# Assignment 4

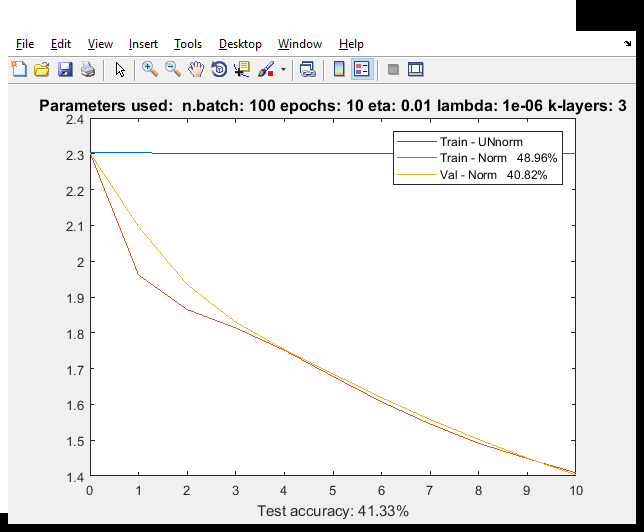
For evaluating the difference between the analytical and numerical gradients, *m*=5 and a sequence of 25 characters from the *book\_data* was considered, obtaining the following differences showing the worst approximation from the different matrices.

|  |  |
| --- | --- |
| Parameter | Worst approximation |
| b | 1.5281e-09 |
| c | 1.0348e-10 |
| W | 1.9793e-07 |
| V | 7.7841e-09 |
| U | 3.1156e-09 |

For this approximation the following equation was used:

After this, the highest value for each case was obtained as worst approximation.

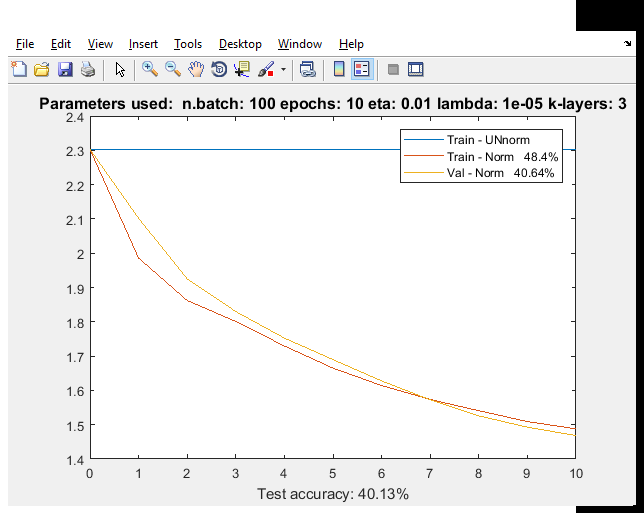
Running a long training of 2 epochs the following graph for the *smooth\_loss* is obtained.

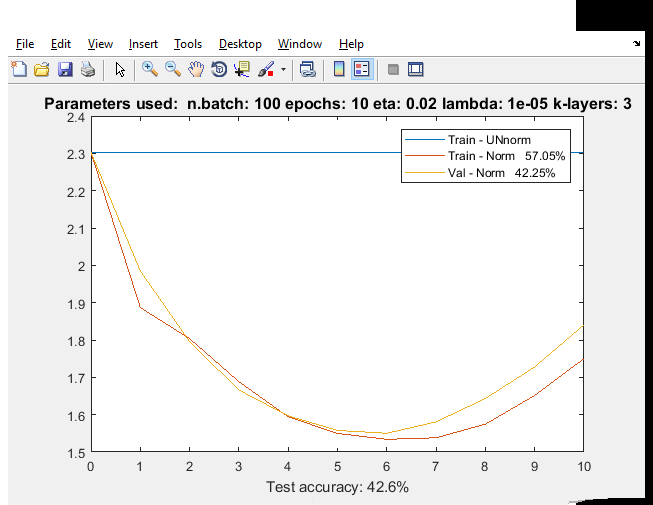


In Figure 1 it can be seen how different is to train a 3-layer network with and without batch normalization; for this test, the eta and lambda found for Assignment 2 were used. (**eta=0.0385** and **lambda=1e-6**) .

In order to find good values for lambda and eta, a coarse and fine search was done. By using a 3-layer network with 50 and 30 nodes for the hidden layers. The range of [1e-7,1] for lambda and eta was used to have a rough estimation of the good parameters to use. In this case the best values were obtained at:

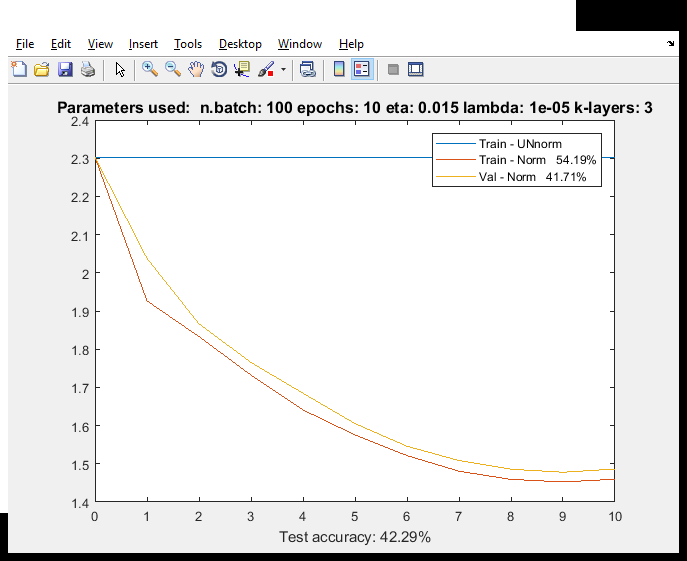
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Coarse search for lambda and eta – from 1e-7 to 1 | | | | |
| eta | lambda | Training Data - Accuracy | Validated Data - Accuracy | Testing Data - Accuracy |
| 0.01 | 1e-6 | 48.98% | 40.82% | 41.33% |
| 0.01 | 1e-5 | 48.4% | 40.64% | 40.13% |
| 0.02 | 1e-5 | 57.05% | 42.25% | 42.6% |

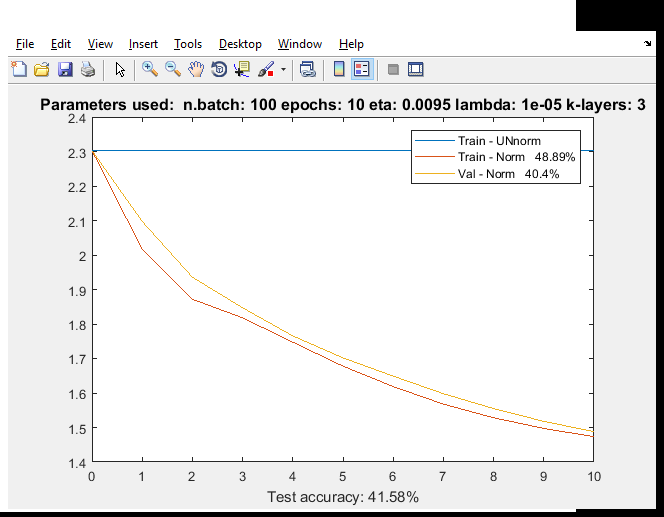


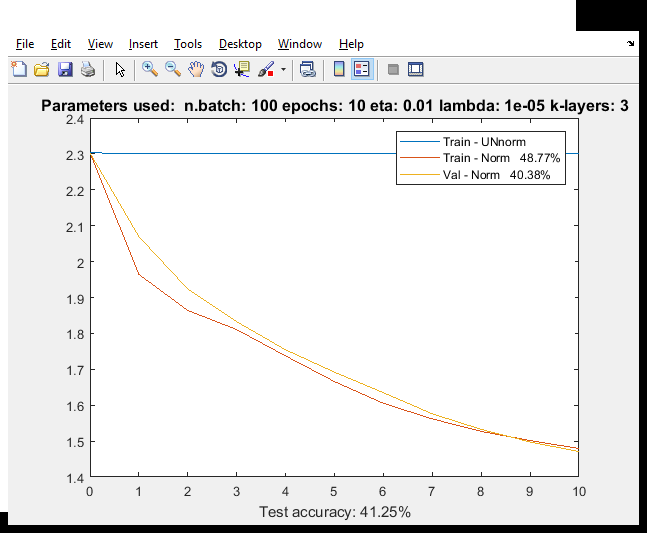


For the fine search, the range of [1e-6, 1e-5] for lambda and [0.001, 0.2] for eta was used. Obtaining the following best values:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fine search for: lambda between [1e-6, 1e-5] and eta – from between [0.001, 0.2] | | | | |
| eta | lambda | Training Data - Accuracy | Validated Data - Accuracy | Testing Data - Accuracy |
| 0.015 | 1e-5 | 54.19% | 41.71% | 42.29% |
| 0.0095 | 1e-5 | 48.89% | 40.4% | 41.58% |
| 0.01 | 1e-5 | 48.77% | 40.34% | 41.25% |



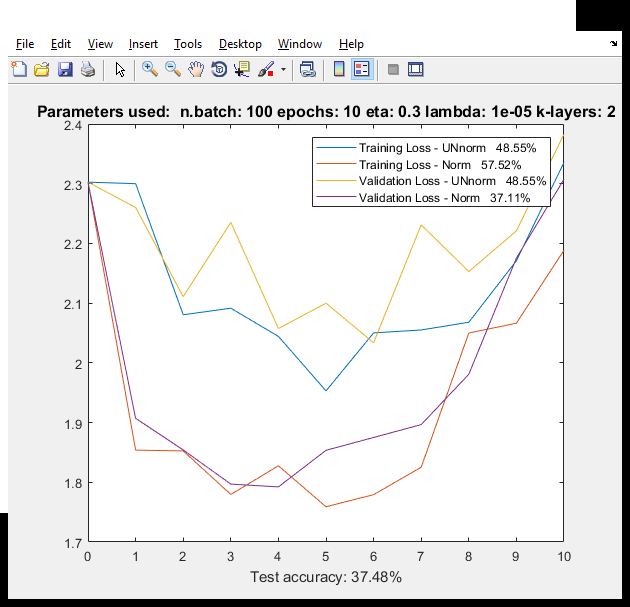




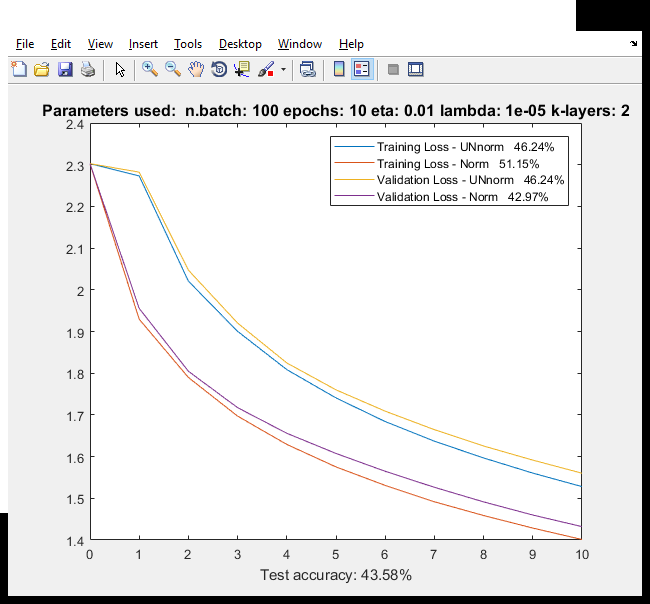
It can be said that good intervals for lambda and eta are [1e-6, 1e-5] for lambda and [0.009, 0.015] for eta. In this case, for best values were chosen **eta: 0.095** and **lambda: 1e-5,** obtaining a test accuracy of 41.58%

By using lambda as 1e-5 and using different learning rates the following results were obtained:

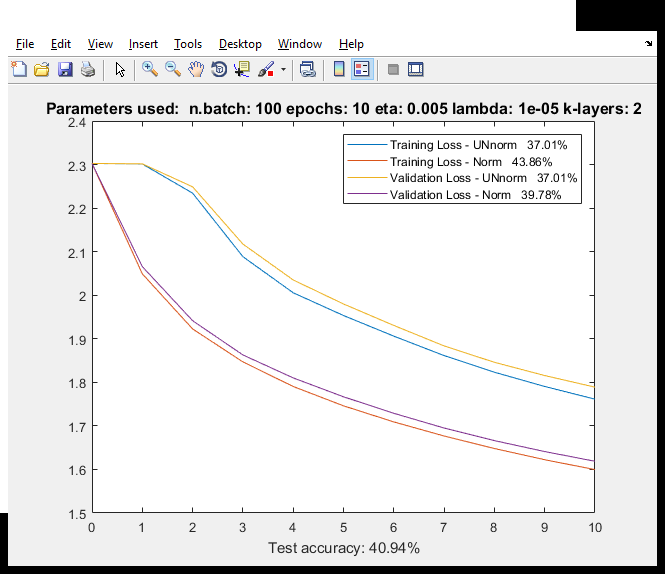
**High learning rate:** Eta: 0.3



**Medium learning rate:** Eta: 0.01

****

**Low learning rate:** Eta: 0.005

****