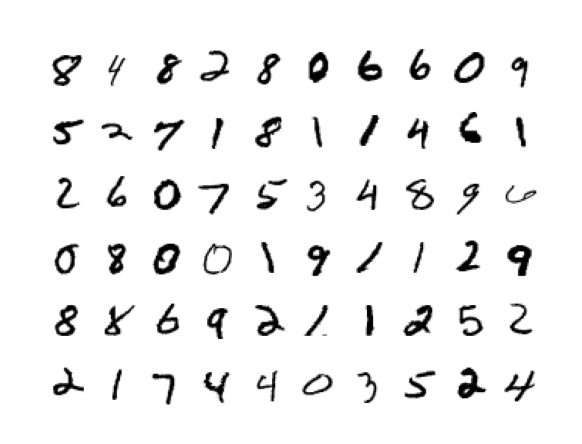
**Computational MR imaging**

**Laboratory 10: Training neural networks with backpropagation**

**Nan Lan**

1. **Neural network Structure:**

**1.1 Plot an example of selected training images.**



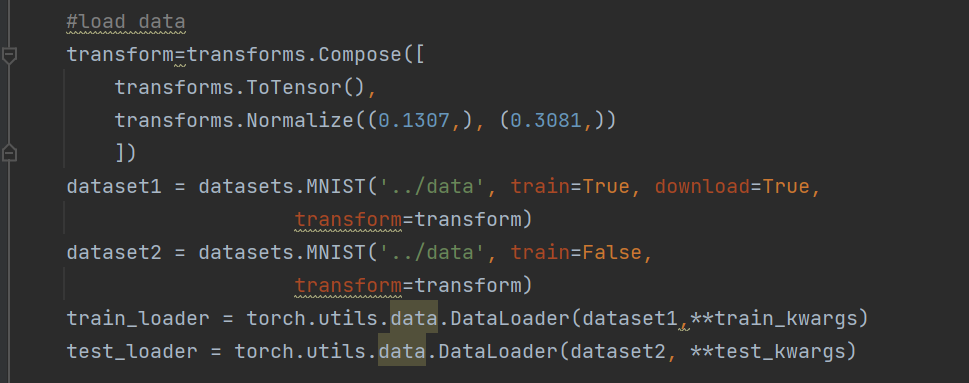
**1.2 Define a fully connected neural network architecture**

1. Hyperparameter setting

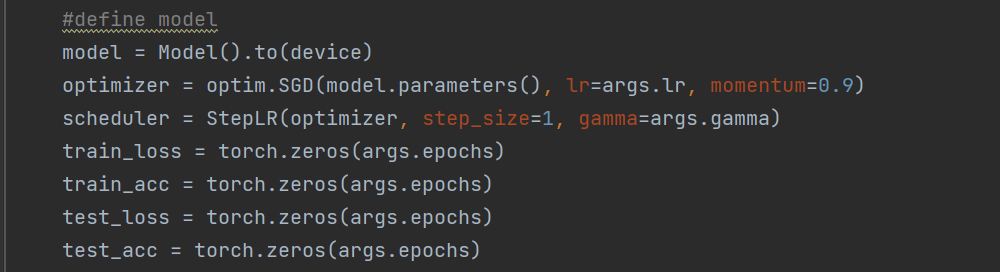




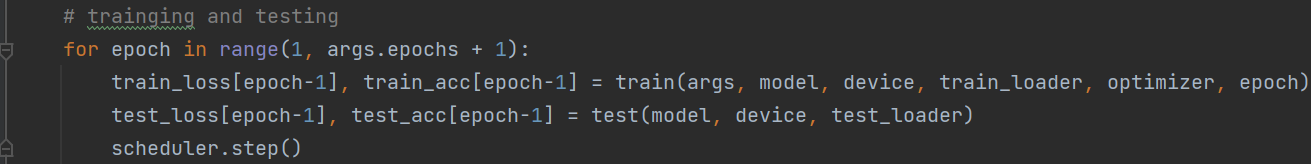
1. Load data



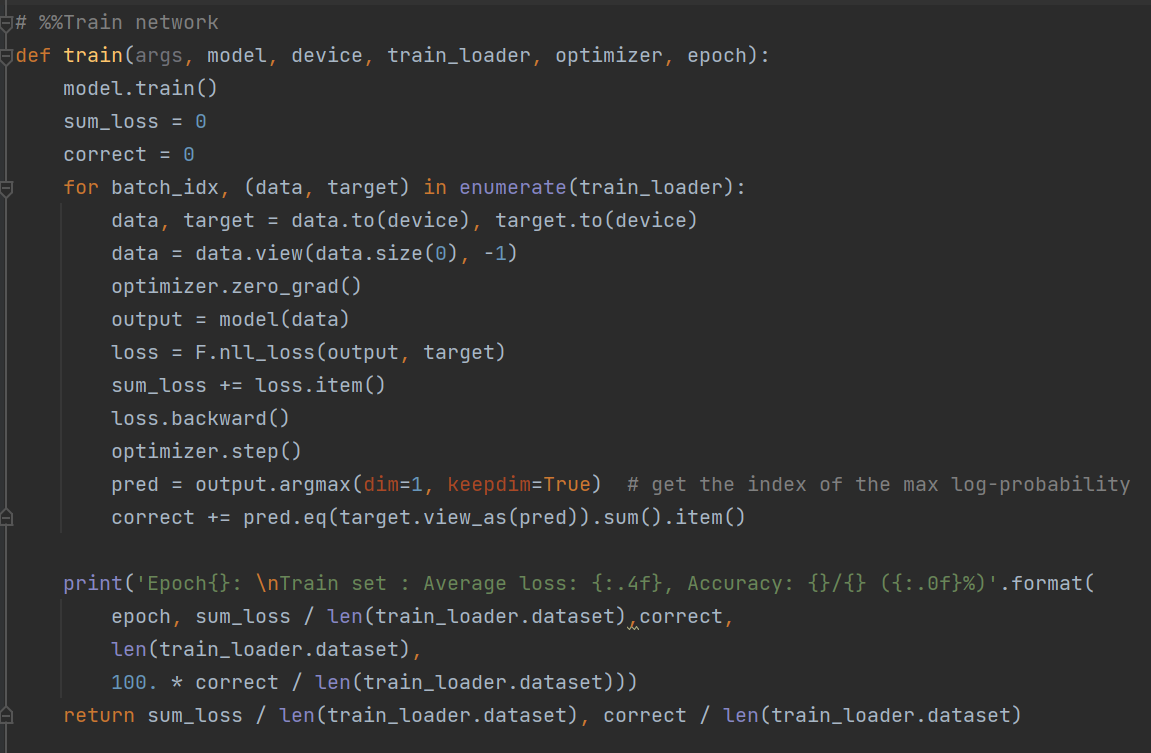
1. Define and Instantiate model

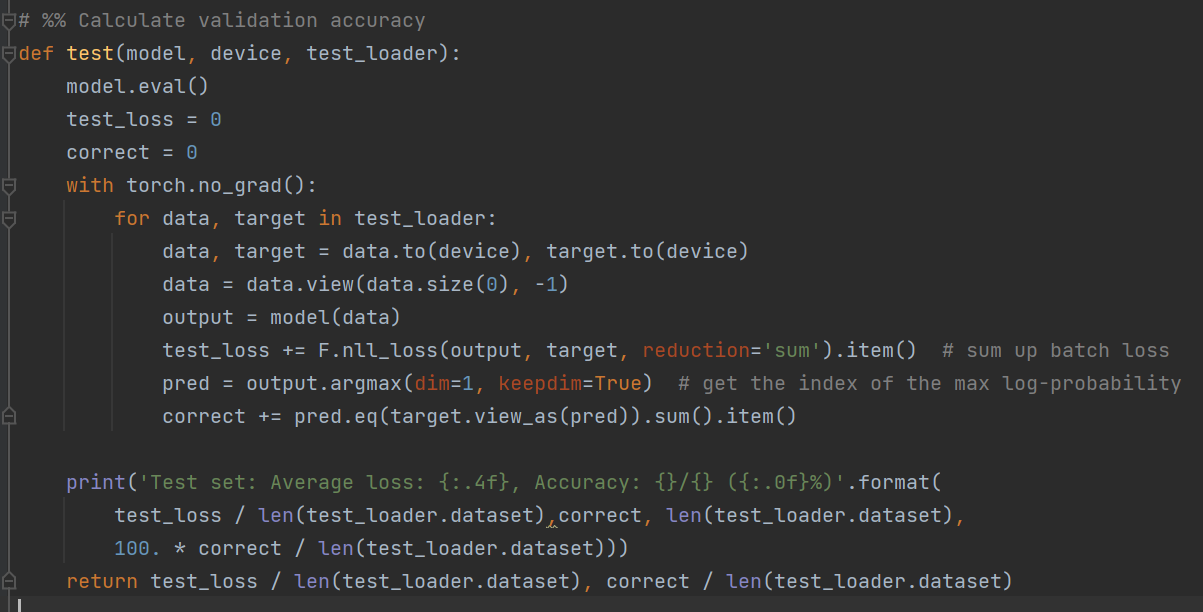


1. Train and test



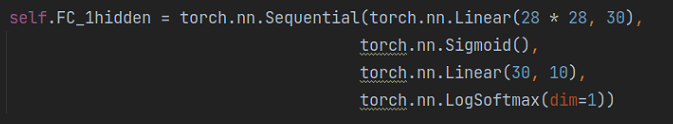
Where the train and test function are defined as followed:



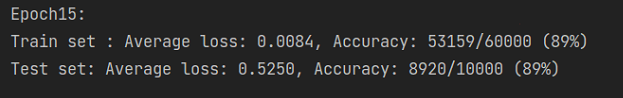


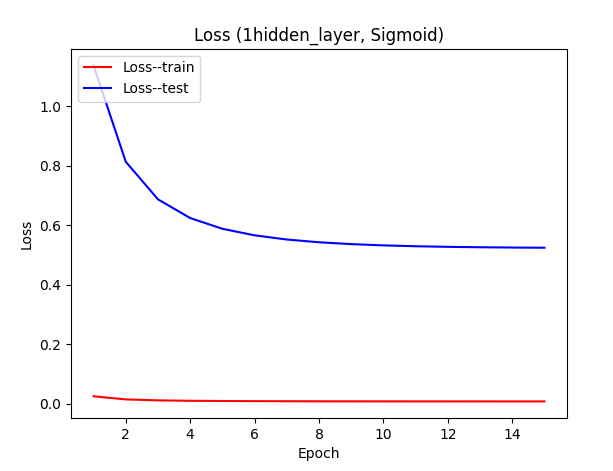
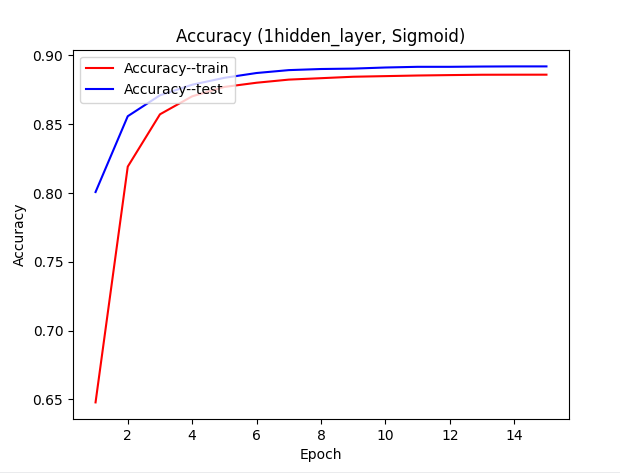
1. **Result and conclusion**
2. **1 hidden layer, Sigmoid**

The definition of model is as followed.



There is the accuracy and loss. The classification effect is ok.

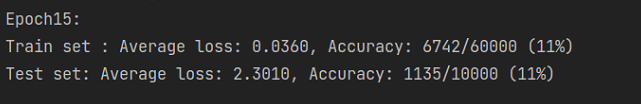


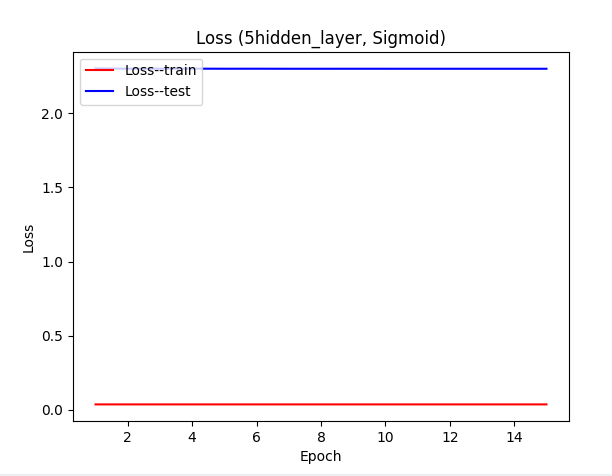
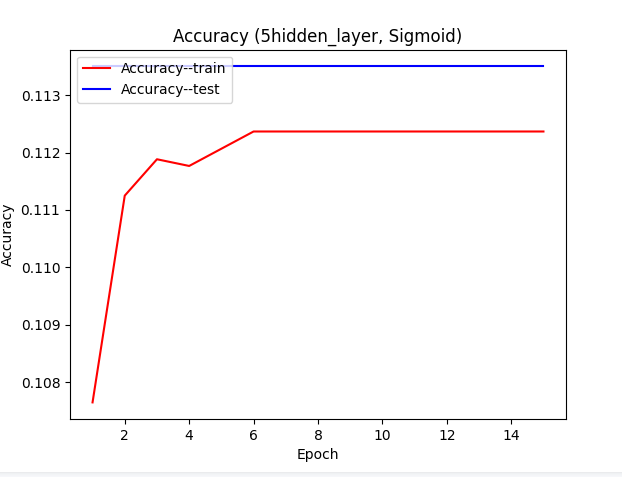


1. **5 hidden layer, Sigmoid**



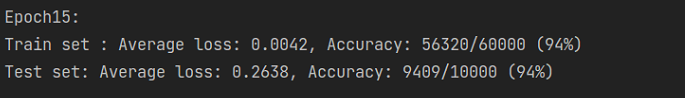
There is the accuracy and loss. The classification effect is super bad due to the overfitting. Batch normalization layers are added after each linear layer to adjust Internal Covariate Shift.

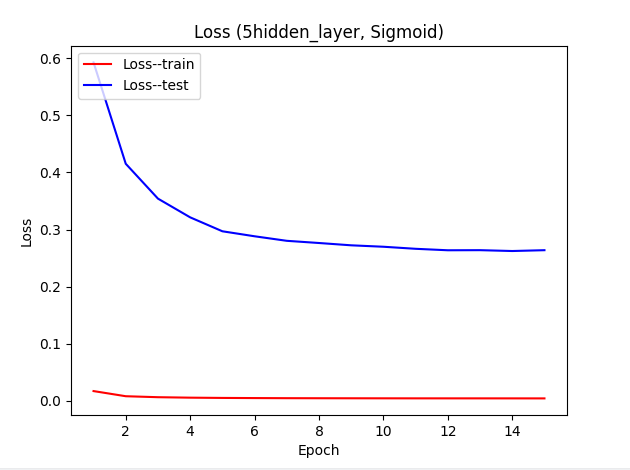
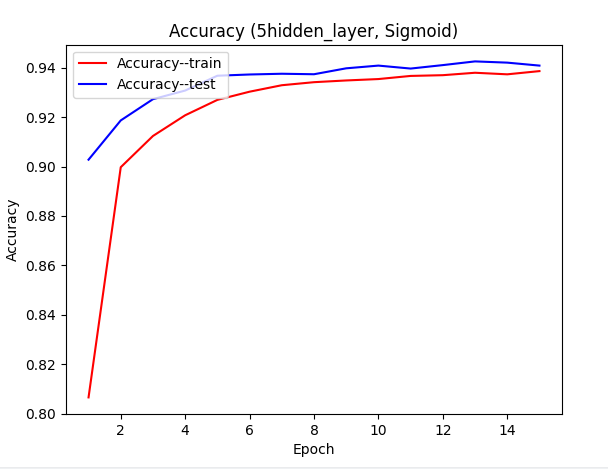




1. **5 hidden layer, Sigmoid(add Batch Normalization after each linear layer)**

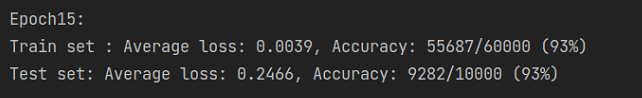
After adding batch normalization layers, the classification effect(94%) turns quite good (Better than the structure of 1 hidden layer with Sigmoid, 89%).

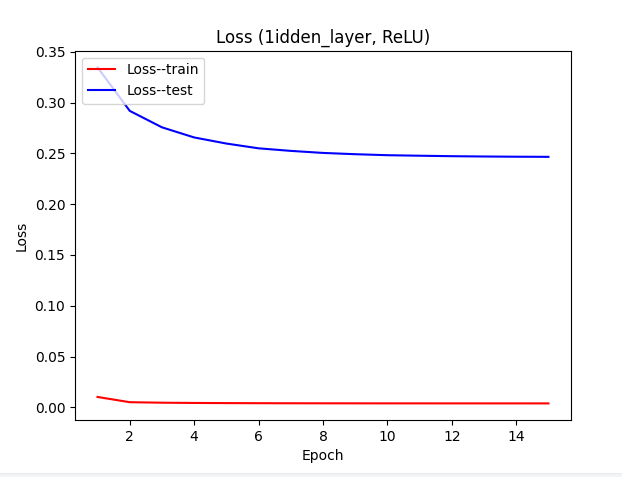
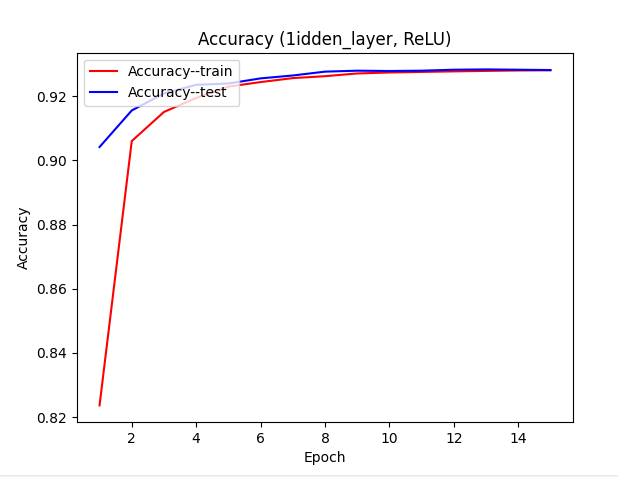




1. **1 hidden layer, ReLU**

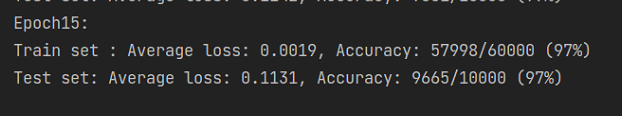
After changing the activation function from Sigmoid to ReLU, the classification effect gets better.

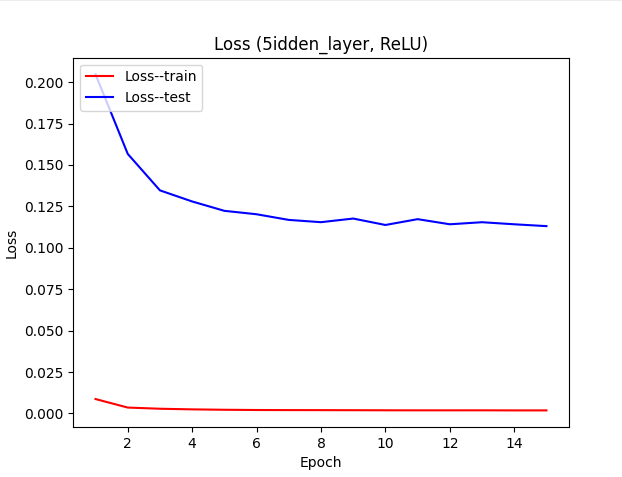
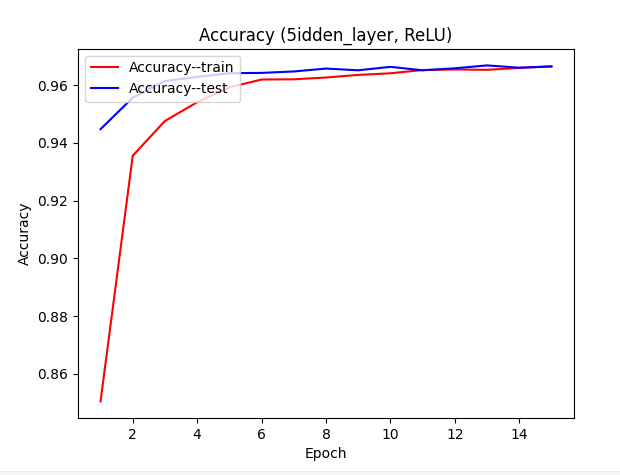




1. **5 hidden layer, ReLU(add Batch Normalization after each linear layer)**

The classification effect is the best in this structure.





**The conclusion is:**

Complicated network tends to get stuck in overfitting.

Under the precondition of fixing overfting, the more complicated the network is, the better the classification effect.

ReLU works better than Sigmoid function. The reason is that Sigmoid function has the problem of gradient vanishing. During backpropagation, the derivative of the sigmoid function is very small. After multiple accumulations in many layers , the gradient is close to 0, and the weight is basically not updated. So the training of the deep network cannot be completed.