

CS6910 Fundamentals of Deep Learning

Assignment 2 Analysis Report

ME18B182 Vasudev Gupta

October 29, 2020

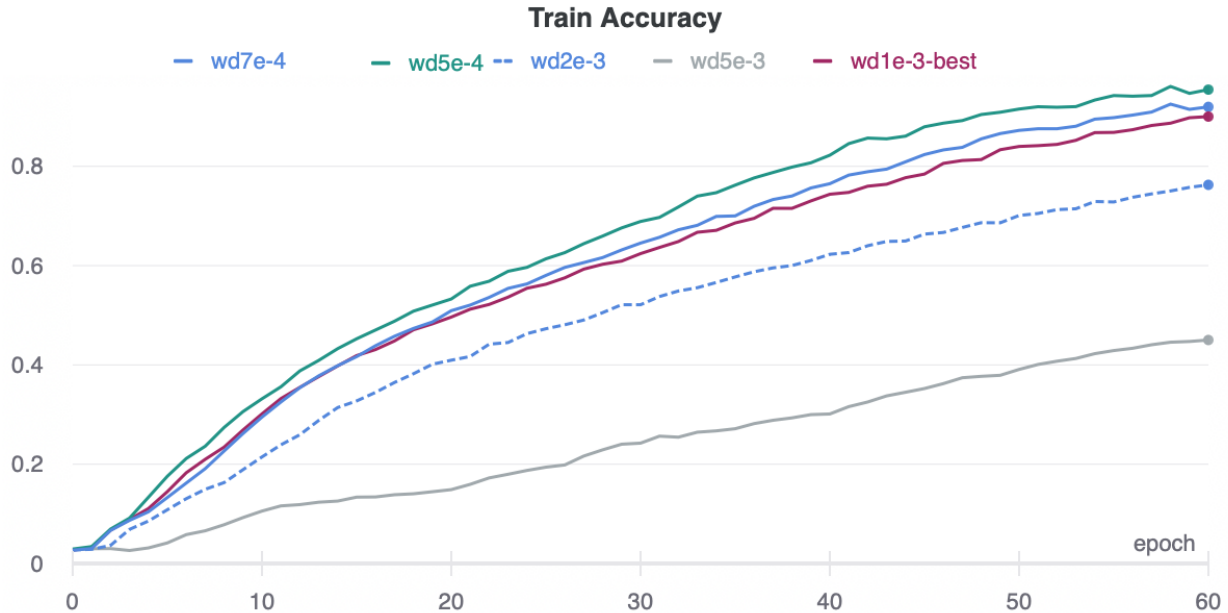
1 Tuning the Regularization parameter

Generally, very deep and complex networks get overfitted on training data and performs very poorly on unseen data. These networks are over-parameterized and loses their ability to generalize. To increase their performance on validation data, its important to remove unimportant weights and introduce some kind of noise during training. Hence, many regularization can be helpful in introducing that noise and making network more robust to noise during inference.

1.1 Weight decay

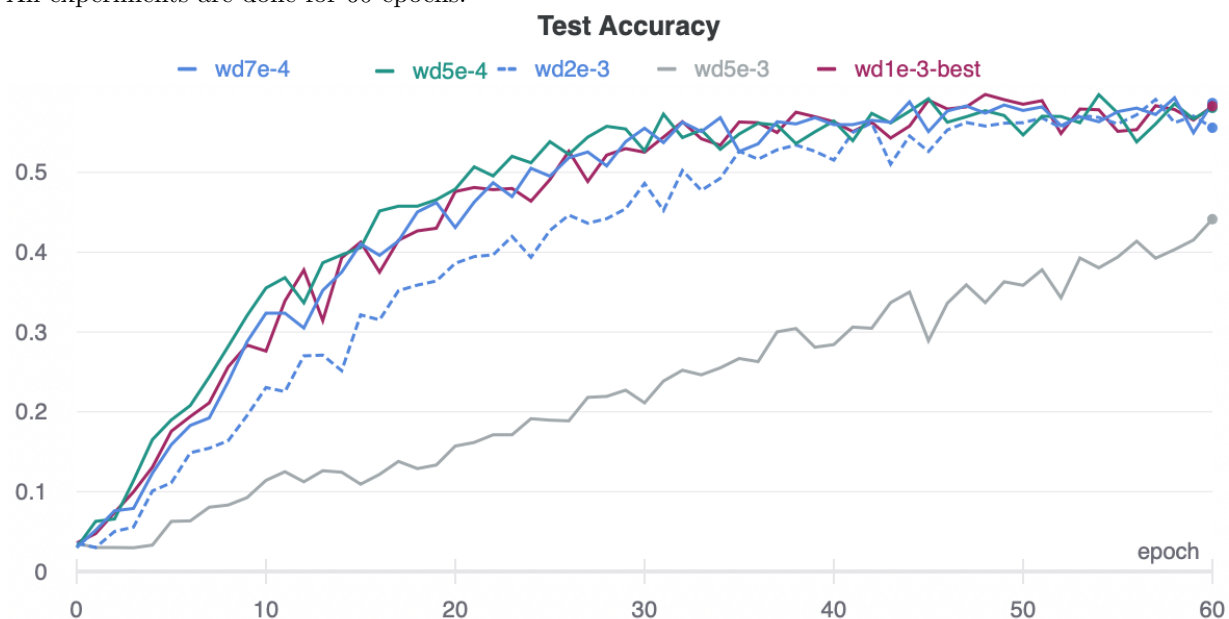
L2 normalization is one sort of regularization which penalizes model's weights by adding l2 norm of weights in the loss function. Weight decay acts as a hyperparameter for keeping l2 norm as regularization technique.

Weight decay	Train Loss	Test Loss	Train Accuracy	Test Accuracy
5e-4	0.1429	2.518	0.9546	0.5808
7e-4	0.2467	2.195	0.9198	0.5801
1e-3	0.3058	2.129	0.9003	0.5829
2e-3	0.7597	1.612	0.7628	0.5559
5e-3	1.855	1.929	0.4502	0.4414



Above figure shows training accuracy for various weight decays. Weight decays are mentioned in figure itself.

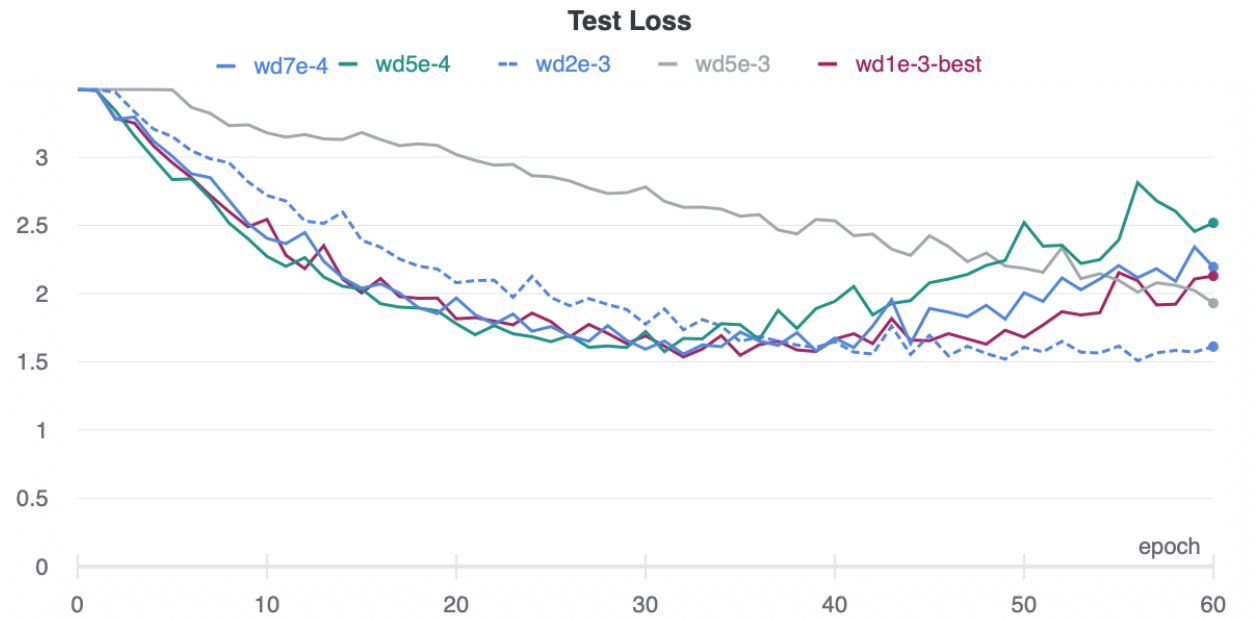
All experiments are done for 60 epochs.



Above figure shows test accuracy for various weight decays. Weight decays are mentioned in figure itself. All experiments are done for 60 epochs.



Above figure shows training loss for various weight decays. Weight decays are mentioned in figure itself. All experiments are done for 60 epochs.



Above figure shows test loss for various weight decays. Weight decays are mentioned in figure itself. All experiments are done for 60 epochs.

2 Conclusion

If weight decay is too large, then model training will be very slow because weights are penalized too much to be near to 0; and hence it is possible that model will underfit. While if weight decay is very small, then model will overfit on the training data and won't generalize on unseen data.