Michael Cruz

November 17, 2023

Programming Assignment 4 Writeup

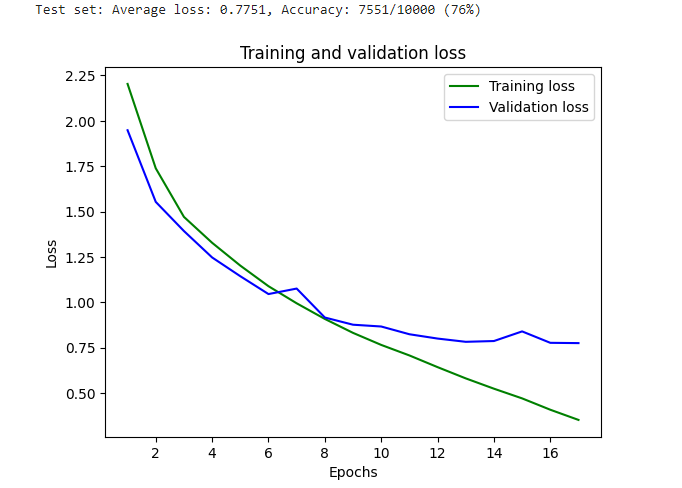
Model 1:

Architecture:

* Convolutional layer 1: 32 3x3 kernels
* Convolutional layer 2: 64 3x3 kernels
* Convolutional layer 3: 128 3x3 kernels
* Fully connected layer 1: 1024 nodes
* Fully connected layer 2: 512 nodes
* Fully connected layer 3: 10 nodes
* Training :
  + The model was trained for 30 epochs with early stopping, which caused the model to stop training at 16 epochs
  + Learning rate: 0.001
  + momentum : 0.09

Analysis:

The first model performed worse than the second model by one percentage in accuracy, however the loss curve for this model indicated that the model is better generalizable and is closer to the global minimum.



Model 2:

Architecture:

* Convolutional layer 1: 32 3x3 kernels
* Convolutional layer 2: 64 3x3 kernels
* Convolutional layer 3: 128 3x3 kernels
* Convolutional layer 4: 512 3x3 kernels
* Fully connected layer 1: 1024 nodes
* Fully connected layer 2: 512 nodes
* Fully connected layer 3: 256 nodes
* Fully connected layer 4: 10 nodes
* Training :
  + The model was trained for 30 epochs with early stopping, which caused the model to stop training at 20 epochs
  + Learning rate: 0.001
  + momentum : 0.09

Analysis:

While empirically this model performed better in terms of accuracy, the loss curve produced by the model is slightly less concave than the previously model. Additionally, the two losses are further apart towards the end, which could be an indication of overfitting.

