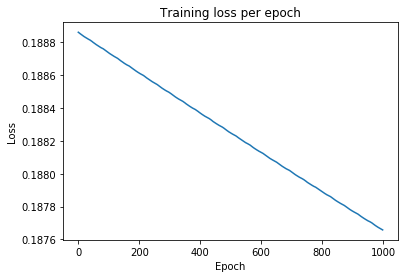
Question 1.f

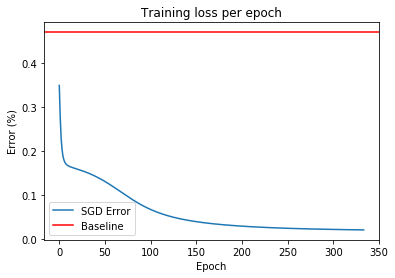
# Observations in all graphs after using MSE Error

1. Gradient descent loss per epoch graph from question 1.a



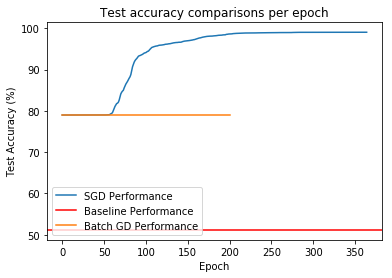
The graph looks the same. The error values are lower as the formula of error calculation is now MSE.

1. Training loss per epoch using SGD and baseline graph from question 1.c



The accuracy on test using MSE is slightly lower (98.4%) compared to that using BCE loss (98.8%). MSE loss values are low as expected compared to BCE loss.

1. Batch GD vs SGD comparison graph from question 1.d



The comparison graph looks similar to the one with Binary cross entropy loss.

Test accuracy with MSE error (99.02%) is better than that with BCE error (98.32).

1. Performance on number of hidden units graph from question 1.e



The graphs expectedly shows decrease in test accuracy with increase in hidden units. This shows clear overfitting.