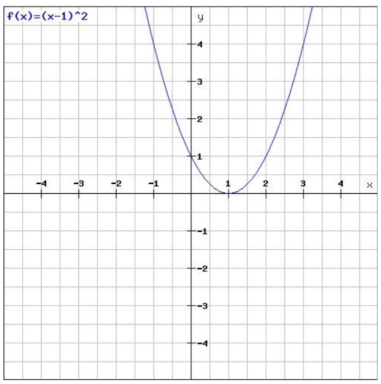
CS 4633

Third Programming Assignment

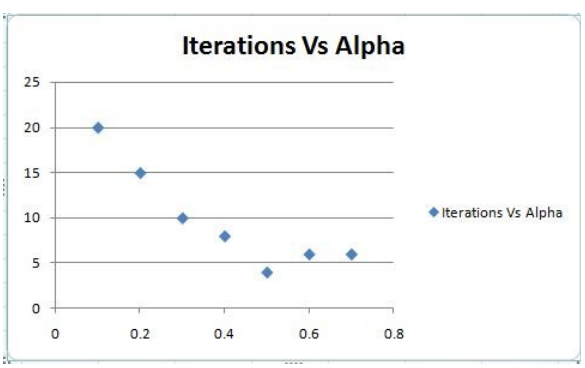
Naveen Neelakandan(nn149)

Part 1

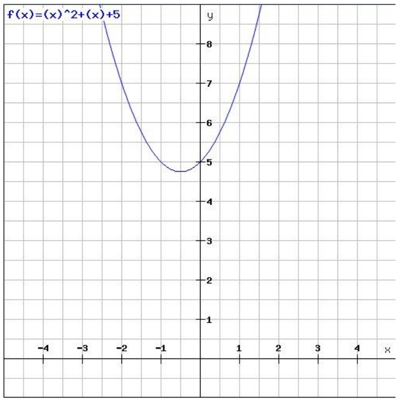
The derivative of f(x) = (x – 1)^2 is 2(x-1). The graph of the function is as follows.



The C program given was run to perform gradient descent to find the minimum of the function. The minimum was found to occur when x=1 and the subsequent value of f(x) = 0. The program was tweaked slightly to include a test for convergence. We assume the value of x has converged if it is not changed by more than 0.001 for 3 consecutive iterations. Using this new test for convergence, we run the program with differing values of alpha (learning rate) to find how it affects convergence. In general, it was found that increasing the value of alpha lead to needing fewer iterations to converge. The graph of the number of iterations needed to converge against the value of alpha is as shown below.

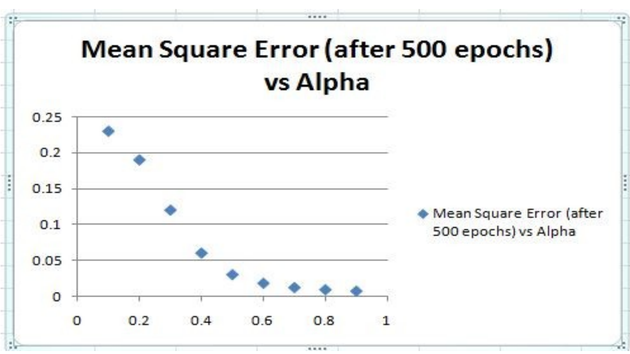


As seen from the graph above, too high a value of alpha though can lead to an increase in the number of iterations to converge due to overstepping. Furthermore, the value converged to is also not optimal. The other function the gradient descent algorithm is run on is X^2 + X +5. The minimum was found to occur when x=-0.5 and the subsequent value of f(x) = 4.75. The graph of the function is as below.

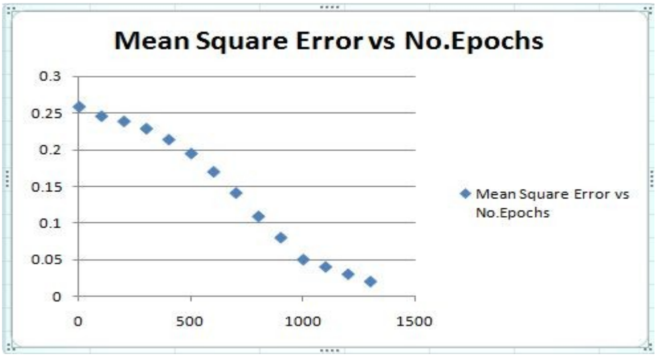


Part 2

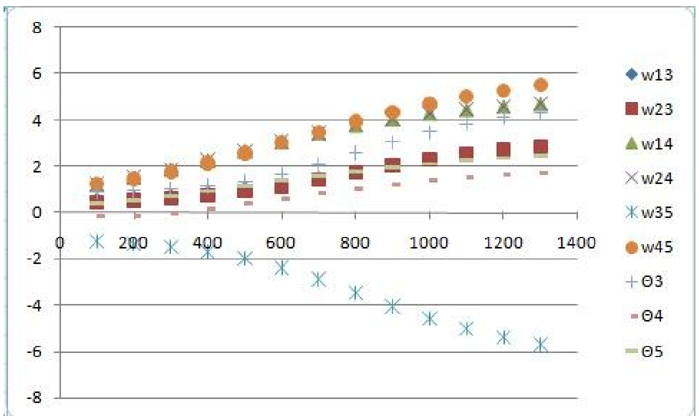
In the second part, the back propagation algorithm was used to train a simple neural network to learn the XOR function. The biasing input is always set to -1. Alpha(learning rate) is modified to test how the value of alpha affects convergence for the neural network. It is found that just like in the gradient descent example, a higher learning rate leads faster convergence in the value of the weights and subsequently the mean square error. However, the issue of overstepping is not as obvious as in gradient descent (at least till the maximum of 1.0). The graph of mean square error after 500 epochs of training vs the learning rate is shown below.



The graph of mean square error vs the number of epochs of training is shown below for a learning rate of 0.2 and a biasing input -1.



The graph and table of mean square error vs the weights is shown below for a learning rate of 0.2 and a biasing input -1.



No.of Epochs	w13	w23	w14	w24	w35	w45	θ3	θ4	θ5
100	0.549	0.455	1.16	1.23	-1.26	1.24	0.861	-0.141	0.391
200	0.62	0.528	1.48	1.52	-1.36	1.45	0.927	-0.119	0.538
300	0.714	0.623	1.83	1.86	-1.5	1.75	1.01	-0.014	0.723
400	0.837	0.751	2.22	2.24	-1.7	2.12	1.141	0.169	0.934
500	1.05	0.922	2.63	2.64	-1.99	2.56	1.34	0.393	1.15
600	1.219	1.14	3.04	3.05	-2.39	3.01	1.64	0.617	1.38
700	1.48	1.42	3.42	3.42	-2.89	3.46	2.06	0.828	1.59
800	1.77	1.72	3.75	3.75	-3.48	3.9	2.56	1.02	1.79
900	2.06	2.03	4.02	4.03	-4.05	4.31	3.05	1.21	1.97
1000	2.32	2.29	4.25	4.25	-4.57	4.67	3.47	1.36	2.13
1100	2.53	2.51	4.42	4.43	-5.01	4.98	3.82	1.5	2.26
1200	2.71	2.69	4.56	4.57	-5.37	5.25	4.1	1.6	2.38
1300	2.85	2.83	4.68	4.68	-5.67	5.47	4.33	1.69	2.48