Neural Networks and Fuzzy Systems (2017)

Coursework 1

Laraib Mahboob

14031206

Abstract

Identification of breast cancer is an open study area. This problem can be solved by “autonomous systems”. One of the approaches can be artificial neural network as this is one of the intelligent tools designed for such systems. This report is about building a neural network to diagnose breast cancer whether it is benign or malignant.

Contents

1. **Introduction:** 3
2. **Background:** 3
3. **Preprocessing:** 3
4. **Network Architecture:** 3
5. **Experimental results and analysis:** 3
   - 5.1. **Effect of changing no of neurons in output layer on accuracy** 4
     - 5.1.1. **Hypothesis** 4
     - 5.1.2. **Results** 4
   - 5.2. **Effect of data distribution on accuracy** 4
     - 5.2.1. **Hypothesis** 4
     - 5.2.2. **Results** 4
   - 5.3. **Effect of changing number of neurons and data distribution on accuracy** 4
     - 5.3.1. **Hypothesis** 4
     - 5.3.2. **Results** 5
   - 5.4. **Effect of changing learning rate on accuracy** 5
     - 5.4.1. **Hypothesis** 5
     - 5.4.2. **Results** 5
6. **References:** 5

1. Introduction:

Breast cancer is very common in women worldwide. In 2017, an estimated 252,710 new cases of invasive breast cancer are expected to be diagnosed in women in the U.S., along with 63,410 new cases of non-invasive (in situ) breast cancer.(*U.S. Breast Cancer Statistics | Breastcancer.org*, no date). Thus many research is being done and is going in this field and neural networks are built to diagnose the cancer. As it is much faster in diagnosing and generating results that is why used most extensively. This report is about building a neural network which takes symptoms as input and responds in type of the breast cancer.

2. Background:

Basic unit of neural network is neuron. An artificial neuron is a device with several inputs and one output that models certain properties of biological neurons. The inputs have a specific weight assigned to them and these inputs are passed to a transfer function which generates an output. An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information.(Gupta, 2013). An artificial neural network is trained on some data. Then given some inputs and shows results according to the trained data.

Breast cancer diagnosis has been approached by various machine learning techniques for many years.(Xin Yao, 1999). One of the approach is based on evolutionary artificial neural networks. In this approach, a feed forward neural network is evolved using an evolutionary programming algorithm.(Xin Yao, 1999).

3. Preprocessing:

The data used to train and test the network is Wisconsin Diagnostic Breast Cancer (WBDC) dataset. This dataset contain a total number of 699 cases. There are 11 columns in which first column is of the ID of the patient followed by nine columns for symptoms, different symptoms are denoted by numbers. The last column is for whether the cancer is benign or malignant. The last column has only two values i.e. 2 for benign and 4 for malignant. There are few columns in this dataset where there is no value but is a '?'. I replaced this '?' with '-1' in order to read the data from the file.

4. Network Architecture:

This network is made through newff function. There is one hidden layer and number of neurons are 20. The neurons in output layer vary for different experiments. Goal is set to 0.01 and the function is set to run for 100 epochs. For training the data function "train" is used and for testing "net" function is used. As neural network shows result not in 2 or 4. A threshold is defined to get results exactly in 2 and 4.

5. Experimental results and analysis:

During experiments changes are made manually. For example for data distribution training and testing data is separated manually. For each experiment weights are selected randomly by neural network.

5.1. Effect of changing no of neurons in output layer on accuracy

5.1.1. Hypothesis

If we change number of neurons in output layer and testing data it affects the accuracy.

Number of Neurons in Output Layer	Accuracy
1	100
2	96.5665

5.1.2. Results

Above table shows that accuracy is affected by changing the number of neurons in output layer. If output neuron is one and the neural network is trained on all the data and then tested on the same data it showed 100% accuracy. But when network is trained on same data with two neurons in output layer it affected accuracy.

5.2. Effect of data distribution on accuracy

5.2.1. Hypothesis

Accuracy decreases as we decrease the training data and increases the testing data.

Training Data	Testing Data	Accuracy
90%	10%	98.5507
80%	20%	98.4612
70%	30%	99.5215
60%	40%	97.8495
50%	50%	97.7077

5.2.2. Results

We see from the above table that our hypothesis was correct as the accuracy does decrease as we increase the testing data and decrease the training data. One possible reason could be that as the training data is less the network cannot predict all the possibilities in testing data. But there is some irregularity in the accuracy because the weights are select by neural network.

5.3. Effect of changing number of neurons and data distribution on accuracy

5.3.1. Hypothesis

There should be very small difference in accuracy of similar data distribution when number of neurons in output layer change.

No of Neurons in Output Layer	Training Data	Testing data	Accuracy

1	90%	10%	98.5507
2	90%	10%	100
1	80%	20%	98.4612
2	80%	20%	100

5.3.2. Results

The difference in results is because of number of neurons in output layer. When same data distribution is applied on different neurons in output layer, it showed different results not only because there are different number of neurons in output layer but also because of the weights.

5.4. Effect of changing learning rate on accuracy

5.4.1. Hypothesis

Learning rate should be inversely proportional to accuracy i.e. when learning rate is decreased accurace increased and vice versa.

Learning Rate	Accuracy
0.05	100
0.04	100
0.03	100

5.4.2. Results

There is no change in the accuracy by changing learning rate. The reason could be weights i.e. weights are selected randomly that's why there is no change in accuracy after changing learning rate.

6. References:

Gupta, N. (2013) 'Artificial Neural Network', *Network and Complex Systems*, 3(1), pp. 24–28.

U.S. Breast Cancer Statistics | Breastcancer.org (no date). Available at: http://www.breastcancer.org/symptoms/understand_bc/statistics (Accessed: 3 December 2017).

Xin Yao, Y. L. (1999) 'Neural Networks for Breast Cancer Diagnosis', *In Proceedings of the 1999 congress on evolutionary computation*, 3, pp. 1760–1767. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.45.3319.