Neural Network Implementation
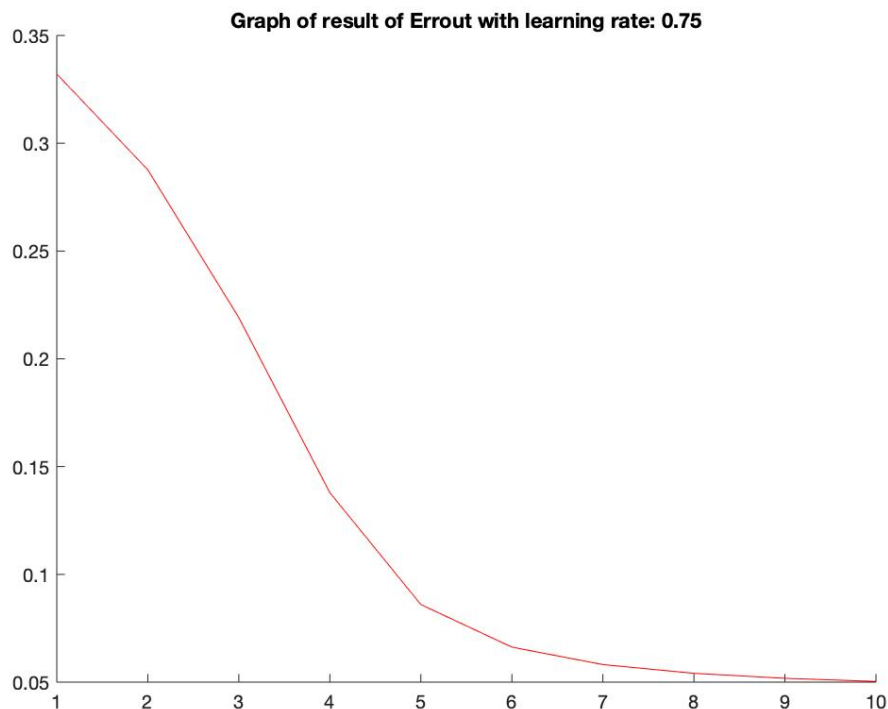## By Marco

A. Difficulties

The first difficulty I got was how to determine the stopping criteria. The stopping criteria will affect the result of the error. For example, the program will keep running, even though the error in the sample and out sample is good enough. So, I have tried to modify the stopping criteria a little bit each time. For example, one of my stopping criteria is that the similarity of an error each time. If the value of error is not changing too much in a range 10 times, the algorithm will stop. So, I checked the curve which is the error in the sample and out sample iteration. So, I found that the changing of error within 0.01 makes us get a good result.

Also, I was stuck in how to implement the batch gradient descent. Due to the previous assignment, we only implement the gradient descent with logistic regression. We can just train all data points at one time. But, in the neural network, it has several layers, we can train our data points at the same time. So, I followed the direction which is training one data point each iteration. And then I calculated the average of error in sample and gradient for updating the weight.
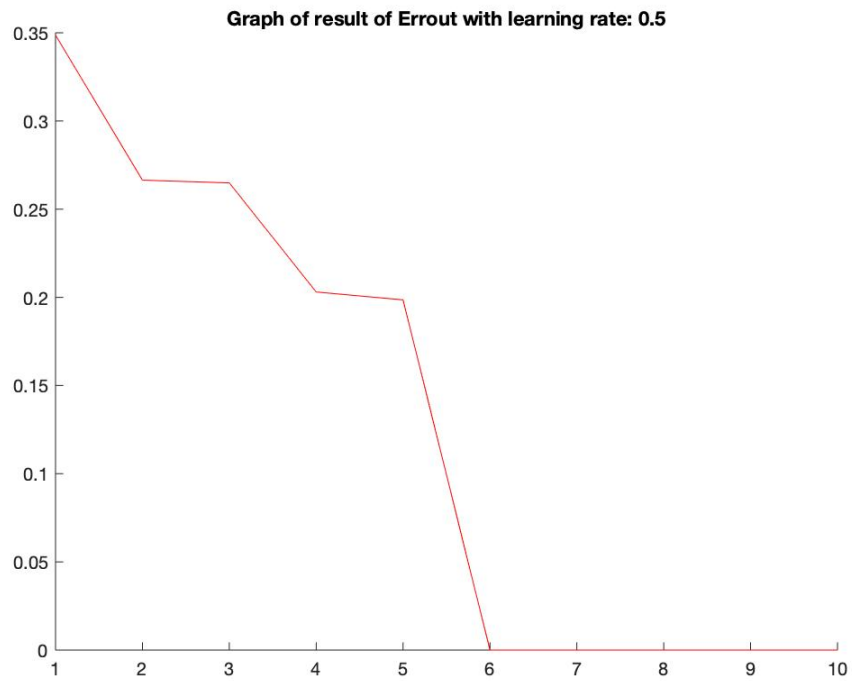
B. Plots of error over iteration for each weight update method:

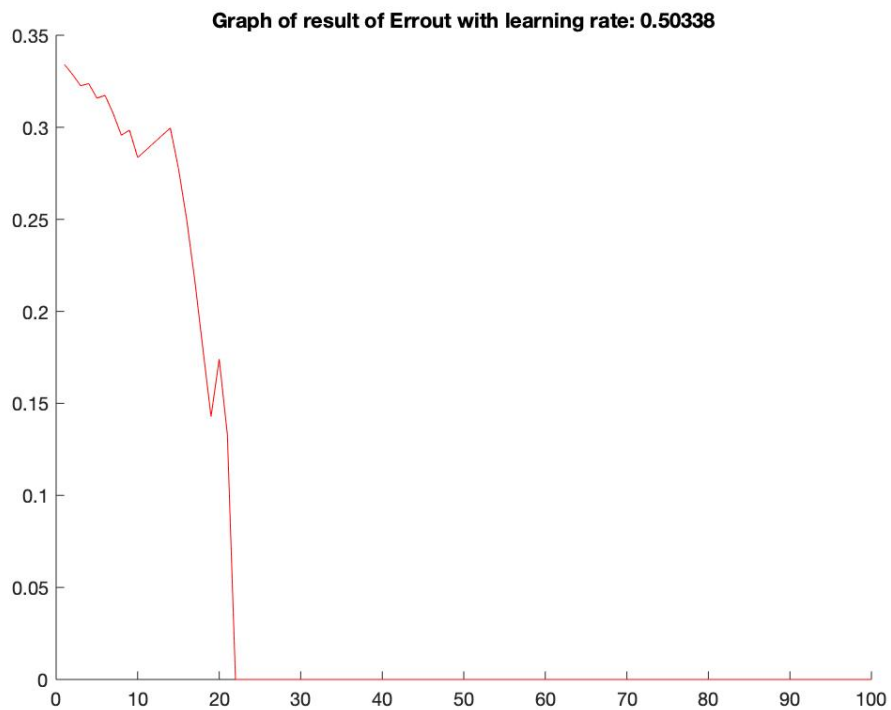The curve represents the tendency of error out of sample over each iteration.
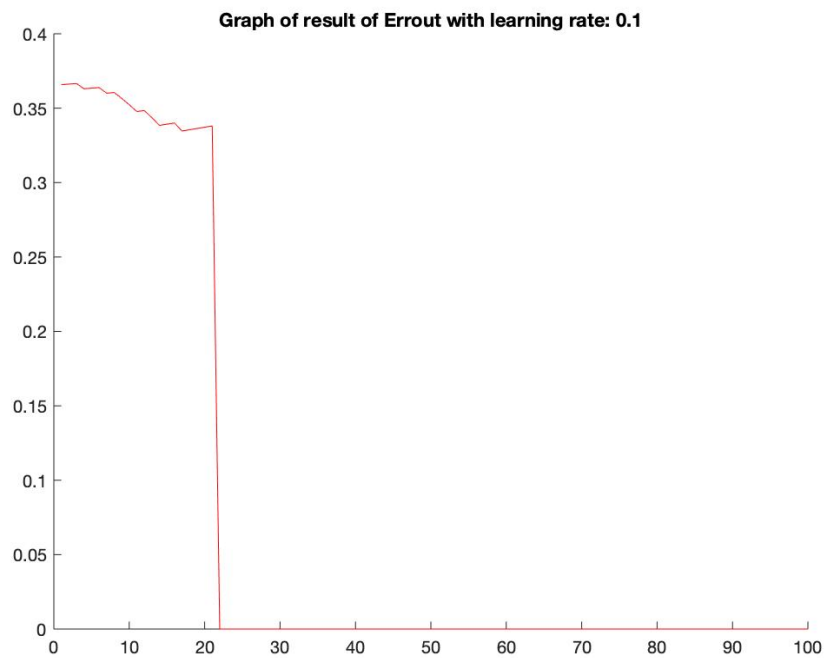
    i.   Batch Gradient Descent

Graph of result of Errout with learning rate: 0.75

ii.   Stochastic Gradient Descent

Graph of result of Errout with learning rate: 0.5

iii.   Variable Learning Rate Gradient Descent

Graph of result of Errout with learning rate: 0.50338

iv.   Steepest Descent (Gradient Descent with Line Search)

Graph of result of Errout with learning rate: 0.1

C. Average of in-sample and test-sample error:

I have run 50 times to get solid average result for each method.

	Batch Gradient Descent	Stochastic Gradient Descent	Variable Learning Rate Gradient	Steepest Descent
In-sample	0.047621	0.1985	0.2439	0.0159
Test-sample	0.041914	0.1438	0.2127	0.0161

D. Map of classification space:

Blue color represents the points which are digital 1. And the green color represents the points which are not digit 1.

Graph of result of data point