1) **MLP and Back Propagation**

e) **Network structure:**

- Linear layer: W_1: (784, 200), b_1: (200,)
- Activation: Sigmoid()
- Linear layer: W_2: (200, 80), b_2: (80,)
- Activation: Sigmoid()
- Linear layer: W_3 : (80, 10), b_3: (10,)
- Activation: Softmax()
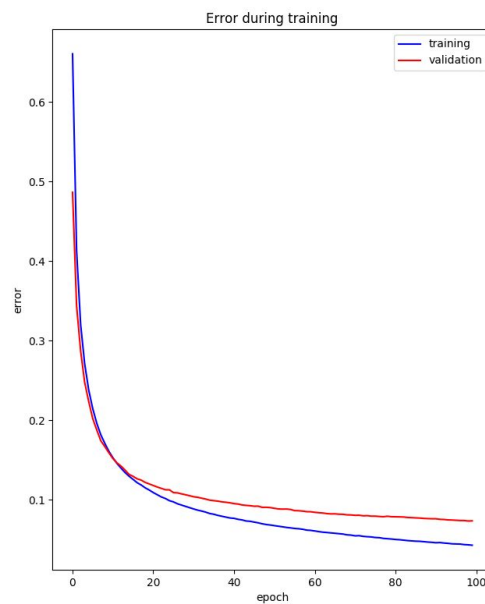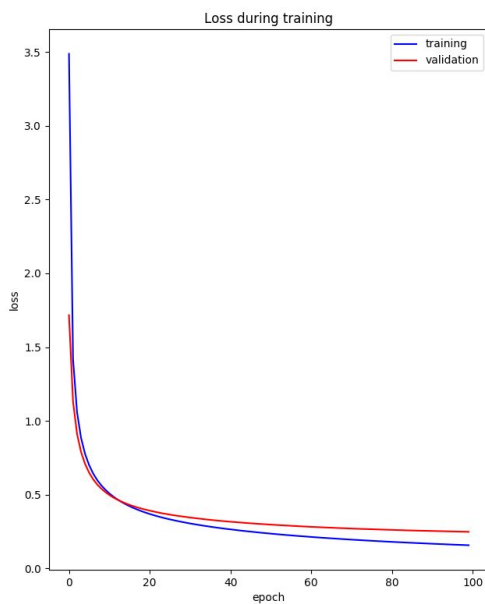- Loss: Cross entropy()

Batch size: 32
Optimizer: SGD
Learning rate: 1e-3
Number of epochs: 100

Results:

- **Training results:**
  Using Sigmoid() as activation for 1st and 2nd layer:

Test loss : 0.2676
Test error: 7.79%