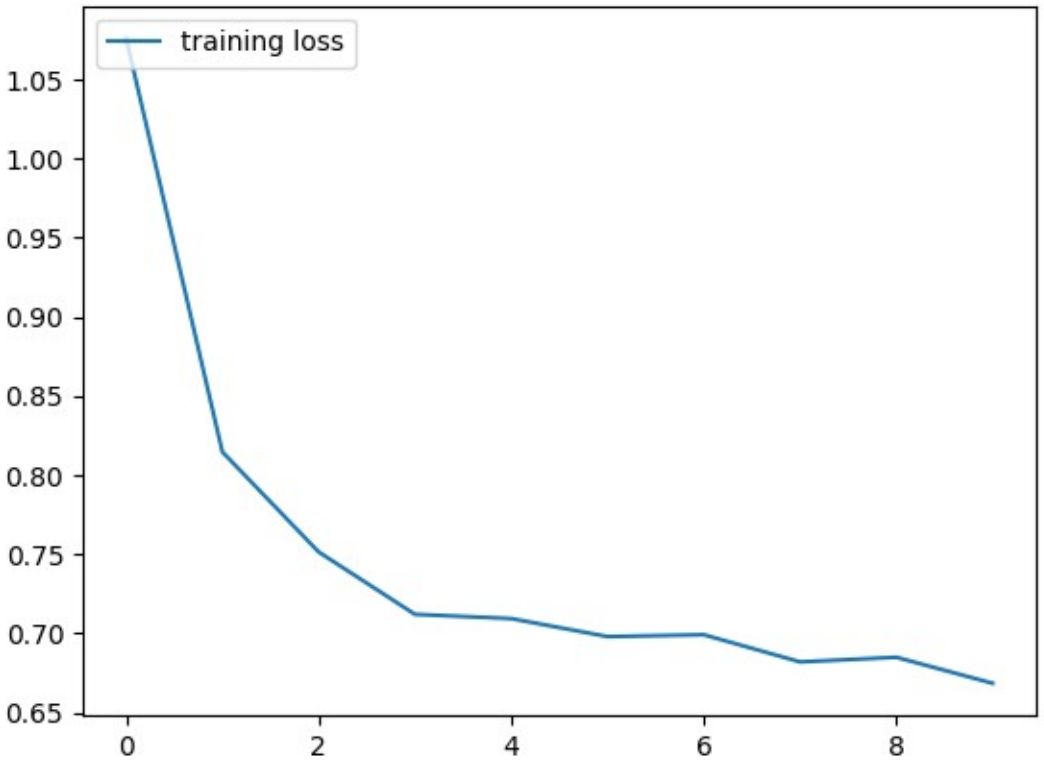
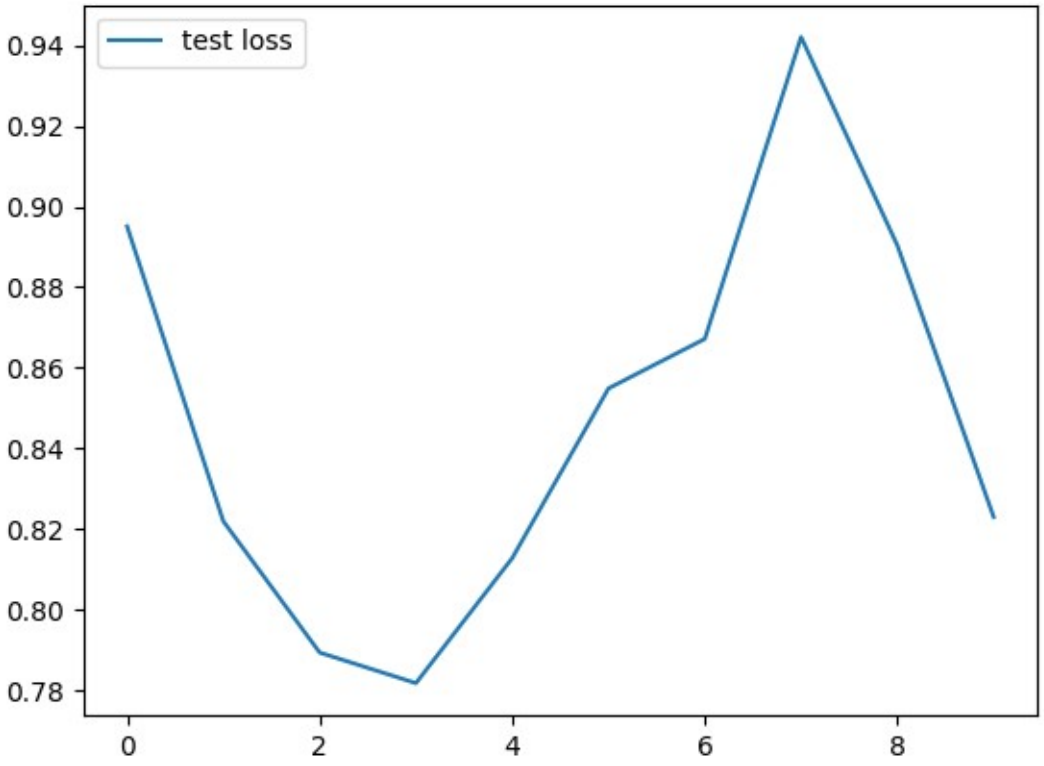
Report:

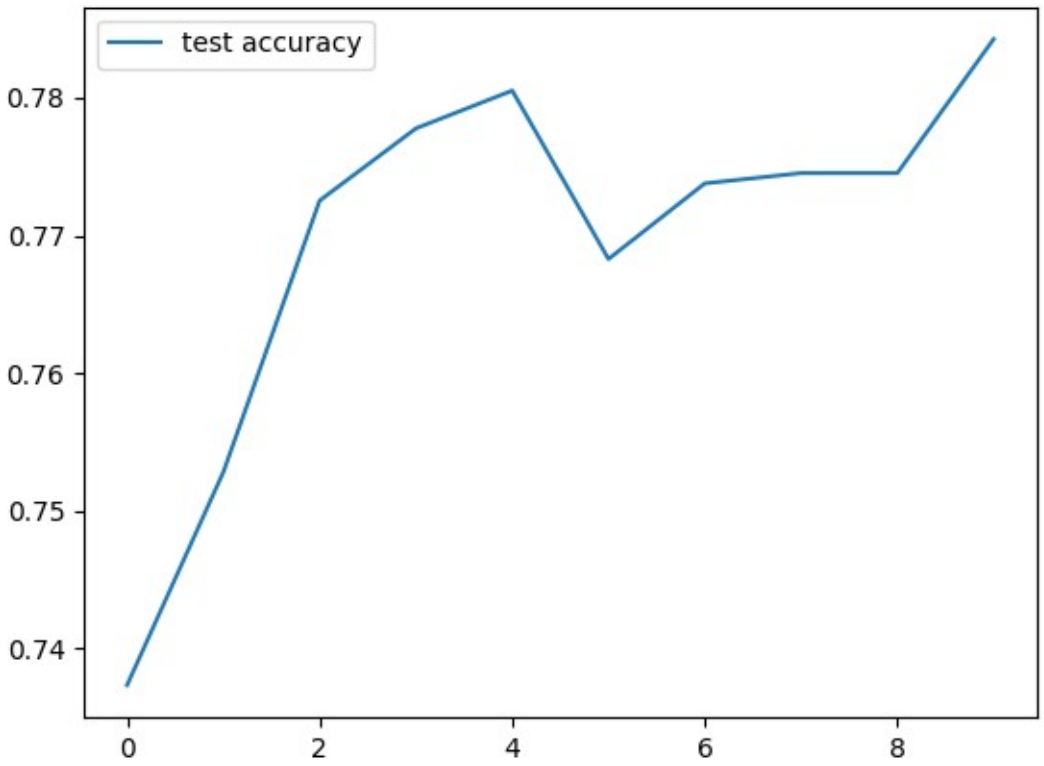
Adam as optimizer: I achieved lower loss and a higher test accuracy with batch size of 30 followed by batch size of 10 and 1. The different batch sizes are tested with hidden unit of 100, but I also tested batch size of 30 with hidden unit of 50 and 250.
SGD as optimizer: I achieved lower loss with batch size of 30 followed by batch size of 10 and 1. However, for batch size of 2 and above, my model failed to learn and only constantly predict the majority class. I suspect that the model has fallen into a local optima and failed to get out.

Adam optimizer:

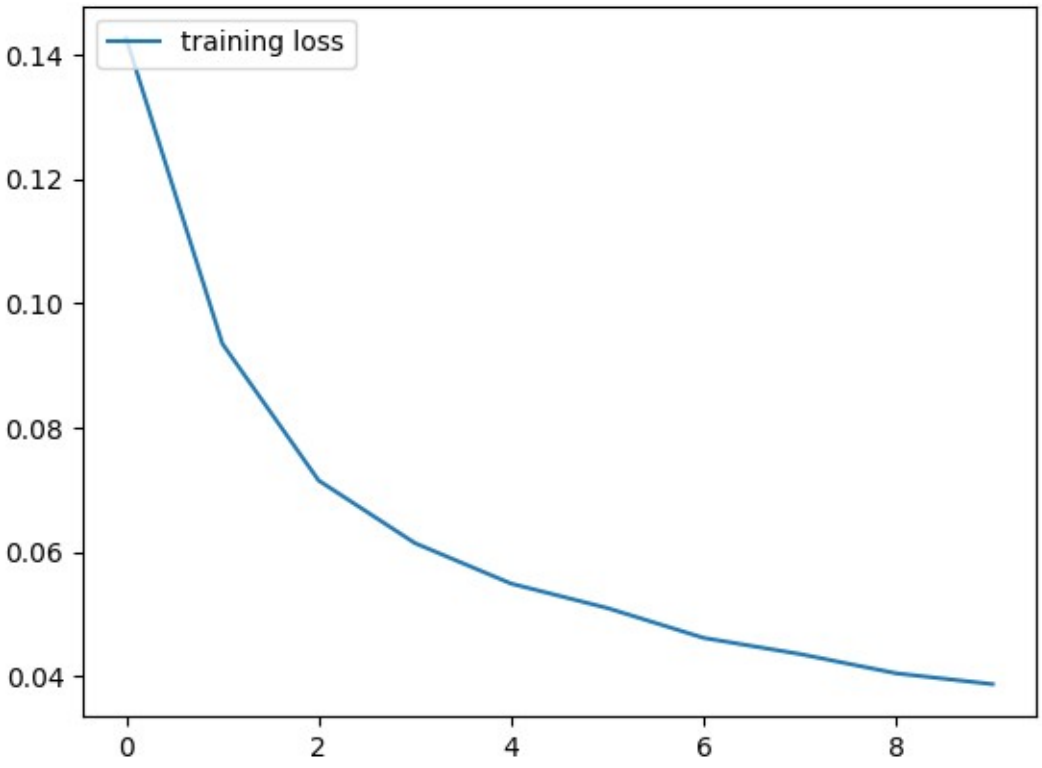
Batch size 1:

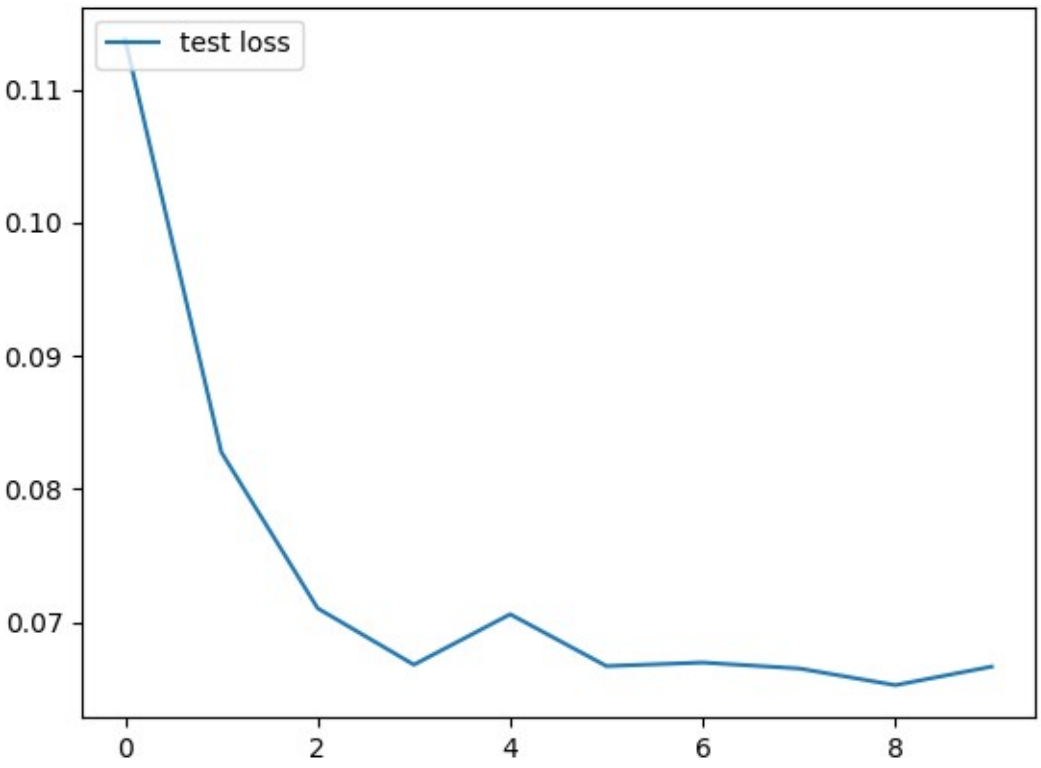


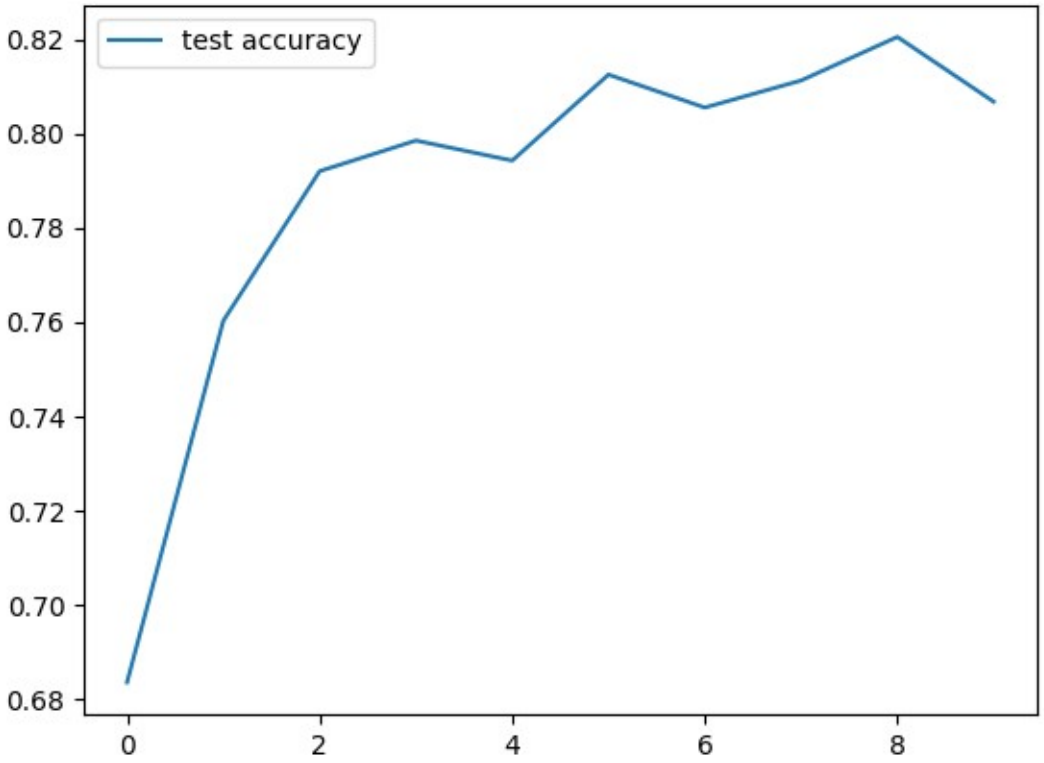

test loss
0.94
0.92
0.90
0.88
0.86
0.84
0.82
0.80
0.78
0
2
4
6
8


test accuracy
0.78
0.77
0.76
0.75
0.74
0
2
4
6
8

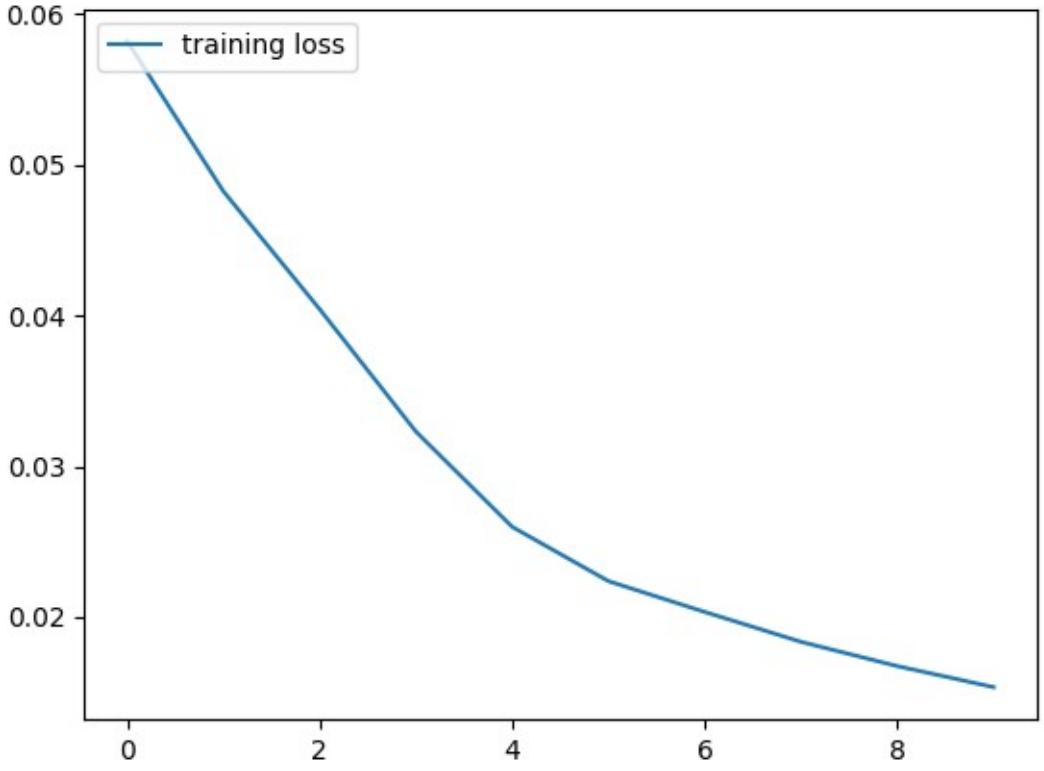
Batch size 10:

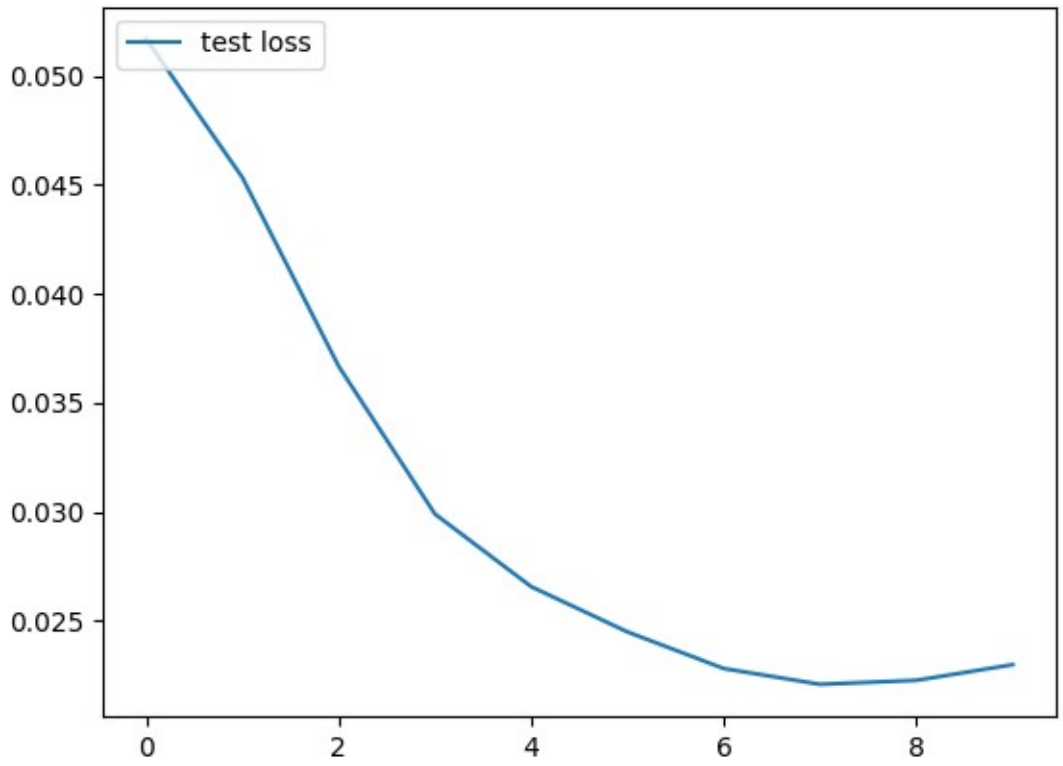


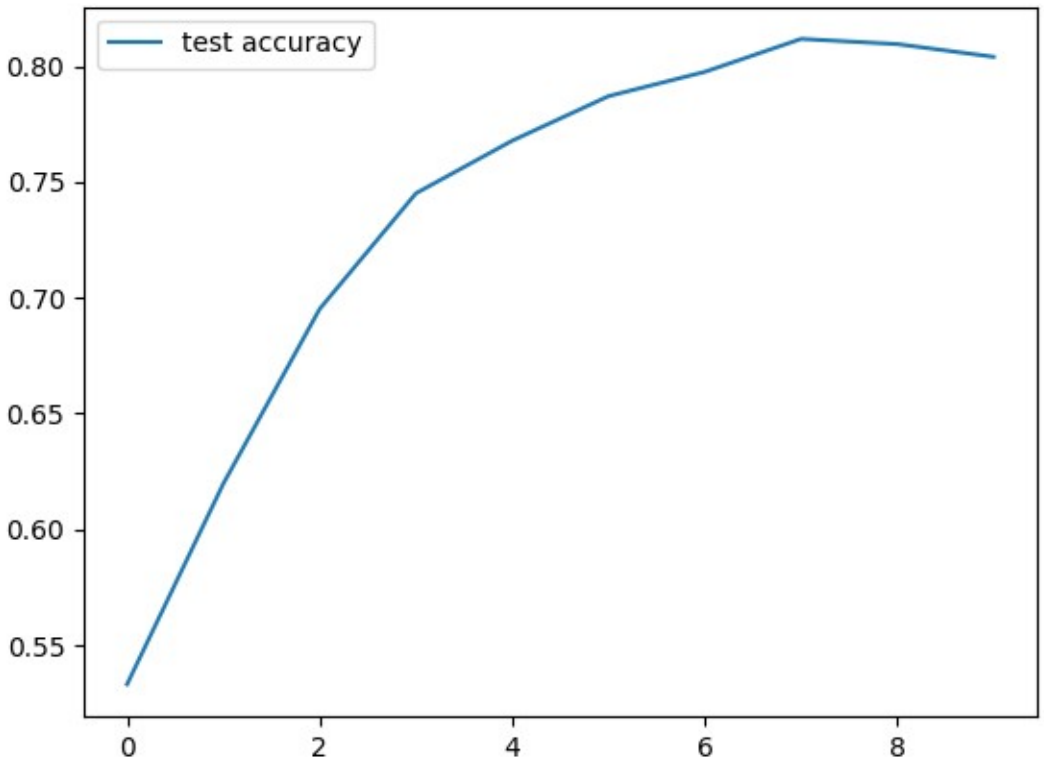




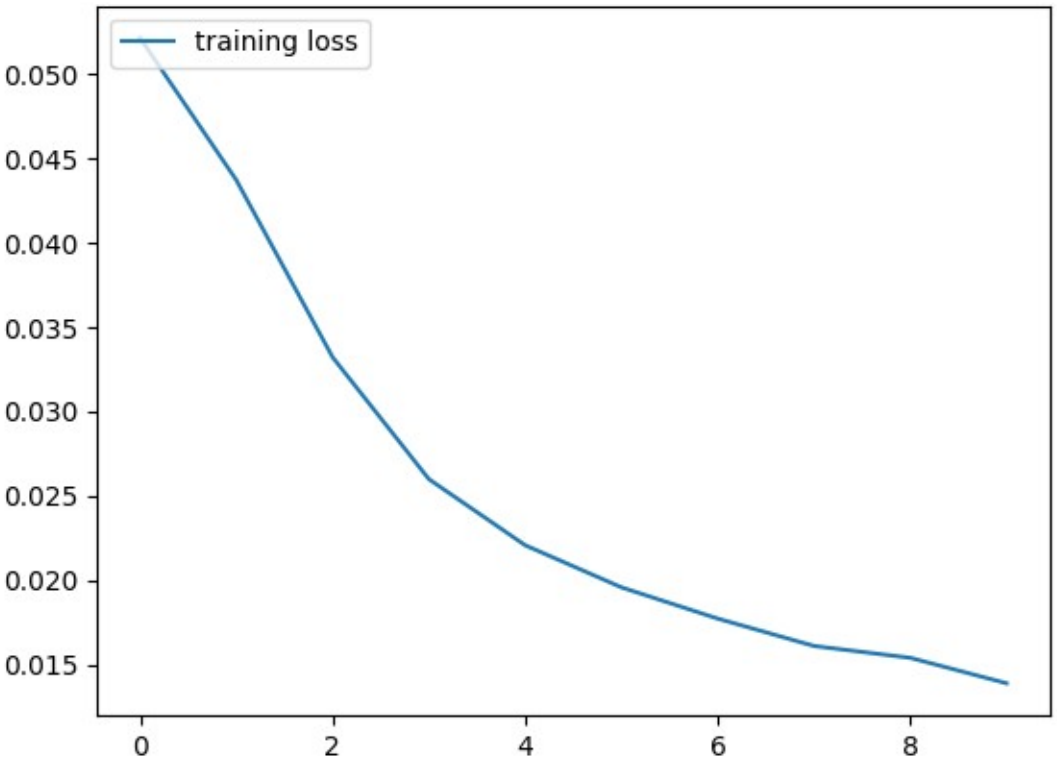
Batch size 30:

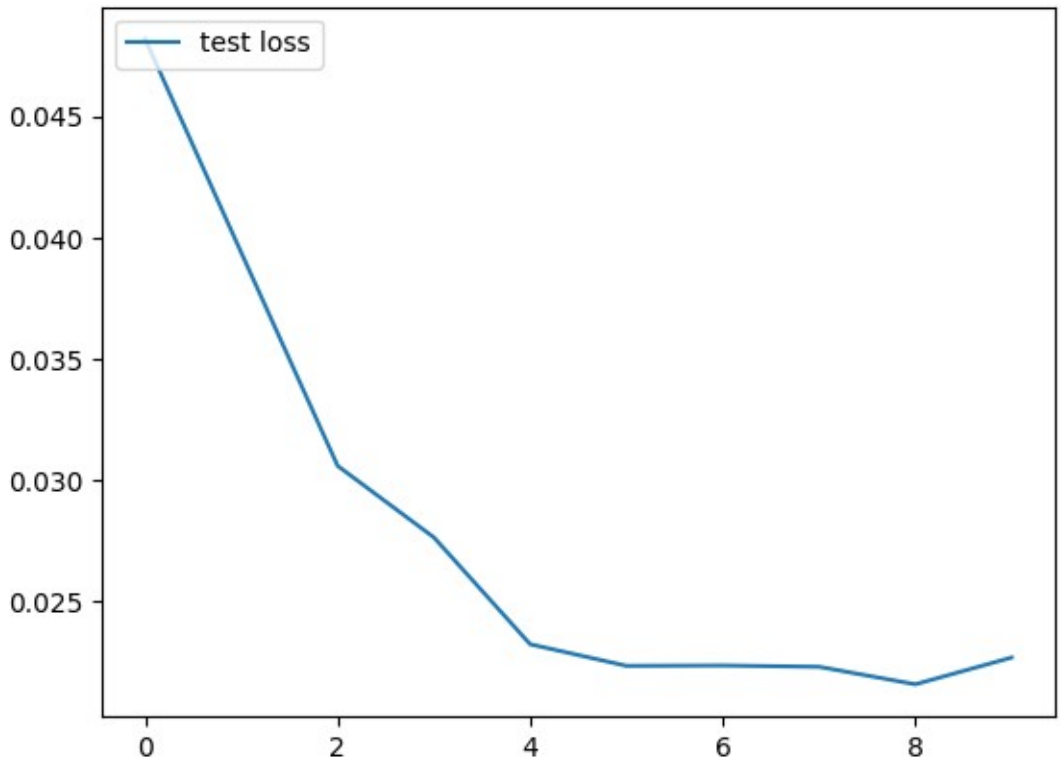


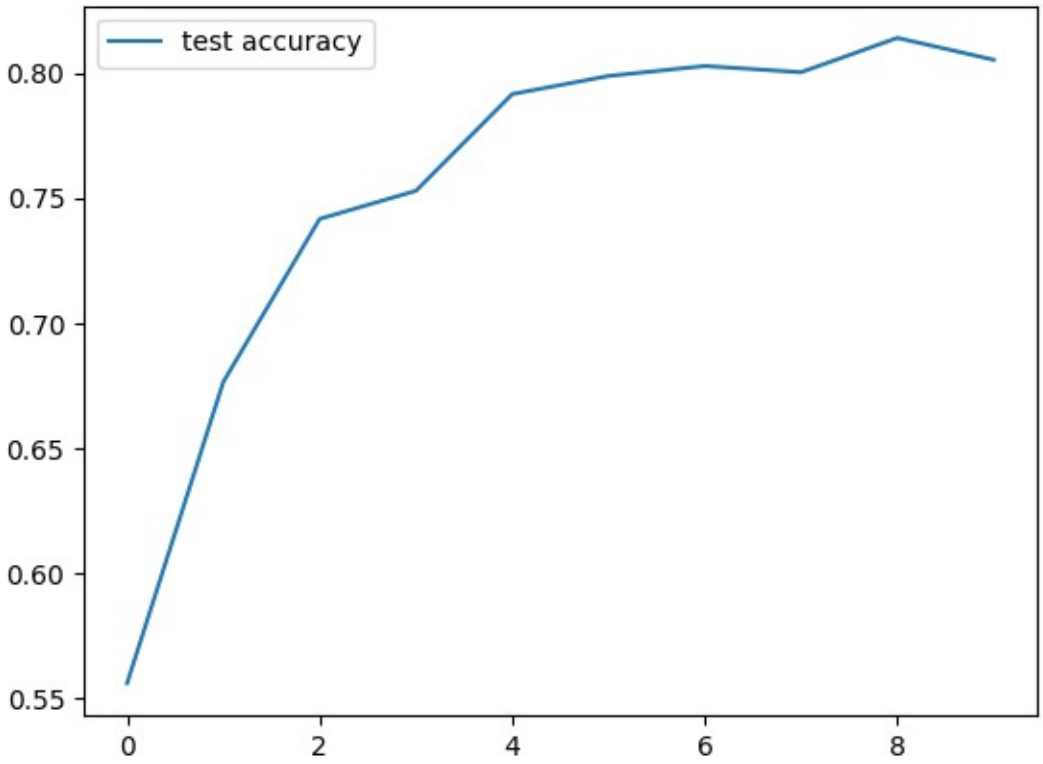

test loss
0.050
0.045
0.040
0.035
0.030
0.025
0
2
4
6
8


test accuracy
0.80
0.75
0.70
0.65
0.60
0.55
0
2
4
6
8

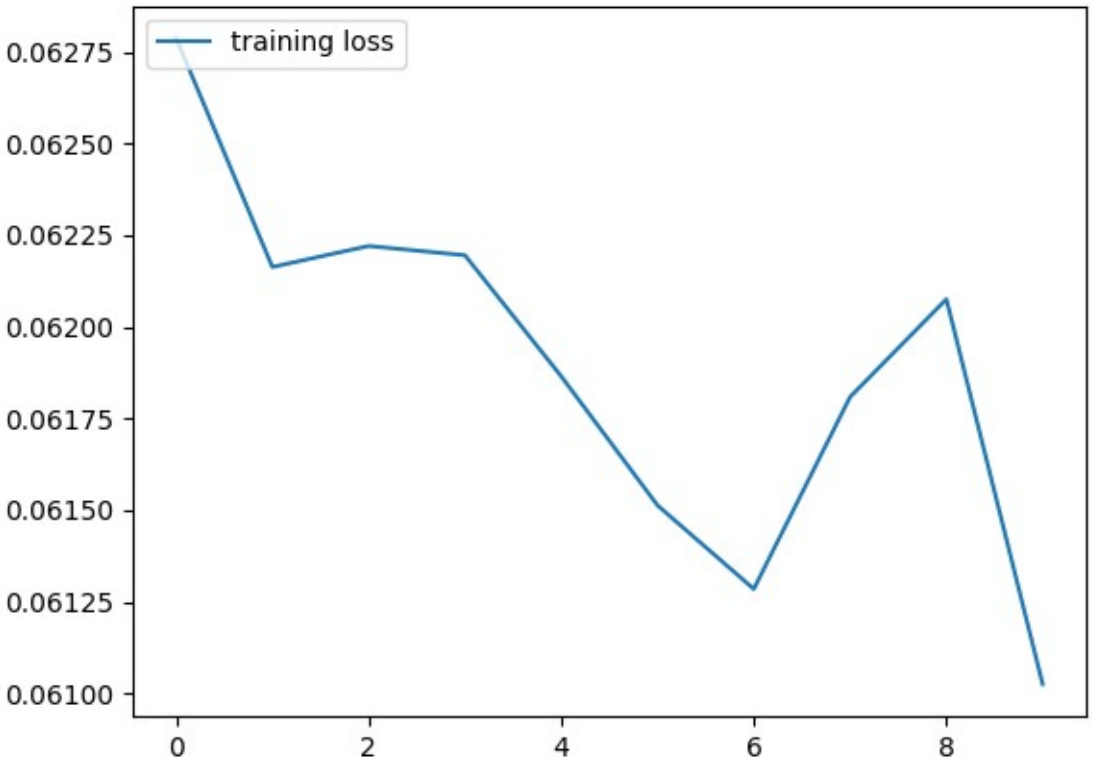
Batch size 30, 50 hidden units

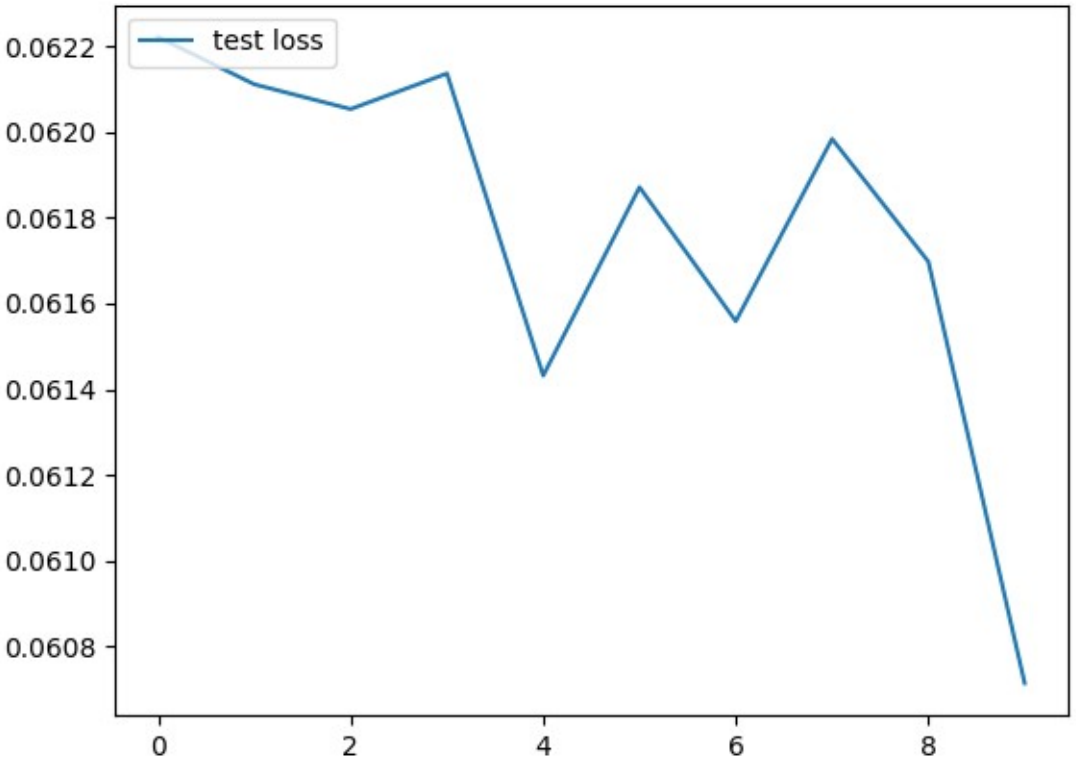


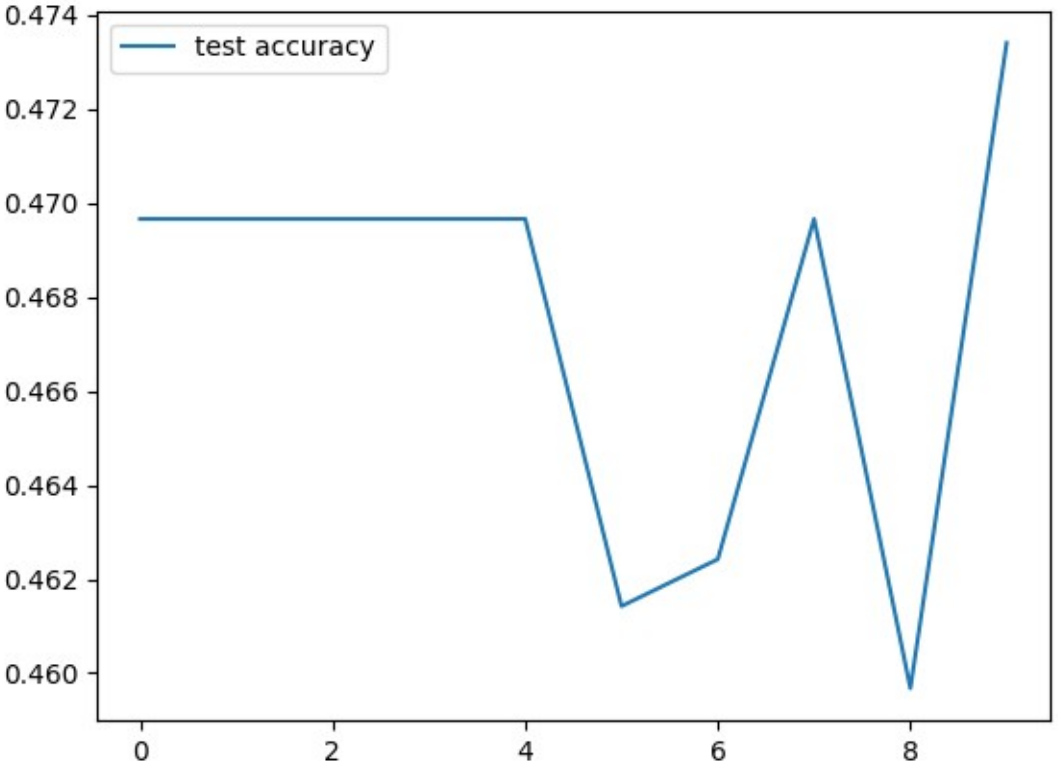




Batch size 30, 250 hidden units

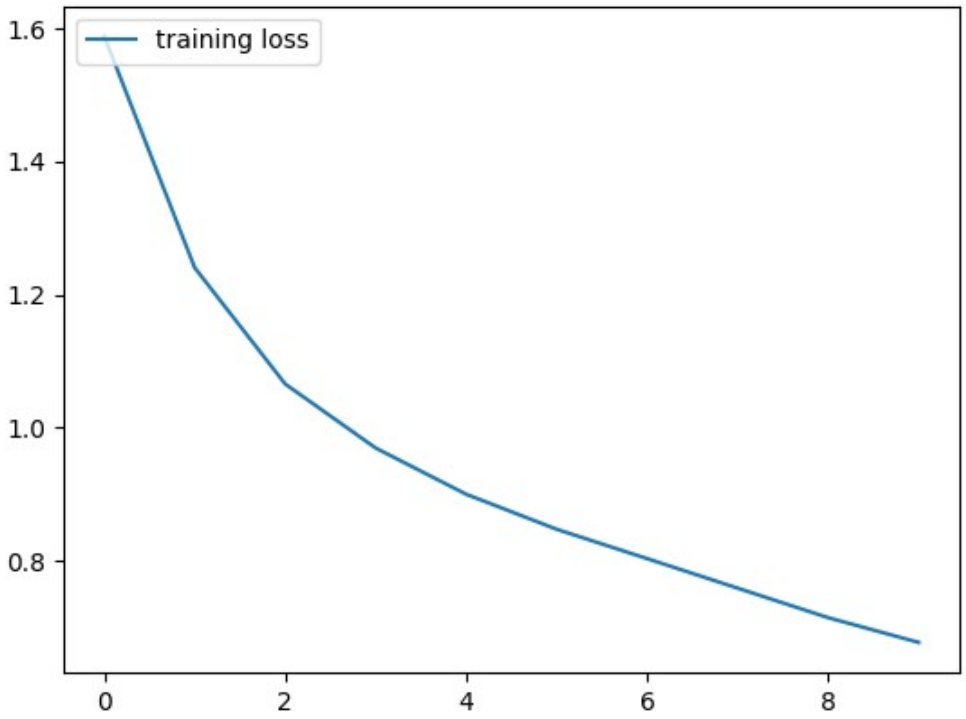


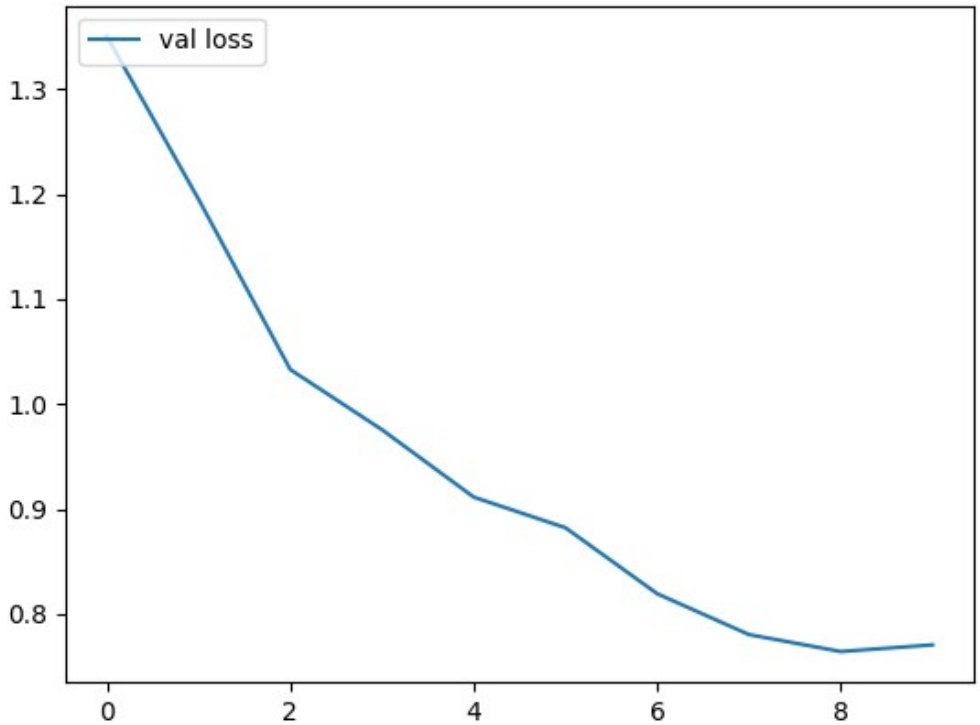

test loss
0.0622
0.0620
0.0618
0.0616
0.0614
0.0612
0.0610
0.0608
0
2
4
6
8

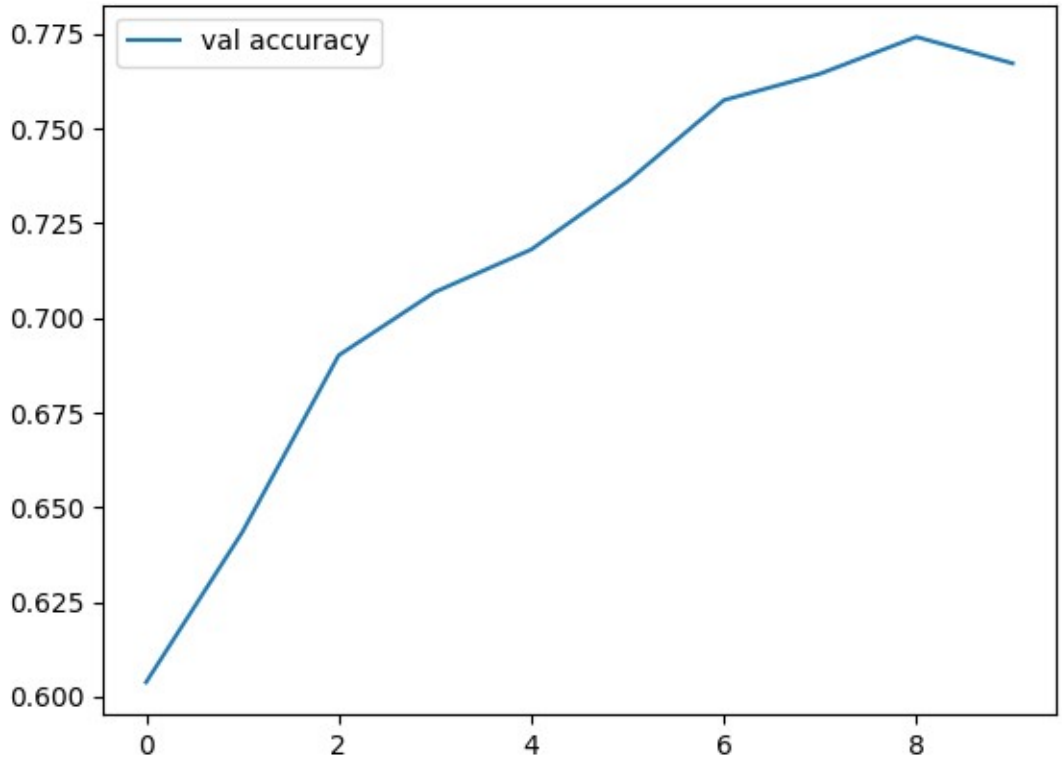

test accuracy
0.474
0.472
0.470
0.468
0.466
0.464
0.462
0.460
0
2
4
6
8

SGD Optimizer:

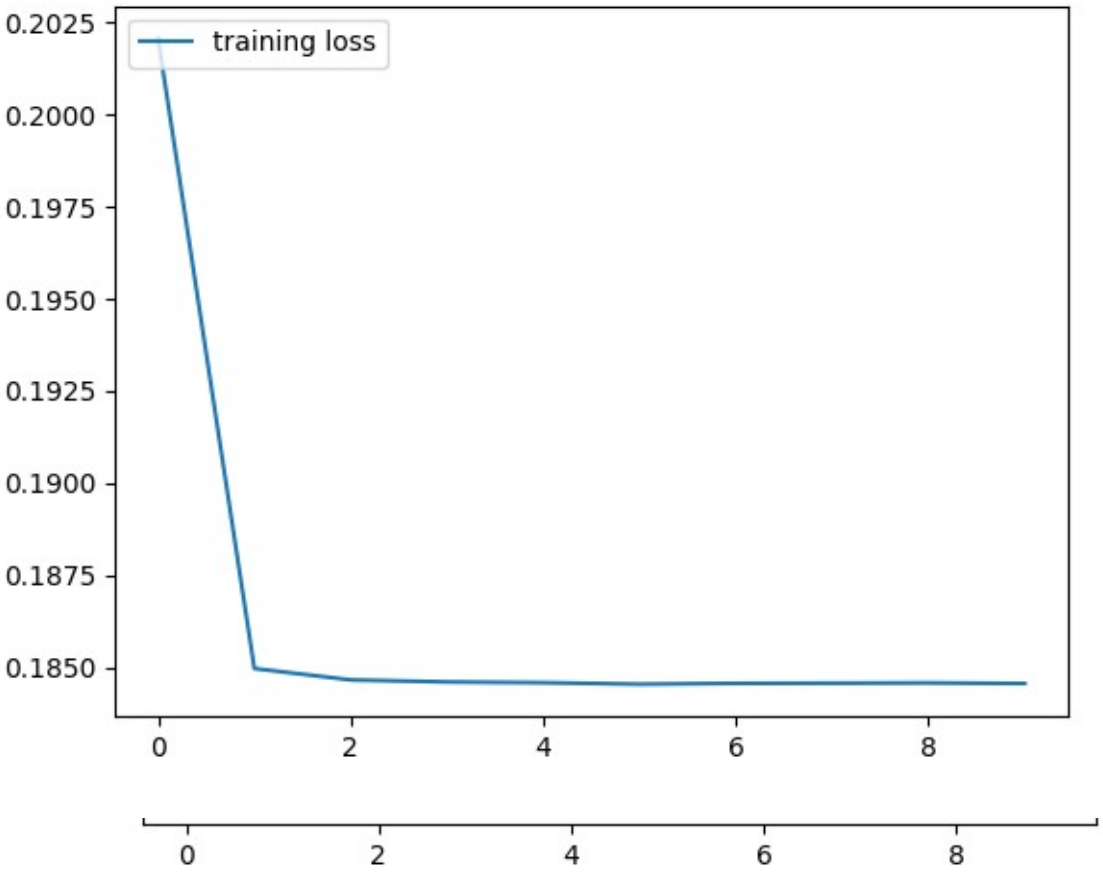
Batch size 1:

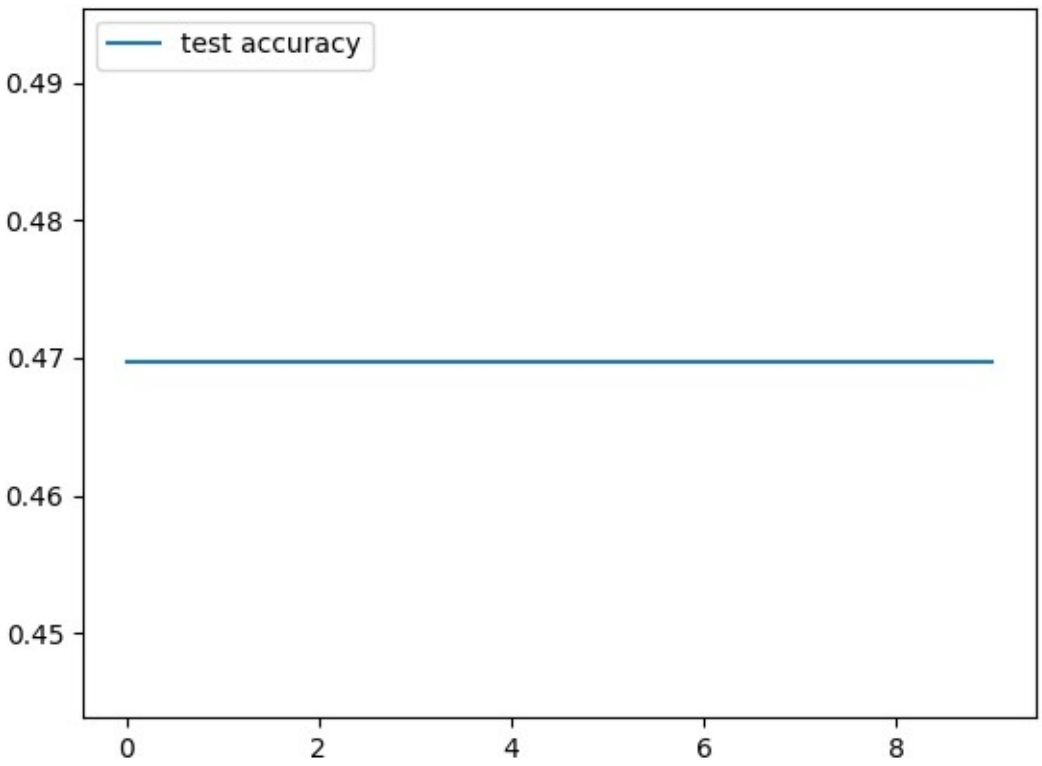




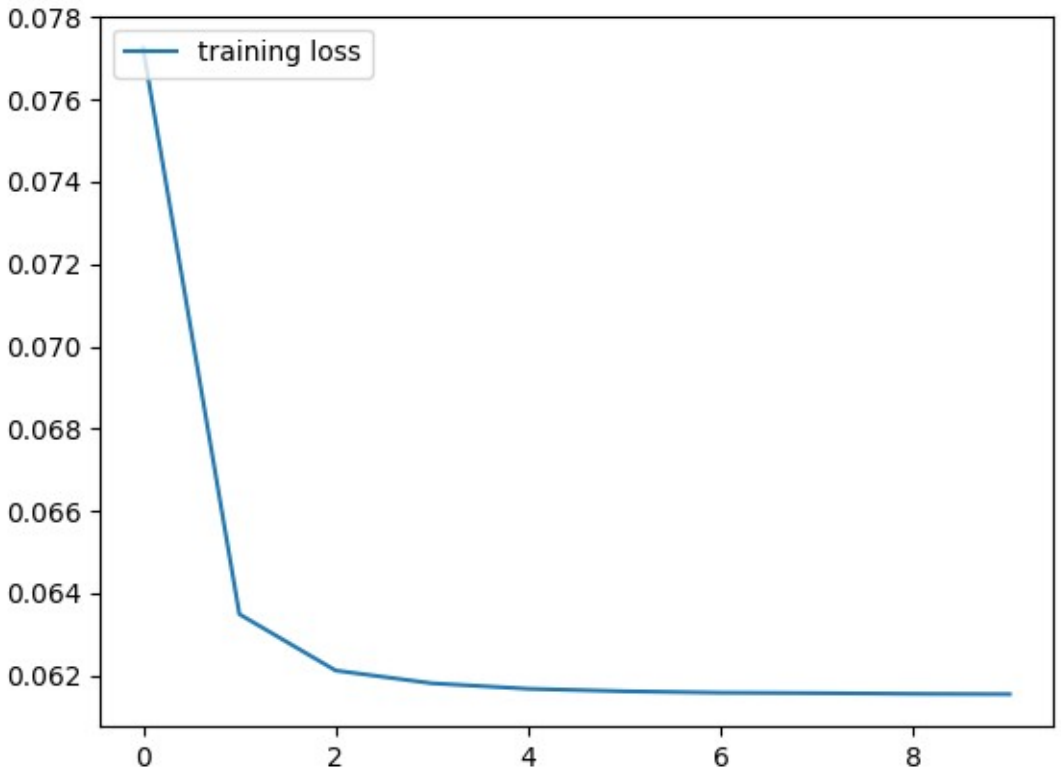


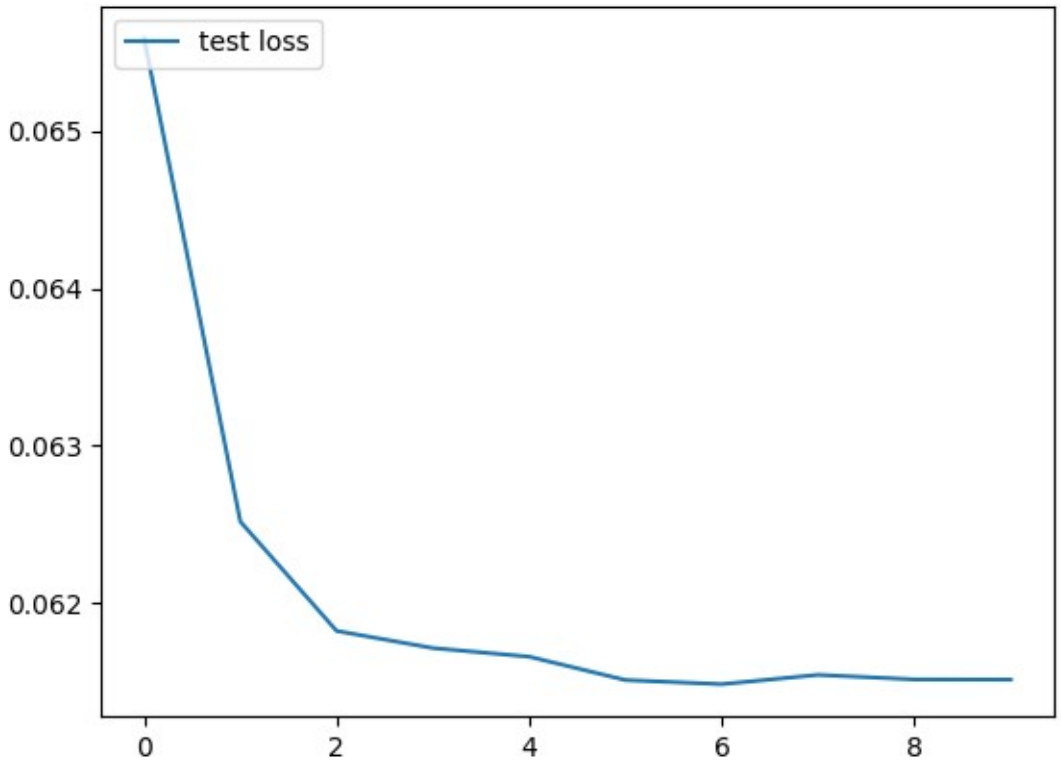
Batch size: 10

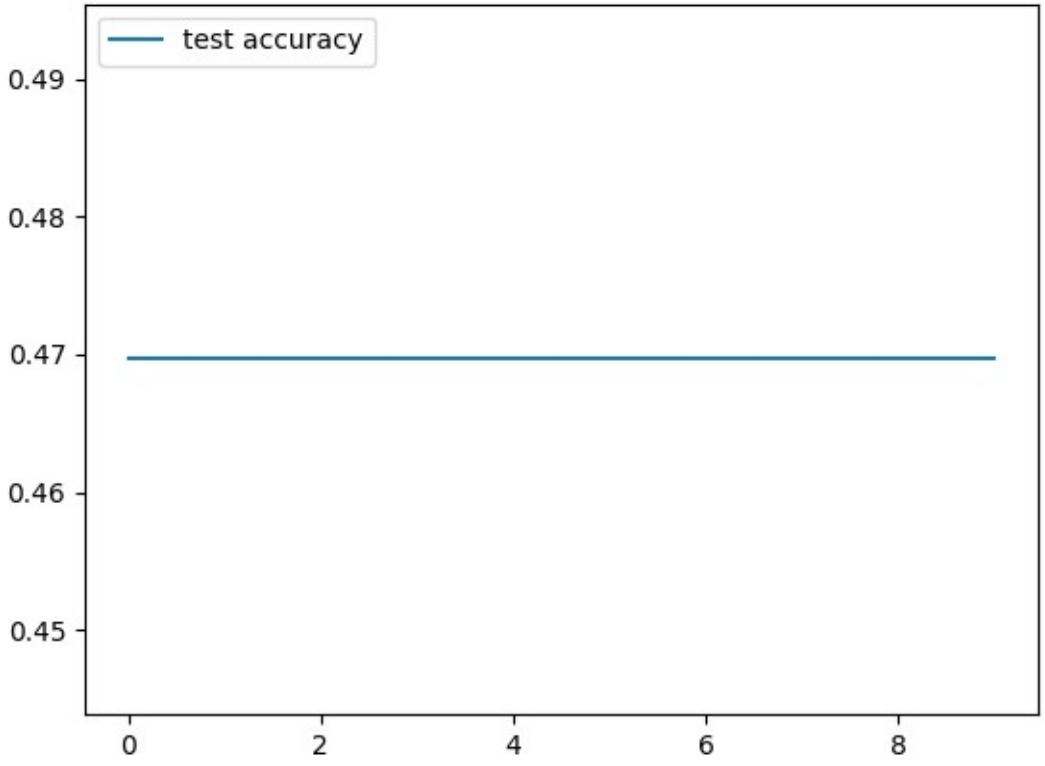

training loss
0.2025
0.2000
0.1975
0.1950
0.1925
0.1900
0.1875
0.1850
0
2
4
6
8
0
2
4
6
8


test accuracy
0.49
0.48
0.47
0.46
0.45
0
2
4
6
8

Batch size: 30






test accuracy
0.49
0.48
0.47
0.46
0.45
0
2
4
6
8