Binary Classification by an Artificial Neural Network

Kumar Sambhav Gupta

Research School of Computer Science, Australian National University

Abstract. In this document we must classify the ionospheric dataset as suitable for further analysis or not. The dataset consisting of 351 instances and 35 attributes are trained on a single layer network then compared with a multilayered network. The final attribute in the dataset is a classification of "good" or "bad" radar return values. Finally, further in the paper a method called as Bimodal distribution removal is attempted to clean the noisy training sets. Furthermore, we dive into trying a convolutional neural network on the dataset and realise the implementation is much better served with an MNIST dataset.

Keywords: Artificial Neural Network (ANN), Binary Classification, Machine Learning, ionosphere, signal -processing, bimodal distribution removal.

1 Introduction

As we have observed in past, neural network have many potential applications in signal processing. Malkoff[1] explains very well in neural network for real-time signal processing how dealing with transient signal with low signal-to-noise ratios (SNR). Golovko, Savitsky and Maniakov[2] also presents a method in signal processing in one of the most interesting and innovative areas of the chaotic time processing in chapter 6, neural networks for signal processing in measurement analysis. Boone, Sigillito and Shaber[3] demonstrate how neural networks can perform as well as trained human experts in detecting certain nodules in radiological signals. We can find many more examples of signal-processing tasks with help of neural networks in July 1988 issue of the IEEE *Transactions on Acoustic, Speech and Signal Processing*[4].

Some methods used in the neural network were inspired by the study materials and research papers provided to the student. Initially, in the method section of the report we consider using a basic implementation of neural network as explained and discussed in the labs of Australian National University. Furthermore, we inspect network reduction techniques and discuss bimodal distribution removal to clean the noisy training set.

2 Method

2.1 Dataset

Dataset used in the report is the ionospheric reading from the UCI machine learning database. The radar data was collected by a system in Goose Bay, Labrador. This system consists of a phased array of 16 high-frequency antennas with a total transmitted power on the order of 6.4 kilowatts. The targets were free electrons in the ionosphere. "Good" radar return is those showing evidence of some type of structure in the ionosphere. "Bad" returns are those that do not, their signals pass through the ionosphere. The dataset has 351 instances and 35 features those I further divided into a set of 300 instances and 51 instances as training set and testing set respectively creating 85-15 split between the training and testing dataset. I was really motivated to use the dataset as there were no missing values in the data and all the 34 features before the "good" or "bad" returns were continuous. The dataset has allowed exploration in the fields of feed-forward networks with minimum no of layers to get the results. I find it is important to remember to keep the network as small as possible to avoid overfitting the data. Initially to test the data and the network I used the simple values, the rows and attributes of a simpler dataset.

2.2 Feedforward network

The network used here are known as feedforward networks, which comprise an input layer of identical processing units (neurons), a hidden layer and an output layer. All units found in any given layer are further connected all units in the layer above. The input layer serves the purpose of feeding the input and does not perform any computation. Hidden layer is completely hidden from the outside world, hence the name hidden layer. Neurons in the hidden layer process the inputs and pass on to the output layer. Training the neural net was done using the Stochastic Gradient Descent as an optimiser.

2.3 Implementation

The radar returns 17 discrete returns, which are composed of a real and an imaginary part hence 34 values per result. These 34 values serve as the input to the neural network. The number of hidden neurons were varied from 3 to 15 and even more for a couple of iterations. As the network is currently used to classify only two classes, good or bad, there is only one output node. This return 0 for bad and 1 for good. Normal back propagation is used to improve the network.

The neural network contains one hidden layer with initially 5 hidden neurons using sigmoid activation function. I have used an optimiser to train the network and used cross-entropy to evaluate the performance. This helps in training with large amount of data in short time. As mentioned above in this feed forward network I have defined a forward function which receives Variable input and produces Variable outputs. In this approach I could analyse and investigate the accuracy of the classification, what accuracy I was able to achieve. The neural network initially was developed from a basic training network with no hidden layers. Furthermore, it was developed into a much more elaborate and guided across the concept of future use to be developed into something bigger. An extensive use of variables and custom functions was effectively the most elegant solution to create a futuristic neural network. As we discuss the results it becomes less apparent why and how such neural networks can be created for simple binary classification and furthermore in future be developed to something bigger.

2.4 Bimodal Distribution Removal

Bimodal distributional removal works around the pattern with errors which are between two peaks. For example, in a standard feedforward neural network, the network is learning *E***[y|x],** the expected value of **y** given **x.** The pattern that are in the higher peak are *not* what the network 'expects' **y** to be in given **x.** If we assume the variance as □, as the network begins to learn, most of the pattern drop, □ drops sharply as well. This is an indication of low variance (□ almost equal to 0.1 and below) indicating the two error peaks. Slade and Gedeon[5] discuss in detail how pattern must not be removed too quickly, as those pattern with midrange errors could eventually be learnt by the network. To decide what peak to remove we calculate the mean error for all the patterns in the training set Ī. To eliminate the lower peak, we subtract pattern with error greater than Ī. This will give us the subset containing the pattern from the high peak.

3 Results

The neural network was capable to perform with a significant accuracy of 64.55% to begin with 100 iterations, the graphs below show the total loss depending on the number of iterations. As we observe the neural network is capable of learning at a very high rate, the losses to begin with are less than 1. The graphs below show how low the losses get by increasing the number of iterations and increase the accuracy. However, just by keep increasing the iterations we ca find the change in the accuracy of the network over a period.

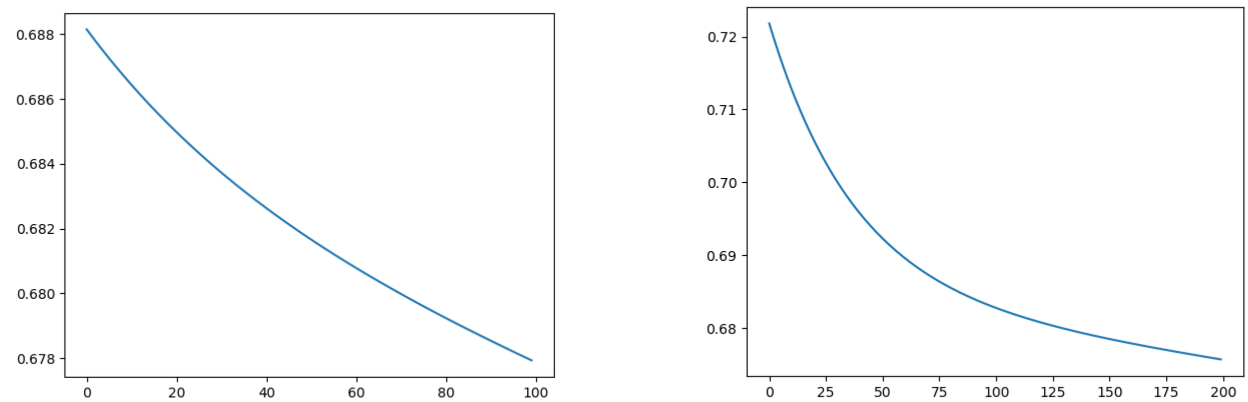
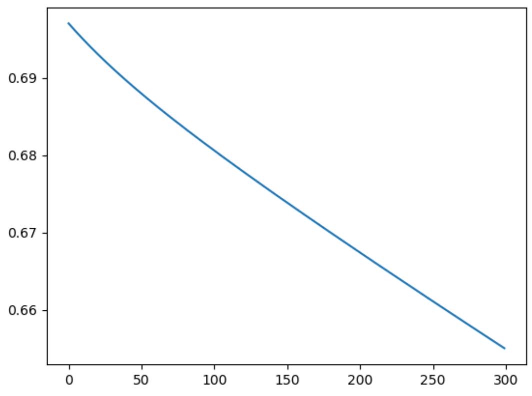


Fig. 1. and Fig. 2. Losses (y axis) vs the number of iterations (x axis). Learning curve of the network. As we can observe , I increased the number of iterations and the accuracy increased.



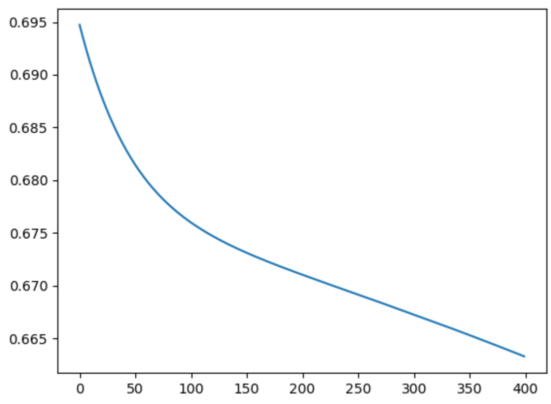


Fig. 3. and Fig. 4. Losses (y axis) vs the number of iterations (x axis). Learning curve of the network. As we can observe , the increase in iterations from 200 to 300 didn't really help the network accuracy instead it decreased the accuracy from that of 200 iterations. This describes us how the number of iterations can only improve the network so much.

The graphs above represent the loses over the number of iterations. After a few iterations I realised the network could be improved using various network reduction techniques. Siestma and Dow[5] explained how the outputs of units can be used in a two-stage pruning process. I worked with other methods introduced in the paper Network Reduction Techniques Tom and Harris[6] but was unable to work out a solution to improve the network. The results were then further tallied by the results in the research paper using the same dataset as mine.

The dataset was used by Eggermont, Kok, and Kosters[7] and the results in the paper were significantly better than the results of the neural network developed . "If we look at the results on the Ionosphere data set we see that using the gain ratio instead of the gain criterion with our refined gp algorithms greatly improves the classification performance. Only our refined gp algorithms using the gain ratio criterion also manages to beat C4.5 regardless of the maximum number of partitions we use. "

4 Future works

To give a pathway into working a Convolutional Neural Network (CNN) we will briefly discuss the convolutional neural networks. As we know CNN is a class of feedforward deep neural networks, most commonly applied to analysing visual imagery. Convolutional Networks are inspired and guided by the biological processes. Further can be discussed in detail. The ionospheric dataset can possibly be used with a sigmoid function to convert the input values to somewhere between 0 to 1. This can in turn be fed to the CNN as a matrix using the input number as the pixel values, this would perhaps lead to a refined output but would highly be used as a classifier.

We can thus, to understand a CNN use the most commonly used MNIST dataset. CNN works with layers and functions called, convolution and max pooling. A brief discussion on convolution layer and polling function. Convolutional layer applies a convolutional operation to the input, passing the result to the next layer. Each convolutional neuron process data only for its receptive field. The convolution operation reduces the number of free parameters, allowing the network to be deeper with fewer parameters. Convolutional networks include local or global pooling layers. These combine the output of neurons of one layer into the single neuron in the next layer.

5 Conclusions

The report demonstrated that the classification of radar signals can be very well be attempted by neural networks. Furthermore, we discovered that the increase in number of hidden neurons and number of epochs significantly increases the accuracy. A multiple layer neural network can be cleared of any noisy inputs with help of bimodal distribution removal. We further discussed how the convolutional neural network technique could be used but then realised the use of convolutional neural networks to derive images and pixel inputs. Interesting ideas of convolutional neural networks to be used as a classifier can be implemented using various libraries but here we just rest our discussions to the limit of using a neural network with improved error reduction techniques.

6 References

[1] Malkoff D. B., "A Neural Network for Real-Time Signal Processing". *Adv. Neural Inform. Process. Syst.* **2,248-257**(1990)

[2] V. Golovko, Y. Savitsky, N. Maniakov, Neural Networks for Signal Processing in Measurement analysis and Industrial Applications: The Case of Chaotic Signal processing. In book Neural Networks for Instrumentation, Measurement and related industrial Applications: NATO book. IOS Press, Amsterdam, (2003), pp. 119-144.

[3]Boone, J. M., Sigillito, V. G., and Shaber, G. S., "Signal Detection Capabilities of Neural Networks: Radiology Applications," Med. Phys. (in press).

[4] "Special Selection on Neural Networks", *IEEE Trans. Acoust. Speech Signal Process*, **36**, 1107-1190 (1988)

[5] Slade, P and Gedeon, TD "Bimodal Distribution Removal," *Proceedings IWANN International Conference on Neural Networks*, Barcelona, 1993. also is Mira, J, Cabestany, J and Prieto, A, *New Trends in Neural Computation*, pp. 249–254, Springer Verlag, *Lecture Notes in Computer Science*, vol. 686, 1993

[6] Harris D. and Gedeon TD, "Network Reduction Techniques" 1991.

[7] J. Eggermont, J. N. Kok, and W. A. Kosters, "Genetic programming for data classification: Partitioning the search space," in Proc. ACM SAC, 2004, pp. 1001–1005