

Learning Rate →		0.05	0.1	0.5
Activation Function				
Sigmoid	Iterations=500	0.2898541935750666	0.28821329020232456	0.39431305915498815
	Iterations=1000	0.28815721579552	0.2840481571679213	0.25899057108401763
	Iterations=1500	0.2861797084751867	0.27927608832547784	0.24513902585184685
Tanh	Iterations=500	39.177540837564116	29.455924343144428	39.18046602794019
	Iterations=1000	29.61220726224655	39.180466024861154	39.18046602794019
	Iterations=1500	29.61208341967985	39.18046602486114	39.18046602794019
ReLu	Iterations=500	0.39653175365814963	0.39653175365814963	0.39653175365814963
	Iterations=1000	0.39653175365814963	0.39653175365814963	0.39653175365814963
	Iterations=1500	0.39653175365814963	0.39653175365814963	0.39653175365814963

- All the values in the cells indicate test error.
- I used the following data set from uci ml repository:
<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer>
- Reasons for choosing the above dataset is
 - It has missing values
 - Attributes are categorical
- The data has 286 instances and 215 (75%) of them were partitioned into training data and the rest 71 (25%) were under testing data.
- From the above data we can conclude that
 - “testError” remains constant with respect to the change in learning rate and iterations if the activation function is ‘ReLu’
 - “testError” is very high using ‘tanh’ activation function and it does not change relatively with learning rate and iterations. (i.e it sometimes increases and sometimes decreases and remains constant with constant learning rate or no.of iterations)
 - “testError” decreases with increase in learning rate and increase in no.of iterations upto some point and decreases thereafter, if the activation function is ‘sigmoid’
- For the given data set under the observations it is clear that sigmoid is the best activation function to use rather than ReLu.
- The reason that sigmoid produce better results than ReLu in this case is due to less number of hidden layers.
- As the number of hidden layers is 2 which is a less quantity there maybe no chance of vanishing gradient and thus sigmoid performs well.
- The best value of “testError” (Observed at 1500 iterations with learning rate=0.52 and 'sigmoid' activation function) is 0.24225440723671607