

# SE395 Programming Assignment1 Report

201811118 이 구

*For all networks (3-layer NN with ReLU, 3-layer NN with LReLU), show the results and compare(ReLU vs LReLU)*

아래의 모든 결과는 각각 epoch 5000, batch size 512 로 학습한 결과이다.

두 모델은 아래와 같이 구성되었다.

Model 1: Linear(784,128)->ReLU()->Linear(128,64)->ReLU()->Linear(64,10)

Model 2: Linear(784,128)->LReLU()->Linear(128,64)->LReLU()->Linear(64,10)

## 1. Show the results

### (a) 10x10 Confusion Matrix (normalized for predicted label)

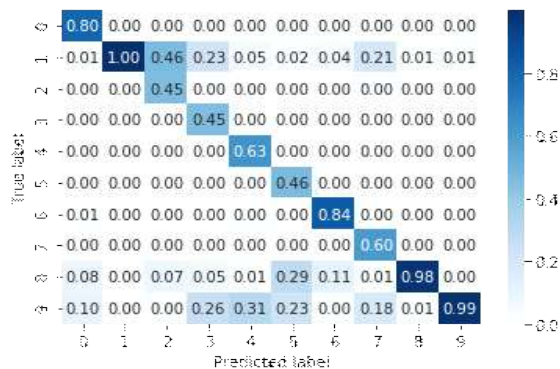


Fig 1. Confusion matrix with Model 1

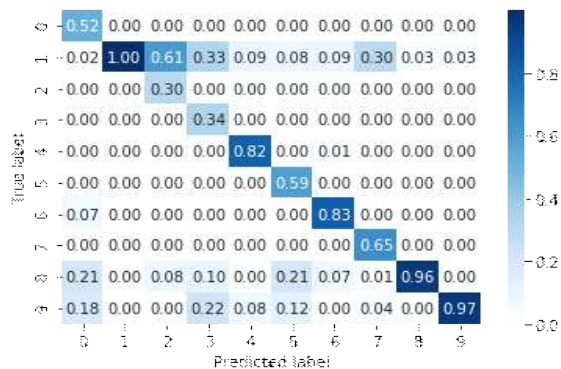
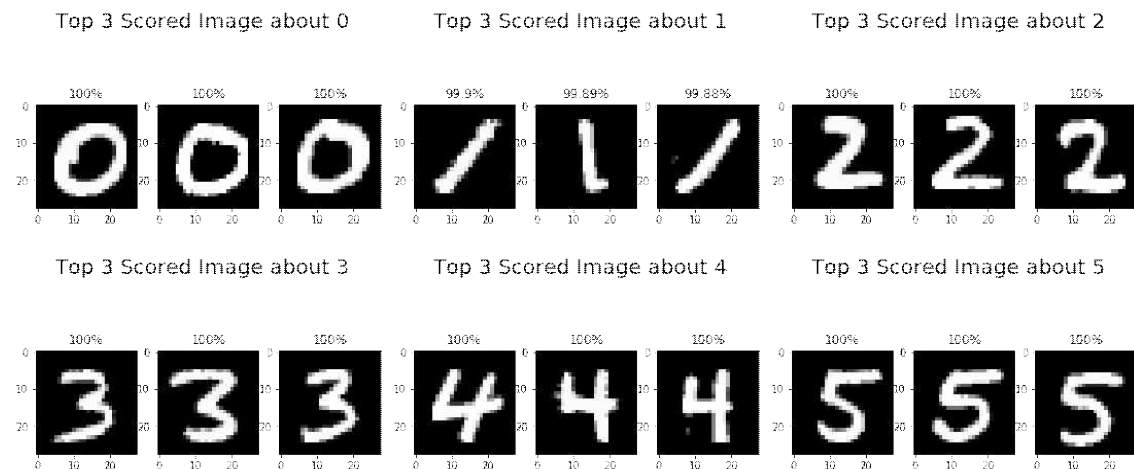


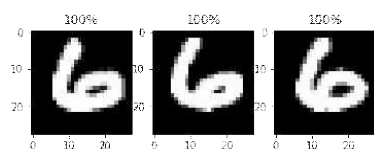
Fig 2. Confusion matrix with Model 2

### (b) Top 3 score images (all classess)

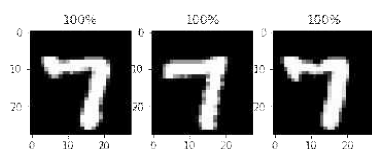
#### - Model 1 Result



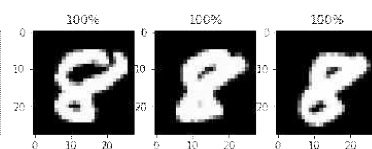
Top 3 Scored Image about 6



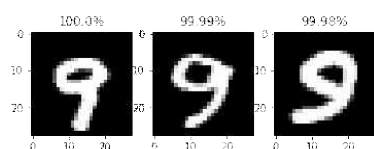
Top 3 Scored Image about 7



Top 3 Scored Image about 8

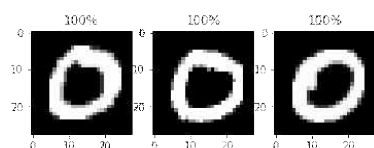


Top 3 Scored Image about 9

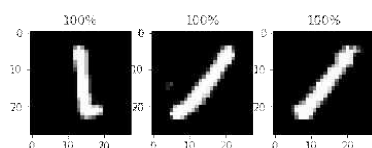


## - Model 2 Result

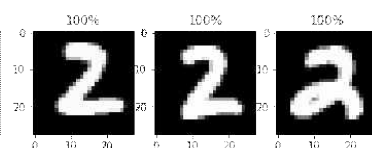
Top 3 Scored Image about 0



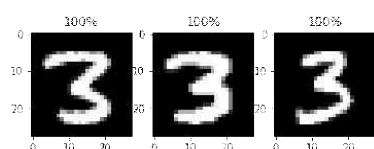
Top 3 Scored Image about 1



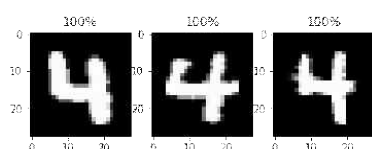
Top 3 Scored Image about 2



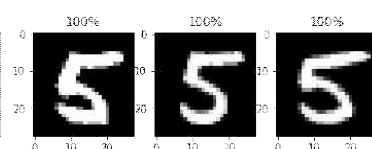
Top 3 Scored Image about 3



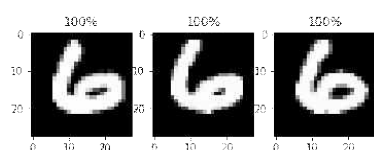
Top 3 Scored Image about 4



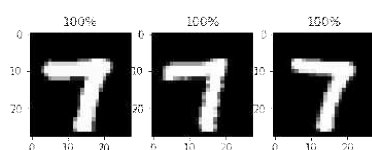
Top 3 Scored Image about 5



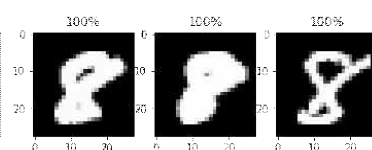
Top 3 Scored Image about 6



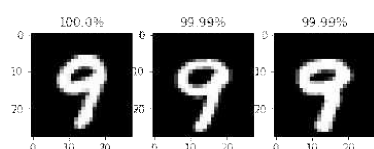
Top 3 Scored Image about 7



Top 3 Scored Image about 8



Top 3 Scored Image about 9



### (c) Training Loss graph

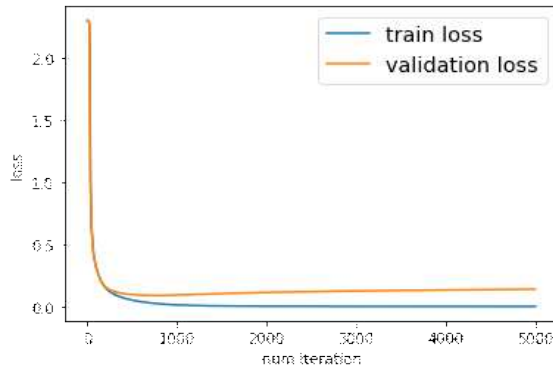


Fig 3. Loss graph with Model 1

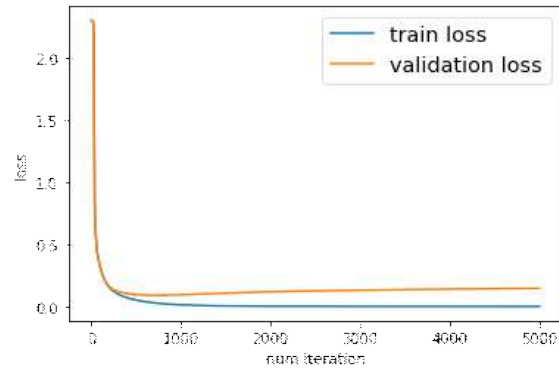


Fig 4. Loss graph with Model 2

## 2. Compare(ReLU vs LReLU)

ReLU function과 LReLU function의 차이는 input 값이 0보다 작거나 같을 때 갖는 gradient 값이다. ReLU function의 경우 input 값이 0보다 작거나 같을 때  $f(x) = 0$  이므로 gradient 역시 0이다. LReLU function의 경우 input 값이 0보다 작을 때  $f(x) = \alpha * x$  이므로 gradient 값은  $\alpha$ 이다. 이때  $\alpha$ 는 hyperparameter로 보통 0.01을 사용한다고 한다.

두 activation function의 차이는 backpropagation 때 발생한다. input 값이 0보다 작을 때, ReLU function을 사용하게 되면 gradient 값이 0이 되므로 더 이상 값의 update가 진행되지 않게 된다. 이로 인해 activation function으로 LReLU를 사용한 모델의 train loss가 더 빠르게 줄어드는 모습을 확인할 수 있다. 하지만 아래에 나타난 것과 같이 모델의 성능에서 큰 차이를 보이지는 않았다.

```
print("Model_ReLU accuracy: ", Model_ReLU.calc_acc(X_test,Y_test))
print("Model_LReLU accuracy: ", Model_LReLU.calc_acc(X_test,Y_test))
```

Model\_ReLU accuracy: 0.9736  
Model\_LReLU accuracy: 0.9764

Fig 5. Model 1, Model 2 accuracy