Scio-machine.com n'est pas un logiciel de modélisation mathématique.

Contact

service@scio-machine.com