ASSIGNMENT 2

Name: Jason Collier
Student Number: 214008258
Date: 15 August 2017

I declare that this is my own, original work.

Signature: _____

IMPLEMENTATION DETAILS

Stopping Conditions:	IF (t < 5000) OR (SSE < 500) OR (accuracy > 80)
Initial Weights:	Weights are initialized to $(-1/\sqrt{f_{anin}}, 1/\sqrt{f_{anin}})$ where fanin is 28 for the weights between the input and hidden layers and 11 for the weights between the hidden and output layers
Training Set size:	1800 patterns
Validation Set size:	100 patterns
Test Set size:	40
Values for η investigated:	1, 0.1, 0.07, 0.01
Activation functions:	Sigmoid activation function
Number of hidden neurons:	10

Write a paragraph on the network morphology, focusing on what inputs are provided to the network and what is given as output.

The network used is a standard feedforward neural network consisting of three layers, the input layer with 27 neurons, the hidden layer with 10 neurons and the output layer with 7 neurons. Each neuron in the input layer has one input signal scaled between 1 and 0, each neuron in the hidden layer has 28 signals (from 27 input neurons and 1 bias) and each output neuron has 11 signals (from 10 hidden neurons and 1 bias). The neural network, once trained using the stochastic gradient descent learning algorithm, will output a 1 from the neuron whose position number corresponds to the fault number and a 0 from the other 6 neurons. For example, for a pattern p with a set of 27 inputs which correspond to a fault number of 4, the trained neural network will output a 1 from the 4^{th} neuron in the output layer and a 0 from the other 6 neurons.

RESULTS

Number of iterations (typically):	5000
Best η value:	0.07
Sum Squared Error (SSE) on Training Set:	515 (η = 0.07, iterations = 5000)
SSE on Validation Set:	551 (η = 0.07)
Number correctly classified on Training Set:	1373 (out of 1800)
Number correctly classified on Validation Set:	71 (out 0f 100)

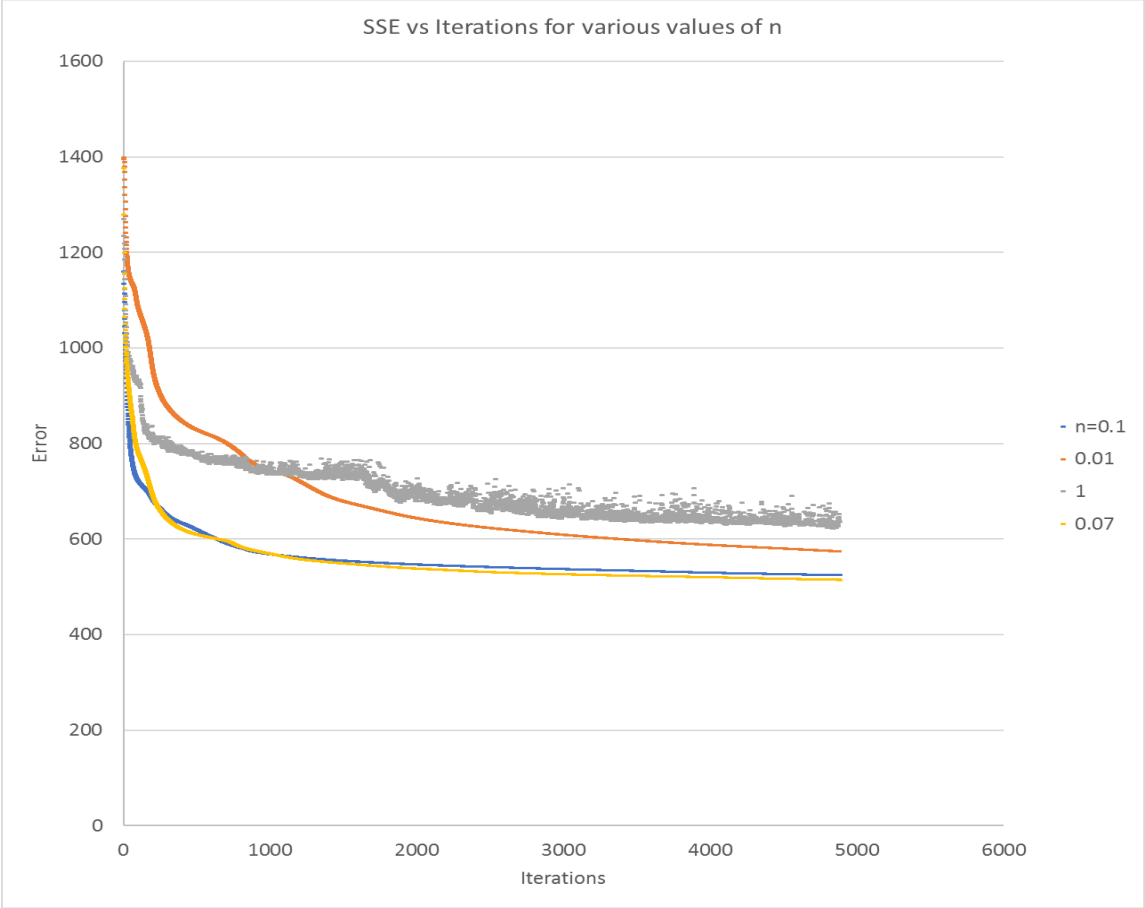

Figure 1: SSE vs iterations for various values of η

INVESTIGATED TECHNIQUES

Additional Techniques Investigated:

Technique	SSE on training set	SSE on test set
Weight Initialization	503	545
Momentum	512	560
Network Architecture		

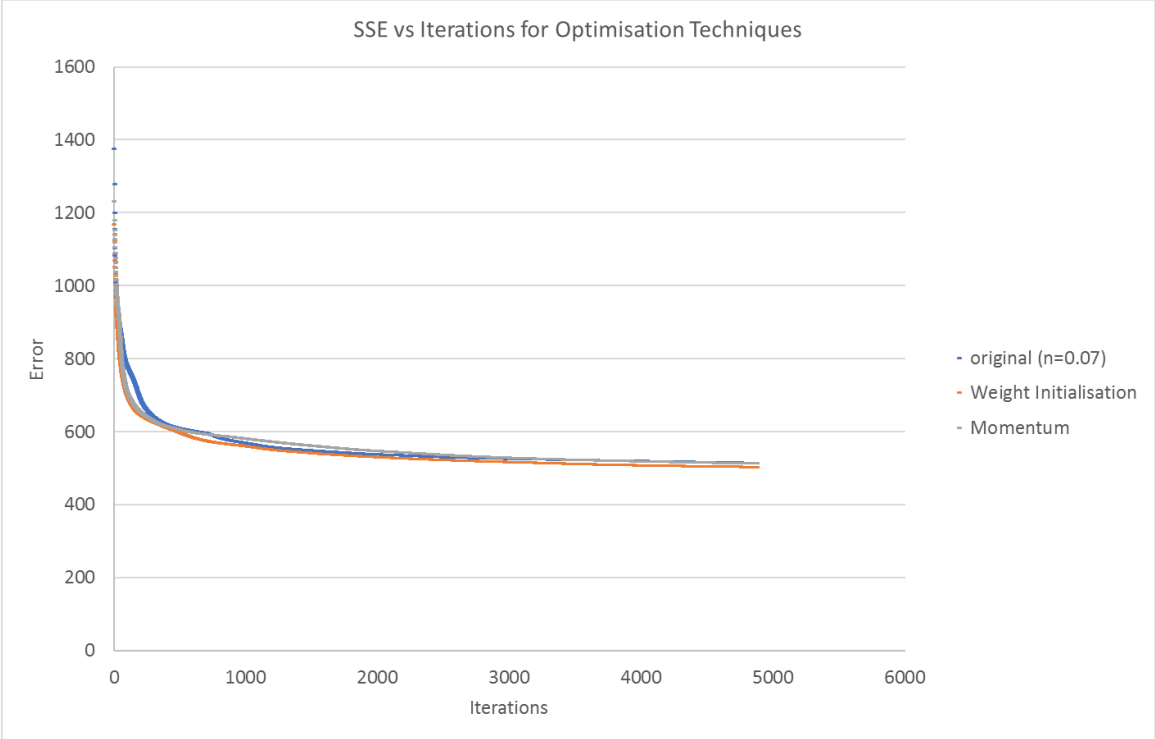

Figure 2: SSE vs iterations for the three techniques

Results on patterns provided without category:

1	1	11	2	21	7	31	7
2	1	12	7	22	3	32	7
3	3	13	6	23	7	33	2
4	4	14	1	24	6	34	7
5	2	15	3	25	7	35	2
6	3	16	7	26	6	36	2
7	7	17	7	27	7	37	7
8	3	18	7	28	2	38	7
9	4	19	5	29	7	39	7
10	1	20	3	30	7	40	6

OBSERVATIONS

Write a few paragraphs on what you have learned about the training of a neural network. How well do you think it solved the problem?

The neural network solved the problem in a satisfactory manner. The most accurate training took place over 65000 iterations and in excess of 3 hours for an overall accuracy of 82 percent and a sum squared error of 450. This approach utilized a learning rate of 0.07, correct weight initialization and a momentum factor. Any further iterations would not be worth it as the changes in sum squared error and accuracy values become negligible. This neural network predicted the outcomes of the validation set with 78 percent accuracy and with a sum squared error of 501.

The results described above show that with real world data, it is unlikely that perfect results will be obtained. The set of parameters, as described in the paragraph above, that produce the lowest error possible, are used to predict the results of the test set.

Of the 4 learning rates tested (1, 0.1, 0.07 and 0.01), it was found that the learning rate of 0.07 was most effective. It had a final sum squared error of 515 and an accuracy of 76 percent on the training set.

In the early years of machine learning, most learning algorithms were optimized through a trial and error method implemented by experienced programmers. Nowadays, certain tried and tested performance enhancement methods can be used. In this case, the weights will be initialized correctly, a momentum factor will implemented and the network architecture will be modified.

The gradient descent algorithm used is very sensitive to the initial weight vectors. If the initial position is close to a local minimum then convergence will occur too fast. In the case of optimization algorithms, weights should be initialized as small values centered uniformly around zero as stated by Wessels and Barnard. This resulted in the best results with a final sum squared error value of 503 and an accuracy of 77.2 percent on the training set.

With regards to momentum, when weights are adjusted after each training pattern as present in stochastic learning, there are fluctuating weights changes that occur over the training set. This problem can be reduced through the addition of a momentum term. The momentum term is used to keep the weight changes occurring at the same rate. The momentum term is simply the previous weight change weighted by a scalar value between 0 and 1. This method resulted in a slightly improved result to the original a with final sum squared error value of 512 and an accuracy of 76 percent on the training set.

If several networks fit the training set equally well, then the simplest network (with the smallest number of weights) will give the best performance. Networks with too many parameters memorize training patterns instead of learning them. Overfitting can be prevented by reducing the size of the network. The complexity of the network must thus be balanced with the goodness-of-fit of the true function through a process called architecture selection. Approaches to architecture selection include regularization, network construction, network pruning etc.