Journal of Engineering Sciences and Information Technology (JESIT) • Vol 8, Issue 3 (2024) • P: 54 - 46
https://journals.ajsrp.com/index.php/jesit
ISSN: 2522-3321(Online) • ISSN: 2522-3321(Print)

Network Classifier on Iris data Using MATLAB

Asst-Lecturer \ Jenan Jader Msaad*[1], Lecturer \ Ruwaidah F. Obaied[1], Asst-Lecturer \ Alaa Majeed Shnin[1]
[1] Al-Furat Al-Awsat Technical University | Iraq

Received: 22/07/2024
Revised: 03/08/2024
Accepted: 26/08/2024
Published: 30/09/2024

* Corresponding author: jenan.jader@atu.edu.iq

Citation: Msaad, J. J., Obaied, R. F. & Shnin, A. M. (2024). Network Classifier on Iris data Using MATLAB. *Journal of engineering sciences and information technology, 8*(3), 46 – 54. https://doi.org/10.26389/AJSRP.N220724

Abstract: Classification is the task of assigning object to one of several predefined categories. Classification analysis is the organization of data in given classes. Also known as supervised classification, the classification uses given class labels to order the objects in the data collection. Classification approaches normally use a training set to train the model where all objects are already associated with known class labels. The classification algorithm learns from the training set and builds a model to classify the target,. The classification analysis would generate a model that could be used to find a class (target).

A neural network consists of patterns represented in terms of numerical values attached to the nodes of the graph and transformations between patterns achieved via simple message-passing algorithms. Certain of the nodes in the graph are generally distinguished as being input nodes or output nodes, and the graph as a whole can be viewed as a representation of a multivariate function linking inputs to outputs. Numerical values (weights) are attached to the links of the graph, parameterizing the input/output function and allowing it to be adjusted via a learning algorithm.

Artificial neural networks (ANNs) consider the good tool to learn and classify patterns such as the biological human brain learning process. It consists of simple elements called neurons, which are operating in parallel (included many neuron units that work in parallel). Connections between neurons have weights emerged with the inputs of the neural to give the certain output. The connection's weights have been adjusted during the learning process by iteratively comparing the output of the network and the required target. The ability of the network to show a good performance in the results depends on the training algorithm. There are different types of neural networks, however, most of researches published in the medical studies used one class of neural networks, the back-propagation (BP).

Keywords: Classification, simple message-passing algorithms, neural network, back-propagation, Artificial neural networks.

مصنف الشبكة على بيانات القزحية باستخدام ماتلاب

المدرس المساعد / جنان جادر مساعد*[1]، المدرس / رويدة فاضل عبيد[1]، المدرس المساعد / الاء مجيد شنين[1]
[1] جامعة الفرات الأوسط التقنية | العراق

المستخلص: التصنيف هو مهمة تعيين كائن إلى واحدة من عدة فئات محددة مسبقًا. تحليل التصنيف هو تنظيم البيانات في فصول معينة. المعروف أيضًا باسم التصنيف الخاضع للإشراف، يستخدم التصنيف ملصقات فئة معينة لطلب الكائنات في جمع البيانات. عادةً ما تستخدم أساليب التصنيف مجموعة تدريب لتدريب النموذج حيث ترتبط جميع الكائنات بالفعل بملصقات فئة معروفة. تتكون خوارزمية التصنيف من مجموعة التدريب وتبني نموذجًا لتصنيف الهدف. سيؤدي تحليل التصنيف إلى إنشاء نموذج جديد يمكن استخدامه للعثور على فئة (هدف).

تتكون الشبكة العصبية من أنماط ممثلة من حيث القيم العددية المرتبطة بعقد الرسم البياني والتحولات بين الأنماط التي تحققت من خلال خوارزميات تمرير الرسائل البسيطة. يتم تمييز بعض العقد في الرسم البياني عمومًا على أنها عقد إدخال أو عقد إخراج، ويمكن النظر إلى الرسم البياني ككل على أنه تمثيل لدالة متعددة المتغيرات التي تربط المدخلات بالمخرجات. يتم إرفاق القيم العددية (الأوزان) بروابط الرسم البياني، وتوضيح وظيفة الإدخال/الإخراج والسماح بتعديلها عبر خوارزمية التعلم.

تعتبر الشبكات العصبية الاصطناعية (ANNS) اداة جيدة جدا لتعلم وتصنيف أنماط مثل عملية تعلم الدماغ البشرية البيولوجية. وهو يتألف من عناصر بسيطة تسمى الخلايا العصبية، والتي تعمل بالتوازي (شملت العديد من وحدات الخلايا العصبية التي تعمل بالتوازي). تدمج الروابط بين الخلايا العصبية الأوزان مع مدخلات العصبية لإعطاء ناتج معين. تم تعديل أوزان الاتصال أثناء عملية التعلم من خلال مقارنة إخراج الشبكة بشكل متكرر والهدف المطلوب. تعتمد قدرة الشبكة على إظهار أداء جيد في النتائج على خوارزمية التدريب. هناك أنواع مختلفة من الشبكات العصبية، ومع ذلك، فإن معظم الأبحاث المنشورة في الدراسات الطبية استخدمت فئة واحدة من الشبكات العصبية، وهي شبكة الانتشار الخلفي (BP).

الكلمات المفتاحية: التصنيف، خوارزميات تمرير الرسائل البسيطة، الشبكة العصبية، الانتشار الخلفي، الشبكات العصبية الاصطناعية.

Introduction:

Classification and species recognition based on individual characteristics and behaviors constitutes a preliminary measure and is an important goal in the behavioral sciences. Current statistical methods do not always yield satisfactory answers. A forward artificial neural network is a computer model inspired by the structure of the human brain. Seen as a group of artificial neurons interconnected with other neurons. The primary goal of this paper is to demonstrate the development process of an artificial neural network-based classifier that classifies the IRIS database.

The problem concerns the identification of iris plant

species on the basis of botanical trait measurements. This paper relates the use of feed forward neural networks towards identifying IRIS plants based on the following measurements: sepal length, sepal width, petal length, and petal width. With this dataset, a neural network (NN) is used to classify the IRIS dataset. EBPA is used to train these Anne. Simulation results demonstrate the effectiveness of the neural system in determining the iris class.

1. The search problem:

in this paper, the classification problem is solved by 3 types of neuron Networks:

1- multilayer perceptron;
2- back-propagation (BP).
3- probabilistic neural network.

These types of networks were soon described in this paper. Both of Such networks have adjustable parameters that affectPerformance.

2. Research hypotheses

The IRIS dataset consist of 4 numeric, predictive attributes and the class features as shown in table (1).

2.1. Attribute Information:

- sepal length in cm
- sepal width in cm
- petal length in cm
- petal width in cm
- class:
  - Iris Setosa
  - Iris Versicolour
  - Iris Virginica

2.2. Missing Attribute Values: None

2.3. Summary Statistics:

Min Max Mean SD Class Correlation
sepal length: 4.3 7.9 5.84 0.83 0.7826
sepal width: 2.0 4.4 3.05 0.43 -0.4194
petal length: 1.0 6.9 3.76 1.76 0.9490 (high!)
petal width: 0.1 2.5 1.20 0.76 0.9565 (high!)

2.4. Class Distribution: 33.3% for each of 3 classes.

3. Importance of research:

To help get a feel for how a simple neural network would work, lets use the following example: we want to classify which of three species of iris (flower) we are looking at. In order to distinguish between them we will use four measurements.

Sepal Length

Sepal Width

Petal Length

Petal Width

In machine learning, these are what we call features. Features can be extracted in a number of ways, for example in the case of classifying positive or negative product reviews, we might decide to count the number of specific tone-related words in a sentence (i.e excellent, terrible, shocking).

In this example, we are making the assumption that the measurements are accurate and that we will receive them in numerical form.

Often these features are much more complicated and can also extracted by the model itself, for example in image classification. I will talk about image classification with CNNs and RNNs in a different post.

The goal of the model is to be able to accurately classify from these features the species of iris. There are three species we will train this model to identify: Iris Setosa, Iris Versicolor, and Iris Virginica.

4. Search Methodology:

The advantage of neural networks lies in the following theoretical aspects. First, neural networks are data driven self-adaptive methods in that they can adjust themselves to the data without any explicit specification of functional or distributional form for the underlying model. Second, they are universal functional approximates in that neural networks can approximate any function with arbitrary accuracy

The Neural Network (NN) considers one of the data mining and machine learning algorithms. There are many types of neural networks such as SOM, Backpropagation and Perception. In this assignment, evaluation of IRIS dataset from UCI machine Learning Repository will be conducted using the NN. The evaluation steps as follow:.

1. The data set is download from the URL https://archive.ics.uci.edu/ml/datasets/iris
2. The main description of this dataset is in figure (1)

Data Set Characteristics:	Multivariate	**Number of Instances:**	150	**Area:**	Life
Attribute Characteristics:	Real	**Number of Attributes:**	4	**Date Donated**	1988-07-01
Associated Tasks:	Classification	**Missing Values?**	No	**Number of Web Hits:**	4277311

Figure (1): The main description of IRIS dataset

5. Results

MATLAB is a high programming language to evaluate the NN on IRIS dataset, we will use the Self Organization Map (SOM) to classify iris flowers into classes topologically, providing insight into the types of flowers and a useful tool for further analysis. Self-organizing maps (SOMs) are very good at creating classifications. Further, the classifications retain topological information about which classes are most similar to others. Self-organizing maps can be created with any desired level of detail. They are particularly well suited for clustering data in many dimensions and with complexly shaped and connected feature spaces. They are well suited to cluster iris flowers. The four flower attributes will act as inputs to the SOM, which will map them onto a 2-dimensional layer of neurons. The following steps are essential in Matlab.

5.1. Save the data into matrix X, figure (2) shows the snapshot of reading data in x using Matlab instruction x = iris_dataset;

```
>> x = iris_dataset

x =

  Columns 1 through 11

    5.1000    4.9000    4.7000    4.6000    5.0000    5.4000    4.6000    5.0000    4.4000    4.9000    5.4000
    3.5000    3.0000    3.2000    3.1000    3.6000    3.9000    3.4000    3.4000    2.9000    3.1000    3.7000
    1.4000    1.4000    1.3000    1.5000    1.4000    1.7000    1.4000    1.5000    1.4000    1.5000    1.5000
    0.2000    0.2000    0.2000    0.2000    0.2000    0.4000    0.3000    0.2000    0.2000    0.1000    0.2000

  Columns 12 through 22

    4.8000    4.8000    4.3000    5.8000    5.7000    5.4000    5.1000    5.7000    5.1000    5.4000    5.1000
    3.4000    3.0000    3.0000    4.0000    4.4000    3.9000    3.5000    3.8000    3.8000    3.4000    3.7000
    1.6000    1.4000    1.1000    1.2000    1.5000    1.3000    1.4000    1.7000    1.5000    1.7000    1.5000
    0.2000    0.1000    0.1000    0.2000    0.4000    0.4000    0.3000    0.3000    0.3000    0.2000    0.4000

  Columns 23 through 33

    4.6000    5.1000    4.8000    5.0000    5.0000    5.2000    5.2000    4.7000    4.8000    5.4000    5.2000
    3.6000    3.3000    3.4000    3.0000    3.4000    3.5000    3.4000    3.2000    3.1000    3.4000    4.1000
    1.0000    1.7000    1.9000    1.6000    1.6000    1.5000    1.4000    1.6000    1.6000    1.5000    1.5000
    0.2000    0.5000    0.2000    0.2000    0.4000    0.2000    0.2000    0.2000    0.2000    0.4000    0.1000
```

Figure (2) : Reading IRIS data in Matlab

5.2. Size of Iris data

> size(x)

ans = 1×2

4 150

The next step is to create a neural network that will learn to cluster. **selforgmap** creates self-organizing

maps for classifying samples with as much detail as desired by selecting the number of neurons in each dimension of the layer. We will try a 2-dimension layer of 64 neurons arranged in an 8x8 hexagonal grid for this example. In general, greater detail is achieved with more neurons, and more dimensions allows for modelling the topology of more complex feature spaces. The input size is 0 because the network has not yet been configured to match our input data. This will happen when the network is trained as shown in figure (3)

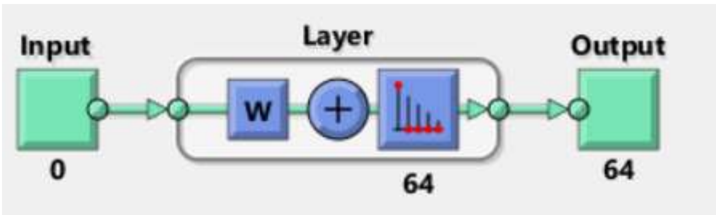

Figure (3): SOM net

Now the network is ready to be optimized with **train**. The Neural Network Training Tool shows the network being trained and the algorithms used to train it. It also displays the training state during training and the criteria which stopped training will be highlighted in green.The buttons at the bottom open useful plots which can be opened during and after training. Links next to the algorithm names and plot buttons open documentation on those subjects as shown in figure (4)

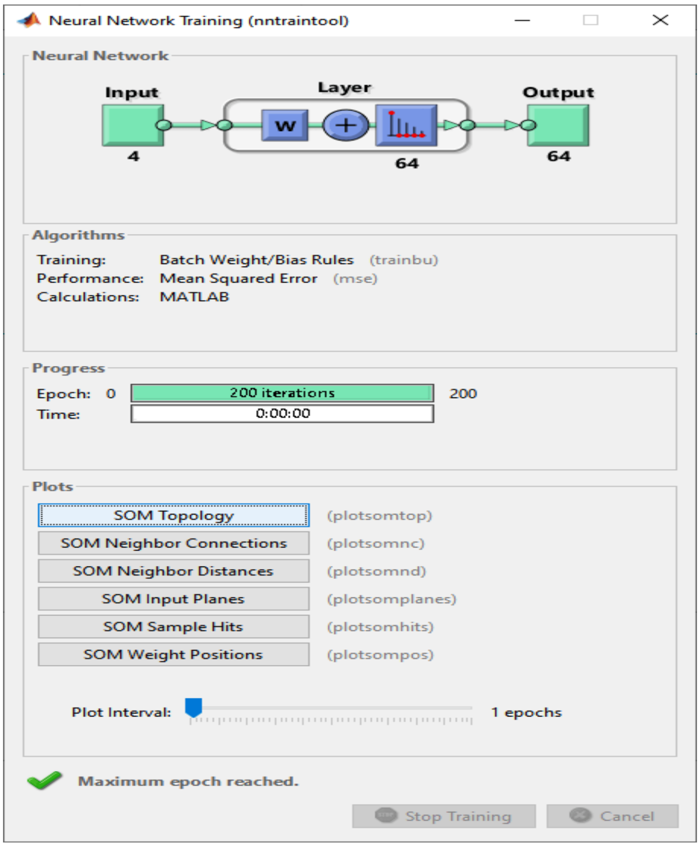

Figure (4): Training Net using SOM neural Network

plotsomtop plots the self-organizing maps topology of 64 neurons positioned in an 8x8 hexagonal grid. Each neuron has learned to represent a different class of flower, with adjacent neurons typically representing similar classes as shown in figure (5).

>plotsomtop(net)

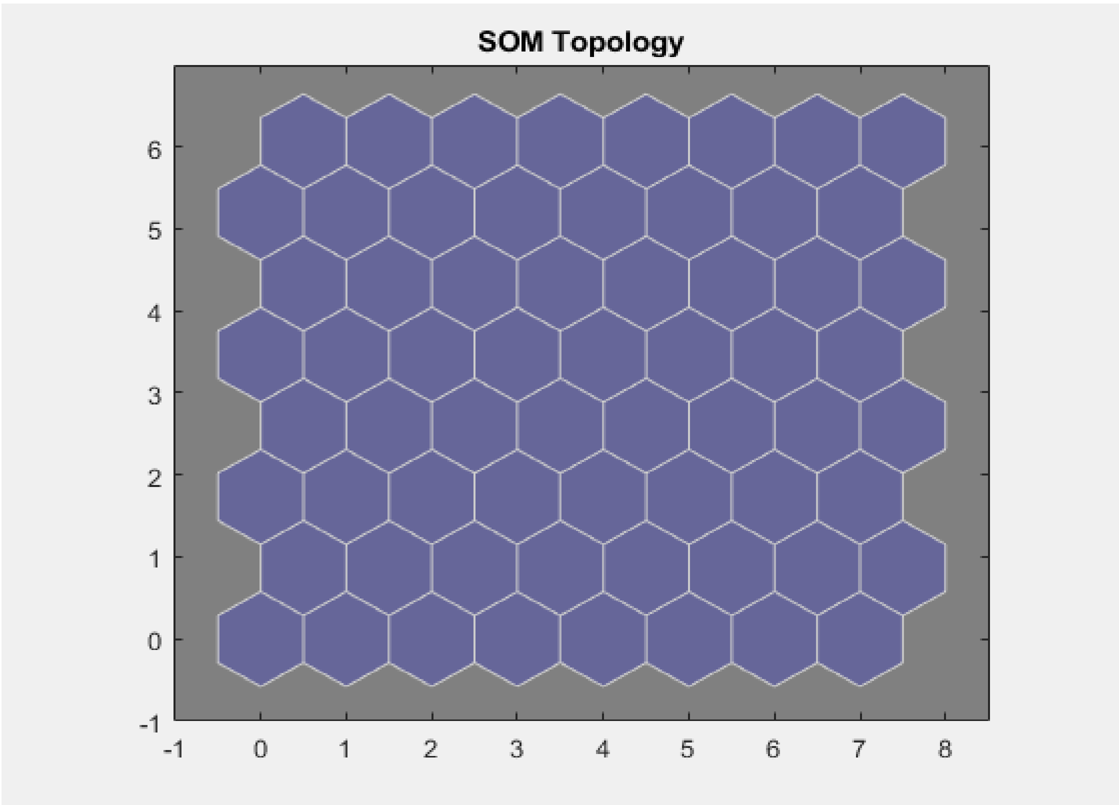

Figure (5): SOM-64 neurons positioned in an 8x8 hexagonal grid

Figure (6) shows SOM neighbor connections.

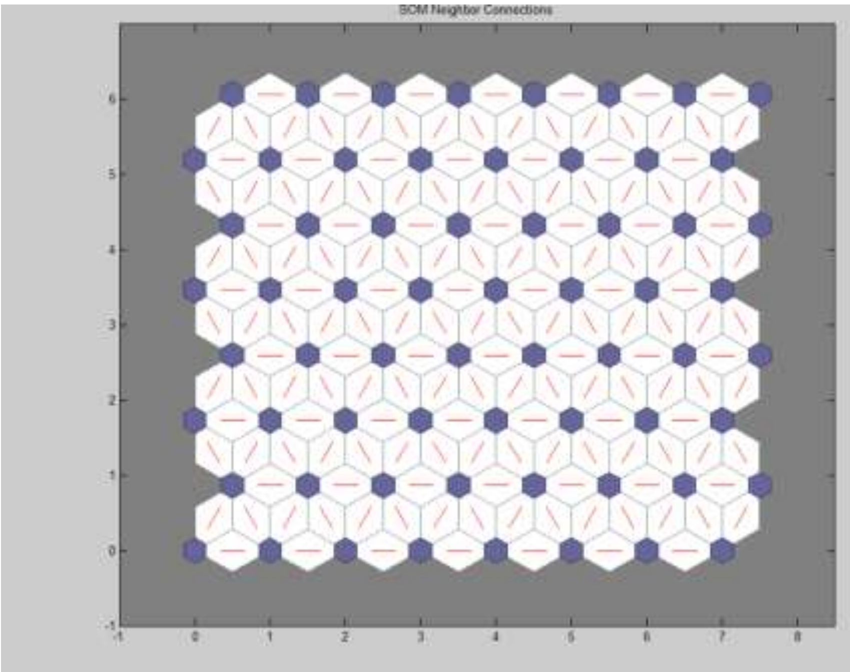

Figure (6): SOM neighbor connections.

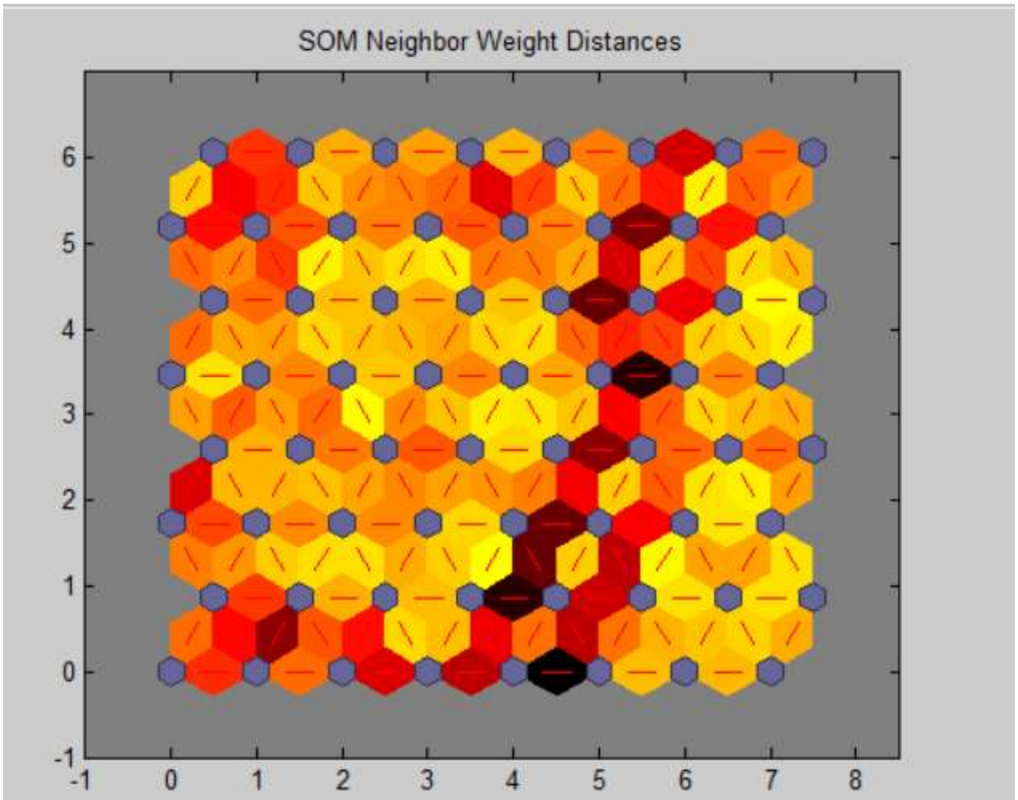

Figure (7): SOM neighbor Weight Distance

Figure (8): SOM Input Planes

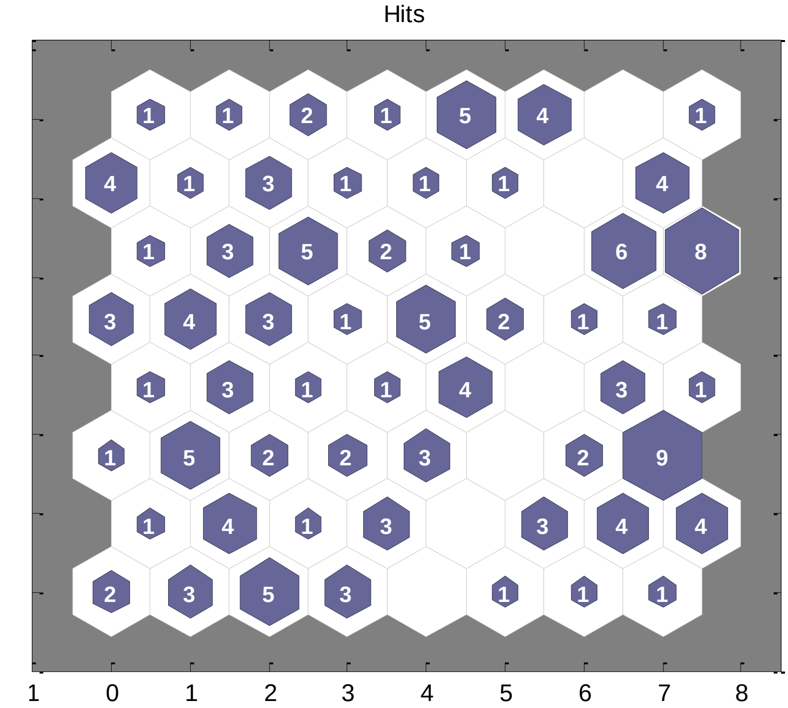

Figure (9): SOM Sample Hits

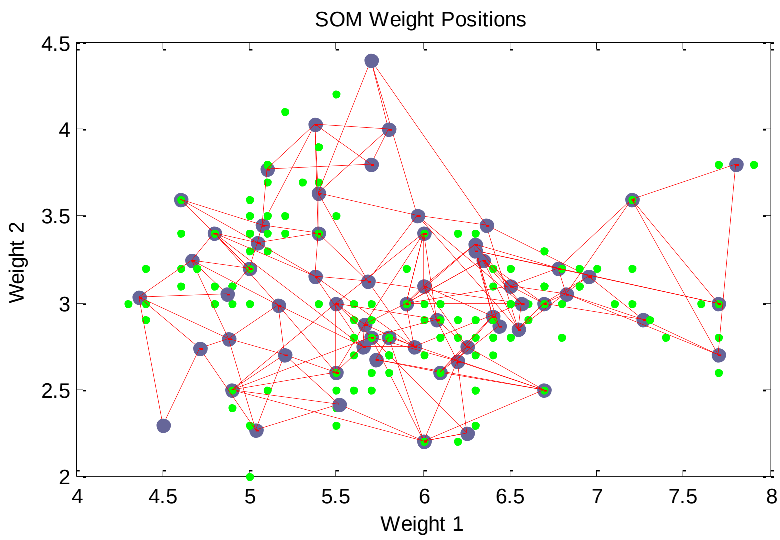

Figure (10): SOM weight Positions

6. Discussion of results

The Iris dataset is one of the most well-known and commonly used datasets in the field of machine learning and statistics. In this article, we will explore the Iris dataset in deep and learn about its uses and applications.

The Iris dataset consists of 150 samples of iris flowers from three different species: Setosa, Versicolor, and Virginica. Each sample includes four features: sepal length, sepal width, petal length, and petal width. It was introduced by the British biologist and statistician Ronald Fisher in 1936 as an example of discriminant analysis.

The Iris dataset is often used as a beginner's dataset to understand classification and clustering algorithms in machine learning. By using the features of the iris flowers, researchers and data scientists can classify each sample into one of the three species.

- Attribute Information:
  - sepal length in cm

- sepal width in cm
- petal length in cm
- petal width in cm
- class:
- Iris Setosa
- Iris Versicolour
- Iris Virginica

3.2. Missing Attribute Values: None

- Summary Statistics:

  Min Max Mean SD Class Correlation

  sepal length: 4.3 7.9 5.84 0.83 0.7826

  sepal width: 2.0 4.4 3.05 0.43 -0.4194

  petal length: 1.0 6.9 3.76 1.76 0.9490 (high!)

  petal width: 0.1 2.5 1.20 0.76 0.9565 (high!)
- Class Distribution: 33.3% for each of 3 classes

The Neural Network Training Tool shows the network being trained and the algorithms used to train it. It also displays the training state during training and the criteria which stopped training will be highlighted in green.The buttons at the bottom open useful plots which can be opened during and after training. Links next to the algorithm names and plot buttons open documentation on those subjects as shown in figure (4) plotsomtop plots the self-organizing maps topology of 64 neurons positioned in an 8x8 hexagonal grid. Each neuron has learned to represent a different class of flower, with adjacent neurons typically representing similar classes as shown in figure (5). In Figure (6) shows SOM neighbor connections.

In Figure 7, we show the distance in the weight of the neighbor SOM in detail, as well as the insertions of Planes , as in Figure 8

7. Conclusin

In this paper we shows how to create and compare neural network classifiers in the Classification, and export trained models to the workspace to make predictions for new data.

Neural network models are structured as a series of layers that reflect the way the brain processes information. The neural network classifiers available in Statistics and Machine Learning Toolbox™ are fully connected, feedforward neural networks for which you can adjust the size of the fully connected layers and change the activation functions of the layers.

Choose among various algorithms to train and validate classification models for binary or multiclass problems. After training multiple models, compare their validation errors side-by-side, and then choose the best model. To help you decide which algorithm to use, see Train Classification Models in Classification Learner App.

This flow chart shows a common workflow for training classification models, or classifiers, in the Classification Learner app.

Workflow in the Classification Learner app. Step 1: Select data and validation. Step 2: Choose classifier options. Step 3: Train a classifier. Step 4: Assess classifier performance. Step 5: Export the classifier.

If you want to run experiments using one of the models you trained in Classification Learner, you can export the model to the Experiment Manager app.

8. Recommendations

We have now successfully built and trained a deep learning model to classify Iris types from their measurements. The model achieved 96% accuracy in the test set, which is very good for a simple model.

I hope this tutorial has given you a better understanding of what goes into building and training a neural network and how to do it.

The model can be made deeper, by adding more layers and adding more data . You can also play with super parameters to see how that changes the results. Perhaps he tried to graph the loss and accuracy while training the model. Another good thing to learn / use is an early stop

References:

[1] Generative benchmark models for mesoscale structure in multilayer networks,2019, M. Bazzi, L. G. S. Jeub, A. Arenas, S. D. Howison, M. A. Porter. arXiv:1608.06196

[2] Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer Series in Statistics. New York, NY: Springer, 2009. https://doi.org/10.1007/978-0-387-84858-7

[3] Classification algorithms on iris datasets, 2020,In book: Exploiting Machine Learning Algorithms (pp.1-9)Publisher: Liaqat Ali & Co

[4] http://docs.roguewave.com/imsl/java/7.0/manual/api/com/imsl/datamining/NaiveBayesClassifierEx3.html , ECE438: Fall 2019, Prof. Boutin https://www.projectrhea.org/rhea/index.php/Naive_Bayes_OldKiwi
https://lnwatson.co.uk/posts/intro_to_nn/

[5] Lewis Watson included in Machine Learning 2023 ,

[6] MATLAB ,2024 ,https://www.mathworks.com/help/stats/classification-learner-app.html?s_tid=CRUX_lftnav

[7] MATLAB ,2024 , https://www.mathworks.com/help/stats/kmeans.html

[8] MATLAB ,2024 , https://www.mathworks.com/help/stats/classification-nearest-neighbors.html