Report

Pre-Report points:

- We took Iris data set to test against all the three activation functions.
- We converted categorical data and into numerical data.
- We converted Iris-setosa into “0”.
- We converted Iris-versicolor into “1”.
- We converted Iris-virginica into “2”.
- We separated the class label column from the remaining data frame and then preprocessed the remaining part of data frame so that preprocessing is not done on the class labels. This way the quality of preprocessing will not deteriorate.

Report:

- The sigmoid activation function has an error of approximately 5.51 on the test data when learning rate is 0.05 and number of iterations are 1000.
- The error of sigmoid activation function is approximately 7.14 on the test data when learning rate is 0.7 and number of iterations are 1000.
- The error of sigmoid activation function is approximately 5.50 on the test data when learning rate is 0.05 and number of iterations are 2000.
- The ReLu activation function has an error of 29 on the test data when learning rate is 0.05 and number of iterations are 1000.
- The error of ReLu activation function is 29 on the test data when learning rate is 0.7 and number of iterations are 1000.
- The error of ReLu activation function is 29 on the test data when learning rate is 0.05 and number of iterations are 2000.
- The tanh activation function has an error of approximately 13.93 on the test data when learning rate is 0.05 and number of iterations are 1000.
- The error of tanh activation function is approximately 80 on the test data when learning rate is 0.7 and number of iterations are 1000.
- The error of tanh activation function is approximately 13.84 on the test data when learning rate is 0.05 and number of iterations are 2000.
- To summarize, the sigmoid activation function has the minimum error regardless of the learning rate and the number of iterations.

This is because

- there can be loss of information in ReLu activation function when x<0. Moreover, the activation's output can be infinite.
- The range of sigmoid function is [0, 1] and I is also easily differentiable. In other words the activation's output can not be infinite.
- When the learning rate is increased, the tanh activation function performs even worse because the range is [-1, 1] and hence error is more in tanh activation function as the step size in gradient is quite large, say for example when learning_rate = 0.9.

All in all, when the number of iterations is increased there is little to no difference in the error of all the activation functions.

Also, when the learning rate is increased to 0.7, the error increases in sigmoid and tanh activation functions with tanh being the worst.