Put Your Code Results Here:

Question(20%):

1. **Which optimizer gave you the best test accuracy? Why do you think it performed better than the others?**

My Momentum optimizer got the best accuracy. I think it worked better because it remembers past updates, so it smooths out the changes and helps the model learn faster and more steadily than the SGD or even Adam.

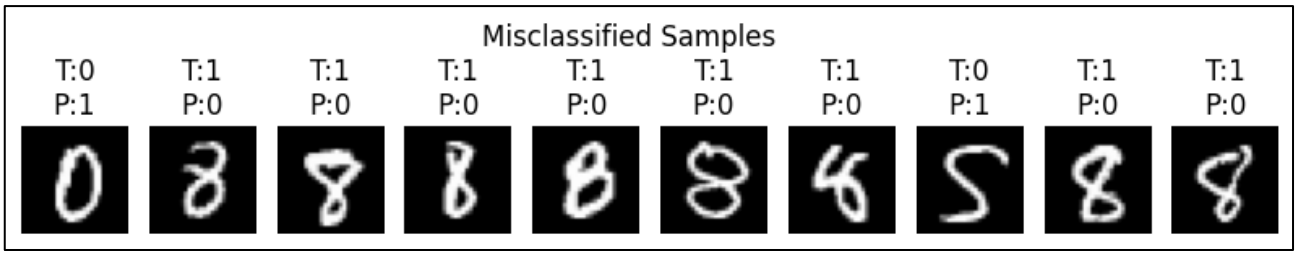
2. **What is the differences in learning stability, convergence speed, or misclassification types across all algorithm? Please explain with examples or observation from your results.**

I notice that Momentum is the most smoothest – its loss went down steadily without too many ups and downs. Adam was pretty good but sometimes wobbled a bit. And the SGD was the worst; it jumped around a lot, which might made it miss some patterns and misclassify more images.

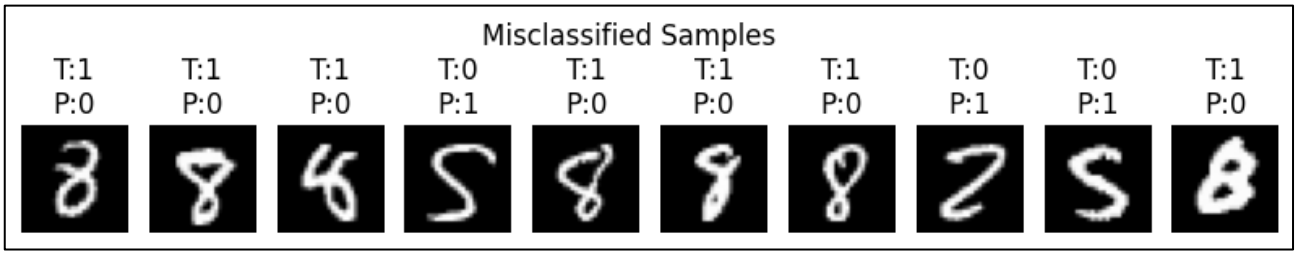
3. **How did your choice of learning rate, batch size, or momentum affect each optimizer? What values worked best in your experiments?**

In my Lab 4, a learning rate of about 0.01 worked best for Momentum, with a batch size of 64 helping to keep things balanced. Also, setting the momentum to 0.9 made it really smooth. In comparison, using the same settings in SGD made it too jumpy, and Adam needed a lower learning rate, like 0.001, to work well.

Mini Batch SGD: Algorithm 7.2：Test Accuracy: 95.16%



Mini Batch SGD with Momentum: Algorithm 7.3：Test Accuracy: 96.16%



Adam Optimizer: Algorithm 7.4：Test Accuracy: 96.06%