

GPT Mastery: A Comprehensive Guide To **Crafting Your Own ChatGPT** **Assignment**

Week 3

Complete the following task in a Colab notebook:-

Exercises:

- 1.) Tune the hyperparameters of the training to beat my best validation loss of 2.2
- 2.) I was not careful with the initialization of the network in this video. (1) What is the loss you'd get if the predicted probabilities at initialization were perfectly uniform? What loss do we achieve? (2) Can you tune the initialization to get a starting loss that is much more similar to (1)?
- 3.) Read the Bengio et al 2003 paper (link above), implement and try any idea from the paper. Did it work?
- 4.) I did not get around to seeing what happens when you initialize all weights and biases to zero. Try this and train the neural net. You might think either that 1) the network trains just fine or 2) the network doesn't train at all, but actually it is 3) the network trains but only partially, and achieves a pretty bad final performance. Inspect the gradients and activations to figure out what is happening and why the network is only partially training, and what part is being trained exactly.

5.)BatchNorm, unlike other normalization layers like LayerNorm/GroupNorm etc. has the big advantage that after training, the batchnorm gamma/beta can be "folded into" the weights of the preceeding Linear layers, effectively erasing the need to forward it at test time. Set up a small 3-layer MLP with batchnorms, train the network, then "fold" the batchnorm gamma/beta into the preceeding Linear layer's W, b by creating a new W_2, b_2 and erasing the batch norm. Verify that this gives the same forward pass during inference. i.e. we see that the batchnorm is there just for stabilizing the training, and can be thrown out after training is done! pretty cool.

SUBMISSION LINK: <https://forms.gle/mfZ6K1DSExjYUWNs6>

NOTE:Please ensure you submit the assignments. While there are no specific deadlines, you must submit at least 4 assignments by the end of this project. Failure to do so will result in not receiving the certificates. This is important as we need to present your work to WnCC for approval of your resume point. Please take this seriously.