

HW3 hyper tuning

EE3 Yan-Hao Chen B04901094

May 2018

1 Introduction

In this experiment, I would fix the model structure, and then modify each hyper parameter out of [learning-rate, transfer function, batch size, epochs, optimizer, and weight initialization.] and try to find the best hyper parameters, and eventually combine all of the best hyper parameter values to get the best result.

2 Method

If not specified, each hyper parameter has its default value, and they are:

1. learning rate = $1e-4$,
2. transfer function = relu,
3. batch size = 512,
4. epochs = 40,
5. optimizer = Adam,
6. kernel initializer = random uniform.

We will then start to tune each hyper parameter one by one, and the below table shows the value we will assign to each hyper parameter

learning rate	1e-2	1e-3	1e-4	1e-5	1e-6		
transfer function	selu	relu	tanh	sigmoid	linear		
batch size	20	30	50	100	200		
epochs	100	200	300	400	500		
optimizer	SGD	RMSprop	Adagrad	Adadelata	Adam	Adamax	Nadam
initializer	lec unif	he unif	glor unif	lec norm	he norm	glor norm	

3 Result

1. learning-rate to accuracy plot is shown in figure 2
2. transfer function to accuracy plot is shown in figure 2
3. .batch size to accuracy plot is shown in figure 4
4. epochs to accuracy plot is shown in figure 5
5. optimizer to accuracy plot is shown in figure 6
6. weight initialization to accuracy plot is shown in figure 7

4 Discussion

1. learning rate cannot be too large to make the training unable to converge.
2. elu, selu ,relu family perform better than other transfer functions.
3. No obvious trend shown in batch size parameter.

5 Conclusion

the best hyper parameter values are shown below:

1. learning-rate = 0.0001
2. transfer function = relu
3. batch size = 100
4. epochs = 400
5. optimizer = Adam
6. weight initialization = random uniform

6 Combine

After assign all the best hyper parameter values into one, we can get our scores shown in figure 8

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 28, 28, 64)	320
max_pooling2d_1 (MaxPooling2D)	(None, 14, 14, 64)	0
dropout_1 (Dropout)	(None, 14, 14, 64)	0
conv2d_2 (Conv2D)	(None, 14, 14, 32)	8224
max_pooling2d_2 (MaxPooling2D)	(None, 7, 7, 32)	0
dropout_2 (Dropout)	(None, 7, 7, 32)	0
flatten_1 (Flatten)	(None, 1568)	0
dense_1 (Dense)	(None, 256)	401664
dropout_3 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 10)	2570
Total params: 412,778		
Trainable params: 412,778		
Non-trainable params: 0		
Train on 54000 samples, validate on 6000 samples		

Figure 1: model structure

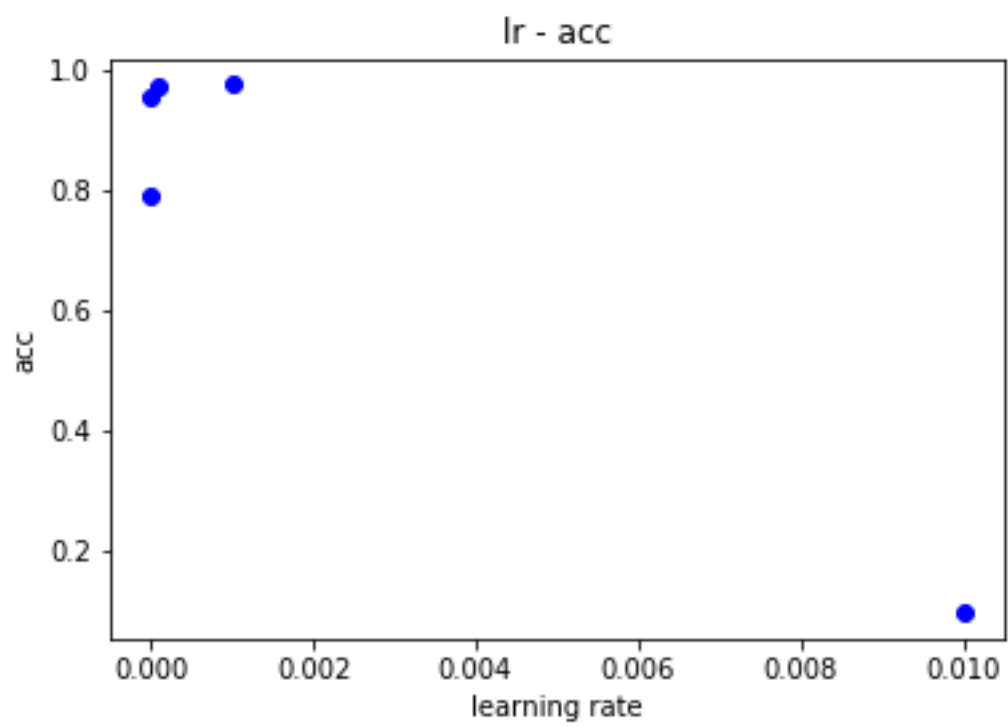


Figure 2: learning rate-accuracy

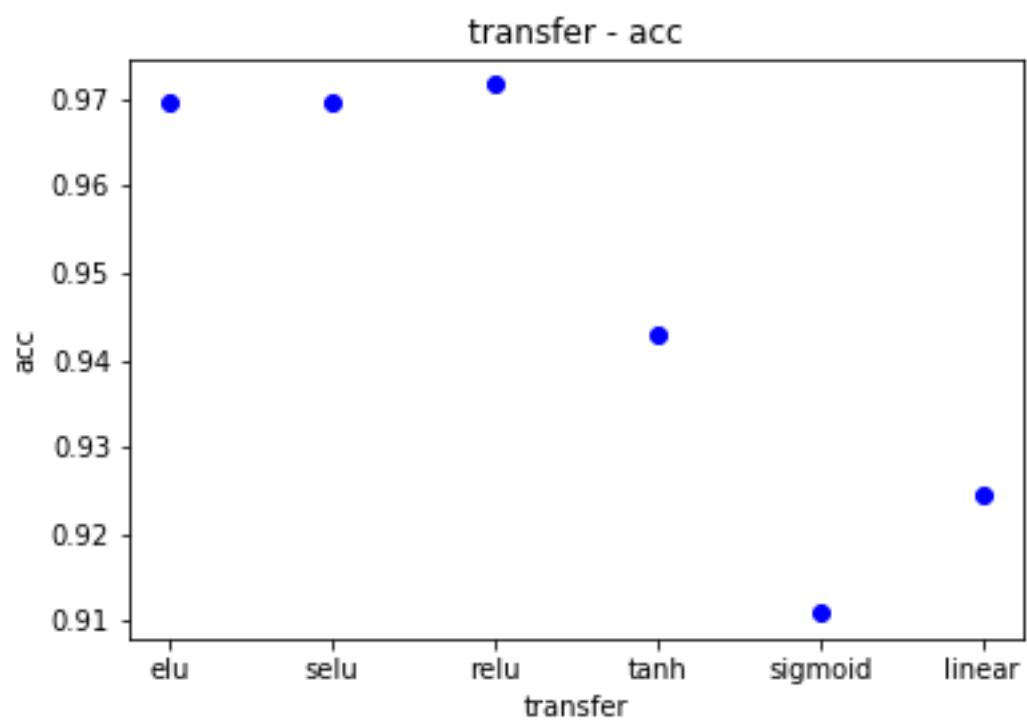


Figure 3: transfer function - accuracy

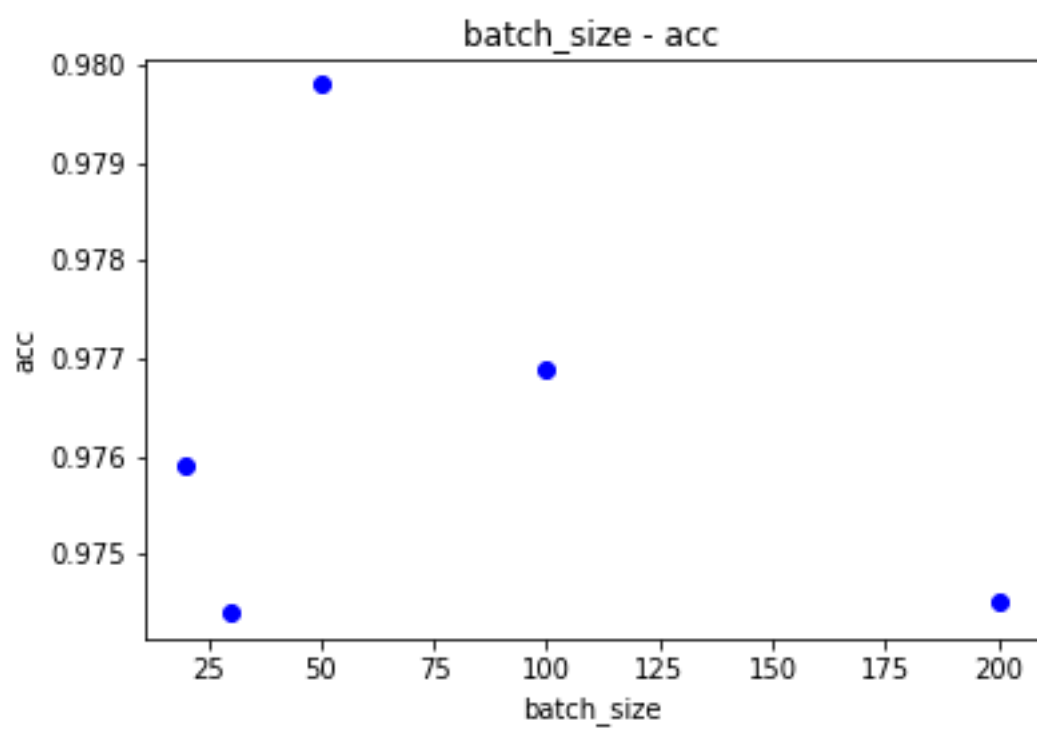


Figure 4: batch size - accuracy

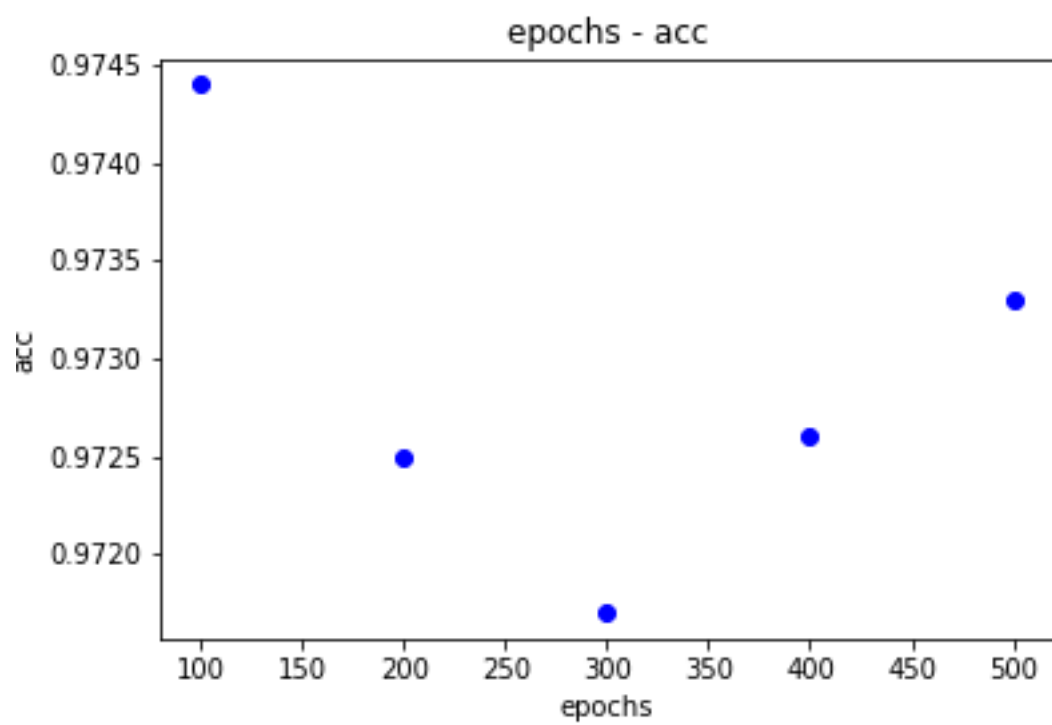


Figure 5: epochs - accuracy

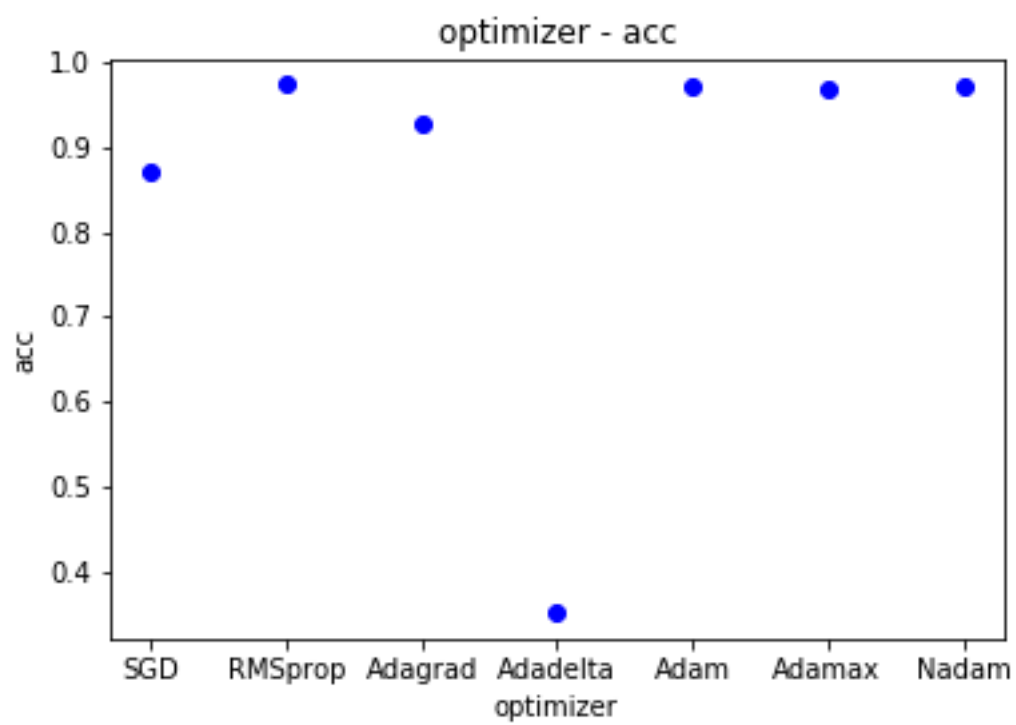


Figure 6: optimizer - accuracy

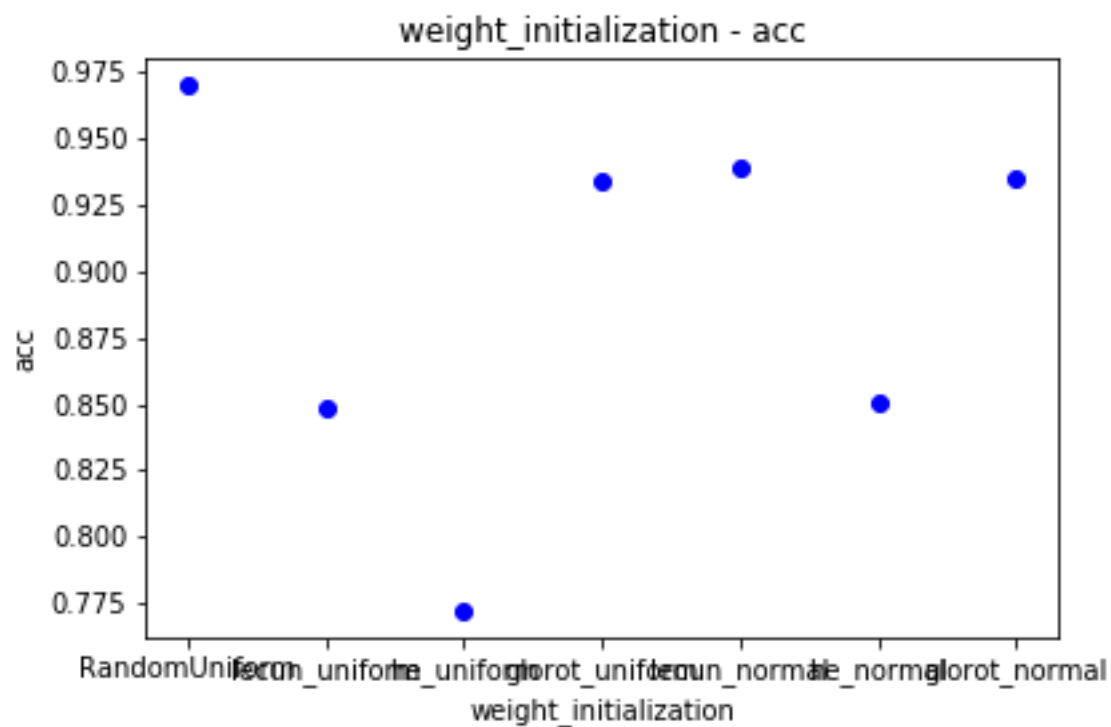


Figure 7: weight initialization - accuracy

```
In [10]: print (best_result)
[0.2247793875939665, 0.9747]
```

Figure 8: score of the combined model