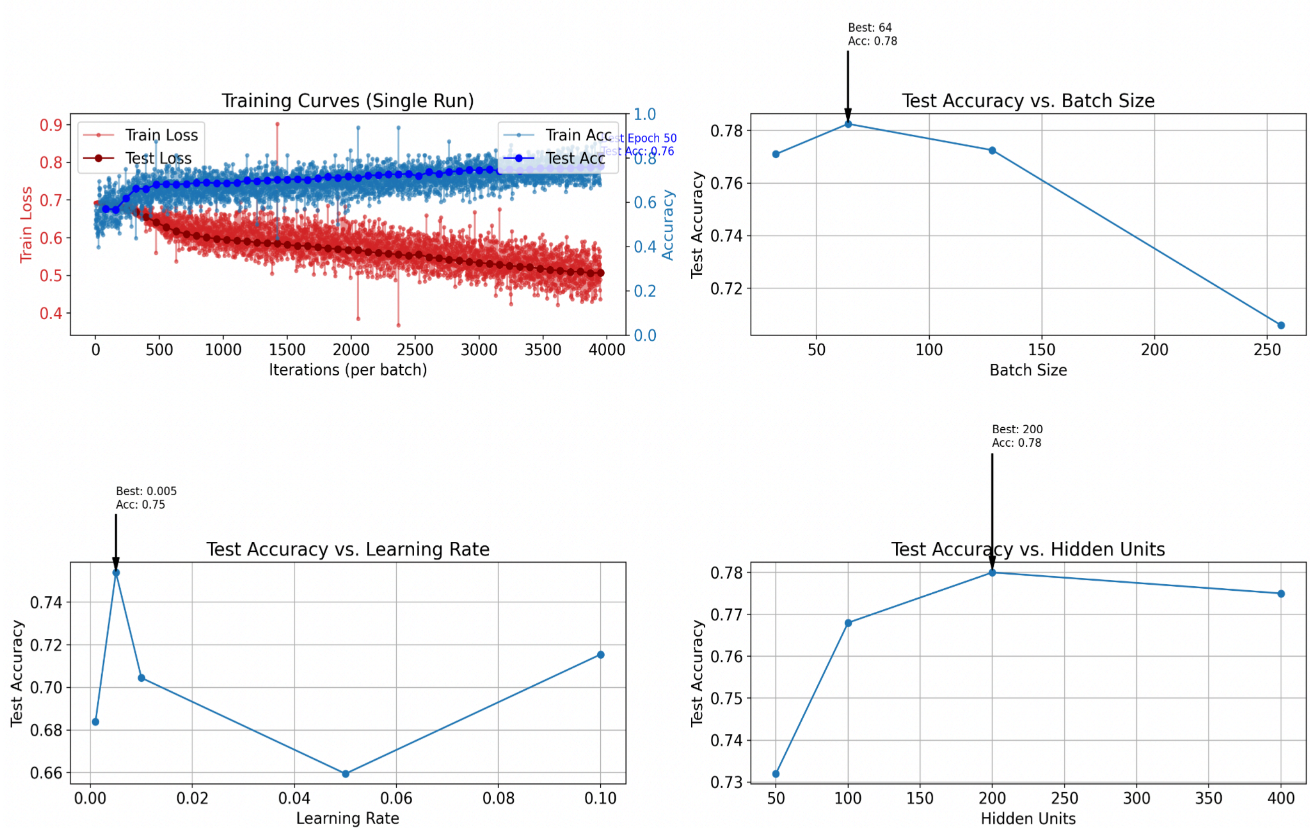
Assignment #2: CIFAR Image Classification using Fully-Connected Network



1. Training Curves

From the top-left chart, both the training loss (light red) and the test loss (dark red) decrease steadily, showing the model's progression toward convergence. The training and test accuracies (light blue and dark blue, respectively) gradually increase, with the final test accuracy reaching around 0.76 near the 4000th iteration. Despite some fluctuations—likely due to mini-batch sampling—the overall trend is positive, indicating that the model learns effectively from the training data while also generalizing relatively well to the test set.

2. Effect of Batch Size on Test Accuracy

In the top-right chart, the test accuracy peaks at approximately 0.78 when the batch size is 64, as highlighted by the “Best: 64” annotation. As the batch size grows larger (e.g., 100, 150, or 250), the test accuracy decreases. This pattern suggests that a smaller to medium batch size (around 64) may provide more frequent gradient updates and better generalization for this particular task.

3. Effect of Learning Rate on Test Accuracy

The bottom-left chart shows that the best performance appears at a learning rate of 0.005, with a test accuracy of 0.75, indicated by the “Best: 0.005” label. When the learning rate is either much lower or higher than this sweet spot—such as 0.001 or 0.04 and beyond—test accuracy dips to around 0.70 or even lower. This illustrates how excessively small learning rates can slow convergence, while excessively large ones can destabilize training.

4. Effect of Hidden Units on Test Accuracy

From the bottom-right chart, using 200 hidden units gives the highest test accuracy of about 0.78, as noted by the “Best: 200” marker. Below or above that number of hidden units (e.g., 50, 100, 300, 400), the model’s test accuracy remains slightly lower (ranging from about 0.73 to 0.77). This suggests that while increasing the hidden layer size can help capture more complex representations, there is an optimal network capacity for this problem—too few units underfit, whereas too many can lead to overfitting or difficult optimization.