CENG499-THE1 Report

## Muhammed Furkan YAĞBASAN

26.03.2019

1 Hyperparameter optimization

1.1 1-layer (0-hidden-layer) network

Below table shows the scores of different hyperparameter settings. In this case only the learning rates are changed. The scores are the training-accuracy results of each setting. learning rate 0.00003 has the best score which is 0.85 training accuracy and it has also the best test accuracy result with 0.66. Its training-validation loss graph is shown below figure.

Layer	Learning Rate							
Activations	0.1	0.03	0.01	0.003	0.001	0.0003	0.0001	0.00003
-	0.37	0.46	0.52	0.69	0.79	0.80	0.83	0.85

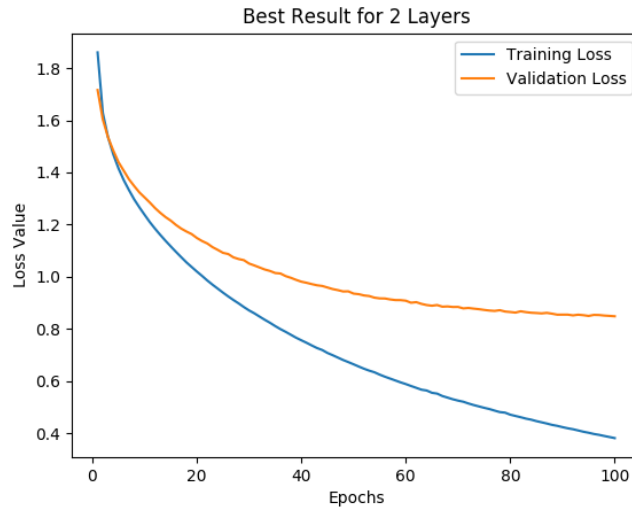Table 1: 1-layer network training accuracy table
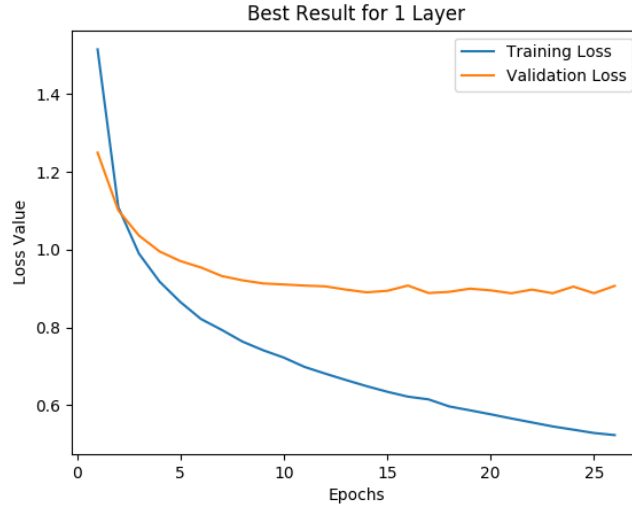
1.2 2-layer (1-hidden-layer) network

In the table below different learning rates and activation functions used to train a 2 layer network. 25 nodes are used in the hidden layer for all settings(Increasing the number of hidden layers also increased the training accuracy for small learning rates but did not improve test accuracy). Sigmoid function had little better results than other settings had in 0.00003 learning rate with 0.69 test accuracy. It did not stop with early stopping and continued till 100 epoches which is set as maximum epoch number. (training-validation loss graph is shown below)

Layer	Learning Rate							
Activations	0.1	0.03	0.01	0.003	0.001	0.0003	0.0001	0.00003
S	0.42	0.48	0.59	0.71	0.81	0.92	0.93	0.92
T	0.39	0.47	0.55	0.65	0.70	0.86	0.95	0.92
R	0.07	0.32	0.43	0.75	0.83	0.88	0.95	0.92

Table 2: 2-layer network training accuracy table

1.3   3-layer (2-hidden-layer) network

In the table below different learning rates and activation functions used to train a 3 layer network. 2 hidden layers are used with 25 nodes for each. Although ST,SR and TR have the close results, ST(first sigmoid, then tanh) has sliently better score with 0.70 test accuracy in 0.00003 learning rate. (training-validation loss graph is shown below)

Layer	Learning Rate							
Activations	0.1	0.03	0.01	0.003	0.001	0.0003	0.0001	0.00003
SS	0.39	0.52	0.63	0.70	0.78	0.90	0.92	0.88
ST	0.34	0.51	0.62	0.68	0.77	0.92	0.93	0.92
SR	0.32	0.45	0.60	0.66	0.74	0.90	0.92	0.90
TS	0.39	0.42	0.51	0.64	0.73	0.85	0.94	0.94
TT	0.34	0.41	0.49	0.62	0.72	0.82	0.92	0.93
TR	0.34	0.44	0.58	0.60	0.72	0.84	0.93	0.92
RS	0.30	0.32	0.48	0.75	0.87	0.92	0.93	0.92
RT	0.26	0.32	0.41	0.67	0.78	0.91	0.96	0.95
RR	0.31	0.09	0.32	0.79	0.82	0.88	0.86	0.90

Table 3: 3-layer network training accuracy table
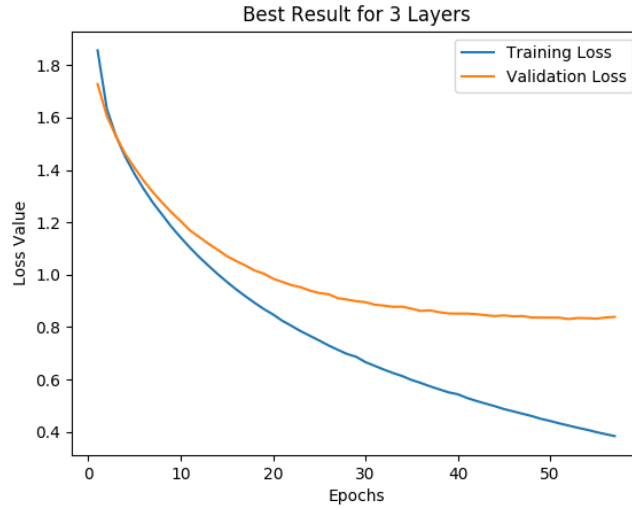
2 Comments

- While training the models, 0.2 of the training set is used as validation set. (Train on 8641 samples, validate on 2161 samples)

- For all trainings, Adam optimizer is used with different learning rates. While activation function was softmax in the output layer, loss function was categorical crossentropy. Moreover, SGD optimizer is also used. It continues to training for many epochs and takes too much time especially for small learning rates. For 0.00003 learning rate and 200 epochs SGD gives 0.60 trainig and 0.56 test accuracy and training and validation losses follows each other so close. For a comparison, with settings: 0.003 learning rate, ST activations, 25 nodes for each hiddel layer, while Adam gives 0.68 training and 0.62 test accuracy, SGD gives 0.87 training and 0.70 test accuracy.

- Data is normilized so that it fits into range -1 to +1. (0,255 to -1,+1)

- Early stop callback function is used to avoid overfitting during training. If the validation loss is not improved for continous 5 epochs, the training is stoped. While training loss keeps decreasing, if validation loss starts to become stable, it means that the model starts to overfit and memorizes the training set.

- After the random initialization of the weights and before the trainig, the test accuracy was araund 0.08 to 0.11. This is an expected result because there are 11 outputs and only 1 correct result. Therefore, randomized result should be around 1/11 statisticly. After the trainig with the settings in the first section of the document , best test accuracy was 0.70 .

- While giving the scores in the tables, training accuracy values are used but validation loss could be a better choice to see actual performance of the model. On the other hand, while deciding the best hyperparameters setting, test accuracy results are also considered, so given graphs belog to settings with best test accuracy results.

- It can be said that the training accuracy is a suitable metric to measure performance in general for the data set used. But for small learning rates, although the training accuracy decreases, test accuracy increases in some settings. On the other hand, if the early stop is not activated, accuracy would not be a suitable metric to measure performance because of overfitting.

4

- Lastly trained the model with SGD, 0.00003 learning rate, 5000 max epoch number, 2 hidden layers with 25 nodes each. It last around 3000 epochs before early stop and gave 0.71 test accuracy. Its training-validation loss graph is below: