SHARIF UNIVERSITY OF TECHNOLOGY

DEEP LEARNING

HW03 Report

Author
M. SHAHBAZI

Professor
Dr. MS BAGHSHAH

April 10, 2019

Attention

In all parts we used accuracy and loss on test data because we thought that it has more information for comparing two different method.

1)MLP

Question 1

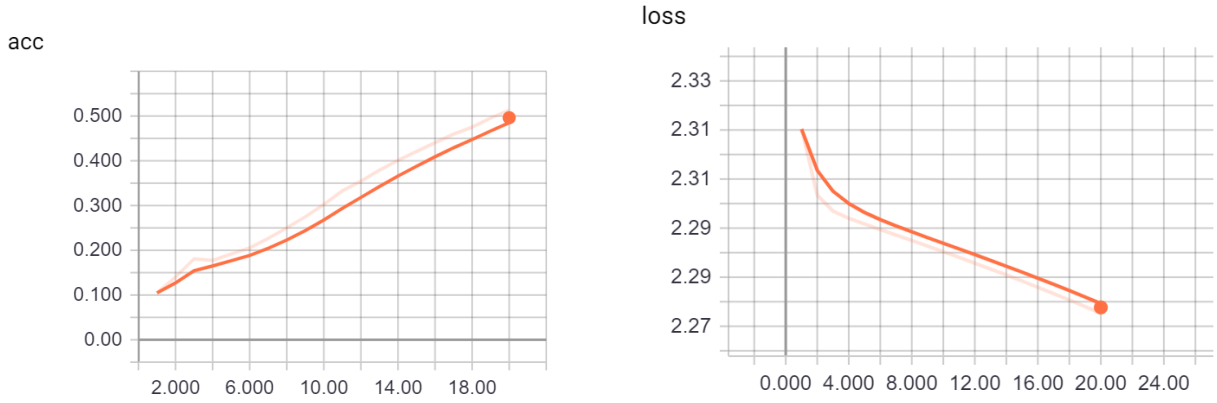


Figure 1: Sigmoid activation function

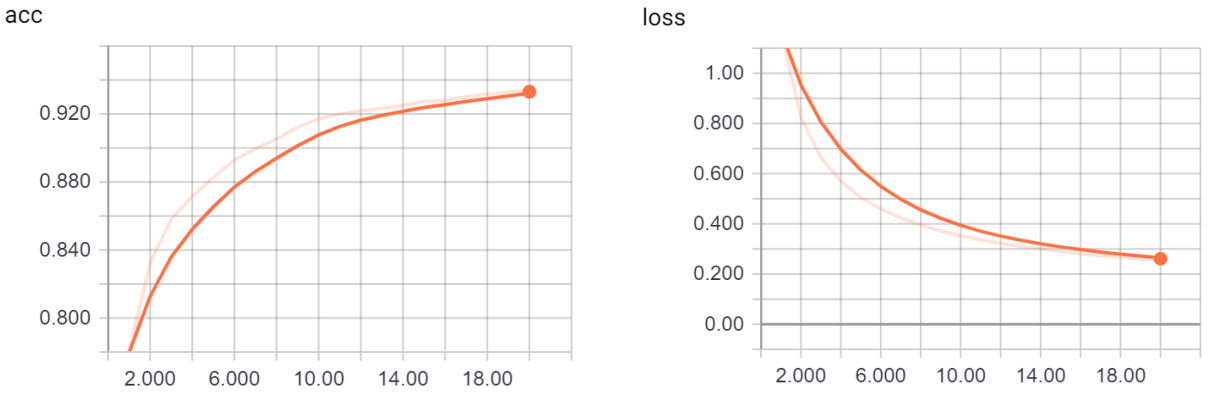


Figure 2: tanh activation function

As we can observe, performance of the tanh activation function is much better than sigmoid. we can infer this result by thinking about the greater gradient of tan around zero.

Question 2

Initialization with zero theoretically causes all neurons of each layer update same as a result there is no learning. But Tensorflow implementation considered this problem and tried to reduce bed effects. We can not track a meaning full pattern in loss and accuracy.

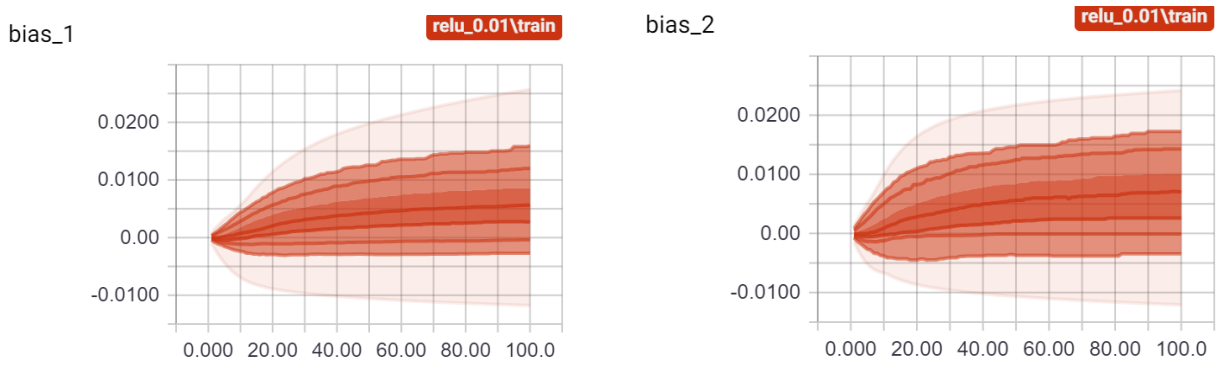


Figure 3: biases of first two layers

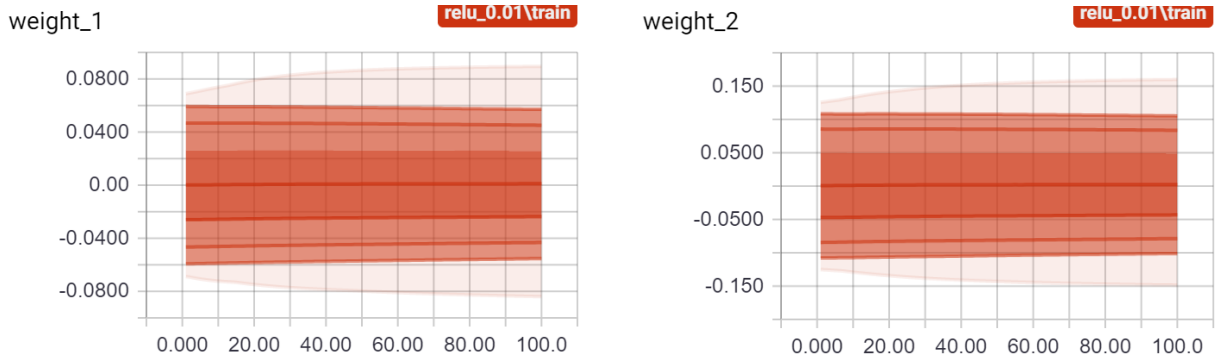
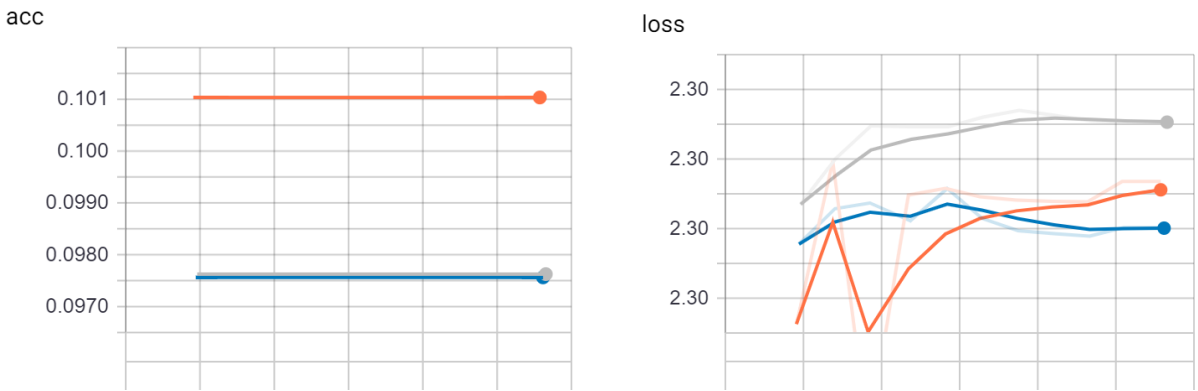


Figure 4: weights of first two layers



Question 3

As we can see, a standard deviation of one is the best case. Because immense weight value may cause saturation of activation function and meager learning rate, very small amount may cause low learning rate again, it is reasonable that choose standard deviation not too much large and not too much little.

Figure 5: stdv = 0.01

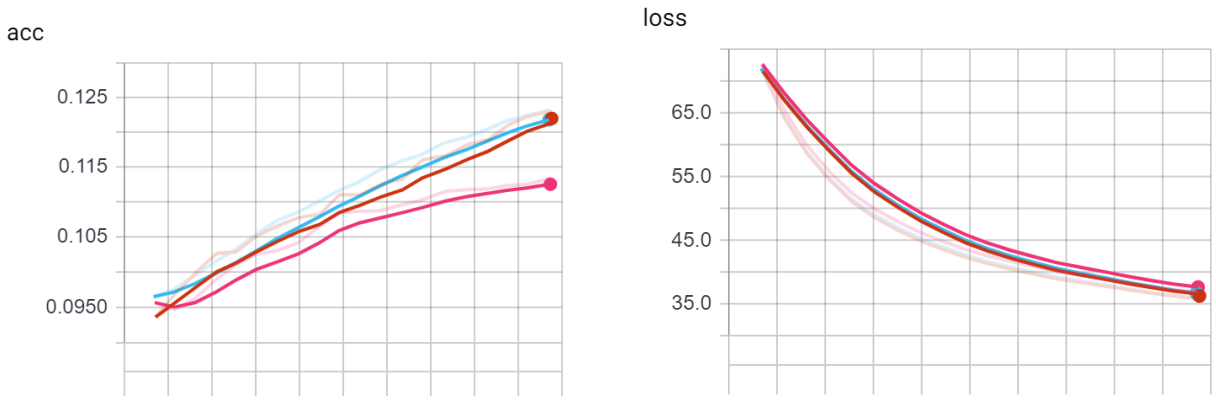


Figure 6: stdv = 10

Figure 7: stdv = 1

Question 4

There is some fact about the model size that we know. First is all larger models have more capacity for learning; therefore if we have a big dataset then larger model will have better performance.
But in another way, if the size of the training data is not enough, then larger models are in danger of overfitting.

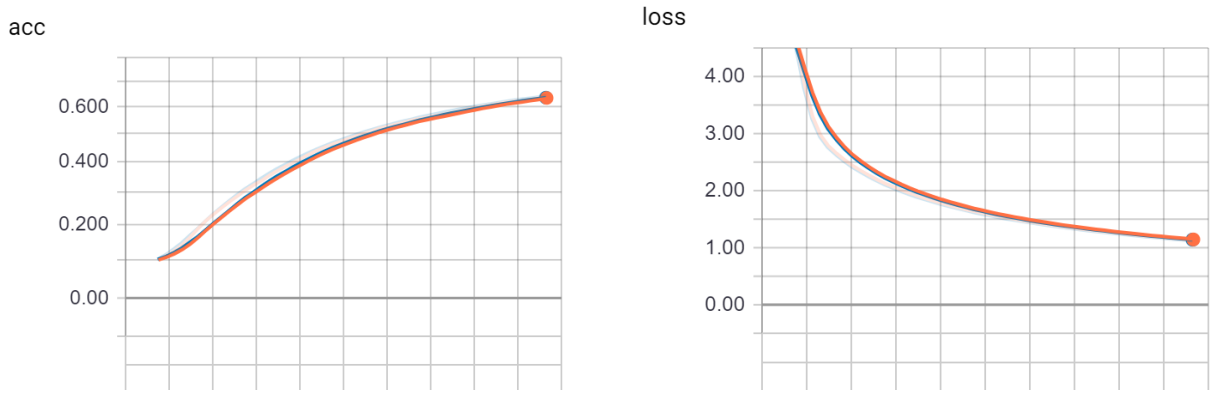


Figure 8: the model with 150 nodes

Figure 9: the model with 500 node

As we can observe larger model has better performance on test and train data.

Question 5

Before looking at plots, we can claim that adding regularization term avoid overfitting and this could be seen by looking at as regularization term increase, different between test and train accuracy at final stages decrease. But we couldn't observe this behavior in plots, and the reason is it seems the model has more capacity because both test and train accuracy increase simultaneously.
There is one more observation. The value of final loss has relative with regularization term, and that is obvious.

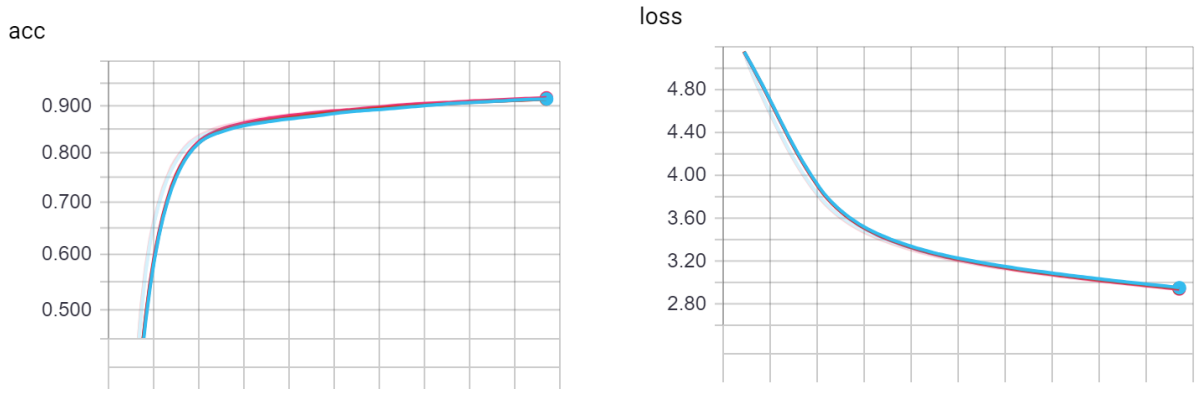


Figure 10: reg. term = 0.1

Figure 11: reg. term = 0

Figure 12: reg. term = 0.0001

CNN Autoencoder

This image is output of t-SNE and different color belongs to different categories of data.

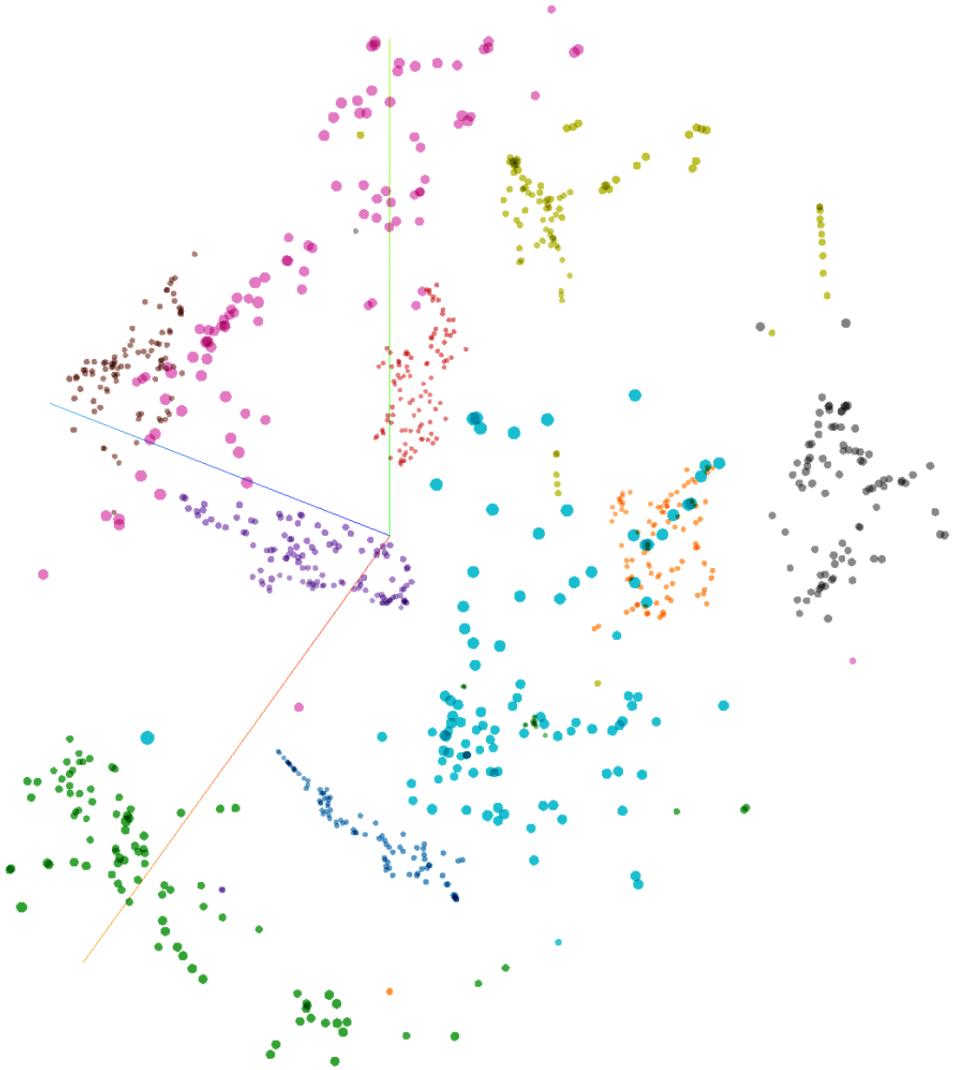


Figure 13: Output of tensorboard