

Neural Network Lab - Report

Minh-Quan DAO, Isaac Gutierrez Payan

1 Introduction

In this lab, a Fully Connected and a Convolutional neural network are implemented to perform a classification task on MNIST dataset. While the two networks' architect are kept unchanged, their hyperparameters are varied to give insight on how each network's classification accuracy is affected.

The code and data files accompanying this report are run on Tensorflow 2.0 and Tensorboard.

2 Fully Connected Neural Network (FCN)

The architect of the FCN used in this lab is shown in Table.1. The "*" in the superscript denotes a hyperparamter. Moreover, this FCN has a single hidden layer, hence being a shallow network.

Layer (type)	Specifications
Flatten	
Dense	$dense_units^*$
Dropout	$dropout_rate^*$
Dense	10

Table 1: FCN architect

Hyperparameter	Value
Learning rate	$1e-3, 1e-2, 1e-1$
$dense_units$	16, 32
Mini batch size	32, 64
$dropout_rate$	0, 0.5

Table 2: FCN hyperparameters value

In addition to two hyperparameters shown in Table.1, there are two more hyperparameters which are the learning rate and the size of a mini batch. The best hyperparameters combination are found by grid search based on their values shown in Table.2.

Due to the limited computing resource, each hyperparameters combination is used to train the FCN for only one epoch. The resulted network's performance is measured by its accuracy on the test set. The comparison of performance among 24 model is summarized in Fig.1-4.

It is clear from Fig.1 and Fig.3 that a smaller learning rate and a larger mini batch result in significantly higher accuracy on test set; while large number of dense layer units or dropout rate do not guarantee improvement on accuracy (Fig.2 and Fig.4).

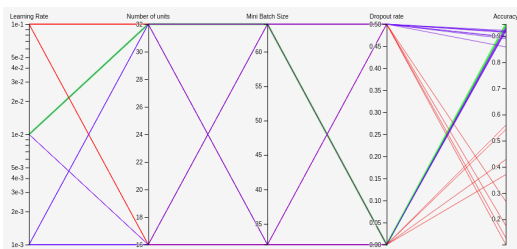


Figure 1: Learning rate effect on FCN

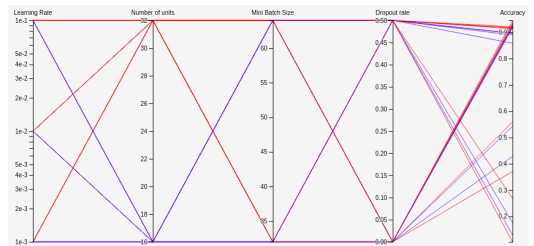


Figure 2: Number of units on hidden layer effect on FCN

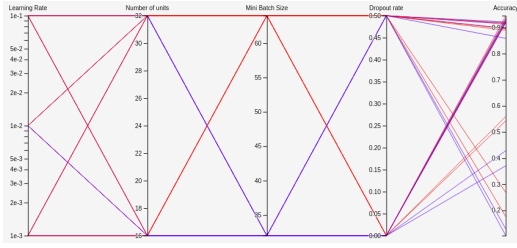


Figure 3: Mini batch size effect on FCN

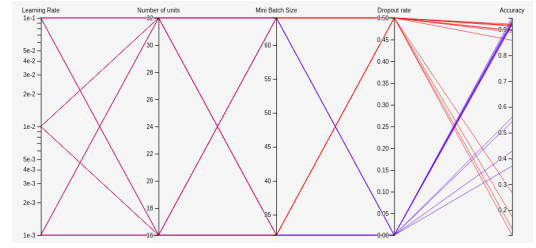


Figure 4: Dropout rate effect on FCN

The best hyperparameters combination is 0.001 learning rate, 0 dropout rate, and 16 units on the hidden layer, and mini batch of 32 samples. The associated trained CNN achieved 91% of accuracy on test set.

3 Convolutional Neural Network (CNN)

The architect of the CNN used in this report and its hyperparameters' value are shown in Table.3 and 4.

Layer (type)	Specifications
Conv2D	32 filters, <i>kernel_size</i> *
Conv2D	64 filters, <i>kernel_size</i> *
MaxPloeing2D	2x2
Dropout	0.25
Flatten	
Dense	<i>dense_units</i> *
Dropout	<i>dropout_rate</i> *
Dense	10

Table 3: CNN architect

Hyperparameter	Value
Learning rate	0.001, 0.01, 0.1
Kernel size	3, 5
<i>dense_units</i>	32, 128
<i>dropout_rate</i>	0, 0.25, 0.5

Table 4: CNN hyperparameters value

Similar to the previous section, the best hyperparameters combination is found based on grid search. The effect of each hyperparameter on network accuracy on test set is summarized in Fig.5 - 8.

It can be seen from Fig.5, a smaller learning rate for Adam optimizer results in higher network accuracy.

On the other hand, the use of larger kernel (Fig.6) or more units on first dense layer (Fig.7) do not show any advantage in term of accuracy. Therefore, a smaller kernel and dense layer can be help reduce the number of trainable parameters and running time, while secure the same level of accuracy.

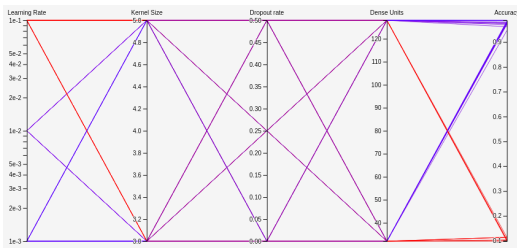


Figure 5: Learning rate effect on CNN

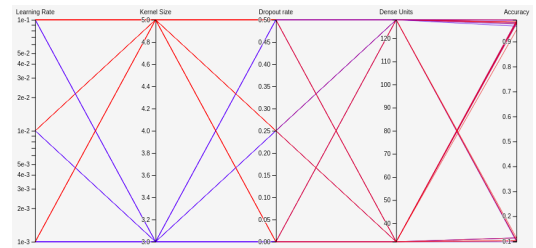


Figure 6: Kernel size effect on CNN

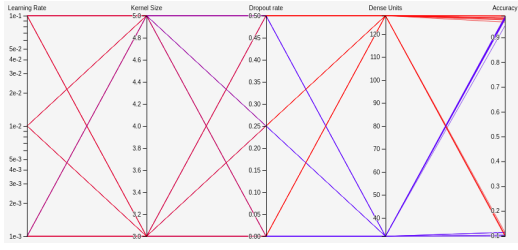


Figure 7: Number of dense units effect on CNN

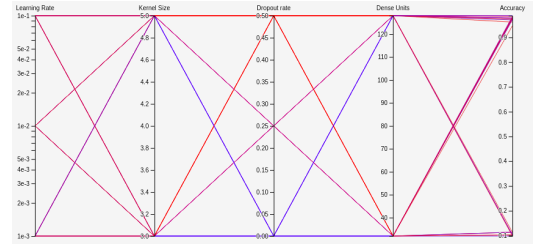


Figure 8: Dropout rate effect on CNN

Finally, the effect of dropout is unclear. With dropout rate set to 0 (no dropout at all), network can still obtain 98% of accuracy.

The best hyperparameters combination is 0.001 learning rate, 3×3 kernel, 0 dropout rate, and 32 units on the first dense layer. The associated trained CNN achieved 98% of accuracy on test set. This level of accuracy make this CNN outperform the best FCN in the previous section.