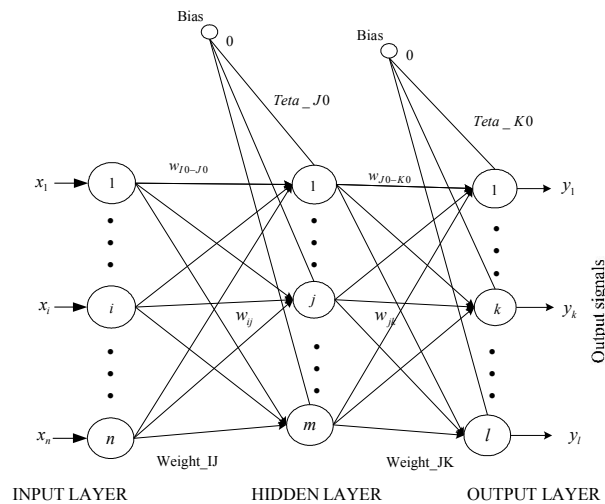


HƯỚNG DẪN SỬ DỤNG PHẦN MỀM MẠNG NƠ RON SPICE-MLP

Cao Thăng 2003 – 2007

Cập nhật 2011 Jan.



1. GIỚI THIỆU

Mục đích chính của tài liệu này là hướng dẫn bạn đọc sử dụng phần mềm Spice-MLP, và không đi sâu vào lý thuyết mạng nơ ron. Nếu muốn tìm hiểu thêm về lý thuyết, các bạn có thể tham khảo các tài liệu về mạng nơ ron có sẵn trên internet hoặc sách chuyên ngành.

Mục đích của các phần mềm Spice-MLP và SpiceSOM là giúp bạn biết cách sử dụng mạng nơ ron một cách cơ bản, nhanh chóng và hiệu quả mà không phải đọc nhiều về lý thuyết mạng nơ ron. Khi bạn đã hiểu rõ từng chức năng của các phần mềm này, bạn có thể dễ dàng tiếp cận với lý thuyết cũng như các tài liệu tham khảo về mạng nơ ron.

Spice-MLP là phần mềm mạng nơ ron 3 lớp, với nhiều đầu vào và nhiều đầu ra. Spice-MLP được viết với mục đích hướng dẫn sinh viên và nghiên cứu sinh học tập và sử dụng mạng nơ ron để mô hình hóa nhiều loại dữ liệu khác nhau. Hiện tại Spice-MLP đang được nhiều bạn trên thế giới sử dụng. Spice-MLP có giao diện với tiếng Việt, tiếng Anh và tiếng Nhật.



Tùy thuộc vào phiên bản Spice-MLP mà bạn đang sử dụng, một số hướng dẫn hoặc giao diện minh họa trong tài liệu này có thể khác với thực tế.

Spice-MLP được viết bởi CAO THANG khi tác giả làm việc tại Soft Intelligence Laboratory, Ritsumeikan University, Japan, 2003-2007 và thường xuyên được cập nhật theo yêu cầu của người sử dụng

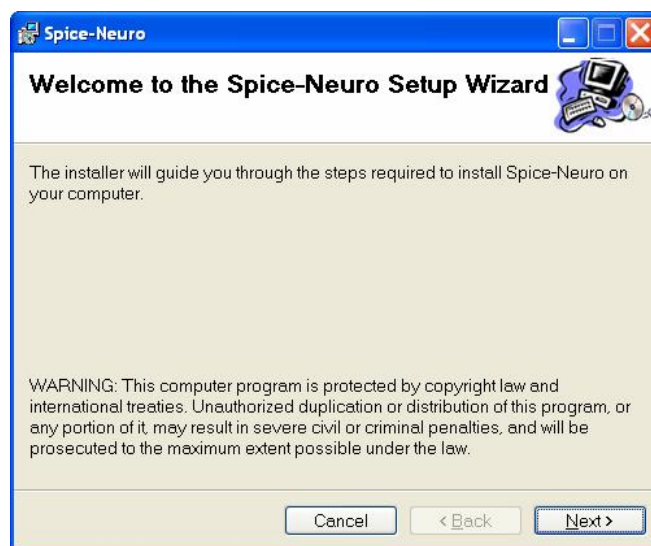
SpiceSOM và Spice-MLP có thể download được tại: download.cnet.com

Để hiểu hơn về mạng nơ ron với các ứng dụng như nhận dạng khuôn mặt, người đi bộ, dự báo chứng khoán, tỷ giá..., các bạn nên đọc thêm tài liệu “Một số ví dụ phân loại dùng SOM và MLP Neural Network” (neural_network_practical_use_vi.pdf).

Nếu có thắc mắc hoặc cần yêu cầu thêm về chức năng của Spice-MLP, bạn có thể liên hệ với tác giả tại <http://spiceneuro.wordpress.com> hoặc spiceneuro AT gmail DOT com. Cảm ơn các bạn.

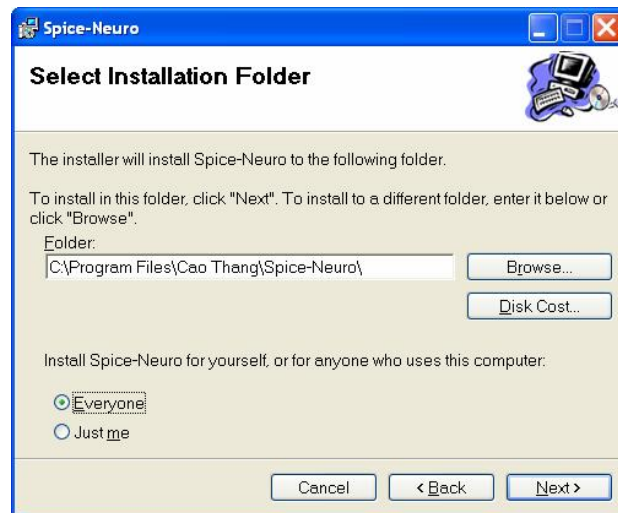
2. CÀI ĐẶT SPICE-MLP

Download file cài đặt của Spice-MLP và chạy setup.exe, trên màn hình hiện ra:



Hình 1. Cài đặt

Chọn Next, sau đó chọn thư mục mà bạn muốn cài đặt Spice-MLP vào, chọn Next và Next tiếp. Spice-MLP sẽ được cài vào thư mục mà bạn đã chọn.



Hình 2. Chọn Folder để cài vào

Lưu ý:

Với Windows version cũ, Nếu sau khi cài đặt, chương trình không chạy, có thể bạn cần cài *Microsoft .NET Framework Redistributable Package 3.5.21022* trước khi cài Spice-MLP.

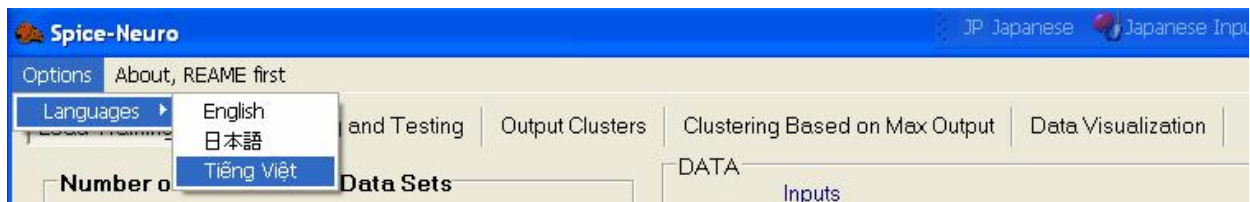
Nếu dữ liệu của bạn dùng dưới dạng MDB format, có thể bạn cần cài thêm *Microsoft Data Access Components*.

3. SỬ DỤNG SPICE-MLP

Chạy Spice-MLP bằng cách click vào biểu tượng Spice-MLP trên desktop hoặc trong “Start → Programs → Cao Thang’s Spice-MLP → Spice-MLP”.



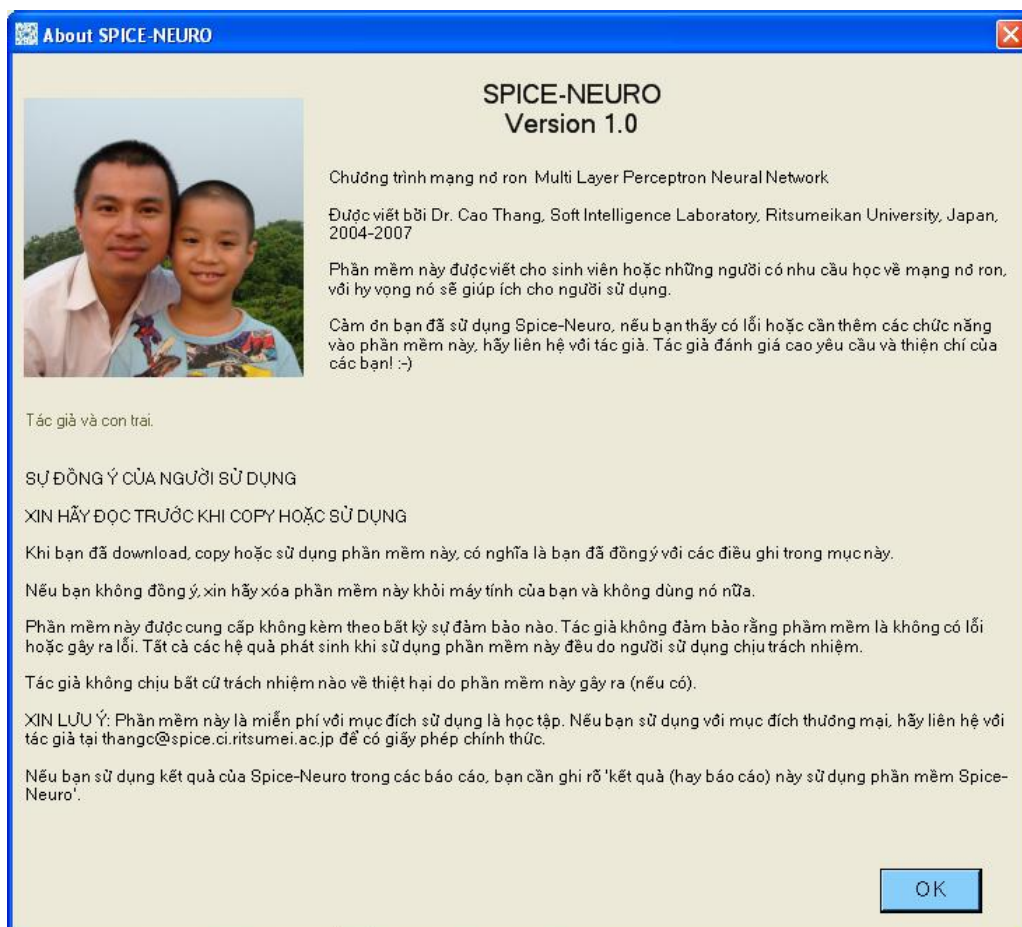
Trên màn hình hiện ra giao diện tiếng Anh, bạn có thể chọn ngôn ngữ tiếng Việt hoặc tiếng Nhật bằng cách chọn “Options → Languages”.



Hình 3. Chọn ngôn ngữ

Trong các giao diện minh họa dưới đây, ngôn ngữ được sử dụng là tiếng Việt.

Menu “Về chương trình. Hãy đọc trước khi sử dụng” là giới thiệu vắn tắt về Spice-MLP và sự đồng ý của người sử dụng. Bạn cần đọc kỹ trước khi sử dụng Spice-MLP.



Hình 4. Về chương trình, ảnh trong đó chụp năm 2007

3.1. Chuẩn bị dữ liệu

Để Spice-MLP đọc được dữ liệu của bạn, bạn cần chuẩn bị dữ liệu của mình theo chuẩn sau.

3.1.1. Dữ liệu dạng file text

Dữ liệu dạng file text cần được chuẩn bị thành các hàng và cột. Đầu tiên là ID, sau đó là đầu vào và tiếp theo là đầu ra. Các giá trị được phân cách bằng dấu phẩy với file CSV (Comma Separated Value File Format) hoặc dấu Tab với file TXT (Tab Separated Value File Format). Bạn có thể dùng MS Excel để biên soạn dữ liệu, sau đó lưu vào file file text hoặc file csv. Ví dụ dữ liệu với 2 đầu vào và 3 đầu ra được tổ chức như bảng 1.

Bảng 1. Dữ liệu text với 2 đầu vào và 3 đầu ra

ID	X0	X1	Y0	Y1	Y2	LABEL
0	0	0	0	0	0	Data 1
1	0	1	1	0	1	Data 2
2	1	0	1	0	1	Data 3
3	1	1	0	1	1	Data 4

Lưu ý:

Dữ liệu phải là dạng số (ngoại trừ nhãn (label) và các ký hiệu đầu vào, đầu ra).

Nếu có dữ liệu trống hoặc null, Spice-MLP sẽ không đọc được. Trong dữ liệu kiểm tra, nếu không có giá trị đầu ra, các bạn cần đặt giá trị ra là 0 hoặc 1, hoặc một giá trị số nào đó.

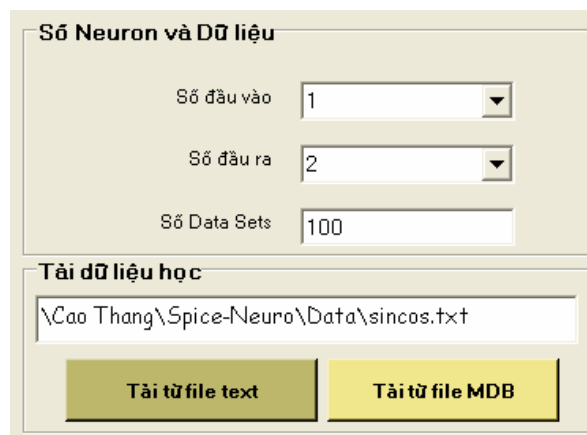
Một số ví dụ về dữ liệu được đặt trong thư mục “Data” của Spice-MLP:

- “Boolean functions.csv” là ví dụ với 4 datasets, 2 inputs và 3 outputs. Đầu vào là các giá trị nhị phân 0 và 1, đầu ra là giá trị của các hàm XOR (Y0), AND (Y1) và OR(Y2) .
- “Herbal data.csv” là ví dụ với 640 datasets, 16 inputs và 33 outputs. Đầu vào là mức độ triệu chứng có giá trị từ 0 tới 1, đầu ra là hệ số các vị thuốc đã được chuẩn hóa trong [0, 1].
- “sincos.csv” là ví dụ với 100 datasets, 1 inputs và 2 outputs. Đầu vào là đối số có giá trị từ 0 tới 2π , được chuẩn hóa trong [0,1] và đầu ra là giá trị hai hàm Sin và Cos của đối số đầu vào đó.
- “iris_for_mlp_4inputs_1output.csv”, “iris_for_mlp_4inputs_3outputs.csv” là dữ liệu của ba loại hoa (Iris setosa, Iris virginica và Iris versicolor), mỗi loại 50 mẫu. Các thuộc tính là độ dài và rộng của đài hoa (sepal) và cánh hoa (petal) tính theo centimeters. Chi tiết tại <http://archive.ics.uci.edu/ml/datasets/Iris>. “iris_for_mlp_4inputs_1output.csv” là dữ liệu 4 đầu vào 1 đầu ra, “iris_for_mlp_4inputs_3outputs.csv” là dữ liệu 3 đầu vào 1 đầu ra.
- “CAD_USD_JPN.csv”, “CAD_USD_JPN_Normalized.csv”, 2489 datasets về tỷ giá CAD \Rightarrow USD, CAD \Rightarrow JPN với 30 đầu vào, 2 đầu ra.

- NASDAQ_5026_data_15inputs_1output.csv là 5026 datasets về chỉ số chứng khoán NASDAQ, với 15 đầu vào và 1 đầu ra.
- Hoabinh_water_level_3_input_1_output.csv là dữ liệu về dự báo lưu lượng nước tương lai trước 10 ngày $Q(t+10)$ của hồ Hòa Bình dựa vào các lưu lượng nước tại thời điểm hiện tại và quá khứ. Dữ liệu có 3 đầu vào gồm lưu lượng nước hiện tại $Q(t)$, lưu lượng nước trước đó 10 ngày $Q(t-10)$ và lưu lượng nước trước đó 20 ngày $Q(t-20)$. Số dữ liệu là 480 mẫu học (từ 1 tới 480) và 90 mẫu kiểm tra (từ 481 tới 570). Dữ liệu này do bạn Phạm Thị Hoàng Nhung, trường ĐH Thủy lợi cung cấp, bạn đọc có thể tham khảo luận văn Master của Phạm Thị Hoàng Nhung (1997) về "khảo sát một số phương pháp học máy tiên tiến, thực hiện việc kết hợp giữa phương pháp học máy mạng neuron với thuật toán gene và ứng dụng vào bài toán dự báo lưu lượng nước đến hồ Hòa Bình". Xin cảm ơn bạn Phạm Thị Hoàng Nhung đã cho phép sử dụng dữ liệu lưu lượng nước hồ Hòa Bình để minh họa trong tài liệu này.

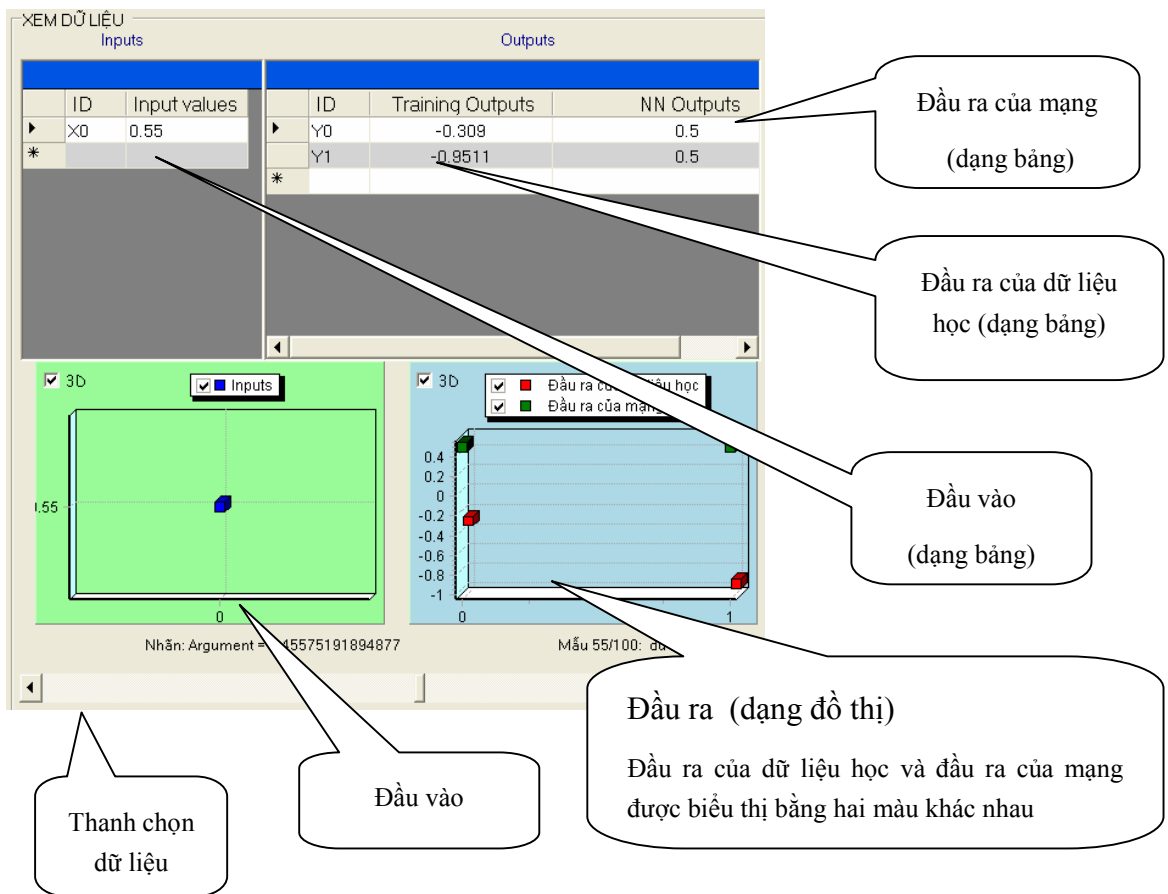
3.2. Load dữ liệu

Giả sử chúng ta dùng dữ liệu là file "sincos.txt" là ví dụ với 100 datasets, 1 đầu vào và 2 đầu ra nói trên. Trong mục "Số Neuron và Dữ liệu" Ta chọn như hình 5:



Hình 5. Chọn tham số để load dữ liệu

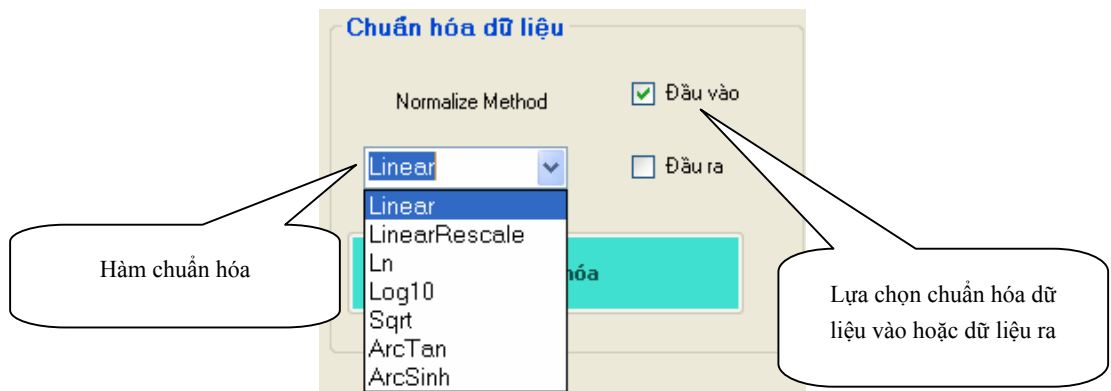
Chọn nút "Tải từ file text", dữ liệu sẽ được tải vào bộ nhớ. Trong mục "XEM DỮ LIỆU" bên phải, bạn có thể xem lại từng dataset của dữ liệu mà bạn vừa load:



Hình 6. Xem dữ liệu

3.3. Chuẩn hóa dữ liệu

Nếu dữ liệu của bạn chưa được chuẩn hóa, bạn có thể dùng chức năng chuẩn hóa dữ liệu như hình 7 sau. Bạn có thể chuẩn hóa dữ liệu đầu vào hoặc đầu ra, hoặc cả hai:



Hình 7. Chuẩn hóa dữ liệu

3.4. Đào tạo mạng.

3.4.1. Chia dữ liệu

Nếu bạn muốn chia dữ liệu làm hai phần, một phần để học và một phần để kiểm tra, bạn dùng chức năng “Chia dữ liệu”. Sau đó bạn cần chọn phần dữ liệu nào để học. Ví dụ sau minh họa việc chia dữ liệu ngẫu nhiên thành hai phần 70% và 30%, dùng 70% để học và 30% để kiểm tra.



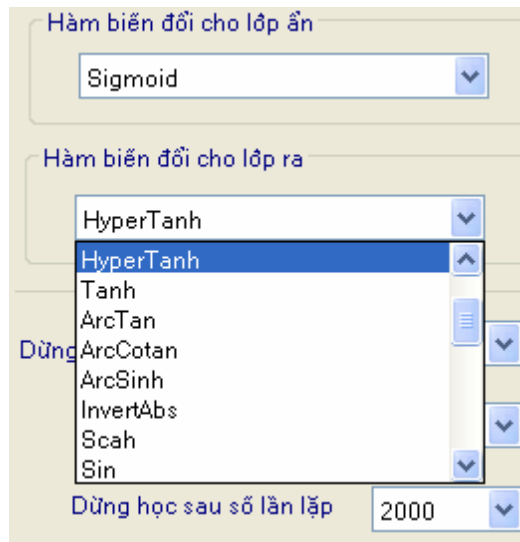
Hình 8. Chia dữ liệu

3.4.2. Chọn dữ liệu học và các tham số

Tiếp theo, bạn cần chọn số neuron cho lớp ẩn (hidden layer), số lần lặp, thời gian học và MSE (Mean of Square Error) yêu cầu. Bạn cũng có thể lựa chọn học thích nghi (tỉ lệ học biến đổi dựa vào MSE học).

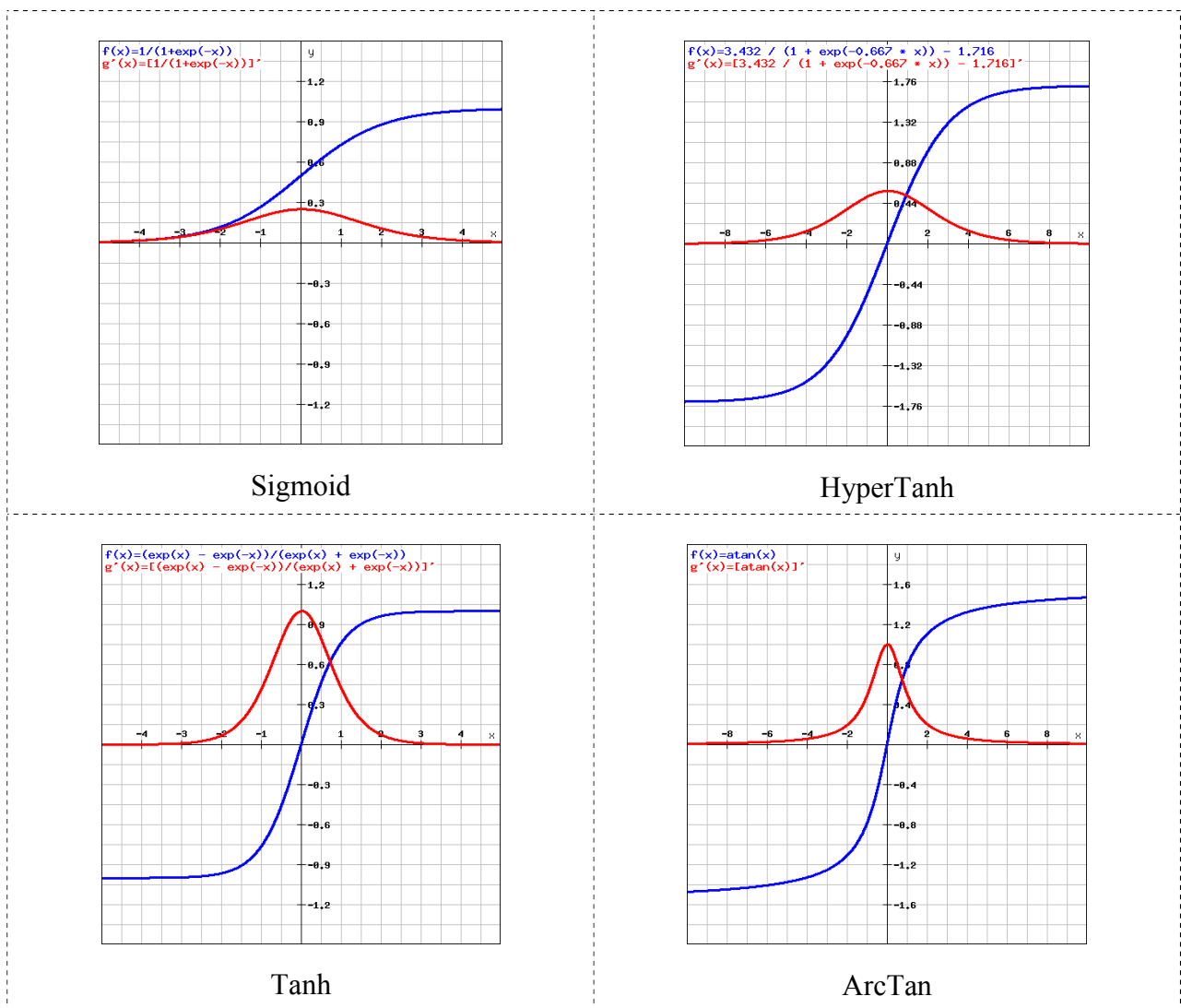


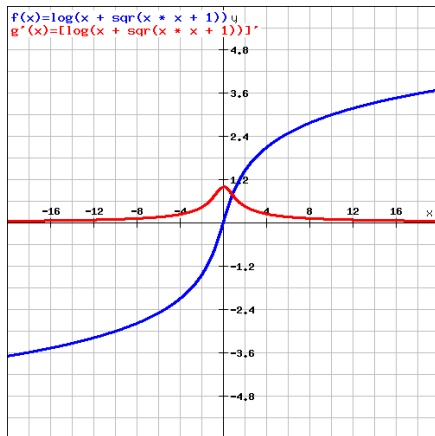
Hình 9. Chọn dữ liệu học và các tham số



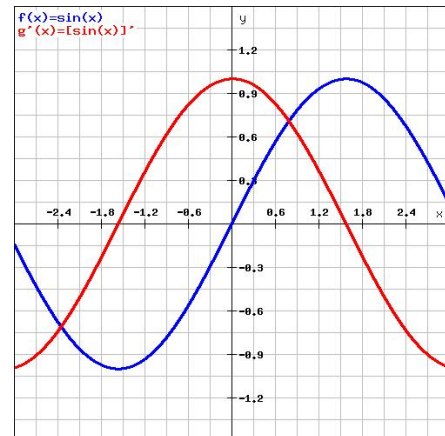
Hình 10. Chọn hàm biến đổi (Activated Functions) cho lớp ẩn và lớp ra

Hiện tại Spice-MLP dùng 27 hàm biến đổi. Công thức và đồ thị của một số hàm đổi được minh họa ở bảng 2 sau.

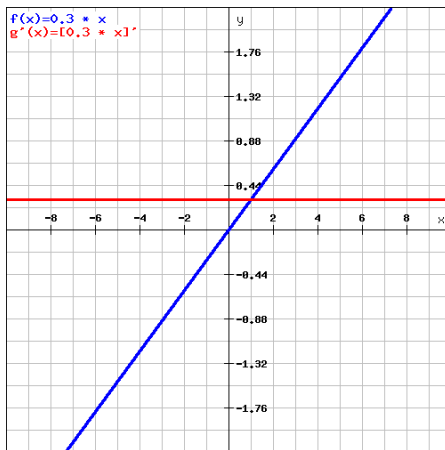




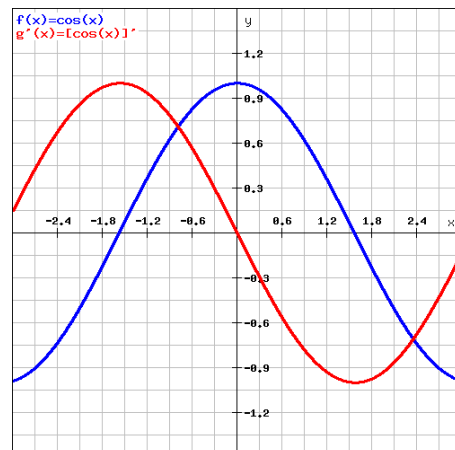
ArcSinh



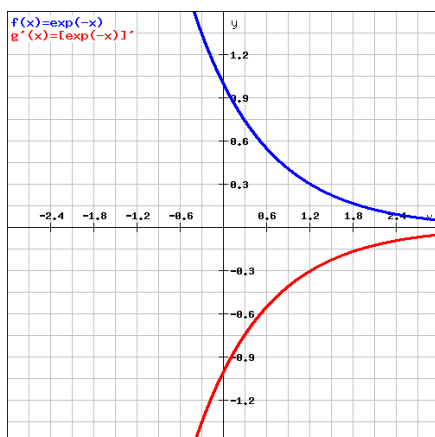
Sin



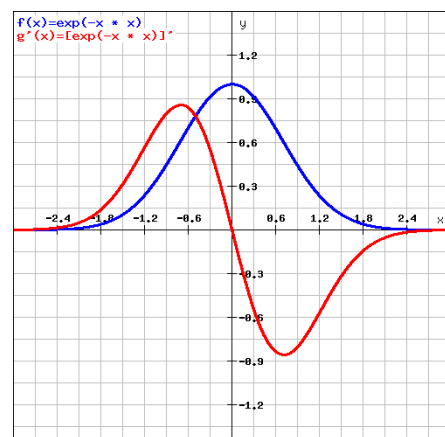
Linear



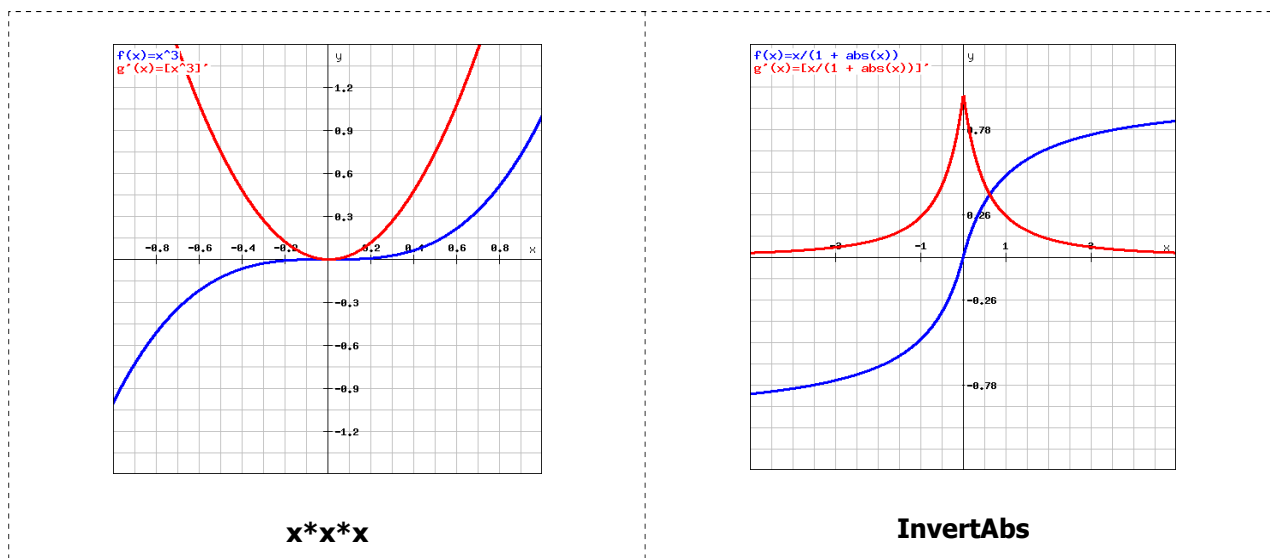
Cos



Exp(-x)



Exp(-x*x)



3.4.3. Đào tạo mạng (training)

Sau khi chọn số neuron cho lớp ẩn, chọn tỷ lệ dữ liệu học và các thông số cần thiết, bạn có thể bắt đầu đào tạo mạng. Sau đây là các nút lệnh chính để đào tạo.



- *Khởi tạo trọng số ban đầu*: khởi tạo (Reset) lại trọng số ban đầu cho các nút mạng.
- *Đào tạo*: đào tạo mạng.
- *Load trọng số mạng từ File nhị phân*: tải trọng số mạng từ file nhị phân có sẵn. Lưu ý, nếu thông số mạng từ file nhị phân có sẵn khác với thông số mạng hiện tại của bạn, có thể chương trình sẽ báo lỗi hoặc đưa ra kết quả sai.
- *Lưu trọng số mạng vào File nhị phân*: lưu trọng số mạng hiện thời vào file nhị phân, mỗi giá trị số được ghi với độ dài 4 bytes. Thứ tự như sau:

Weight_IJ: I0J0, I0J1, ...

Weight_TetaJ: Teta_J0, Teta_J1, ...

Weight_JK: J0K0, J0K1, ...

Weight_TetaK: Teta_K0, Teta_K1, ...

- *Lưu trọng số mạng vào File text*: lưu trọng số mạng hiện thời vào file text để người sử dụng có thể kiểm tra giá trị của từng nút mạng.

Ví dụ trọng số của mạng với 2 inputs, 2 hidden và 3 outputs được lưu trong file text như bảng 3.

Bảng 3. Nội dung file text chứa trọng số của mạng với 2 inputs, 2 hidden và 3 outputs

```

Weights of the NN. Activated Function was Hyperbolic Tangent
Saved on 2/4/2008 2:06:11 PM

Weight_IJ: array contain the weights in the first connections between input and hidden layers

          J0          J1
I0      2.26013784930634000000    3.92545735150867000000
I1      2.26013782334056000000    3.92545726757851000000

Weight_TetaJ: array contain the weights from Bias of neurons in hidden layers
Teta_J0  2.75109071396534000000
Teta_J1  0.85361332621648900000

Weight_JK: array contain the weights in the second connections between hidden and output layers

          K0          K1          K2
J0      -2.41085860219597000000    2.00917738577701000000    -0.40168121642706000000
J1      2.40321691280414000000    -1.07731644733275000000    1.32590046568425000000

Weight_TetaK: array contain the weights from Bias of neurons in Output layers
Teta_K0  1.85485947371567000000
Teta_K1 -1.98609819228695000000
Teta_K2 -0.13123871869928800000

```

Sau khi đào tạo mạng xong, thông tin về lần học cuối cùng của mạng được hiển thị ở box bên phải phía dưới.

Thông tin của lần học cuối cùng (4/29/2009 10:28:57 PM) Hàm biến đổi cho lớp ẩn: Hyperbolic Tanh. Hàm biến đổi cho lớp ra: Hyperbolic Tanh Tỷ lệ học cuối cùng: 0.006545 Giá trị MSE của Dữ liệu học: 0.00353937150872786 Giá trị MSE của Dữ liệu kiểm tra: 0.00369602710152686 Số lượng dữ liệu đã học: 70, (70% of 100) Số lần lặp: 5000

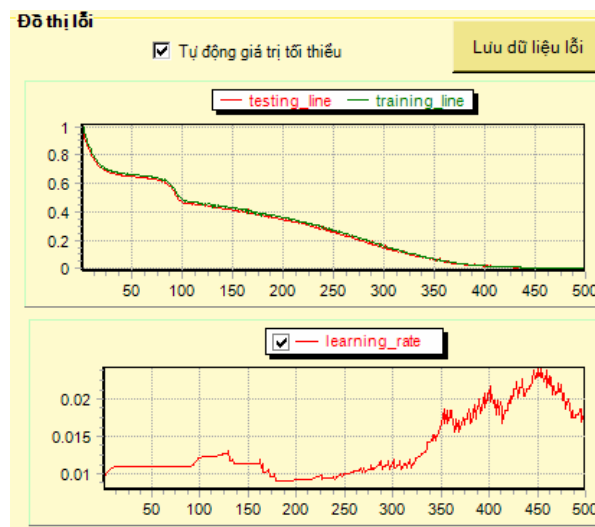
Hình 11. Thông tin lần học cuối

3.4.4. a. Kiểm tra đồ thị lỗi

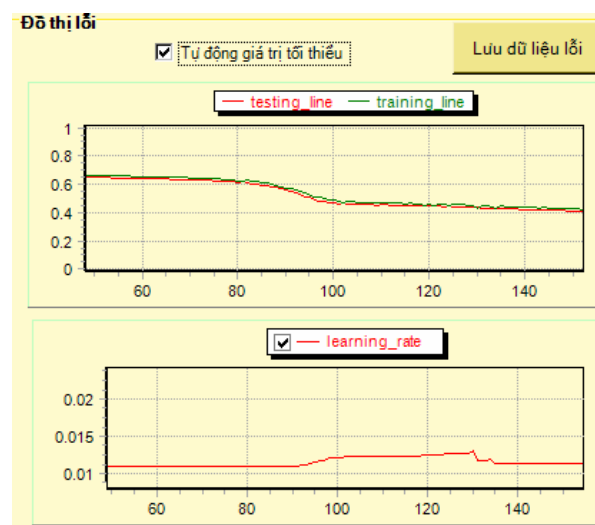
Sau khi đào tạo mạng, bạn có thể kiểm tra lỗi trong quá trình học bằng checkbox “Xem chi tiết đồ thị lỗi”. Bạn cũng có thể lưu lại dữ liệu lỗi vào file text với nút lệnh “Lưu dữ liệu lỗi”. Bảng 4 minh họa một tập dữ liệu lỗi được lưu trong file text.

Nếu bạn chọn checkbox “Xem chi tiết đồ thị lỗi” trước khi đào tạo mạng, đồ thị lỗi sẽ hiện ra online ngay khi bạn đang tiến hành đào tạo, tuy nhiên nếu dùng chức năng này thời gian đào tạo sẽ tăng lên đáng kể do máy tính phải thêm việc vẽ đồ thị cùng với đào tạo mạng.

Khi xem đồ thị bằng checkbox “Xem chi tiết đồ thị lỗi”, bạn có thể xem chi tiết cho từng lần lặp bằng cách click chuột vào đồ thị. Hình 13 minh họa đồ thị lỗi và đồ thị tỷ lệ học (learning rate) trong khoảng lần lặp thứ 50 và 150.



Hình 12. Đồ thị lỗi xem bằng checkbox “Xem chi tiết đồ thị lỗi”



Hình 13. Chi tiết đồ thị lỗi

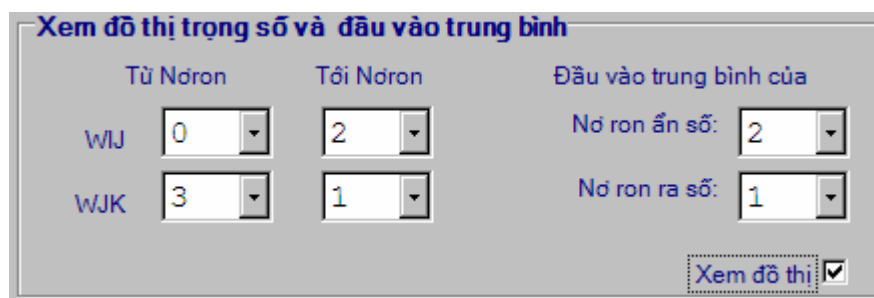
Bảng 4: dữ liệu lỗi được lưu trong file text

SPICE-MLP by Cao Thang			
Training Error			
Thông tin của lần học cuối cùng (4/30/2009 10:57:49 PM)			
Hàm biến đổi cho lớp ẩn: Hyperbolic Tanh.			
Hàm biến đổi cho lớp ra: Hyperbolic Tanh			
Tỷ lệ học cuối cùng: 0.007843			
Giá trị MSE của Dữ liệu học: 0.00323196838546706			
Giá trị MSE của Dữ liệu kiểm tra: 0.00359799661294161			
Số lượng dữ liệu đã học: 70, (70% of 100)			
Số lần lặp: 5000			
Iterations	TrainingError	TestingError	LearningRate
1	1.006932778	0.999370971	0.00918
2	0.989821872	0.990941336	0.00918
3	0.976928521	0.986117834	0.00918
4	0.969271326	0.981868921	0.00918
5	0.963313559	0.977746958	0.00918
6	0.958254105	0.973941811	0.00918
7	0.953665223	0.971387056	0.00918
8	0.949320804	0.967659064	0.00918
9	0.945233933	0.964940659	0.00918
10	0.941177193	0.961143105	0.00918
11	0.936883505	0.956635027	0.00918
...
4995	0.003240048	0.003576933	0.007842627
4996	0.003238211	0.003618928	0.007842627
4997	0.003234612	0.003602332	0.007842627
4998	0.003227556	0.003564963	0.007842627
4999	0.003221500	0.003589969	0.007842627
5000	0.003231968	0.003597997	0.007842627

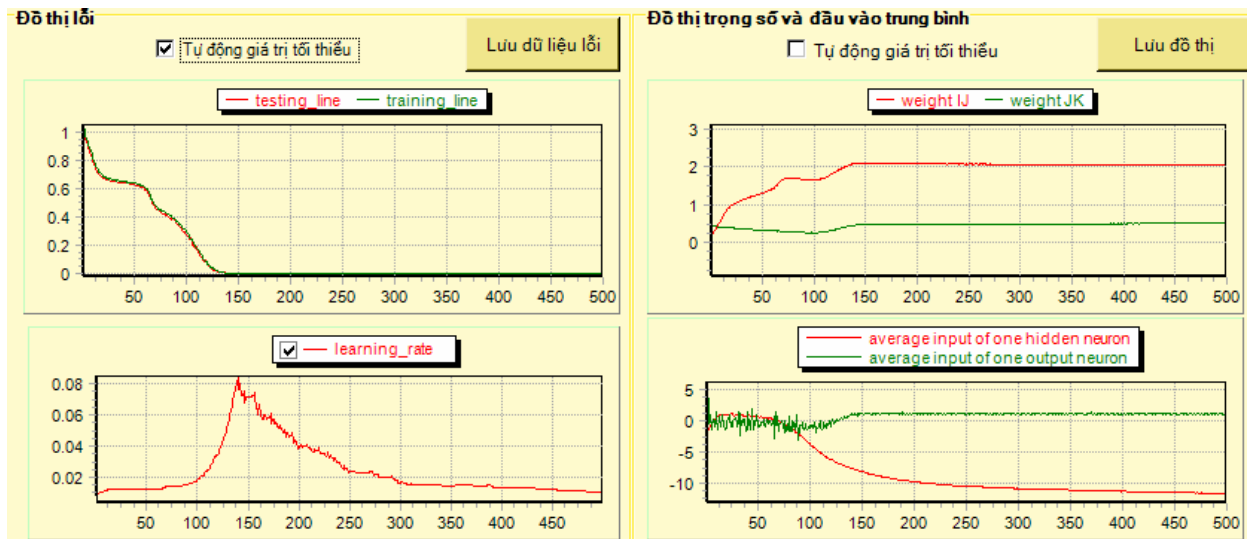
3.4.4. b. Kiểm tra đồ thị trọng số và đầu vào trung bình của một nơ ron

Spice-MLP cho phép bạn kiểm tra sự biến đổi của một vài trọng số và đầu vào trung bình của một số nơ ron trong quá trình học. Bên trái hình 14 minh họa sự lựa chọn xem trọng số nơ ron 0 của lớp vào tới nơ ron 2 của lớp ẩn WIJ[0][2], và trọng số từ nơ ron 3 của lớp ẩn tới nơ ron 1 của lớp ra WJK[3][1]. Bên phải hình 14 minh họa sự lựa chọn xem đầu vào trung bình của nơ ron ẩn số 2 và nơ ron ra số 1.

Nếu bạn chọn checkbox “Xem đồ thị” trước khi đào tạo mạng, đồ thị trọng số và đầu vào sẽ hiện ra online ngay khi bạn đang tiến hành đào tạo, tuy nhiên cũng như chọn checkbox “Xem chi tiết đồ thị lỗi”, nếu dùng chức năng này thời gian đào tạo sẽ tăng lên đáng kể do máy tính phải thêm việc vẽ đồ thị cùng với đào tạo mạng.



Hình 14. Chọn trọng số và giá trị trung bình của một số nơ ron trong khi học



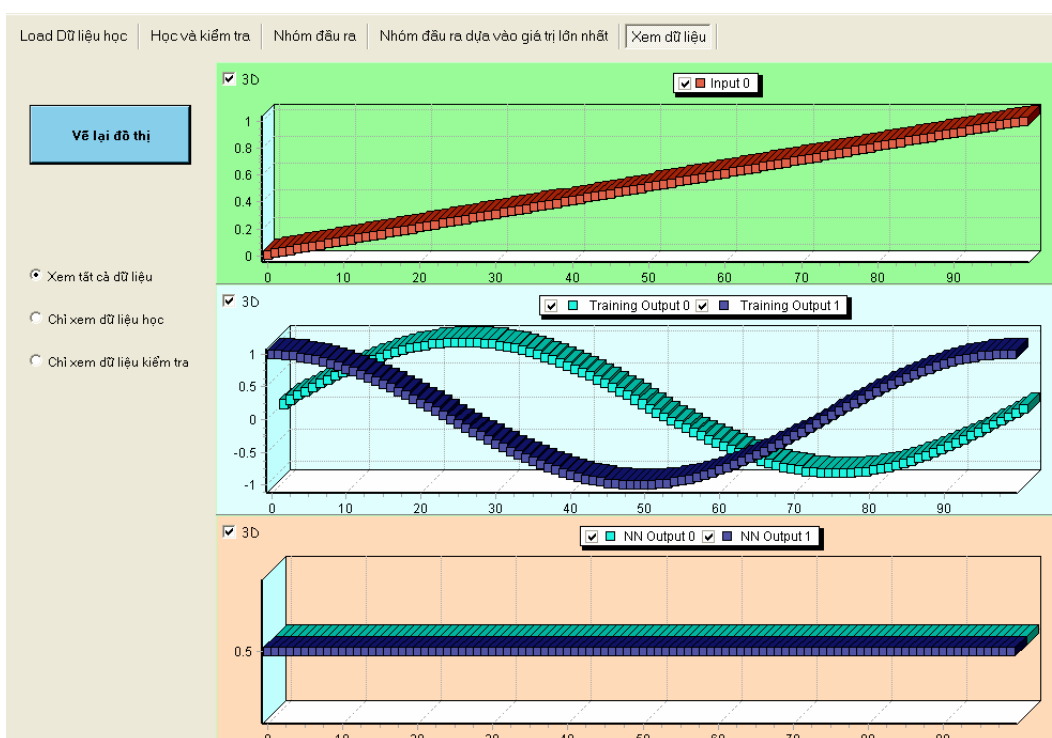
Hình 15. Đồ thị lỗi, đồ thị trọng số và giá trị trung bình đầu vào của một số nơ ron trong khi học

Hình 15 minh họa đồ thị lỗi, đồ thị của hai trọng số và giá trị trung bình đầu vào của hai nơ ron (đã chọn như trên hình 14) trong khi học. Bạn dễ dàng nhận thấy khi mạng hội tụ (khi training error giảm gần tới 0) thì trọng số và đầu vào trung bình của các nơ ron đã chọn cũng hội tụ về một giá trị nhất định.

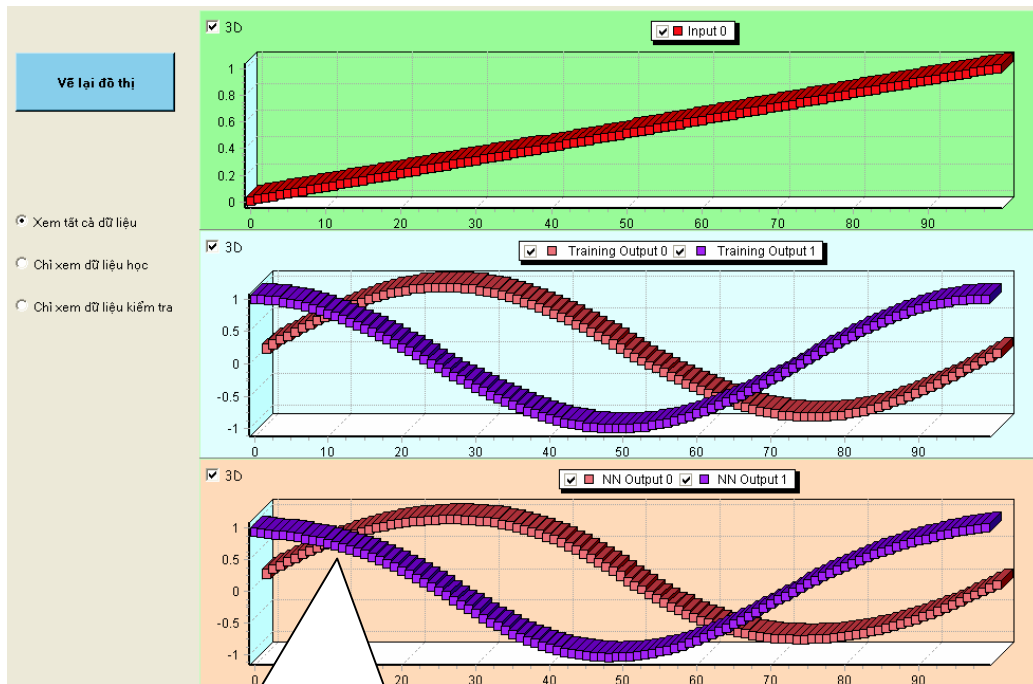
3.4.5. Xem đồ thị dữ liệu và đồ thị đầu ra của mạng

Trong phần “Load dữ liệu” đã trình bày ở trên, bạn có thể xem lại từng dataset của dữ liệu mà bạn vừa load cũng như giá trị ra của mạng. Trong Tab “Xem dữ liệu” bạn có thể xem đồ thị của toàn bộ dữ liệu nếu số lượng đầu vào và đầu ra không lớn.

Bạn có thể xem toàn bộ dữ liệu, chỉ xem dữ liệu học hoặc chỉ xem dữ liệu kiểm tra. Các hình sau minh họa toàn bộ dữ liệu đầu vào, đầu ra của hàm Sin và Cos trong khoảng $[0, 2\pi]$, đầu ra của mạng chưa được học và đầu ra của mạng đã được học. Các bạn dễ dàng nhận thấy khi chưa được học, đầu ra của mạng là xấp xỉ 0.5 cho cả hai đầu ra. Tuy nhiên sau khi học với 70% dữ liệu, đầu ra thực tế của mạng đều xấp xỉ đầu ra của dữ liệu học, nghĩa là đều có dạng hai hàm Sin và Cos.

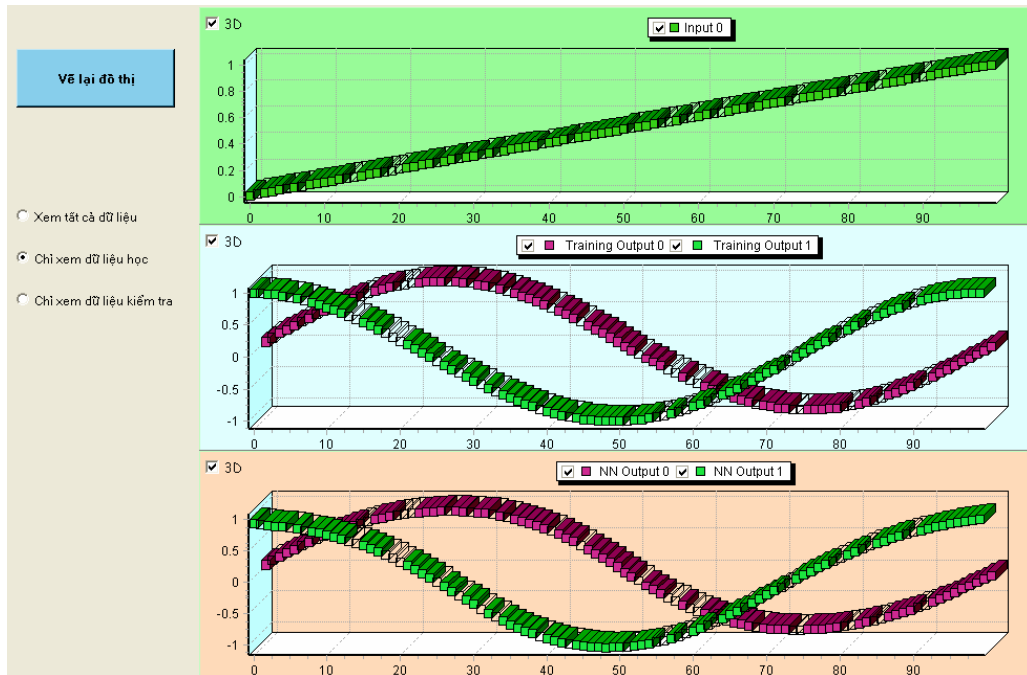


Hình 16. Toàn bộ dữ liệu đầu vào, đầu ra và đầu ra của mạng ban đầu (chưa được học)

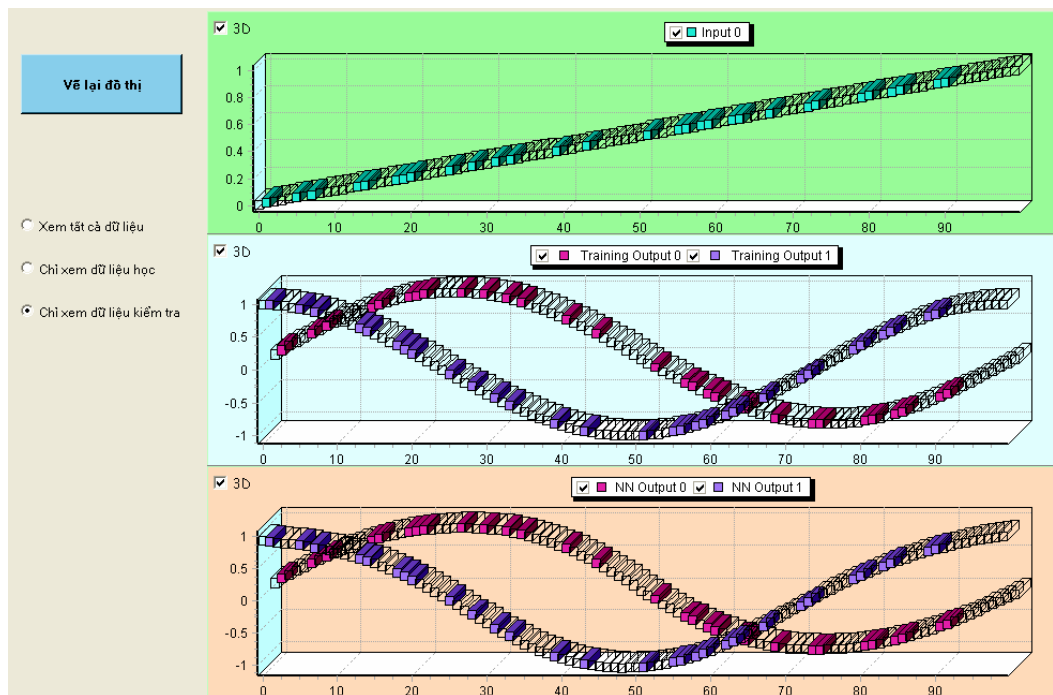


Đầu ra thực tế của mạng đã học gần giống với đầu ra của dữ liệu học

Hình 17. Toàn bộ dữ liệu đầu vào, đầu ra và đầu ra của mạng đã được học



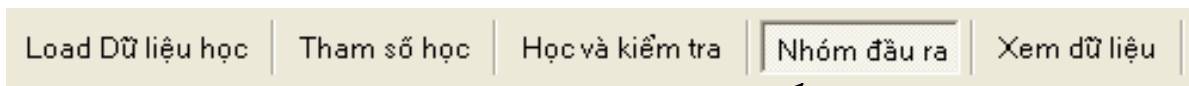
Hình 18. Chỉ xem dữ liệu học



Hình 19. Chỉ xem dữ liệu kiểm tra

3.4.6. Nhóm dữ liệu dựa trên đầu ra của mạng

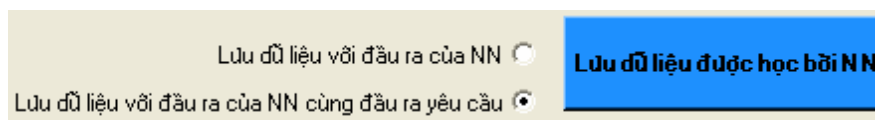
Bạn có thể nhóm dữ liệu để kiểm tra lại dữ liệu của mình. Spice-MLP cho phép bạn nhóm dữ liệu thành 3 nhóm dựa trên một đầu ra (do bạn chọn).



Nhóm đầu ra thành 3 nhóm dữ liệu dựa trên một đầu ra

3.4.7. Lưu dữ liệu đã được mô hình hóa bởi mạng NN

Sau khi đào tạo mạng, bạn có thể lưu dữ liệu đã được mô hình hóa bởi NN. Bạn có thể lưu dữ liệu với đầu ra của mạng, hoặc lưu cùng với đầu ra yêu cầu như được minh họa bởi hình 21 sau.



Hình 21. Lưu dữ liệu đã được mô hình hóa bởi mạng NN

4. Kết luận

Tài liệu này hướng dẫn bạn sử dụng phần mềm mạng nơ ron Spice-MLP. Sau khi vận dụng, các bạn có thể hiểu rõ hơn về hoạt động của mạng nơ ron đa lớp (Multi-Layer Neural Network). Cũng như Spice-SOM (phần mềm mạng nơ ron tự tổ chức), Bạn có thể dùng Spice-MLP để mô hình hóa nhiều loại dữ liệu khác nhau và có thể áp dụng trong các bài toán thực tế như nhận dạng, nhóm, clustering, hỗ trợ quyết định Tác giả hy vọng Spice-MLP giúp ích trong việc nghiên cứu và học tập của các bạn.

Sau khi đọc tài liệu này, các bạn có thể tìm đọc tài liệu [neural_network_practical_use_vi.pdf](#) trong đó có minh họa phân loại ảnh khuôn mặt, ảnh người đi bộ, ảnh xe hơi, dự báo chứng khoán và một số ví dụ khác.

Xin cảm ơn các bạn đã đọc tài liệu này và sử dụng Spice-MLP. Nếu các bạn cần thêm các chức năng của Spice-MLP, xin hãy liên lạc với tác giả tại <http://spiceneuro.wordpress.com> hoặc spiceneuro AT gmail DOT com Tác giả rất mong các ý kiến đóng góp của các bạn.