Yue Zhou

27 May 2020

Assignment 2 bonus

DD2424

1.Optimize the performance of the network

1.1. Grid search on the length of the cycles and the number of the cycles.

(lambda=10**(-2.82638))

Accuracy(test)	ns=1epoch	ns=2epoch	ns=3epoch	ns=4epoch	ns=5epoch	ns=6epoch	ns=7epoch	ns=8epoch
1cycle	47.72	50.1	50.54	51.12	51.15	51.9	51.79	51.8
2cycles	50.08	51.38	51.7	51.91	51.46	52.2	52.41	52.16
3cycles	50.87	51.68	51.93	52.21	51.77	51.98	52.55	52.58
4cycles	51.42	52.3	52.35	51.97	51.81	52.15	52.75	52.15
5cycles	51.57	51.83	52.21	52.05	52.05	52.04	52.61	52.72

When the ns increases and cycles increase, the accuracy on test data is increasing overall.

So I choose the hyper parameters which achieve the highest accuracy—na=7epoch, 4cycles.

1.2 Apply more hidden nodes

(lambda=10**(-2.82638), cycle=2, ns=7epoch)

Hidden nodes	50	75	100	150	200	300
Test accuracy	52.01	53.24	53.38	53.38	52.01	52.02

We can see from the results that increasing hidden nodes can improve the accuracy at some extend, but we will encounter overfitting if the hidden node are too many.

1.3 Apply drop out to training

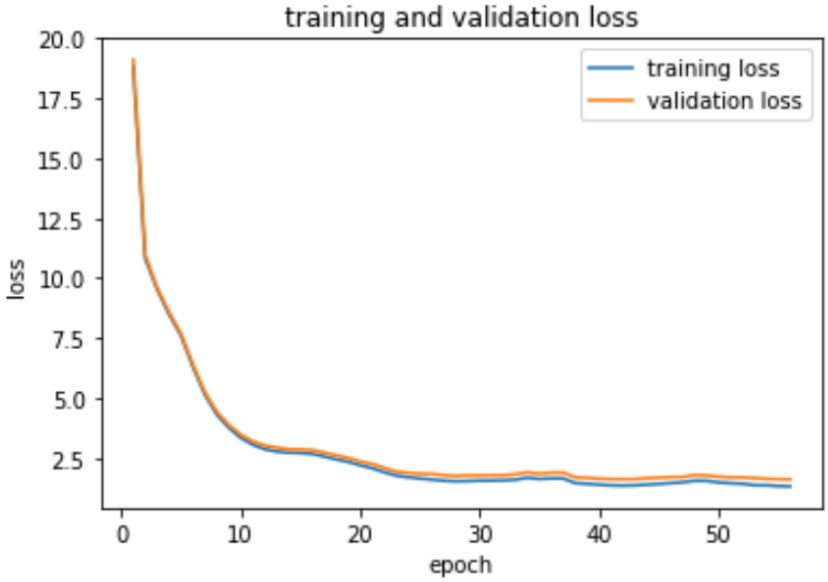
(lambda=10**(-2.82638), cycle=2, ns=7epoch, hidden nodes=100)

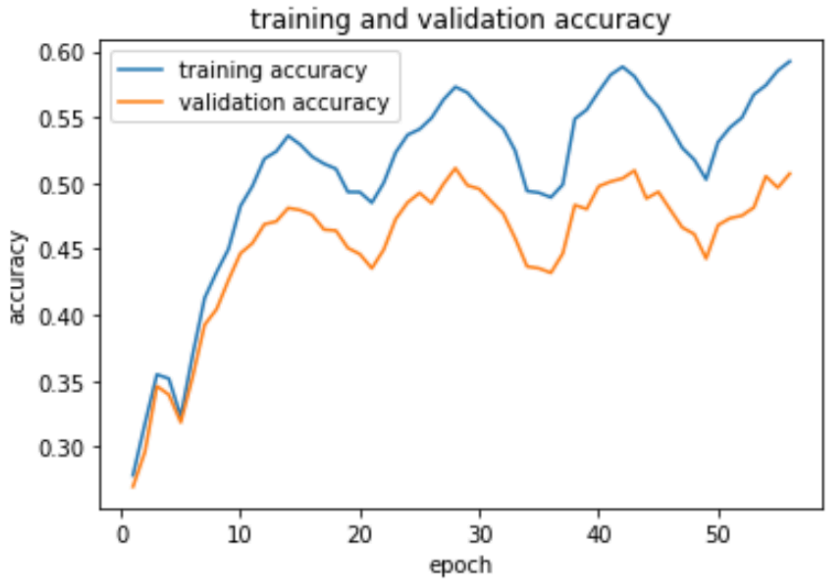
Drop out rate	0	0.2	0.3	0.5
Training accuracy	71.81	57.30	52.87	45.94
Test accuracy	53.38	54.17	53.41	50.91

We can see that apply proper dropout can increase the performance of test data, but decrease the performance of training data, because we add more redundancy and noise to the data at training time, thus reduces overfitting.

1.4 The best accuracy achieved

-lambda=10**(-2.82638), cycle=4, ns=7epoch, hidden nodes=100, dropout=0.2



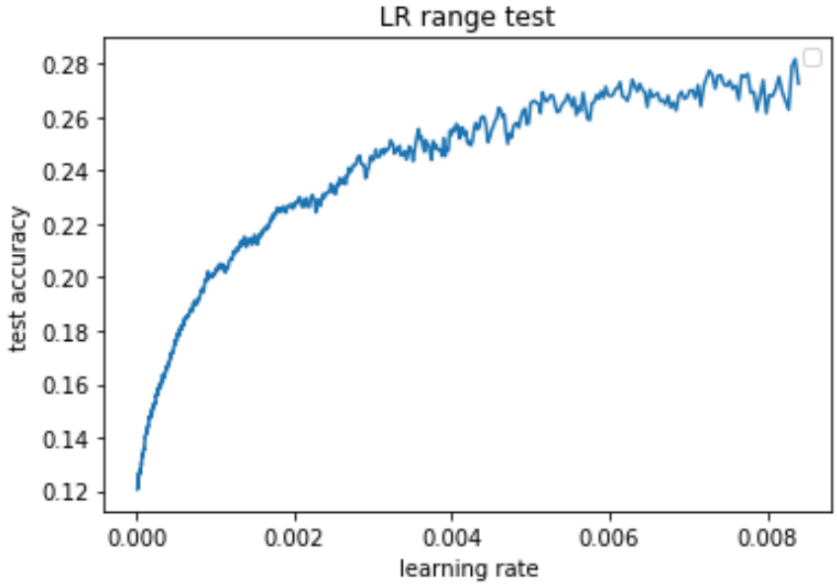


accuracy	
Training	59.24
Validation	50.72
Test	54.83

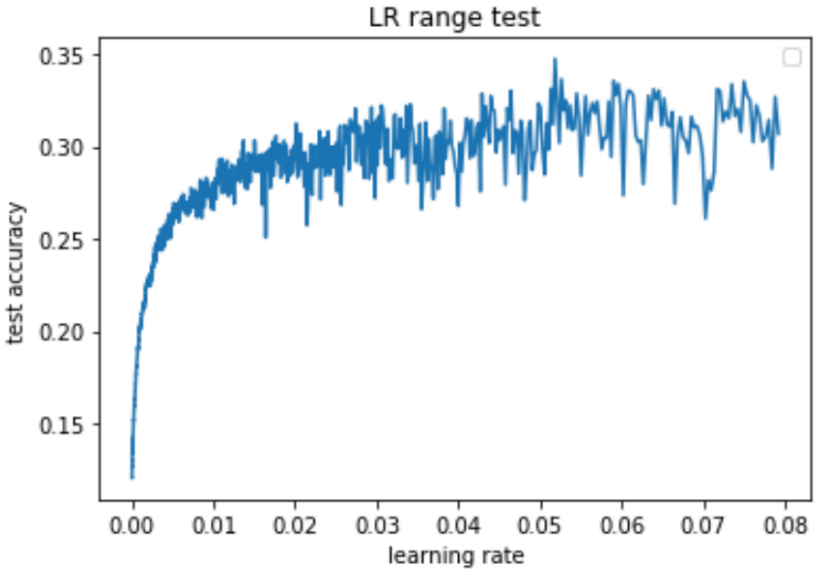
2. Learning rate range test

I use LR range test to find the best boundary for cyclic learning rate. I run my model for several epochs while letting the learning rate increase by multiply by factor 1.004 after every iteration. The other parameters are the same in 1.4. The initial value for learning rate is 1e-5, and run for 3-5 epochs. Then plot the test accuracy versus learning rate.

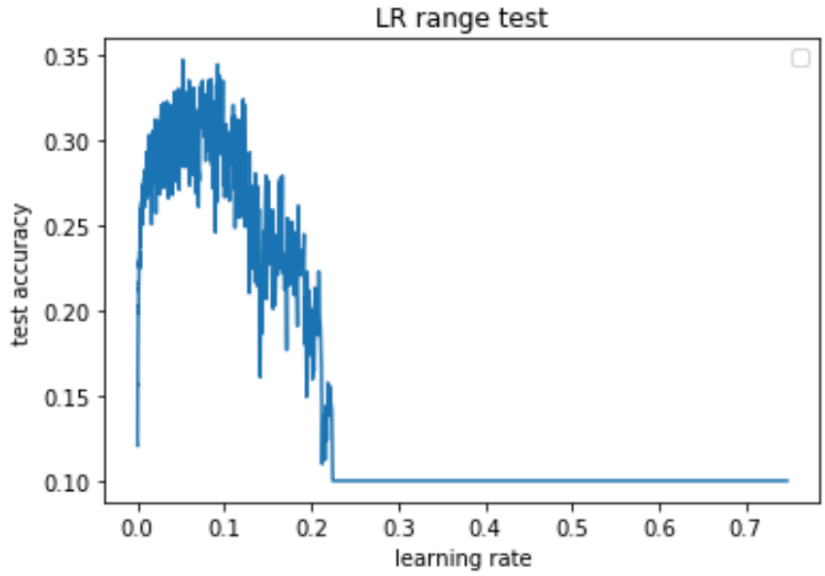
3 epoch:



4 epoch:



5 epoch:



From these results, we take learning rate values when the accuracy starts to increase and when the accuracy slows, became ragged, or starts to fall. I chose max learning rate=0.1 and min learning rate=0.02.