Report For Homework Two

Tianyi Ma 2017011408

Tsinghua University

Abstract. In this experiment, an MLP network and a CNN network are constructed. They are run with or without the batch normalization and the dropout unit to reveal their effects. The differences between the two network's performances on this task are also discussed in the report.

1 Architecture

1.1 The basic architecture and parameter setting of the network

The depth of the network remains the same as the initial setting. The mlp network contains one hidden layer and the cnn network contains two convolution layers. The size of the convolution kernels, pooling kernels, and the linear layers are shown in 1. And for the mlp network, the hidden layer has an input of 3072 dimensions and the output of 1024 dimensions. Other hyper-parameters remain unchanged.

Layers	Type	Input Size	Output Size	Size
1	Conv2d	3×32 × 32	32×30 × 30	32×3 × 3
2	BN	32×30 × 30	32×30 × 30	-
3	Relu	32×30 × 30	32×30 × 30	-
4	Dropout	32×30 × 30	32×30 × 30	-
5	MaxPool2D	32×30 × 30	32×15 × 15	32×2 × 2
6	Conv2d	32×15 × 15	128×12 × 12	128×4 × 4
7	BN	128×12 × 12	128×12 × 12	-
8	Relu	128×12 × 12	128×12 × 12	-
9	Dropout	128×12 × 12	128×12 × 12	-
10	MaxPool2D	128×12 × 12	128×6 × 6	128×2 × 2
11	Linear	4608	10	-

Table 1. Basic setting of CNN

2 Experiments

2.1 The parameter setting of the forward function

In model.py, two lines are about the parameter-setting of the forward function. In the first line, I filled the parameters with "is_train = True, reuse = False". In

the second line, I filled it with "is_train = False, reuse = True". The first line is set for the training steps so the training permission should be given and the parameter does not need to be reused. The second line is for the valid and test steps so the parameters should be fixed and the permission of using the trained parameters should be given. Also, when training the network, we implement BN and dropout layers. However, when the network is testing with the valid collection or test collection, the gates of the BN layers and dropout layers should be closed. That is also the reason why we should set the is_train as False when testing the network.

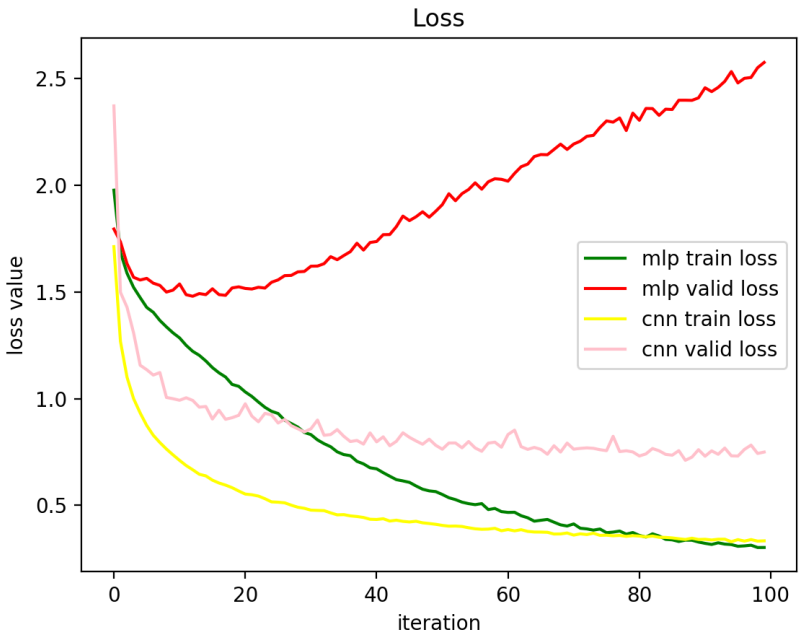

Fig. 1. Loss values in every iteration

2.2 The loss values of the mlp network and cnn network

We trained the mlp network and the cnn network with the architecture mentioned in section 1. The dropout rate is 0.5 and the learning rate is 0.001. Other hyper-parameters are as same as those of the initial setting. The loss values are depicted in 1.

Model	Train Acc.	Valid Acc.	Test Acc.
MLP	90.35%	52.40%	52.06%
CNN	88.19%	76.87%	76.71%

Table 2. Top accuracy rates

2.3 The accuracy rate values of the mlp network and cnn network

Same as the section before, the two networks are trained under the same settings and the accuracy rates against every iteration are depicted in 2 and the best accuracy rate values are shown in 2. The CNN network can gain a higher accuracy and lower loss value on the valid and test collection. Also, the CNN network is less likely to overfit than the MLP network. The reason for this phenomenon is that the CNN network's connection between each neural is more sparse than the MLP networks so it can be more easy to train.

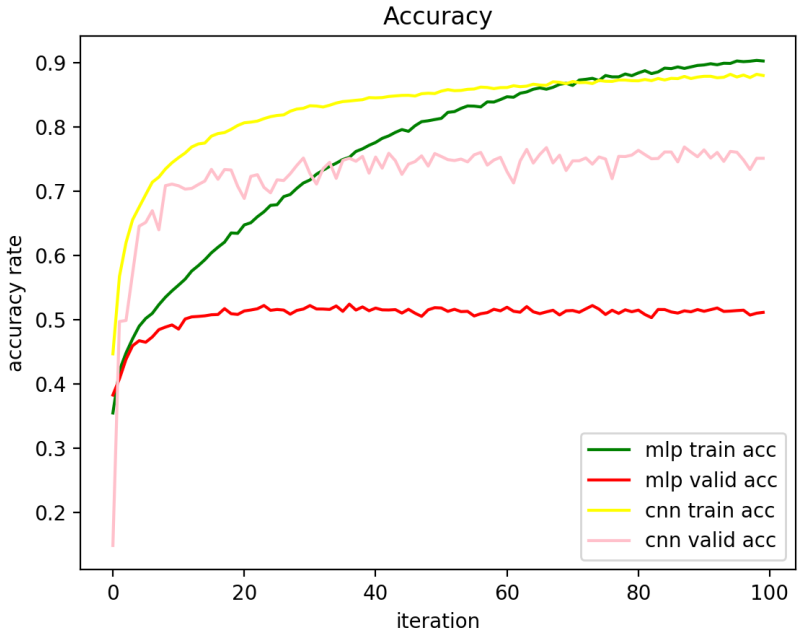

Fig. 2. Accuracy rate values in every iteration

2.4 The effect of the Batch normalization layers

With the batch normalization layers removed and other settings unchanged, the mlp network and cnn network are trained. The loss and accuracy rates are depicted in 3. The effects of the batch normalization layers can be surmised from the result. For the mlp network, the BN layers can enhance the performance of the network since it regularizes the data and makes it easier for the mlp network to learn. Conversely, for the cnn network, the performance of the network drops a bit when the BN layer added. The reason for that might be that the CNN network's connection between neural networks is more sparse than the mlp network and it can be more easy to train. Thus, it is not necessary to regularize the data since it can cause the loss of the information.

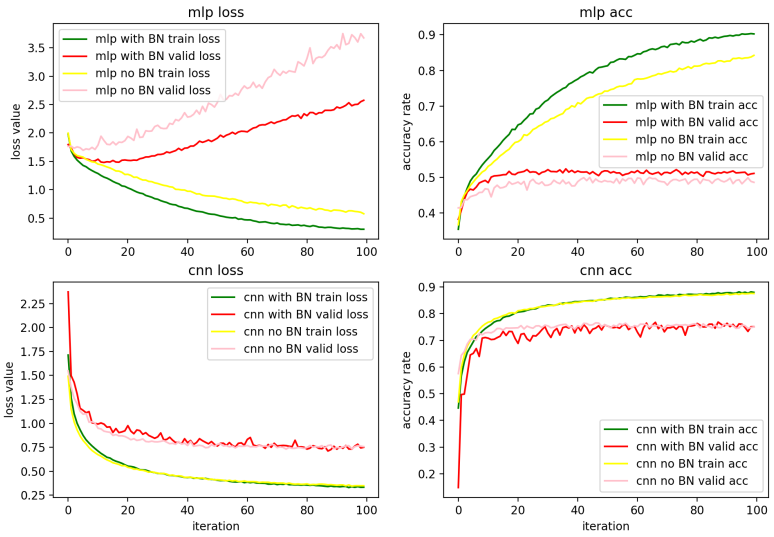

Fig. 3. Result with or without batch normalizaition

2.5 The effect of the dropout rate

To learn the effect of the dropout date to the mlp and the cnn networks, we conduct multiple experiments. The networks are trained for multiple times with different dropout rates. We set the rate as 0.9, 0.7, 0.5, 0.3, 0.1, 0 and trained the network to see their performances. The accuracy rates and loss values are depicted in 4 and top accuracy values are shown in 3. The dropout layers are added to the networks to alleviate over-fitting and an appropriate dropout rate

can enhance the performance of the network on the valid/test collections significantly. However, if the rate is too high or too low, the performances of the networks shall drop. From the result, we can surmise that 0.5 is the best rate for both mlp network and cnn network. We can also learn that as the dropout rate decrease, the learning accuracy increase. Meanwhile, the valid accuracy does not change a lot along with the change in the dropout rate. This phenomenon indicates that the dropout layers are capable to prevent or alleviate over-fitting.

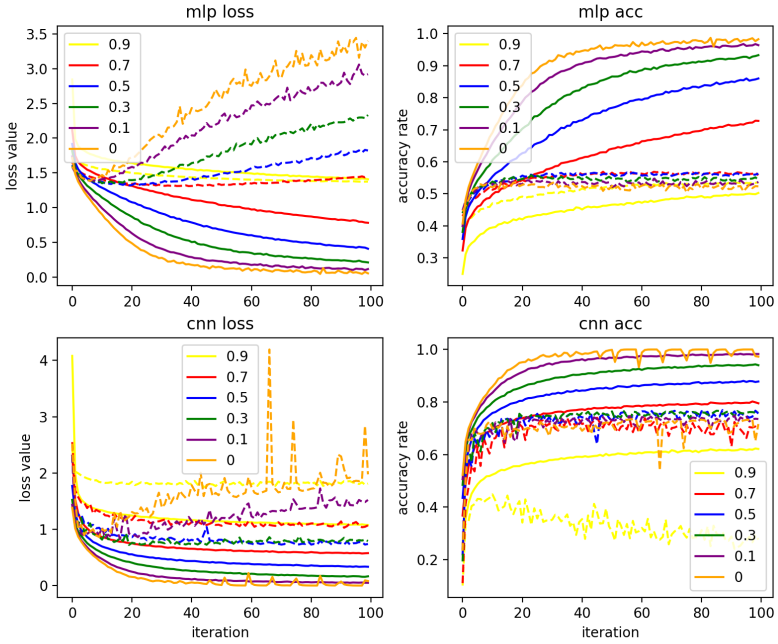

Fig. 4. Result with different dropout rate (dot line is validation result)

2.6 The function of the valid collection

In the experiments conducted above, we divided the data sets into three parts: the training set, the valid set, the test set. During the training of the network, the training set and validation set are visible to us since for every epoch, we can check the loss and acc. rate of the two sets. Conversely, the test set is invisible for us. We should only test the network with the test set as we finish the training process. Why do we need a valid set? We can check the valid loss and valid

Model	Dropout Rate	Train Acc.	Valid Acc.
MLP	0.9	50.24%	53.54%
MLP	0.7	72.81%	56.90%
MLP	0.5	86.01%	56.62%
MLP	0.3	93.30%	55.81%
MLP	0.1	96.78%	54.69%
MLP	0.0	98.70%	53.70%
CNN	0.9	62.31%	44.85%
CNN	0.7	80.21%	74.02%
CNN	0.5	88.04%	76.93%
CNN	0.3	94.34%	76.68%
CNN	0.1	98.36%	74.85%
CNN	0.0	100.00%	73.53%

Table 3. Top accuracy rates in different dropout rate

accuracy rate in every epoch. As long as the loss rate begins to increase while the accuracy rate is dropping, we should stop the training process since the network is going to get over-fitting on the training set. Or we can tune the dropout rate to alleviate overfitting. This procedure can prevent the network from overfitting. Then, we test the network with the test set to estimate the performance of the network. Noted that we fix the parameters when using the valid and test set, they can help us to learn the extensibility of our network. If the network reaches a high accuracy rate on the training set while a very low accuracy rate on the valid set, we can surmise that our network has low scalability since it can only adapt training set.

3 Conclusion

3.1 The conclusion of the experiments

After conducting this homework experiment, I gain a better knowledge of the effect of the dropout layers and batch normalization layers. The BN layers can regularize the data and make it easier for the network to learn from and the dropout layers can alleviate over-fitting. I also learn the difference between the training set and the valid set. Former is used to train the network and the latter is used to help us to estimate if the network is over-fitting. What is more, I learn to use the TensorFlow architecture in this homework.