CSI 5138
Introduction: Deep Learning & RL

Homework 2

Date: 2020/10/13

NAME: Yunfeng Li

Student Number: 300139547

Experiment Description

The purpose of this experiment is to design three different classifiers for two datasets, respectively. The experiment is carried out per the procedures below.

1. Design three classifiers (Soft-Max Regression, MLP, and CNN) for the MNIST dataset. Comparing the performance of these classifiers.
2. Adding normalizing methods to observe the improvement of models.
3. Design three classifiers (Soft-Max Regression, MLP, and CNN) for CIFAR-10 dataset. Comparing the performance.
4. Add the normalizing method for CNN model of CIFAR-10.

Results Analysis

This section will be separated into two parts by the different dataset. And since the MNIST dataset is more efficient than CIFAR-10 in the code execution, we will spend more time to analyze this dataset.

MNIST Dataset

This dataset only has grey pictures which means less information than colorful pictures which generally have three channels. Three models are defined as below.

- **Soft-max Regression**

Simply, we only need to have one layer with the activation function of Soft-max.

- **MLP**

Input Layer	Hidden Layer 1	Hidden Layer 2	Output Layer
3 nodes	512 nodes	512 nodes	10 nodes

Table 1

This model will have two hidden layers which both have 512 nodes. And the activation function is relu.

- **CNN**

In this model, we set two convolution layers and two full connection layers. To be specific, the model is designed as below.

Conv_2d	Max_pool2d	Conv_2d	Max_pool2d	Layer 1	Layer 2
In_channel=1 Out_channel=16 Kernel_size=3	(2, 2)	In_channel=16 Out_channel=32 Kernel_size=3	2	(32*7*7, 512)	(512, 10)

Table 2

- **Other parameters settings**

Since we will compare performances of different models, the same parameters in three models are set as same value. And the setting is below.

Parameters	Value
Batch Size	512
Learning Rate	0.1
Epochs	10
Dropout Rate	0.5
Optimizer	SGD
Loss Function	Cross Entropy

Table 3

- **Results**

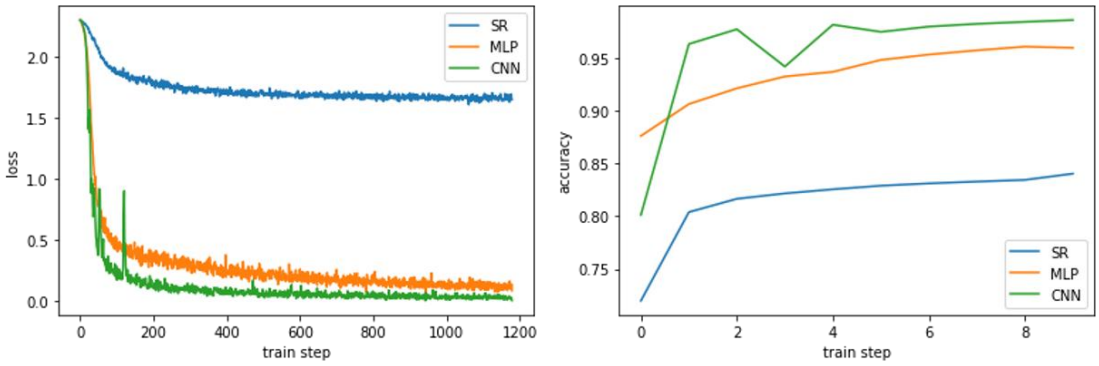

Figure 1

From plotted diagrams, the coverage of MLP and CNN is faster than the Soft-max regression model. And for the accuracy, CNN perform the highest score in these three models and the MLP is smaller than CNN. But they both works well than soft-max regression significantly. And reasons for these situations are acceptable:

a) The soft-max regression has the easiest structure in all three models. And no matter the complexity or the size, soft-max is relatively small. This will lead to a low accuracy and slow convergence.

b) Although the results of CNN and MLP is not significantly different, the reason why CNN has a better performance may involve the features' quality. After convolution layers, features having more intuitive information will be left and finally works a better accuracy.

Depth of MLP

In this part, we will examine how depth of MLP will influence the model. Two models whose depth of layers is the only difference are trained. From the figure below, both models have convergences and regarding to the accuracy, the MLP network with deeper layers will converge faster which makes sense, since more layers means more information will be remained. And we can find that 5 layers model has a higher accuracy finally.

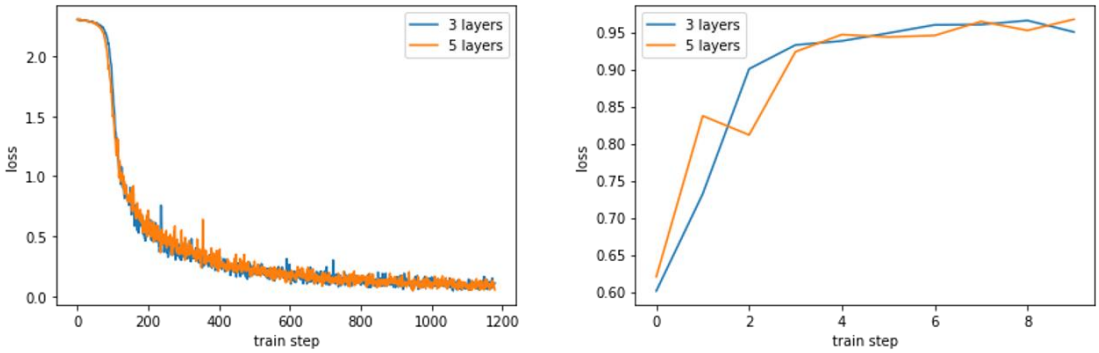

Figure 2. Depth of MLP

Depth of CNN

In this section, different convolution layers are applied to observe the performance of models. The model with 3 convolution layers outperforms another model slightly. This maybe because we only add one layer. Besides, the model with 2 layers has already perform well enough regarding to the accuracy. From the accuracy, 3 layers CNN have a higher accuracy than 2 layers CNN, which because features after 3 convolution layers will be more informative.

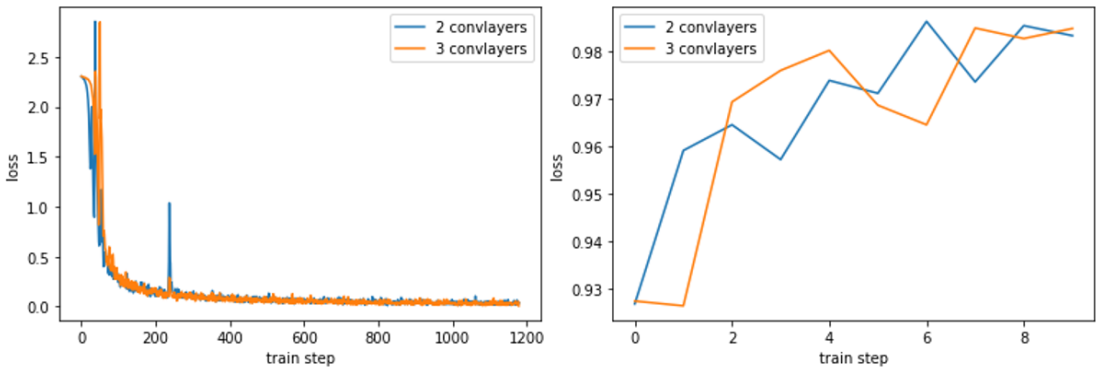

Figure 3. Depth of CNN

Dropout Influence

In this part, the dropout method will be applied to examine the performance of three models. This method aims to eliminate the overfitting phenomenon. We set all parameters same and observe how dropout influence the results in the figures below.

Soft-max Regression: It is obvious that dropout will damage the soft-max model. The accuracy of model with dropout is relatively low. The explanation is probably that the dropout will delete nodes by a probability. However, the soft-max only has one layer which means the output

nodes will be delete. Finally, the output will be changed. That is the reason why dropout makes the model worse.

MLP: The dropout has a slightly impact on the MLP model. That maybe because dropout is only simply applied. The model with dropout also shows a good performance and the accuracy is closed to the model without dropout.

CNN: Since the convolution layer can not have dropout, the only way to apply it is in full connection layers. The figures show that dropout has litter influence on the model. But we can also observe that the model with dropout will converge faster than the model without dropout.

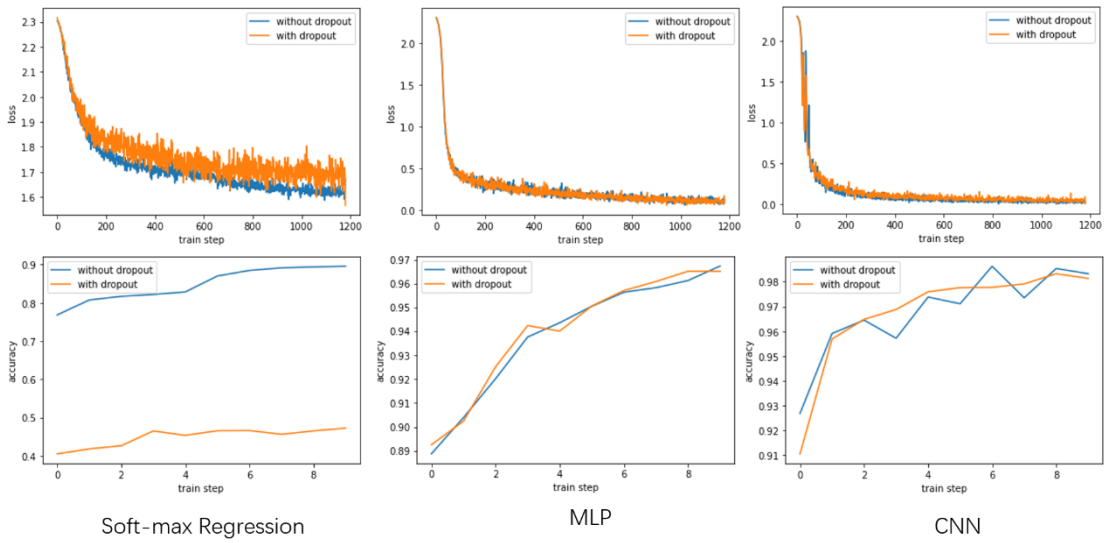

Figure 4

CIFAR-10 Dataset

The CIFAR-10 dataset is a collection of images that are commonly used. This dataset also has ten classes which are type of goods. Generally, models which are trained by this dataset have lower accuracy than MNIST dataset. And the training time is limited to hardware of computers. In this section, we try to compare these three models and then add normalizing methods to improve the accuracy. In order to compare in same level, models are designed with same parameters as below.

- **Soft-max Regression**

This model is designed in a easy structure which only have input and output layers.

- **MLP**

Input Layer	Hidden Layer 1	Hidden Layer 2	Output Layer
3*32*32 nodes	512 nodes	512 nodes	10 nodes

Table 4

This model will have two hidden layers which both have 512 nodes. And the activation function is relu.

- **CNN**

In this model, we will have two convolution layers and three full connection layers. To be specific, the model is designed as below.

Conv_2d	**Max_pool2d**	**Conv_2d**	**Max_pool2d**	**Layer 1**	**Layer 2**	**Layer 3**
In_channel=3 Out_channel=6 Kernel_size=5	(2, 2)	In_channel=6 Out_channel=16 Kernel_size=5	(2, 2)	(16*5*5, 120)	(120, 84)	(84, 10)

Table 5

- **Other parameters settings**

Parameters	**Value**
Batch Size	4
Learning Rate	0.001
Epochs	10
Optimizer	SGD
Loss Function	Cross Entropy

Table 6

Results

In this part, the image shown below is the accuracy of three models. It is not surprising that CNN outperforms other two models. MLP is better than soft-max regression which also makes sense, since soft-max has the simplest network. From this image, the highest accuracy of all models is almost 60% which performs a probability of wrong classification. And the reason for all these is described as below.

1. Soft-max regression has a low accuracy which maybe because there are only input and output layers. This simple network will converge in very early epochs.
2. The CNN works better than MLP which maybe because the convolution layer as same as the CNN model in MNIST dataset. The image after convolution operation will remain more informative features and train the model more accurate.
3. The reason why the accuracy of CNN model only scores 60% is that the network we design is not deep enough. It has been proved that with more than 10 convolution layers, the accuracy will even more than 80%. Since the limitation of computer CPU and GPU, we will not attempt here.

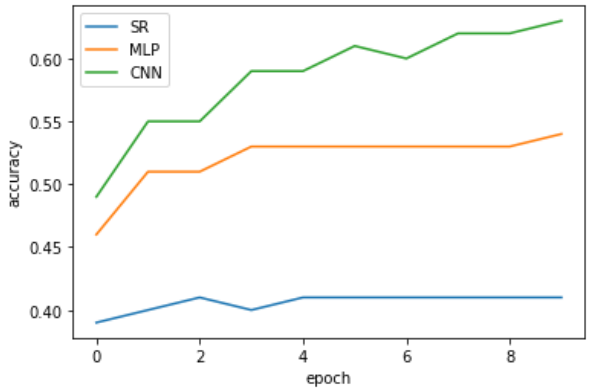

Figure 5

Influence of Normalization

It is obvious that Normalization method will make the model have a higher accuracy. Thinking of training time, in this part, we will only focus on the CNN model. Both dropout and batch normalizing will be applied to the network and comparing their accuracy. To be specific, in this experiment, batch normalizing is added to each convolution layers. Since the dropout can not works in convolution layers, it will be added in full connection layers.

Generally, normalizing method will solve the overfitting phenomenon and improve the accuracy. However, from the figure below, CNN model with batch normalizing and dropout have a low accuracy than the one who is without normalizing method. The reason why this happened maybe that the batch size in training data is 4 which is relatively small and probably lead to overfitting in model without normalizing. After normalizing method, the model has been eliminated a little overfitting and finally has a lower accuracy. Moreover, we can find that model with normalizing will converge faster which is also because of normalizing.

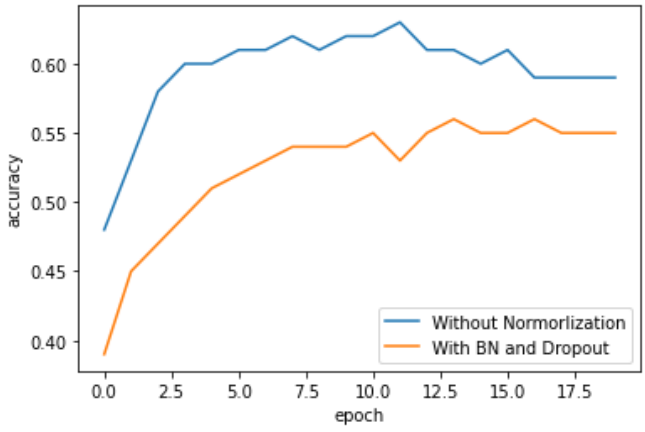

Figure 6

Summarization

To summarize, in this experiment, we have two datasets and we design three classifiers for each dataset. After that, comparing performances of different models and exploring the model structure by setting different values to observe accuracy and loss of each model. According to the experiment, CNN generally works best than MLP and Soft-max regression, and MLP is better than Soft-max regression. Furthermore, the MNIST datasets is simpler than CIFAR-10.