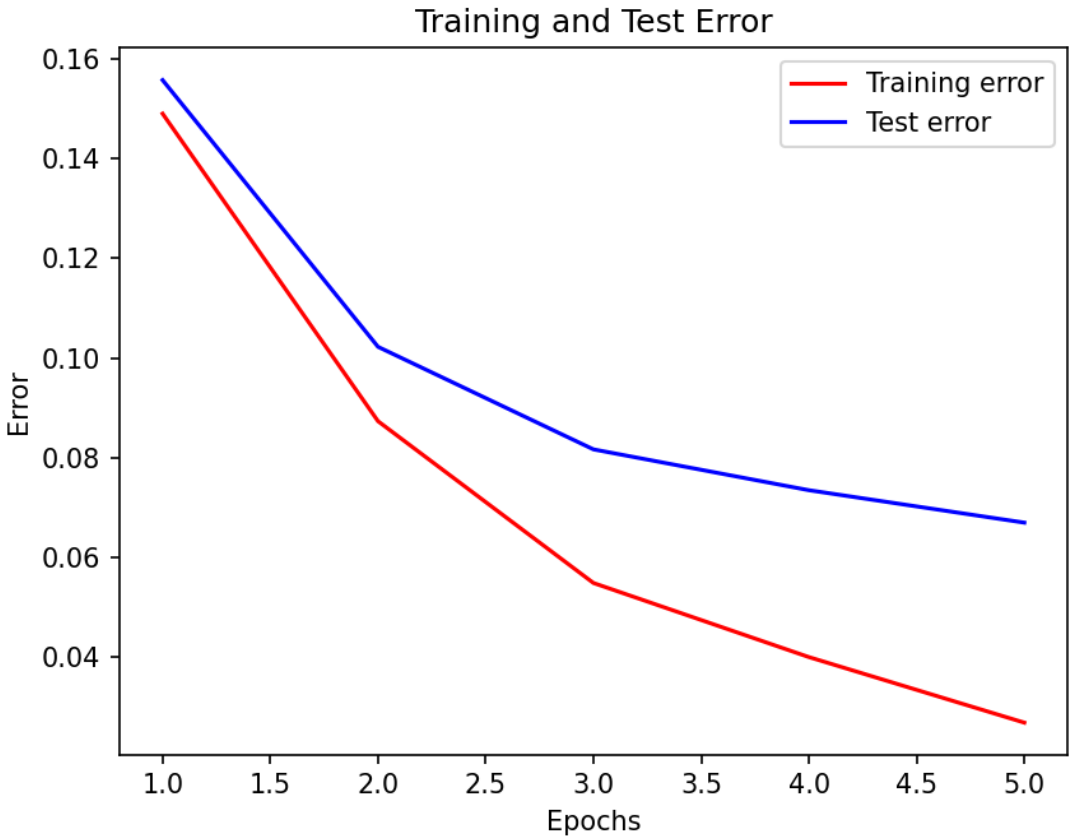
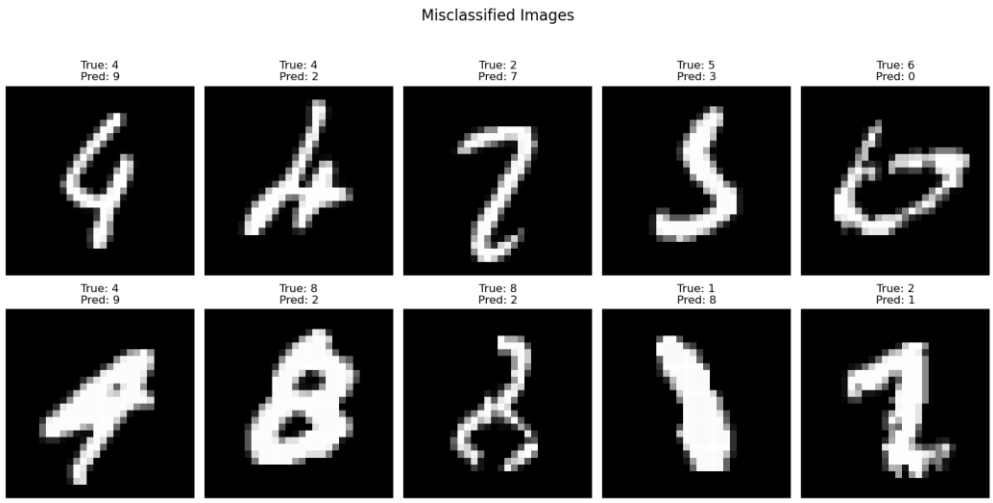
Task1

Train and Test plot:

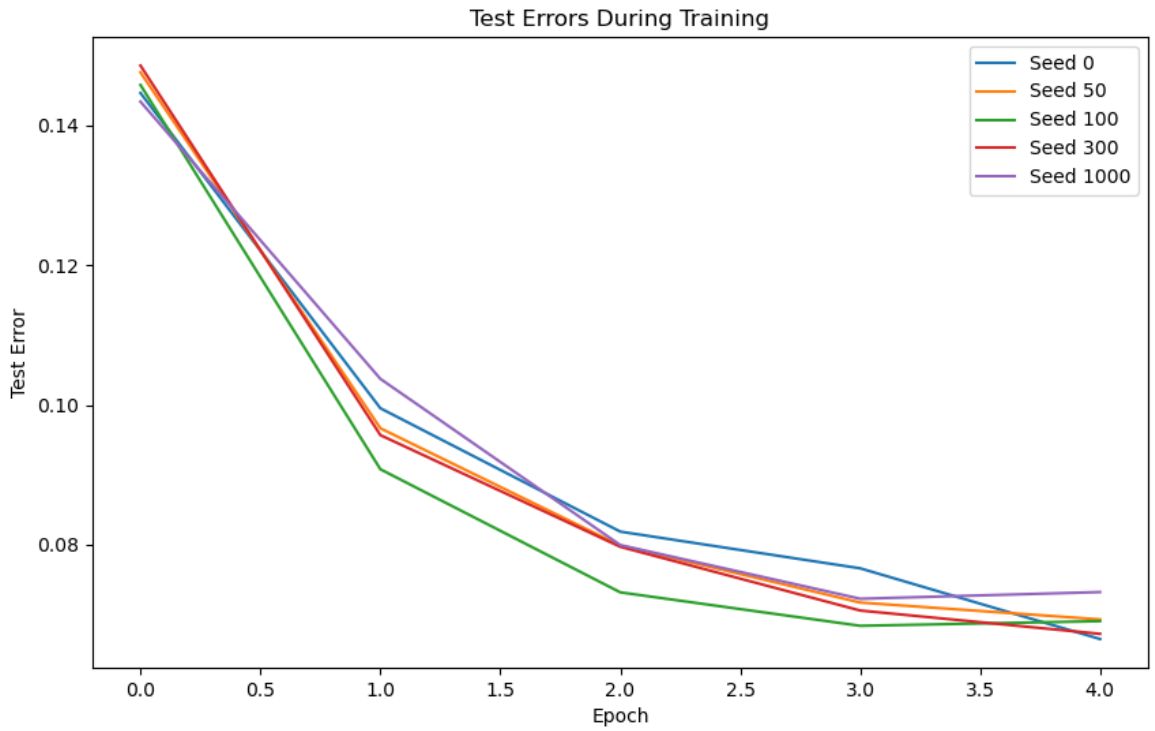


Test error: 0.06696

Misclassified images:



Task2



Mean final test error: 0.0691

Standard deviation of final test error: 0.0023

The computed variance is very small, indicating that the model is robust to the choice of seed number. This robustness is also evident in the plot, where the results remain consistent and similar across different seed values.

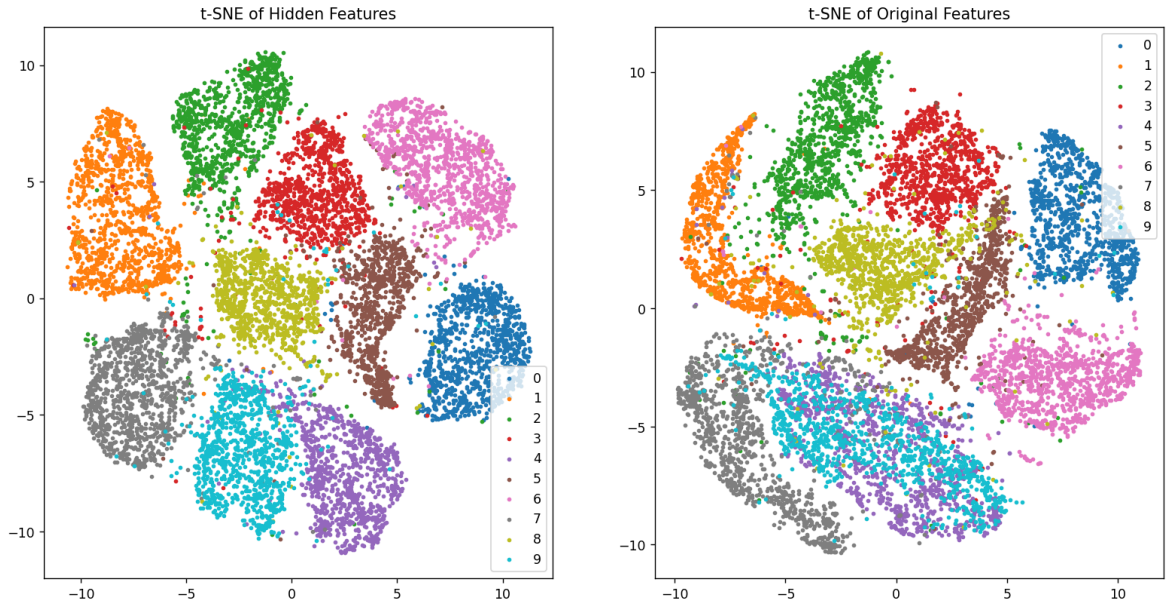
Task3

seed	Minimum e_validation	Correspond e_test
0	0.0816	0.0741
50	0.0882	0.0811
100	0.0741	0.0728
300	0.0745	0.0743
1000	0.0800	0.0729

Task4

Batch size	Hidden size	Learning rate	Error test
50	500	0.0001	0.170339
50	750	0.0001	0.150730
50	500	1e-05	0.417958
50	750	1e-05	0.382232
100	500	0.0001	0.208526
100	750	0.0001	0.187195
100	500	1e-05	0.567590
100	750	1e-05	0.500939
200	500	0.0001	0.251734
200	750	0.0001	0.228616
200	500	1e-05	0.871929
200	750	1e-05	0.753530

Task5



In the plot of the original features, the colours representing the numbers are mixed together, with no clear separation between them—for instance, the purple (4) and light blue (9) are entirely combined. Conversely, in the plot of the hidden features, we observe a much clearer separation of colours. While not perfect, the separation is significantly improved.

The learned model yields good results, with distinct separation for most digits. However, for digits with similar shapes, such as 4 and 9, there remains some overlap. In addition, for some cases there were a misconception of the digit itself, therefore the dots, which represent the digits, in different place then we have predicted them to be.