Assignment 3

(207548231) (319081667)

**Task1:**

Attached is the Train – Test loss graph over training epochs.

![תמונה שמכילה טבלה

התיאור נוצר באופן אוטומטי]()

Test Error of the network when finished: 2.17%

Accuracy of the network on the 10000 test images: 97.83 %

6 Misclassified images:



**Task2:**

Test error for each seed:



Mean test error for last epoch: 0.000839

Standard deviation for last epoch: 0.0001146

Based on the Standard deviation we received; we can observe that all errors at the last epoch are very close to the average. Therefore, our model can be considered as a Robust to the choice of a seed number, as the Std is considerably very small.

In other words, our model is not overly sensitive to the particularities of the data in each subset, and its performance is more likely to generalize to new, unseen data.‏

**Task3:**

Using 5 Seeds and 20 epochs we get:

Minimum validation error: 0.00077234

Minimum validation error epoch: 6

Minimum validation error seed: 1

Test error for minimum validation error: 0.0006534

**Task4:**

Attached are 3 tables of the which corresponds to the best ­ for every combination.

**for learning rate = 0.001:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hidden\_size/  batch\_size | 10 | 100 | 500 | 1000 |
| 3 | **0.0006628** | 0.000716 | 0.00080542 | 0.0009623 |
| 20 | 0.0019211 | 0.0021345 | 0.002070248 | 0.0021009 |
| 100 | 0.0030899 | 0.0030671 | 0.0034900 | 0.0032776962 |
| 350 | 0.00440865 | 0.004342957 | 0.0047386 | 0.004722894 |
| 400 | 0.006848371 | 0.0066136 | 0.0069807 | 0.0067984 |

**for learning rate = 0.01:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hidden\_size/  batch\_size | 10 | 100 | 500 | 1000 |
| 3 | 0.0011414 | 0.0011294 | 0.0014453 | 0.0015187 |
| 20 | 0.0022848 | 0.00249221 | 0.002732829 | 0.00289702 |
| 100 | 0.003750661 | 0.0041062 | 0.0042510 | 0.003924456 |
| 350 | 0.005196747 | 0.004259754 | 0.0051234729 | 0.0059238827 |
| 400 | 0.00743455 | 0.00679217 | 0.0070944 | 0.007339676 |

**for learning rate = 0.1:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hidden\_size/  batch\_size | 10 | 100 | 500 | 1000 |
| 3 | 0.00187592 | 0.00169168 | 0.0017709 | 0.00182021 |
| 20 | 0.0026856 | 0.0026045054 | 0.003456465 | 0.0037977 |
| 100 | 0.0039530898 | 0.00427927 | 0.00433243 | 0.00397887 |
| 350 | 0.0050248 | 0.0055429 | 0.00624288 | 0.0068001 |
| 400 | 0.0080643 | 0.00899144 | 0.00768185 | 0.0088616 |

**Task5:**

תמונה שמכילה טבלה

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, מפה

התיאור נוצר באופן אוטומטי

Based on our results, it appears that the hidden features obtained from the first layer (zi) have resulted in better-defined clusters compared to the original input features (xi). This can be observed from the tSNE plot where the points in the zi plot are well-separated and do not overlap as much as the points in the xi plot.

The colors used to differentiate the points in the plot based on their labels, which are the different digits, also highlight the effectiveness of the learned model. The distinct colors in the zi plot indicate that the model has been able to identify and capture the unique features of each digit, resulting in well-separated clusters.

Based on the tSNE plots of the hidden features (zi) and the input features (xi), it appears that the learned model has been able to effectively extract the relevant features from the input images and use them to differentiate between the different digits. The well-separated clusters in the zi plot indicate that the learned model has successfully learned to encode the input images into a more compact representation.