CS764: Assignment 7

Report

Siddharth Saha
170100025

1 Lifting!

1.1 Questions

Q1 Explain the purpose and location of the batch normalization layer, and the dropout layer. What is the reason for duplicating the layer. What if we used three blocks?
Ans The purpose of batch normalization layer is to reduce the effect of choosing random initial weights and provide some degree of regularization by adding a little noise to the system.[1] The original BatchNorm papers suggests that $Wu + b$ is more likely to have a symmetric, non-sparse distribution, that is "more Gaussian". Thus, normalizing it beforehand is likely to produce activations with a stable distribution.[2]

The use of the dropout layer is to prevent overfitting by ignoring a random subsample of the weights given by the dropout probability.[3] The dropout is advised to be done in front of every linear projection in the original paper.[4] The reason for duplicating the layer is to provide two different configurable "views" of the learning problem. If we used three blocks, the system will have more weights to tweak. The provided dataset would become insufficient and the system may tend to overfit.

Q2 Provide details of your training. What experiments did you conduct? Show charts.
Ans I have trained the given model using the following values of the parameters:

1. Epochs: 1 and 2

2. Optimizer: Adam with learning rate=0.001

3. Batch size: 64

4. 80% Training set and 20% Validation set

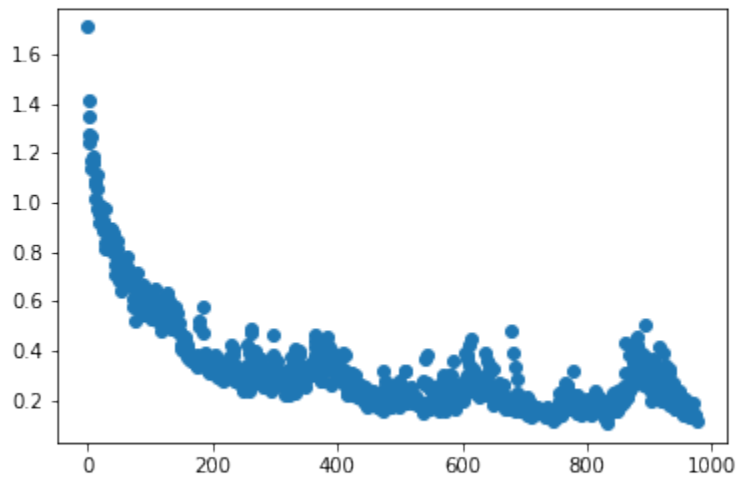The resulting charts for mean per joint position error in the training phase are as follows:
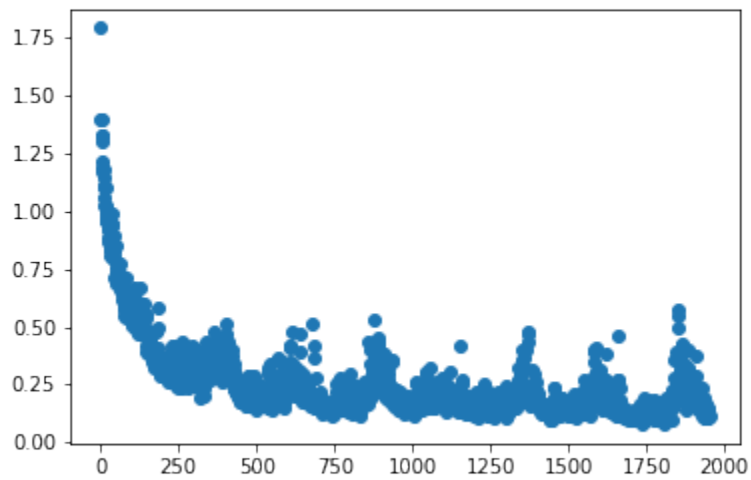
Figure 1: Epochs: 1, Batchsize: 64

Figure 2: Epochs: 2, Batchsize: 64

Comments: Running it for 2 epochs shows signs of overfitting

References

[1] https://towardsdatascience.com/batch-normalization-8a2e585775c9

[2] https://www.reddit.com/r/MachineLearning/comments/67gonq/d_batch_normalization_before_or_after_relu/

[3] https://machinelearningmastery.com/dropout-for-regularizing-deep-neural-networks

[4] https://www.cs.toronto.edu/~hinton/absps/JMLRdropout.pdf