Artificial Intelligence

Neural Networks

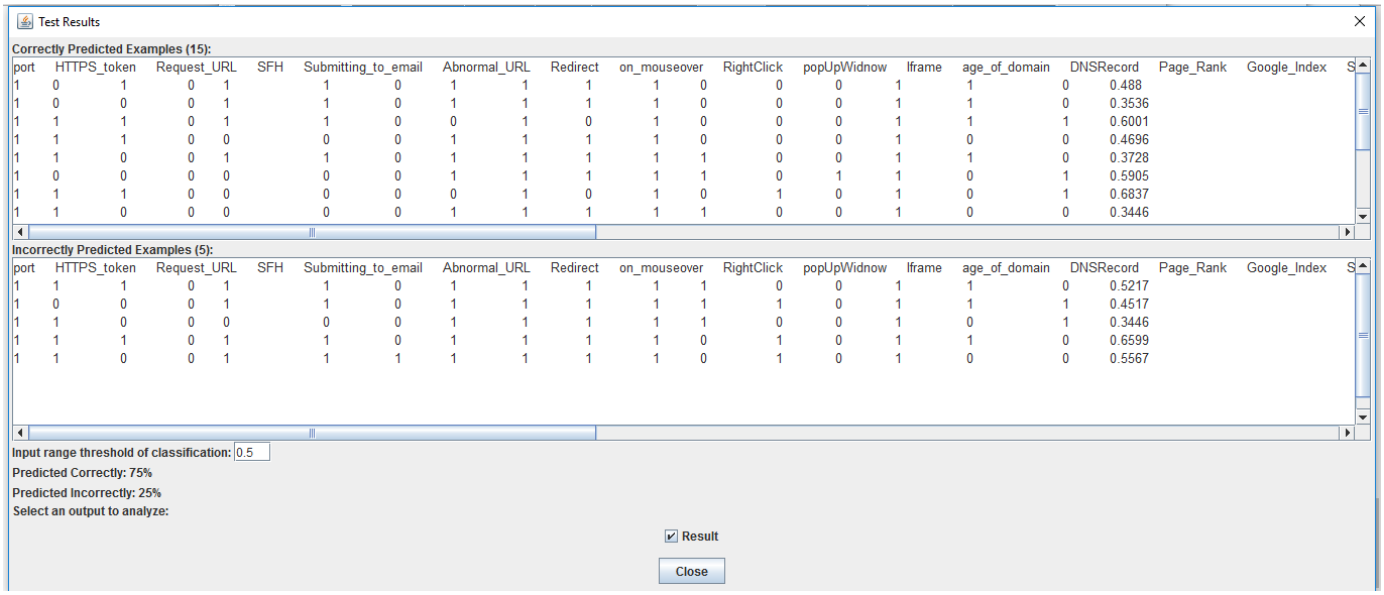
Task Report 1

Harun GÜLEÇ

Bialystok,2017

Task 1

At this task, we have to prepare a data for neural network and design the network. This data should be at least ten objects and each object should has at least three attributes. Numeric attributes are better than others. So, I chose phishing website dataset which I found from the internet[1]. Then I selected three objects randomly and saved them seperate file. I created a neural network which has one hidden layer with one neuron, to add data properly. After, I chose 33% of data to add test sets and I trained network with 50x step. The training results shown below.



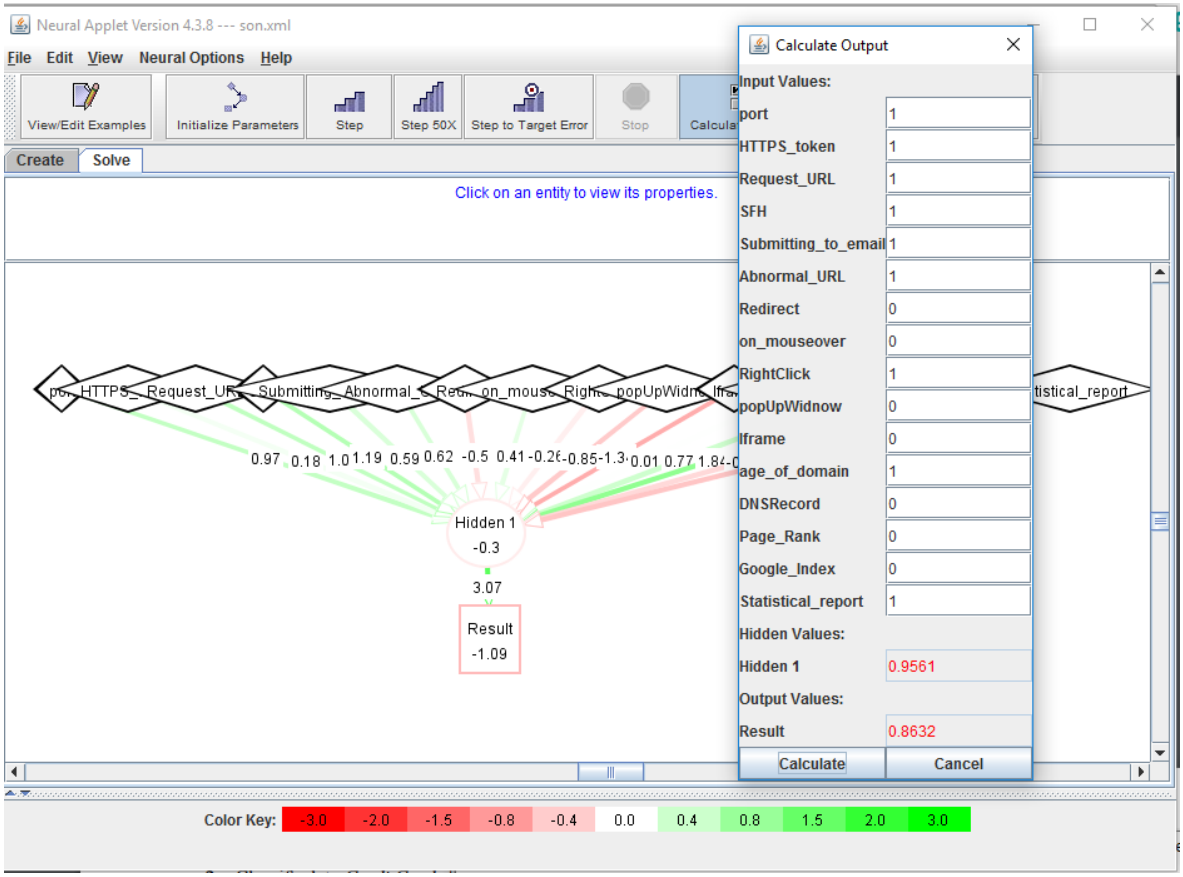
As we can see obviously on the above picture, our network has 75% success and our threshold is 0,5. Now, we can calculate other 3 data that we select before randomly. Our data is ;

```
(a)1,1,1,1,1,1,0,0,1,0,0,1,0,0,0,1,Result->1
(b)1,0,1,0,0,0,0,1,1,1,1,1,0,0,0,0,Result->1
(c)1,1,0,0,1,1,0,1,1,1,1,0,1,0,1,1,Result->0
```

This data means in order of port, HTTPS token, Request URL, SFH, Submitting to email, Abnormal URL, Redirect, on_mouseover, Right Click, pop up window, Iframe, age of domain, DNS record, page rank, Google index, statistical report and result.

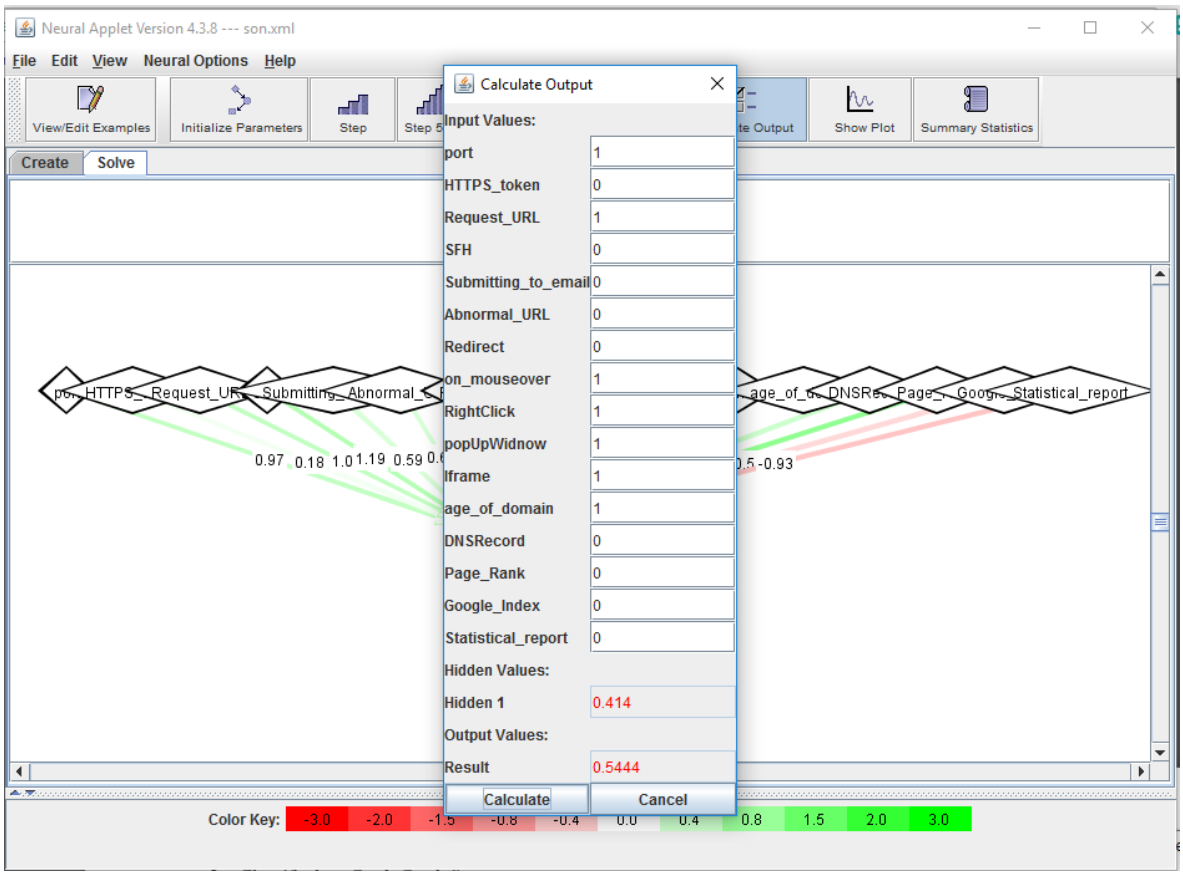
Firsly, I calculated data (a) and its results shown below.

[1] https://archive.ics.uci.edu/ml/datasets/Phishing+Websites

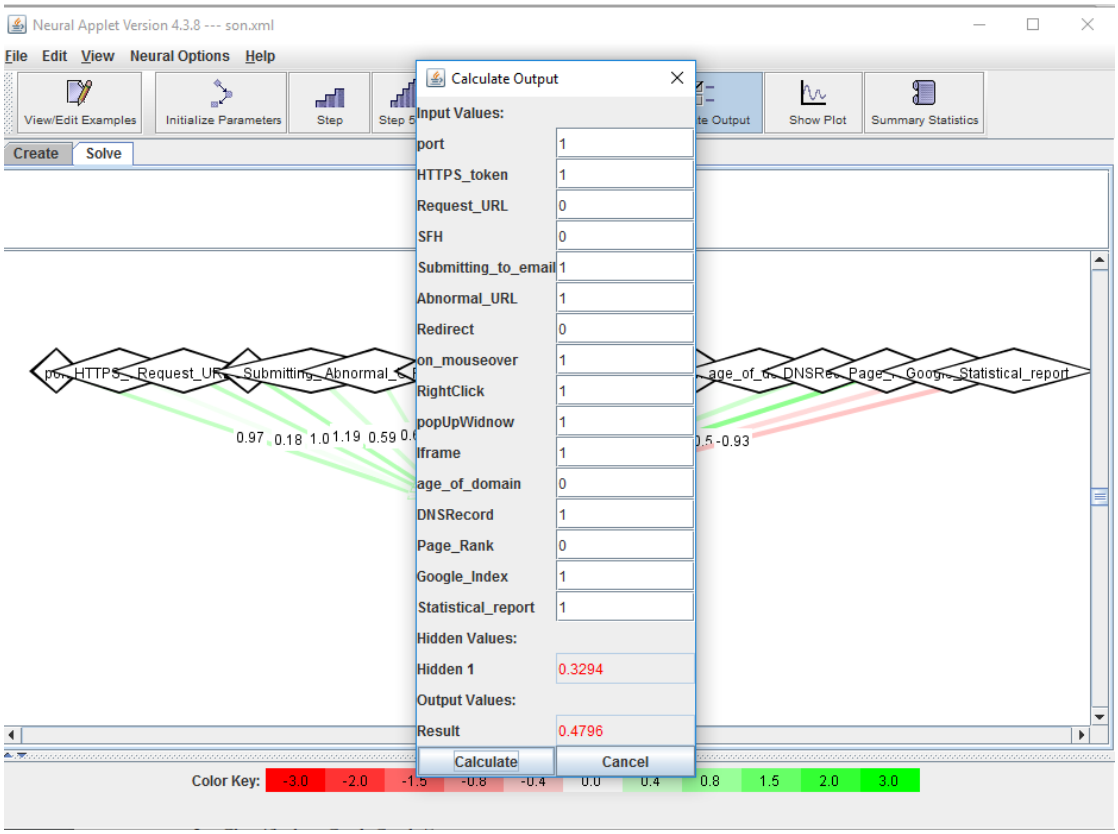


When I calculated it, result is 0,8632. It is higher than our threshold, so it is positive. In the other words, predicted result is 1 and true.

Next calculation for data (b) and its result shown below;

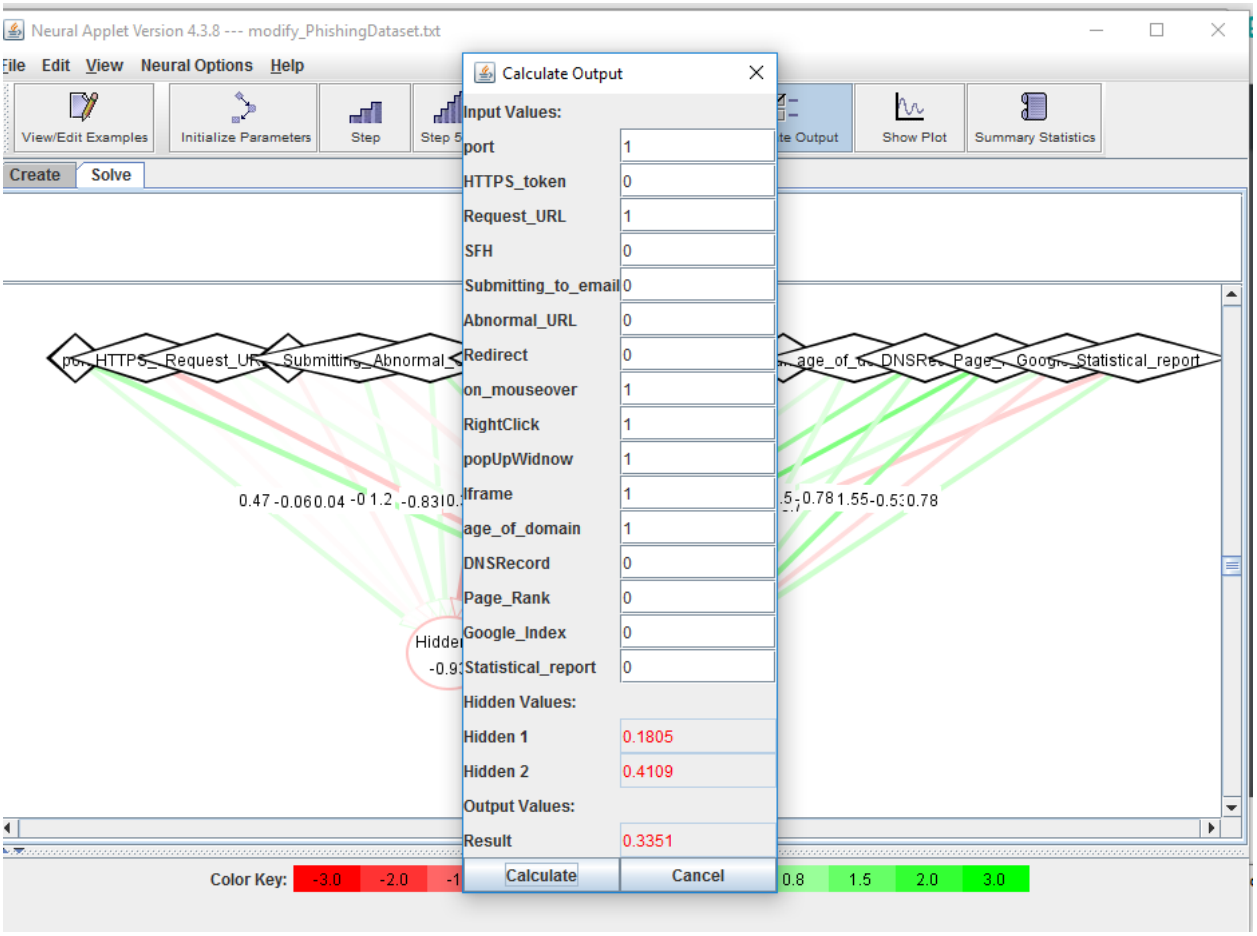


The result is 0,5444 and higher than threshold. It is again positive and true.
Finally, I calculate last data (c). The results shown below;

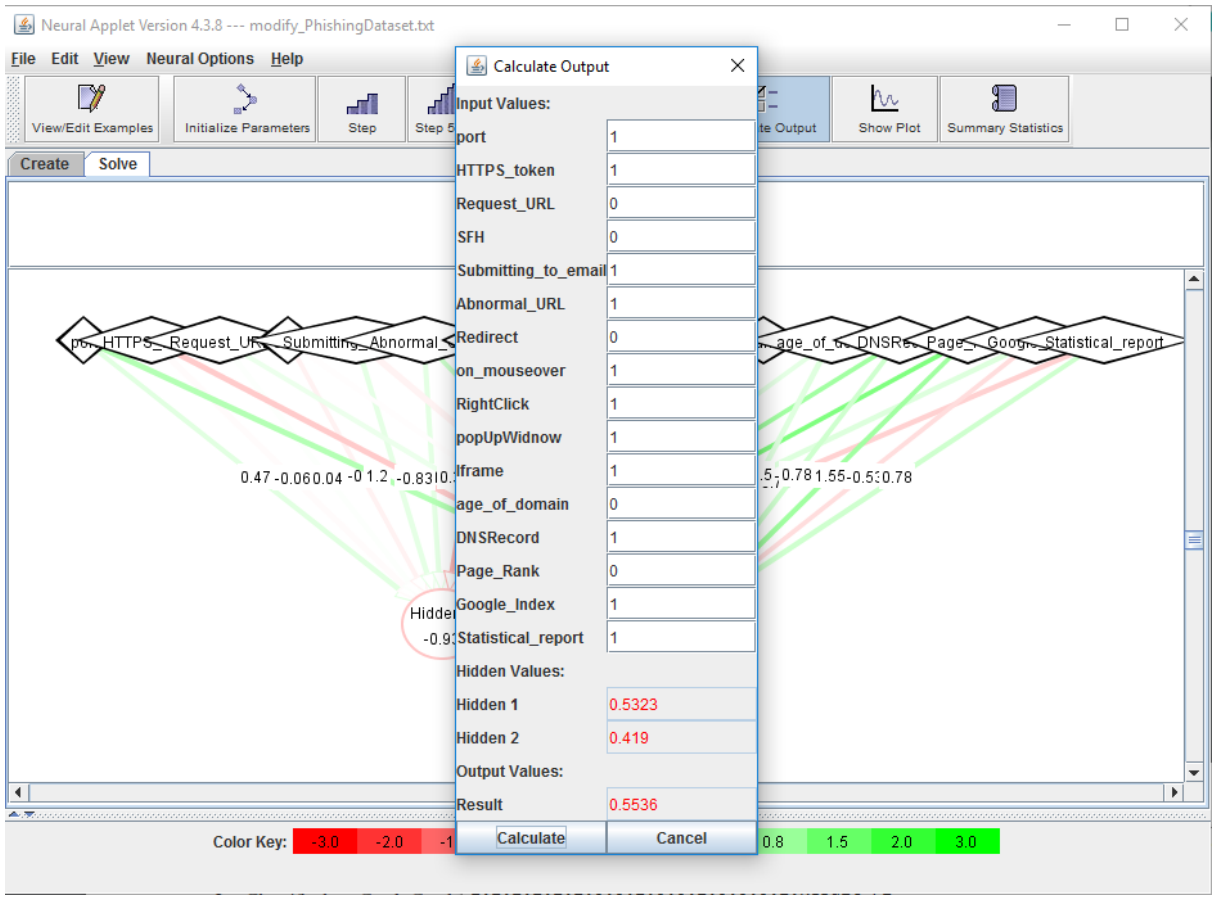


This time, result is lower than threshold. It means negative or 0. Also, expected result is 0 so, it is true.

After that, I repeat same calculation on diffrent networks. One of them is one hidden layer with two neurons and other one hidden layer with three neurons. Networks have 75% success again. But the network which has two neuron cannot calculate correctly our data as shown on below picture. So it is not a good network design for our problem. We should use one neuron or three neuron network designes to get more stable responses.



Data (b) : Result should be positive but predicted result negative.



Data (c) : Result should be negative but predicted result positive.

Task 2

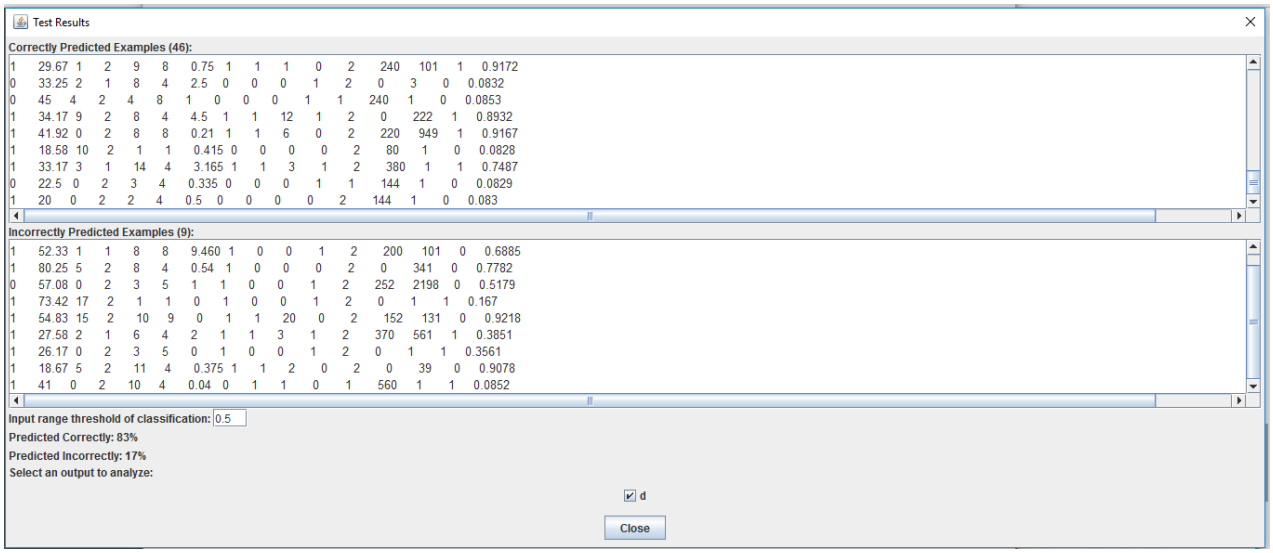
At this task, we have to classify the given data[2] with diffrent options and evaluate the results.Firstly, I loaded the data and divide randomly the data into the training and test sets in the ratio 2 : 1. First network has one hidden layer with one neuron and use sigmoid function to calculate output. I trained the network and network has 36% classification success. It is too low. Then, I trained the network using Tanh function. This time, I achieved 63% classification success. This is good percentage but we should try other funcitons.So, I chose exponantial function and again I get 63% success. Lastly, I chose linear function and I achieved 65% classification success. In conclusion we can say that Linear function is better than other functions for this dataset and network design.

Our second experiment is about normalized inputs. To apply this I selected normalize input options and again trained the network. After this training, I achieved 83% success. This is the best percentage. It is shown below.

For third experiment, I used one hidden layer with three neurons network design. I applied design and then trained the network. New classification success is 47%. This is not so good.

Last experiment is about number of hidden layers. I increased hidden layers so network has 2 hidden layers and each layer has 2 neurons. After training operations, I checked classification result and it is 56%.

In conclusion, network with normalized inputs has the highest classification success for this data. On the other hand, Linear function has better result than other functions.



[2] The data is placed in file CreditCards.txt and it concerns bank clients for which decisions to issue credit cards or not are made.