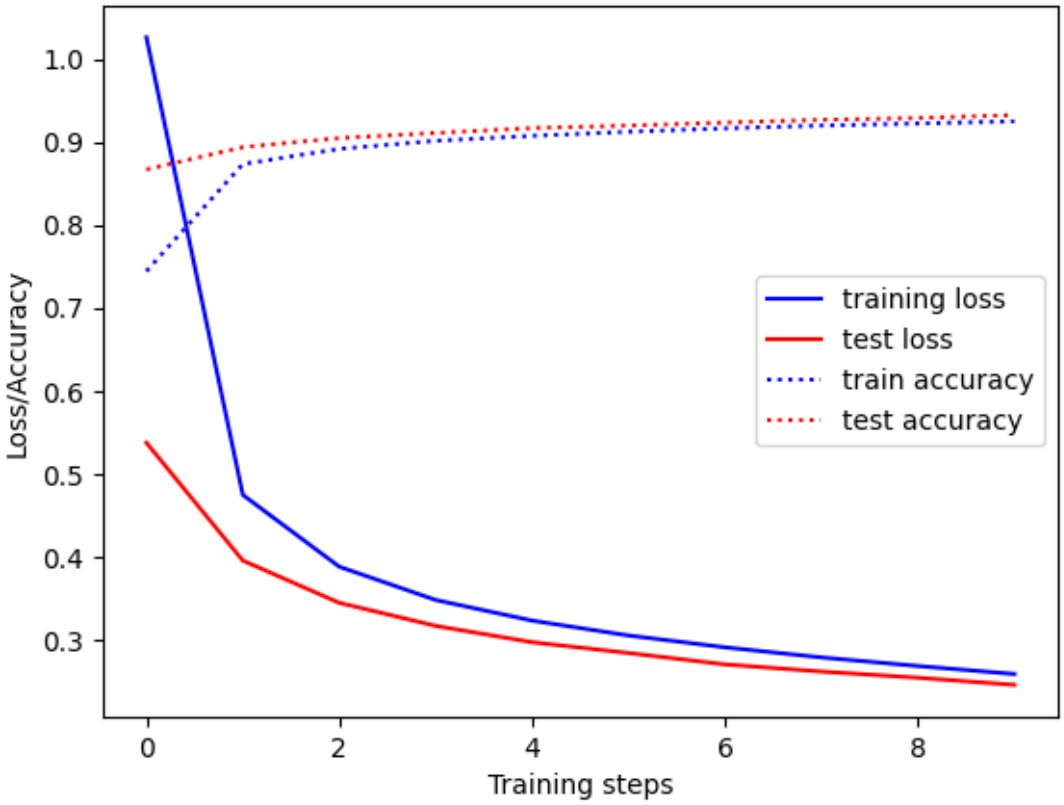
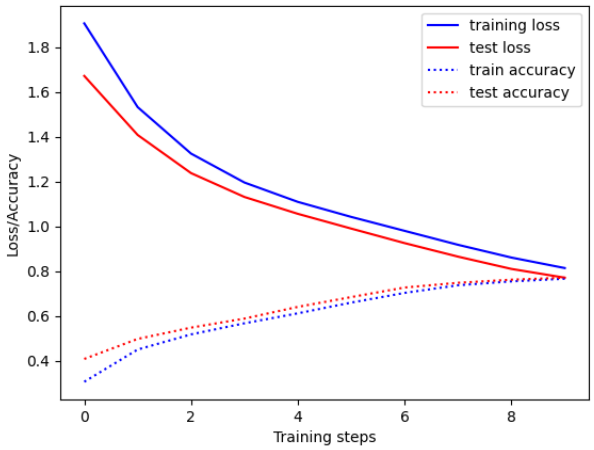
# Variations

- **Original (2 hidden layers, 256 units each, learning rate=0.01, momentum 0.0, batch size = 32)**



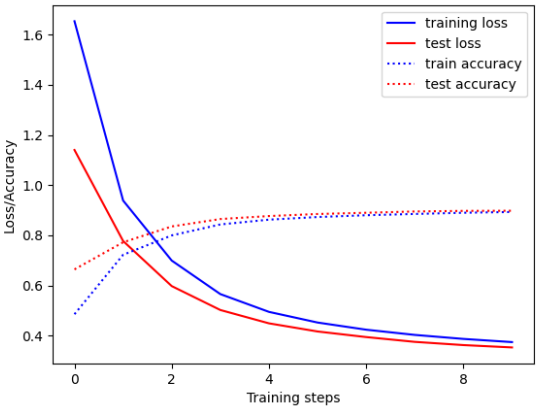

  - Training loop took:1894.734375 seconds

- **Only one hidden layer with 4 units (rest remains the same)**



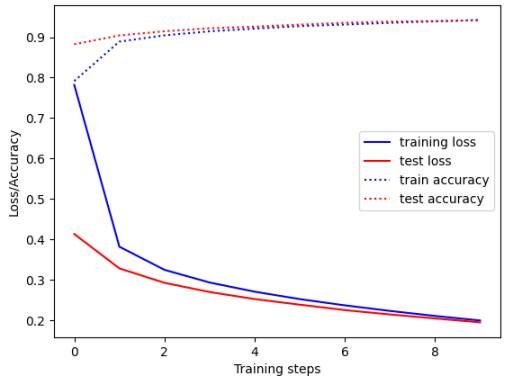

  - performs worse but accuracy rises to 0.77 which is quite good
  - train& test loss do not decrease as much
  - Training loop took:1604.5 seconds -> faster

- **Only one hidden layer with 12 units**



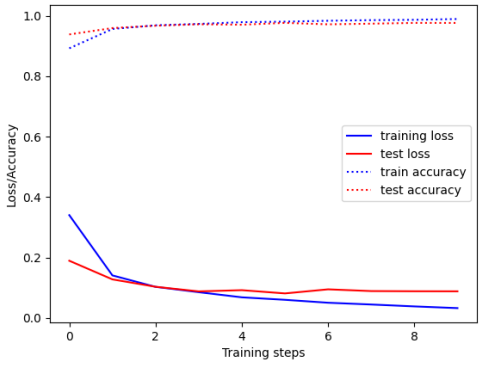

  - accuracy really good 0.89
  - loss functions look good
  - Training loop took:1625.453125 seconds -> faster
  - interesting that we can achieve comparable results to the original with only 12 and one single layer

- **momentum 0.5**



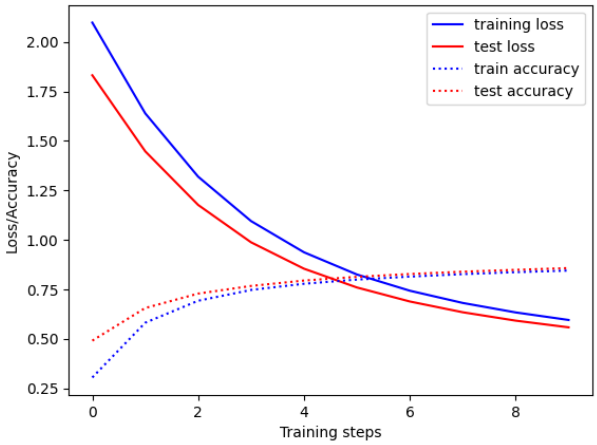

  - Training loop took:1988.671875 seconds -> slower
  - train and test accuracy are continuously improving
  - accuracy 0.94 -> momentum helps learning process

- **momentum 0.99**



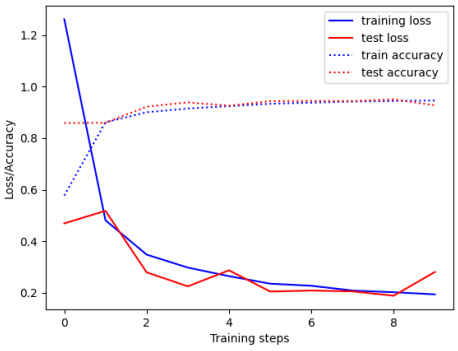

  - learns faster (but maybe too much faster)
  - accuracy really high: 0.97
  - Training loop took:1971.984375 seconds -> much slower
  - train accuracy gets at the end better than the test accuracy
  - train loss improves, but test loss not so much after a certain point

- **learning rate = 0.0001**



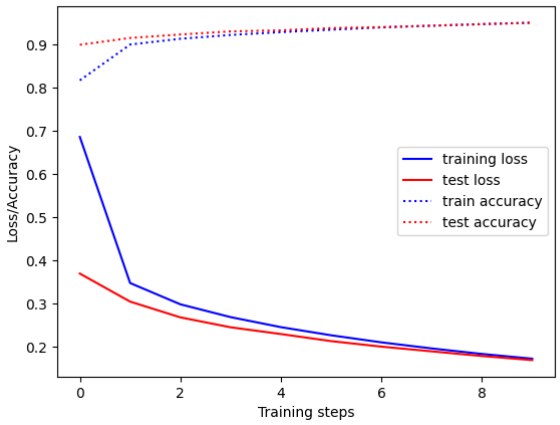

  - accuracy does not grow as fast (as expected)
  - performs still good
  - losses do not decrease as much
  - Training loop took:1895.6875 seconds

- **learning rate = 0.2**



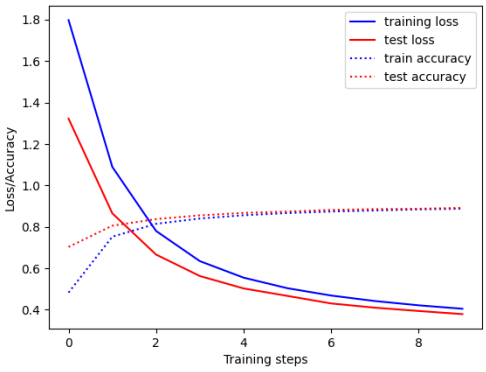

  - Training loop took:1892.6875 seconds
  - especially the test loss fluctuates very much
  - learning rate is too big

- **batch size smaller (16)**




  - Training loop took:4602.6125 seconds -> very slow
  - train and test accuracy are continuously improving

- **batch size bigger (128)**




  - Training loop took:504.2375 seconds -> much faster
  - train and test accuracy do not improve as much after a certain point
  - training and testing loss do not decrease as fast

| Number | hidden layers | units per layer | learning rate | momentum | batch size | computation time (s) | test accuracy (%) |
|---|---|---|---|---|---|---|---|
| Original (1) | 2 | 256 | 0.1 | 0 | 32 | 1894.73 | 92 |
| 2 | 1 | 4 | 0.1 | 0 | 32 | 1604.50 | 77 |
| 3 | 1 | 12 | 0.1 | 0 | 32 | 1625.45 | 89 |
| 4 | 2 | 256 | 0.1 | 0.5 | 32 | 1988.67 | 94 |
| 5 | 2 | 256 | 01 | 0.99 | 32 | 1971.98 | 97 |
| 6 | 2 | 256 | 0.0001 | 0 | 32 | 1895.68 | 88 |
| 7 | 2 | 256 | 0.2 | 0 | 32 | 1892.68 | 91 |
| 8 | 2 | 256 | 0.1 | 0 | 16 | 4602.61 | 94 |
| 9 | 2 | 256 | 0.1 | 0 | 128 | 504.23 | 89 |